Monotonicity Inference in Theory and Practice

Larry Moss

NASSLLI'2025

June 23-25, 2025

Entertainment to initiate this topic: a 3 minute video on monotonicity

https://www.youtube.com/watch?v=-zBPHuBGZAE&feature=youtu.be

This was part of a competition for non-technical videos on mathematics for the general public.

Monotonicity is a big part of the Natural Logic Program

Show that significant parts of inference in natural language can be done automatically, using surface forms as much as possible.

To work in decidable, and "light" logical systems.

To make connections with all relevant fields of logic.

To re-work semantics in light of computational linguistics and cognitive science.

Let's give up on grammar, at least on the very strict forms of grammar which we have seen so far!

Backdrop

Monotonicity is a big part of the Natural Logic Program

Show that significant parts of inference in natural language can be done automatically, using surface forms as much as possible.

To work in decidable, and "light" logical systems.

To make connections with all relevant fields of logic.

To re-work semantics in light of computational linguistics and cognitive science.

In another direction

Since 2018, neural machine learners have been able to do some forms of inference at human level, or better.

Thus, we have a watershed moment in the history of logic.

This calls for a response.

Inference by substitution

(1) some$^{\uparrow}$ dog$^{\uparrow}$ hit$^{\uparrow}$ some$^{\uparrow}$ cat$^{\uparrow}$

(2) some$^{\uparrow}$ dog$^{\uparrow}$ kissed$^{\downarrow}$ no$^{\uparrow}$ cat$^{\downarrow}$

(3) most$^{\uparrow}$ dog$^{=}$ hit$^{\downarrow}$ no$^{\uparrow}$ cat$^{\downarrow}$

(4) no$^{\uparrow}$ dog$^{\downarrow}$ hit$^{\uparrow}$ no$^{\downarrow}$ cat$^{\uparrow}$

(5) at-most two$^{\uparrow}$ dog$^{\downarrow}$ chased$^{\uparrow}$ at-most three$^{\downarrow}$ cats$^{\uparrow}$

knowledge base for nouns, transitive verbs, determiners, and numbers

dog $\leq$ animal	kiss $\leq$ touch	every $\leq$ most
cat $\leq$ animal	hit $\leq$ touch	most $\leq$ some
poodle $\leq$ dog	thrash $\leq$ hit	one $\leq$ two
ragdoll $\leq$ cat	hit vigorously $\leq$ hit	two $\leq$ three
$\vdots$	hit lightly $\leq$ hit	three $\leq$ four
$\vdots$	$\vdots$	no $\leq$ at-most one

(6) some$^{\uparrow}$ dog$^{\uparrow}$ hit$^{\uparrow}$ some$^{\uparrow}$ cat$^{\uparrow}$

(7) some$^{\uparrow}$ dog$^{\uparrow}$ kissed$^{\downarrow}$ no$^{\uparrow}$ cat$^{\downarrow}$

(8) most$^{\uparrow}$ dog$^{=}$ hit$^{\downarrow}$ no$^{\uparrow}$ cat$^{\downarrow}$

$\Rightarrow$ some dog thrashed no ragdoll

(9) no$^{\uparrow}$ dog$^{\downarrow}$ hit$^{\uparrow}$ no$^{\downarrow}$ cat$^{\uparrow}$

$\Rightarrow$ at most one poodle touched no animal

(10) at-most two$^{\uparrow}$ dog$^{\downarrow}$ chased$^{\uparrow}$ at-most three$^{\downarrow}$ cats$^{\uparrow}$

knowledge base for nouns, transitive verbs, determiners, and numbers

dog $\leq$ animal	kiss $\leq$ touch	every $\leq$ most
cat $\leq$ animal	hit $\leq$ touch	most $\leq$ some
poodle $\leq$ dog	thrash $\leq$ hit	one $\leq$ two
ragdoll $\leq$ cat	hit vigorously $\leq$ hit	two $\leq$ three
$\vdots$	hit lightly $\leq$ hit	three $\leq$ four
$\vdots$	$\vdots$	no $\leq$ at-most one

Put the arrow on chased in

No dog (chased no cat)

Put the arrow on chased in

No dog (chased no cat)

$=$ Every dog (chased some cat)

Put the arrow on chased in

No dog (chased no cat)

$=$ Every dog (chased some cat)

Moss 2012: no$^{\uparrow}$ dog$^{\downarrow}$ chased$^{\downarrow}$ no$^{\downarrow}$ cat$^{\uparrow}$.

The arrow on chased$^{\downarrow}$ is wrong.

An algebraic expression like

$$z - (v + w)$$

is increasing in z, and decreasing in v and w.

If we assume

- $z_1 \leq z_2$
- $v_2 \leq v_1$
- $w_2 \leq w_1$

Then we are entitled to conclude

$$z_1 - (v_1 + w_1) \leq z_2 - (v_2 + w_2)$$

We had

$$z - (v + w)$$

We would write

$$\frac{v^{\downarrow} \quad w^{\downarrow} \quad z^{\uparrow}}{(z - (v + w))^{\uparrow}}$$

The responsible parties here are the facts that

$+ : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$	is increasing (monotone) in both arguments
$- : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$	is increasing in the first argument and decreasing (antitone) in the second argument

An algebraic expression like

$$z - (v + |w|)$$

is increasing in z, and decreasing in v, and
there's nothing we can say about w.

If we assume

- $z_1 \leq z_2$
- $v_2 \leq v_1$
- $w_2 = w_1$

Then we are entitled to conclude

$$z_1 - (v_1 + w_1) \leq z_2 - (v_2 + w_2)$$

We had

$$z - (v + |w|)$$

We would write

$$\frac{v^{\downarrow} \quad w^{=} \quad z^{\uparrow}}{(z - (v + w))^{\uparrow}}$$

We just saw

$$\frac{v^{\downarrow} \quad w^{=} \quad z^{\uparrow}}{(z-(v+w))^{\uparrow}}$$

and it follows that

$$\frac{v^{\downarrow} \quad w^{=} \quad z^{=}}{(z-(v+w))^{\uparrow}}$$

We have strengthened the hypotheses and thus weakened the statement:
If we assume

- $z_1 = z_2$
- $v_2 \leq v_1$
- $w_2 = w_1$

Then we are once again entitled to conclude

$$z_1 - (v_1 + w_1) \leq z_2 - (v_2 + w_2)$$

. . . because algebra is first-order; we need higher-order stuff

Some$^{\uparrow}$ dog$^{\uparrow}$ and$^{\uparrow}$ no$^{\uparrow}$ cat$^{\downarrow}$ sleeps$^{=}$

Abstract some words:

$s(x, y, z) = [\![\mathit{Some}\ x\ \mathit{and}\ \mathit{no}\ y\ z]\!]$ in some underlying model

Our intuitions tell us that
s is decreasing in x, decreasing in y, and
there's nothing we can say about z.

If we assume

- $x_1 \leq x_2$
- $y_2 \leq y_1$
- $z_1 = z_2$

Then we are entitled to conclude

$$[\![\mathit{Some}\ x_1\ \mathit{and}\ \mathit{no}\ y_1\ z_1]\!] \leq [\![\mathit{Some}\ x_2\ \mathit{and}\ \mathit{no}\ y_2\ z_2]\!]$$

. . . because algebra is first-order; we need higher-order stuff

Some$^{\uparrow}$ dog$^{\uparrow}$ and$^{\uparrow}$ no$^{\uparrow}$ cat$^{\downarrow}$ sleeps$^{=}$

Abstract some words:

$s(x, y, z) = [\![\mathit{Some}\ x\ \mathit{and\ no}\ y\ z]\!]$ in some underlying model

Our intuitions tell us that
s is decreasing in x, decreasing in y, and
there's nothing we can say about z.

If we assume

- $x_1 \leq x_2$, but where?
- $y_2 \leq y_1$, but where?
- $z_1 = z_2$, but where? And what about the Dets?

Because the model is hypothetical,

Some x_1 and no y_1 z_1 implies Some x_2 and no y_2 z_2

A lexicon

Observe the 3 *np* types (!) and the $+/-/\cdot$ notation

item	category	type
Fido, Felix	NP	e
cat, dog	N	$n = pr = e \to t$
swim, run	$\mathrm{IV} = \mathrm{S}\backslash\mathrm{NP}$	pr
chase, see, hit, kiss	$\mathrm{TV} = \mathrm{IV}/\mathrm{NP}$	$e \to pr$
every	$\mathrm{DET} = \mathrm{NP}/\mathrm{N}$	$pr \xrightarrow{-} np^{+}$
some	NP/N	$pr \xrightarrow{+} np^{+}$
no	NP/N	$pr \xrightarrow{-} np^{-}$
most	NP/N	$pr \xrightarrow{\cdot} np^{+}$
who	$(\mathrm{N}\backslash\mathrm{N})/(\mathrm{S}/\mathrm{NP})$	$(np^{+} \xrightarrow{+} t) \xrightarrow{+} (pr \xrightarrow{+} pr)$
didn't	IV/IV TV/TV	$pr \xrightarrow{-} pr$ $(e \to pr) \xrightarrow{-} (e \to pr)$
and	$\mathrm{X}/(\mathrm{X}\backslash\mathrm{X})$	$x \xrightarrow{+} (x \xrightarrow{+} x)$
one, two, three	NUM	num
more than	$\mathrm{DET}/\mathrm{NUM}$	$num \xrightarrow{-} (pr \xrightarrow{+} np^{+})$
less than	$\mathrm{DET}/\mathrm{NUM}$	$num \xrightarrow{+} (pr \xrightarrow{-} np^{-})$
if ... then ...	$(\mathrm{S}\backslash\mathrm{S})/\mathrm{S}$	$t \xrightarrow{-} (t \xrightarrow{+} t)$

Markings vs. Polarities

Markings are what words have in the lexicon:

$$\text{every} : pr \xrightarrow{-} np^{+}$$

We indicate markings by $+$, $-$, and $\cdot$.

Poiarities arrows $\uparrow$, $\downarrow$ and $=$
indicate "in sentence" monotonicity/antitonicity/neither assertions.

They are indicating the same things, but markings are permanent features of the lexical items.

Polarities are what we want to calculate in this work.

Rules

$$\frac{u^{d} : x \xrightarrow{m} y \quad v^{md} : x}{(uv)^{d} : y}\ > \qquad \frac{u^{md} : x}{u^{d} : (x \xrightarrow{m} y) \xrightarrow{+} y}\ \mathrm{T} \qquad \frac{u^{md} : u \xrightarrow{n} y \quad v^{md} : y \xrightarrow{n} z}{(uv)^{d} : (x \xrightarrow{mn} z)}\ \mathrm{B}$$

$$\frac{u^{md} : e \rightarrow b}{u^{d} : np^{m} \xrightarrow{+} b}\ \mathrm{K} \qquad \frac{u^{d} : x \xrightarrow{m} y}{u^{d} : x \xrightarrow{\cdot} y}\ \mathrm{M} \qquad \frac{u^{=} : x \xrightarrow{m} y}{u^{d} : x \xrightarrow{m} y}\ \mathrm{W}$$

	$\uparrow$	$\downarrow$	$=$
$+$	$\uparrow$	$\downarrow$	$=$
$-$	$\downarrow$	$\uparrow$	$=$
$\cdot$	$=$	$=$	$=$

d can be $\uparrow$ or $\downarrow$ or $=$.

m can be $+$, $-$, or $\cdot$.

There are two operations that combine these notation,
and we'll see one of them soon.

What one does with the polarization rules

One looks for a tree with the following features:

- The leaves must have a type from the lexicon.
- The leaves may be polarized $\uparrow$, $\downarrow$, or $=$; it's our choice.
- The type at the bottom must be t.
  The polarity at the bottom must be $\uparrow$.
- At each non-leaf node, one of the rules of the system must be matched.

The goal

If we take the sentence at the root of the tree,
then all polarity arrows at the leaves
should match our intuitions.

Example

Example

$$
\dfrac{
\dfrac{\text{no}^{\uparrow} : pr \xrightarrow{-} np^{-} \qquad \text{dog}^{\downarrow} : pr}{\text{no dog}^{\uparrow} : pr \xrightarrow{-} t}\,{>}
\qquad
\dfrac{
\dfrac{\text{chased}^{\uparrow} : e \to pr}{\text{chased}^{\downarrow} : np^{-} \xrightarrow{+} pr}\,\mathrm{K}
\qquad
\dfrac{\text{no}^{\downarrow} : pr \xrightarrow{-} np^{-} \qquad \text{cat}^{\uparrow} : pr}{\text{no cat}^{\downarrow} : np^{-}}\,{>}
}{\text{chased no cat}^{\downarrow} : pr}\,{>}
}{\text{no dog chased no cat}^{\uparrow} : t}\,{>}
$$

Rules

$$
\dfrac{u^{d} : x \xrightarrow{m} y \qquad v^{md} : x}{(uv)^{d} : y}\,{>}
\qquad
\dfrac{u^{md} : x}{u^{d} : (x \xrightarrow{m} y) \xrightarrow{+} y}\,\mathrm{T}
\qquad
\dfrac{u^{md} : u \xrightarrow{n} y \qquad v^{md} : y \xrightarrow{n} z}{(uv)^{d} : (x \xrightarrow{mn} z)}\,\mathrm{B}
$$

$$
\dfrac{u^{md} : e \to b}{u^{d} : np^{m} \xrightarrow{+} b}\,\mathrm{K}
\qquad
\dfrac{u^{d} : x \xrightarrow{m} y}{u^{d} : x \xrightarrow{\cdot} y}\,\mathrm{M}
\qquad
\dfrac{u^{=} : x \xrightarrow{m} y}{u^{d} : x \xrightarrow{m} y}\,\mathrm{W}
$$

Example

$$\frac{\text{no}^{\uparrow} : pr \xrightarrow{-} np^{-} \quad \text{dog}^{\downarrow} : pr}{\text{no dog}^{\uparrow} : pr \xrightarrow{-} t} > \qquad \frac{\text{no}^{\downarrow} : pr \xrightarrow{-} np^{-} \quad \text{cat}^{\uparrow} : pr}{\text{no cat}^{\downarrow} : pr \xrightarrow{-} t} >$$

item	type
cat, dog	$pr = e \to t$
no	$pr \xrightarrow{-} np^{-} = (e \to t) \xrightarrow{-} np^{-}$

van Benthem's rule

$$\frac{u^{d} : x \xrightarrow{m} y \quad v^{md} : x}{(uv)^{d} : y} >$$

	$\uparrow$	$\downarrow$	$=$
$+$	$\uparrow$	$\downarrow$	$=$
$-$	$\downarrow$	$\uparrow$	$=$
$\cdot$	$=$	$=$	$=$

In the left application, we have $d = \uparrow$ and $m = -$, so that $md = \downarrow$.
In the right application, we have $d = \downarrow$ and $m = -$, so that $md = \uparrow$.

The arrow in the conclusion is inherited from the function above, not the argument.

van Benthem's groundbreaking contributions

Work on the logical side of monotonicity/polarity stems from the seminal contributions of Johan van Benthem.

His central insight

Polarities can be computed from derivation trees in categorial grammar (CG).

References

Essays in Logical Semantics, 1986

Language in Action: Categories, Lambdas, and Dynamic Logic, 1991

"A Brief History of Natural Logic", 2008

see also the 1991 PhD dissertation by Victor Sánchez Valencia Studies on Natural Logic and Categorial Grammar

The K rule in this example

Example

$$\frac{\text{chased}^{\uparrow} : e \to (e \to t)}{\text{chased}^{\downarrow} : np^{-} \xrightarrow{+} (e \to t)}\ \mathrm{K}$$

item	type
chased	$e \to pr = e \to (e \to t)$

K

$$\frac{u^{md} : e \to b}{u^{d} : np^{m} \xrightarrow{+} b}\ \mathrm{K}$$

	$\uparrow$	$\downarrow$	$=$
$+$	$\uparrow$	$\downarrow$	$=$
$-$	$\downarrow$	$\uparrow$	$=$
$\cdot$	$=$	$=$	$=$

b must be a boolean type:
t is boolean, and if β is boolean, so is $\sigma \to \beta$.

We take $m = -$ and $d = \downarrow$.

Example

$$
\cfrac{\cfrac{\mathsf{some}^{\uparrow} : pr \xrightarrow{+} np^{+} \qquad \mathsf{dog}^{\uparrow} : pr}{\mathsf{some\ dog}^{\uparrow} : pr \xrightarrow{+} t}\,{>} \qquad \cfrac{\cfrac{\cfrac{\mathsf{chased}^{\downarrow} : e \xrightarrow{+} pr}{\mathsf{chased}^{\uparrow} : np^{-} \xrightarrow{+} pr}\,\mathrm{K} \qquad \cfrac{\mathsf{no}^{\uparrow} : pr \xrightarrow{-} np^{-} \qquad \mathsf{cat}^{\downarrow} : pr}{\mathsf{no\ cat}^{\uparrow} : np^{-}}\,{>}}{\mathsf{chased\ no\ cat}^{\uparrow} : pr}\,{>}}{\mathsf{some\ dog\ chased\ no\ cat}^{\uparrow} : t}\,{>}
$$

Rules

$$
\frac{u^{d} : x \xrightarrow{m} y \qquad v^{md} : x}{(uv)^{d} : y}\,{>} \qquad \frac{u^{md} : x}{u^{d} : (x \xrightarrow{m} y) \xrightarrow{+} y}\,\mathrm{T} \qquad \frac{u^{md} : u \xrightarrow{n} y \qquad v^{md} : y \xrightarrow{n} z}{(uv)^{d} : (x \xrightarrow{mn} z)}\,\mathrm{B}
$$

$$
\frac{u^{md} : e \to b}{u^{d} : np^{m} \xrightarrow{+} b}\,\mathrm{K} \qquad \frac{u^{d} : x \xrightarrow{m} y}{u^{d} : x \xrightarrow{\cdot} y}\,\mathrm{M} \qquad \frac{u^{=} : x \xrightarrow{m} y}{u^{d} : x \xrightarrow{m} y}\,\mathrm{W}
$$

	$\uparrow$	$\downarrow$	$=$
$+$	$\uparrow$	$\downarrow$	$=$
$-$	$\downarrow$	$\uparrow$	$=$
$\cdot$	$=$	$=$	$=$

Example

$$\dfrac{\dfrac{\text{most}^{\uparrow} : pr \xrightarrow{\cdot} np^{+} \qquad \text{dogs}^{=} : pr}{\text{most dogs}^{\uparrow} : pr \xrightarrow{+} t}\,{>} \qquad \dfrac{\dfrac{\text{chased}^{\uparrow} : e \xrightarrow{+} pr}{\text{chased}^{\uparrow} : np^{+} \xrightarrow{+} pr}\,\mathrm{K} \qquad \dfrac{\text{most}^{\uparrow} : pr \xrightarrow{\cdot} np^{+} \qquad \text{cats}^{=} : pr}{\text{most cats}^{\uparrow} : np^{+}}\,{>}}{\text{chased most cats}^{\uparrow} : pr}\,{>}}{\text{most dogs chased most cats}^{\uparrow} : t}\,{>}$$

Example

$$\dfrac{\text{Fido}^{\uparrow} : e \qquad \dfrac{\text{didn't}^{\uparrow} : pr \xrightarrow{-} pr \qquad \dfrac{\dfrac{\text{chase}^{\downarrow} : e \xrightarrow{+} pr}{\text{chase}^{\downarrow} : np^{+} \xrightarrow{+} pr}\,\mathrm{K} \qquad \dfrac{\text{every}^{\downarrow} : pr \xrightarrow{-} np^{+} \qquad \text{cat}^{\uparrow} : pr}{\text{every cat}^{\downarrow} : np^{+}}\,{>}}{\text{chase every cat}^{\downarrow} : pr}\,{>}}{\text{didn't chase every cat}^{\uparrow} : pr}\,{>}}{\text{Fido didn't chase every cat}^{\uparrow} : t}\,{<}$$

Example

$$
\dfrac{
\mathsf{F}^{\uparrow} : e
\qquad
\dfrac{
\dfrac{\mathsf{ch}^{\uparrow} : e \xrightarrow{+} pr}{\mathsf{ch}^{\uparrow} : np^{+} \xrightarrow{+} pr}\,\mathrm{K}
\qquad
\dfrac{
\mathsf{some\ cat}^{\uparrow} : np^{+}
\qquad
\dfrac{\mathsf{and} : np^{+} \xrightarrow{+} (np^{+} \xrightarrow{+} np^{+}) \qquad \mathsf{some\ arm}^{\uparrow} : np^{+}}{\mathsf{and\ some\ arm} : np \xrightarrow{+} np}\,>
}{\mathsf{some\ cat\ and\ no\ armadillo}^{\uparrow} : np}\,<
}{\mathsf{chased\ some\ cat\ and\ some\ armadillo}^{\uparrow} : pr}\,>
}{\mathsf{Fido\ chased\ some\ cat\ and\ some\ armadillo}^{\uparrow} : t}\,<
$$

some cat and no armadillo

Example

$$
\dfrac{\mathsf{F}^{\uparrow} : e \qquad \dfrac{\dfrac{\dfrac{\mathsf{ch}^{=} : e \xrightarrow{+} pr}{\mathsf{ch}^{=} : np \xrightarrow{+} pr}\,{\scriptstyle \mathrm{K}}}{\mathsf{ch}^{\uparrow} : np \xrightarrow{+} pr}\,{\scriptstyle \mathrm{W}} \qquad \dfrac{\dfrac{\mathsf{some\ cat}^{\uparrow} : np^{+}}{\mathsf{some\ cat}^{\uparrow} : np}\,{\scriptstyle \mathrm{M}} \qquad \dfrac{\mathsf{and} : np \xrightarrow{+} (np \xrightarrow{+} np) \qquad \dfrac{\mathsf{no\ arm}^{\downarrow} : np^{-}}{\mathsf{no\ arm}^{\downarrow} : np}\,{\scriptstyle \mathrm{M}}}{\mathsf{and\ no\ arm} : np \xrightarrow{+} np}\,{\scriptstyle >}}{\mathsf{some\ cat\ and\ no\ armadillo}^{\uparrow} : np}\,{\scriptstyle <}}{\mathsf{chased\ some\ cat\ and\ no\ armadillo}^{\uparrow} : pr}\,{\scriptstyle >}}{\mathsf{Fido\ chased\ some\ cat\ and\ no\ armadillo}^{\uparrow} : t}\,{\scriptstyle <}
$$

We use (M) twice
in order conjoin some cat and no armadillo.

In effect, we are weakening the type of some cat from np^{+} to np,
and also weakening the type of no cat from np^{-} to np

We do this in order to conjoin them,
since the type of and is $x \xrightarrow{+} (x \xrightarrow{+} x)$.

Note also that we used (W) in order to change the polarity of chased from $=$ to $\uparrow$.

fewer than three cats swam

Example

$$\dfrac{\dfrac{\dfrac{\text{fewer than}^{\uparrow} : \mathit{num} \xrightarrow{+} (\mathit{pr} \xrightarrow{-} \mathit{np}^{-}) \quad \text{three}^{\uparrow} : \mathit{num}}{\text{fewer than three}^{\uparrow} : \mathit{pr} \xrightarrow{-} \mathit{np}^{-}}> \quad \text{cats}^{\downarrow} : \mathit{pr}}{\text{fewer than three cats}^{\uparrow} : \mathit{np}^{-}}> \quad \text{swam}^{\downarrow} : \mathit{pr}}{\text{fewer than three cats swam}^{\uparrow} : t}>$$

We are going to move from sets to preorders in semantics

In order to talk about inference by replacement,
we need sensible notions of "bigger" and "smaller".

It's clear enough how to do this for "simple" semantic objects like intransitive verbs, but others (complex NPs), it's not that easy.

It's better to change the way we do formal semantics by moving from sets to preorders as much as possible.

Definition

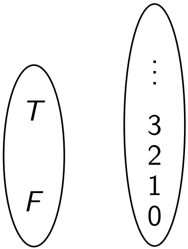
A preorder is a pair $\mathbb{P} = (P, \leq)$ consisting of a set P together with a relation $\leq$ which is reflexive and transitive.

This means that the following hold:

- $p \leq p$ for all $p \in P$.
- If $p \leq q$ and $q \leq r$, then $p \leq r$.

The preorder $\mathbb{2}$ of truth values: $\{\mathsf{T},\mathsf{F}\}$, with $\mathsf{F} \le \mathsf{T}$.

The preorder $\mathbb{N}$ of natural numbers: $0 \le 1 \le 2 \le \cdots$



$\mathbb{2}$ $\qquad$ $\mathbb{N}$

Going up in these pictures means getting “bigger” in the preorder.

Definition

For any preorder $\mathbb{P}$, there is an opposite preorder $\mathbb{P}^{op}$.
Its domain is the same domain as that of $\mathbb{P}$.

$$p \leq q \text{ in } \mathbb{P}^{op} \quad \text{iff} \quad q \leq p \text{ in } \mathbb{P}$$

Definition

For any preorder $\mathbb{P}$, there is an flattened version $\mathbb{P}^\flat$.
Its domain is the same domain as that of $\mathbb{P}$.

$$p \leq q \text{ in } \mathbb{P}^\flat \quad \text{iff} \quad p = q$$

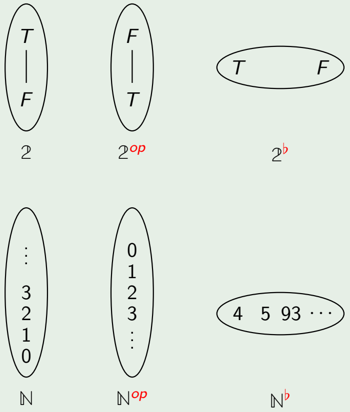
Definition

For any preorders $\mathbb{P}$ and $\mathbb{Q}$, there is a product preorder $\mathbb{P} \times \mathbb{Q}$.
Its domain is the cartesian product the given domains.

$$(p, q) \leq (p', q') \text{ in } \mathbb{P} \times \mathbb{Q} \quad \text{iff} \quad p \leq p' \text{ in } \mathbb{P}, \text{ and } q \leq q' \text{ in } \mathbb{Q}$$

Going up in the pictures means $\geq$

Example



Example

$$\begin{tikzcd}[column sep=small] & (T,T) \arrow[dl, dash] \arrow[dr, dash] & \\ (T,F) \arrow[dr, dash] & & (F,T) \arrow[dl, dash] \\ & (F,F) & \end{tikzcd}$$

$\mathbb{2} \times \mathbb{2}$

$$\begin{tikzcd}[column sep=small] & (F,F) \arrow[dl, dash] \arrow[dr, dash] & \\ (T,F) \arrow[dr, dash] & & (F,T) \arrow[dl, dash] \\ & (T,T) & \end{tikzcd}$$

$(\mathbb{2} \times \mathbb{2})^{op}$

Example

$$\begin{tikzcd}[column sep=small] & (F,T) \arrow[dl, dash] \arrow[dr, dash] & \\ (F,F) \arrow[dr, dash] & & (T,T) \arrow[dl, dash] \\ & (T,F) & \end{tikzcd}$$

$\mathbb{2}^{op} \times \mathbb{2}$

$$\begin{tikzcd}[column sep=small] & (T,F) \arrow[dl, dash] \arrow[dr, dash] & \\ (F,F) \arrow[dr, dash] & & (T,T) \arrow[dl, dash] \\ & (F,T) & \end{tikzcd}$$

$\mathbb{2} \times \mathbb{2}^{op}$

Monotone $f : \mathbb{P} \to \mathbb{Q}$

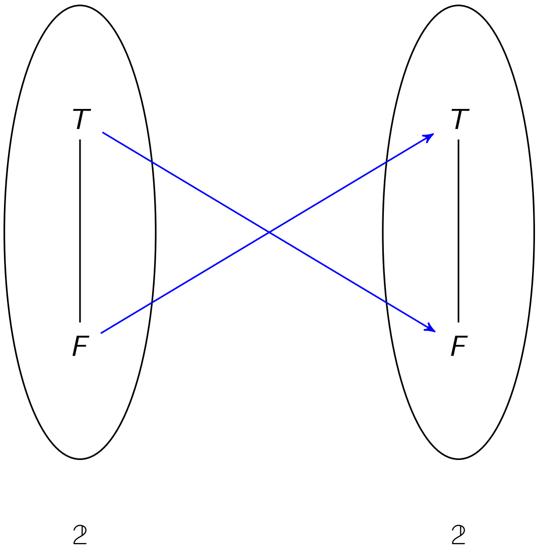
If $p \leq q$ in $\mathbb{P}$, then $f(p) \leq f(q)$ in $\mathbb{Q}$.

We write $f : \mathbb{P} \xrightarrow{+} \mathbb{Q}$.

Antitone $f : \mathbb{P} \to \mathbb{Q}$

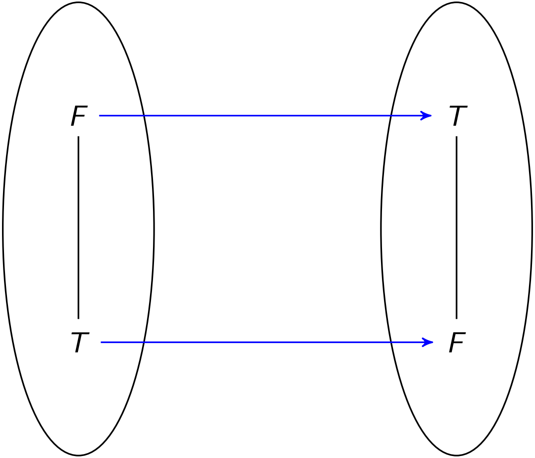
If $p \leq q$ in $\mathbb{P}$, then $f(q) \leq f(p)$ in $\mathbb{Q}$.

We write $f : \mathbb{P} \xrightarrow{-} \mathbb{Q}$.


T
F
T
F
$\mathbb{2}$
$\mathbb{2}$

Negation is antitone

This is the same as a monotone function from $\mathbb{2}^{op}$ to $\mathbb{2}$



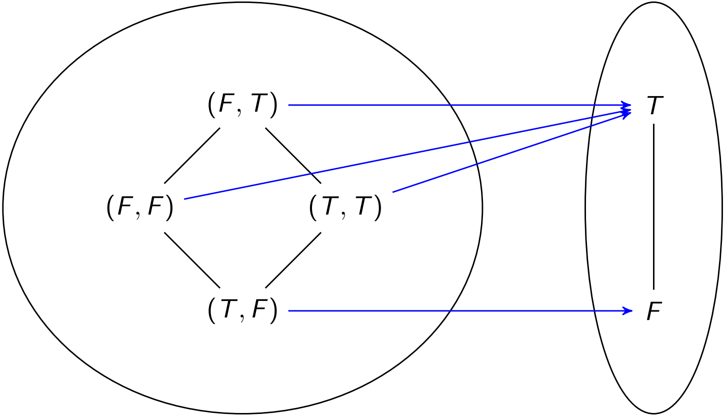
$\mathbb{2}^{op}$ $\qquad$ $\mathbb{2}$

We write these as

$$\neg\colon \mathbb{2} \xrightarrow{-} \mathbb{2}$$

$$\neg\colon \mathbb{2}^{op} \xrightarrow{+} \mathbb{2}$$

Standard implication $\rightarrow$ is a monotone function, but how?



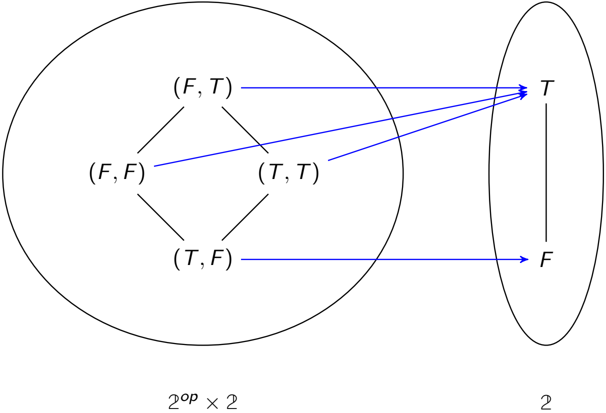
$\mathbb{2}^{op} \times \mathbb{2}$

$\mathbb{2}$

Standard implication $\rightarrow$ is a monotone function, but how?

We write this as

$$\rightarrow : (2^{op} \times 2) \xrightarrow{+} 2$$



The monotonicity summary of logical implication $\rightarrow$

$\rightarrow$ is antitone in its first argument,
and monotone in its second argument.

This should be familiar.
In a mathematical result like

Theorem

$(\textit{If}\ \underline{\quad\downarrow\quad},\ \textit{then}\ \underline{\quad\uparrow\quad})^{\uparrow}.$

We get a better result by weakening the hypothesis,
or by strengthening the conclusion.

The monotonicity summary of logical implication $\rightarrow$

$\rightarrow$ is antitone in its first argument,
and monotone in its second argument.

$$\rightarrow : (2^{op} \times 2) \xrightarrow{+} 2$$

curried forms

$$\rightarrow : 2^{op} \xrightarrow{+} (2 \xrightarrow{+} 2)$$
$$\rightarrow : 2 \xrightarrow{-} (2 \xrightarrow{+} 2)$$

Definition

For any preorder $\mathbb{P}$ and any set X,
we have a new preorder called $X \to \mathbb{P}$.

The domain of this preorder is the function set

$$X \to P$$

The order on $X \to \mathbb{P}$ is the pointwise order:

$$f \leq g \text{ iff for all } x \in X,\ fx \leq_{\mathbb{P}} gx.$$

Monotone function f

If $p \leq q$ in $\mathbb{P}$, then $f(p) \leq f(q)$ in $\mathbb{Q}$.
We write $f : \mathbb{P} \xrightarrow{+} \mathbb{Q}$.

Antitone function f

If $p \leq q$ in $\mathbb{P}$, then $f(q) \leq f(p)$ in $\mathbb{Q}$.
We write $f : \mathbb{P} \xrightarrow{-} \mathbb{Q}$, or equivalently $f : \mathbb{P}^{op} \xrightarrow{+} \mathbb{Q}$.

Arbitrary function f with no special properties

We write $f : \mathbb{P} \xrightarrow{\cdot} \mathbb{Q}$.
So this means "in general, neither monotone nor antitone."

It is equivalent to $f : \mathbb{P}^{\flat} \xrightarrow{+} \mathbb{Q}$.

Every function from a flat preorder is automatically monotone.

For sets X, Y, there is a set $X \to Y$ of all functions from X to Y.

Moving to preorders, this idea becomes more involved.

Given preorders $\mathbb{P}$ and $\mathbb{Q}$, we have

- A preorder $\mathbb{P} \xrightarrow{+} \mathbb{Q}$ of all monotone functions
- A preorder $\mathbb{P} \xrightarrow{-} \mathbb{Q}$ of all antitone functions
- A preorder $\mathbb{P} \xrightarrow{\cdot} \mathbb{Q}$ of all functions

The order on each of these is the pointwise order.

We need to change the entire architecture of CG.

Standard CG

Define

$$k : \text{Syntactic categories} \to \text{Semantic types}$$

by

$$\begin{array}{lcl} k(\mathrm{S}) & = & t \\ k(\mathrm{N}) & = & e \to t \\ k(\mathrm{NP}) & = & (e \to t) \to t \\ k(X/Y) & = & k(Y) \to k(X) \\ k(X\backslash Y) & = & k(Y) \to k(X) \end{array}$$

Then in a model we choose a set S_e for e, we always take $S_t = \{T, F\}$ and $S_{X/Y}$ is

$$(S_Y \to S_X) = \text{the set of all functions from } S_{k(X)} \text{ to } S_{k(X)}$$

We'll see soon how this works in the preorder enrichment.

It used to be that the "e,t" type for NP was $(e \to t) \to t$.

Now we have 9 semantic types:

$$\begin{array}{lll}
(e \xrightarrow{+} t) \xrightarrow{+} t & (e \xrightarrow{-} t) \xrightarrow{-} t & (e \xrightarrow{\cdot} t) \xrightarrow{+} t \\
(e \xrightarrow{+} t) \xrightarrow{-} t & (e \xrightarrow{+} t) \xrightarrow{\cdot} t & (e \xrightarrow{\cdot} t) \xrightarrow{-} t \\
(e \xrightarrow{-} t) \xrightarrow{+} t & (e \xrightarrow{-} t) \xrightarrow{\cdot} t & (e \xrightarrow{\cdot} t) \xrightarrow{\cdot} t
\end{array}$$

Which do we actually want for NP now??

More semantic types brings more power and more responsibility

We want to interpret semantic types by preorders, not sets.

For t, we want $\mathbb{2}$.

For e, we want an arbitrary flat preorder $\mathbb{P} = \mathbb{P}^\flat$.

By flatness $e \xrightarrow{+} t$, $e \xrightarrow{-} t$ and $e \xrightarrow{\cdot} t$ give the same preorder.

We call this pr, for property.

We want to interpret semantic types by preorders, not sets.

For t, we want $\mathbb{2}$.

For e, we want an arbitrary flat preorder $\mathbb{P} = \mathbb{P}^\flat$.

By flatness $e \xrightarrow{+} t$, $e \xrightarrow{-} t$ and $e \xrightarrow{\cdot} t$ give the same preorder.

We call this *pr*, for property.

Notation

$$\begin{aligned} np^+ &= pr \xrightarrow{+} t \\ np^- &= pr \xrightarrow{-} t \\ np &= pr \xrightarrow{\cdot} t \end{aligned}$$

Preordered CG

Define

$$k : \text{Syntactic categories} \to \text{Semantic types}$$

by

$$\begin{array}{lcl} k(\mathrm{S}) & = & t \\ k(\mathrm{N}) & = & e \to t \\ k(\mathrm{NP}) & = & (e \to t) \xrightarrow{+} t \end{array}$$

For the types σ corresponding to categories
like transitive or intransitive verbs, we have choices for $\mathbb{P}_\sigma$.

Anyways, when we form function types, we always have three choices:

$$\mathbb{P}_\sigma \xrightarrow{+} \mathbb{P}_\sigma$$
$$\mathbb{P}_\sigma \xrightarrow{-} \mathbb{P}_\sigma$$
$$\mathbb{P}_\sigma \xrightarrow{\cdot} \mathbb{P}_\sigma$$

If $f : \mathbb{P} \xrightarrow{+} \mathbb{Q}$ and $g : \mathbb{Q} \xrightarrow{+} \mathbb{R}$, then $g \circ f : \mathbb{P} \xrightarrow{+} \mathbb{R}$.

If $f : \mathbb{P} \xrightarrow{+} \mathbb{Q}$ and $g : \mathbb{Q} \xrightarrow{-} \mathbb{R}$, then $g \circ f : \mathbb{P} \xrightarrow{-} \mathbb{R}$.

If $f : \mathbb{P} \xrightarrow{-} \mathbb{Q}$ and $g : \mathbb{Q} \xrightarrow{+} \mathbb{R}$, then $g \circ f : \mathbb{P} \xrightarrow{-} \mathbb{R}$.

If $f : \mathbb{P} \xrightarrow{-} \mathbb{Q}$ and $g : \mathbb{Q} \xrightarrow{-} \mathbb{R}$, then $g \circ f : \mathbb{P} \xrightarrow{+} \mathbb{R}$.

If either was $\xrightarrow{\cdot}$, the composition would also be $\xrightarrow{\cdot}$.

Not to miss a generalization

$$\frac{f : \mathbb{P} \xrightarrow{m} \mathbb{Q} \quad g : \mathbb{Q} \xrightarrow{n} \mathbb{R}}{g \circ f : \mathbb{P} \xrightarrow{mn} \mathbb{R}}$$

We use a "multiplication" operation on the markings

$$m, n \mapsto mn$$

given in the chart:

mn	$+$	$-$	$\cdot$
$+$	$+$	$-$	$\cdot$
$-$	$-$	$+$	$\cdot$
$\cdot$	$\cdot$	$\cdot$	$\cdot$

For all preorders $\mathbb{P}$ and $\mathbb{Q}$, and $f_1, f_2 : P \to Q$, and all $p_1, p_2 \in P$:

1. If $f_1, f_2 : \mathbb{P} \xrightarrow{+} \mathbb{Q}$, and $f_1 \leq f_2$, and $p_1 \leq p_2$, then $f_1(p_1) \leq f_2(p_2)$.
2. If $f_1, f_2 : \mathbb{P} \xrightarrow{-} \mathbb{Q}$, and $f_1 \leq f_2$, and $p_2 \leq p_1$, then $f_1(p_1) \leq f_2(p_2)$.
3. If $f_1, f_2 : \mathbb{P} \xrightarrow{\cdot} \mathbb{Q}$, and $f_1 \leq f_2$, and $p_2 = p_1$, then $f_1(p_1) \leq f_2(p_2)$.
4. If $f_1, f_2 : \mathbb{P} \xrightarrow{+} \mathbb{Q}$, and $f_2 \leq f_1$, and $p_2 \leq p_1$, then $f_2(p_2) \leq f_1(p_1)$.
5. If $f_1, f_2 : \mathbb{P} \xrightarrow{-} \mathbb{Q}$, and $f_2 \leq f_1$, and $p_1 \leq p_2$, then $f_2(p_2) \leq f_1(p_1)$.
6. If $f_1, f_2 : \mathbb{P} \xrightarrow{\cdot} \mathbb{Q}$, and $f_2 \leq f_1$, and $p_2 = p_1$, then $f_2(p_2) \leq f_1(p_1)$.

7-9. If $f_1, f_2 : \mathbb{P} \xrightarrow{\cdot} \mathbb{Q}$, and $f_1 = f_2$, and $p_1 = p_2$, then $f_1(p_1) = f_2(p_2)$.

Pattern

$$\frac{f^{d} : \mathbb{P} \xrightarrow{m} \mathbb{Q} \quad p : \mathbb{P}^{md}}{(f(p))^{d} : \mathbb{Q}} >$$

We combine markings and polarities as in the table below:

md	$+$	$-$	$\cdot$
$\uparrow$	$\uparrow$	$\downarrow$	$=$
$\downarrow$	$\downarrow$	$\uparrow$	$=$
$=$	$=$	$=$	$=$

Rules to build trees (explanation below)

Rules: x, y are semantic types, b is a boolean type

$$\frac{u^{d} : x \xrightarrow{m} y \quad v^{md} : x}{(uv)^{d} : y} >$$

$$\frac{u^{d} : x \xrightarrow{m} y \quad v^{md} : y \xrightarrow{n} z}{(uv)^{d} : (x \xrightarrow{mn} z)} \textsc{B}$$

$$\frac{u^{md} : x}{u^{d} : (x \xrightarrow{m} y) \xrightarrow{+} y} \textsc{T}$$

$$\frac{u^{md} : e \to b}{u^{d} : np^{m} \xrightarrow{+} b} \textsc{K}$$

$$\frac{u^{d} : x \xrightarrow{m} y}{u^{d} : x \xrightarrow{\cdot} y} \textsc{M}$$

$$\frac{u^{=} : x \xrightarrow{m} y}{u^{d} : x \xrightarrow{m} y} \textsc{W}$$

The $>$ rule is function application.

The B-rule is function composition, backwards.

The T-rule is polarized form of type raising a la Montague.

In the K-rule, b has to be a boolean type.
The soundness of the rule is connected to Keenan and Faltz' Justification Theorem.

The M- and W-rules are forms of weakening.

Definition

Caba is the category of complete atomic boolean algebras, with morphisms the maps $f : A \to B$ such that

$$f(\bigwedge S) = \bigwedge_{a \in S} f(a)$$

for all $S \subseteq A$, and also such that $f(-a) = -f(a)$ for all $a \in A$.

We have an adjunction

$$\begin{tikzcd} \mathrm{Set} \arrow[r, bend left, "2^{2^X}"] & \mathrm{Caba} \arrow[l, bend left, "U"] \end{tikzcd}$$

Keenan and Faltz' "Justification Theorem"

Equivalent to the characterization of free complete atomic boolean algebras

Definition

Caba is the category of complete atomic boolean algebras, with morphisms the maps $f : A \to B$ such that

$$f(\bigwedge S) = \bigwedge_{a \in S} f(a)$$

for all $S \subseteq A$, and also such that $f(-a) = -f(a)$ for all $a \in A$.

Theorem (Justification Theorem)

Let $\mathbb{B}$ be any Caba, and let

$$f : E \to U\mathbb{B}$$

be any Set function.

Then there is a unique Caba-morphism

$$f_\star : 2^{2^E} \to \mathbb{B}$$

such that

$$Uf_\star \circ \eta_E = f$$

where $\eta_E : E \to 2^{2^E}$ is the Montague lift.

Notation

$\mathbb{2}^{\mathbb{2}^E}$ is $E \to \mathbb{2} \to \mathbb{2}$.
$(\mathbb{2}^{\mathbb{2}^E})^{+}$ is $(E \to \mathbb{2}) \xrightarrow{+} \mathbb{2}$.
$(\mathbb{2}^{\mathbb{2}^E})^{-}$ is $(E \to \mathbb{2}) \xrightarrow{-} \mathbb{2}$.

For $m = +$ and $m = -$, we have a restriction map

$$r_m : (\mathbb{2}^{\mathbb{2}^E} \xrightarrow{+} \mathbb{B}) \xrightarrow{+} ((\mathbb{2}^{\mathbb{2}^E})^{m} \xrightarrow{+} \mathbb{B}). \tag{1}$$

Let

$$s_m : (E \xrightarrow{\cdot} \mathbb{B}) \to ((\mathbb{2}^{\mathbb{2}^E})^{m} \xrightarrow{+} \mathbb{B}) \tag{2}$$

be given by

$$s_m f = f_\star r_m$$

That is, $s_m f Q = f_\star Q$, for all $f : E \xrightarrow{\cdot} \mathbb{B}$ and all $Q \in (\mathbb{2}^{\mathbb{2}^E})^{m}$.

$$\begin{tikzcd}
 & E \overset{\cdot}{\to} \mathbb{B} \arrow[dl, "\text{Keenan-Faltz}"'] \arrow[d, "s_+"] \arrow[dr, "s_-"] & \\
2^{2^E} \overset{+}{\to} \mathbb{B} \arrow[r, "r_+"'] & (2^{2^E})^+ \overset{+}{\to} \mathbb{B} \arrow[r, shift right, "\mathrm{DUAL}"'] & (2^{2^E})^- \overset{+}{\to} \mathbb{B} \arrow[l, shift right, "\mathrm{DUAL}^{-1}"']
\end{tikzcd}$$

Lemma

s_+ is monotone, and s_- is antitone:

$$s_+ : (E \overset{\cdot}{\to} \mathbb{B}) \overset{+}{\to} ((2^{2^E})^+ \overset{+}{\to} \mathbb{B})$$
$$s_- : (E \overset{\cdot}{\to} \mathbb{B}) \overset{-}{\to} ((2^{2^E})^- \overset{+}{\to} \mathbb{B})$$

Moreover, $s_- = \mathrm{DUAL} \circ s_+$, and $s_+ = \mathrm{DUAL}^{-1} \circ s_-$.

$$
\begin{tikzcd}
 & E \xrightarrow{\cdot} \mathbb{B} \arrow[dl, "\text{Keenan-Faltz}"'] \arrow[d, "s_+"] \arrow[dr, "s_-"] & \\
2^{2^E} \xrightarrow{+} \mathbb{B} \arrow[r, "r_+"'] & (2^{2^E})^+ \xrightarrow{+} \mathbb{B} \arrow[r, bend right, "\mathrm{DUAL}"'] & (2^{2^E})^- \xrightarrow{+} \mathbb{B} \arrow[l, bend right, "\mathrm{DUAL}^{-1}"']
\end{tikzcd}
$$

This justifies the rule

$$
\frac{f^{md} : e \to b}{(s_m f)^d : \mathrm{np}^m \xrightarrow{+} b}\ \mathrm{K}
$$

Fact that also comes up

$$(\mathbb{P}^{op} \xrightarrow{+} \mathbb{Q}^{op}) = (\mathbb{P} \xrightarrow{+} \mathbb{Q})^{op}$$

Example (a polarized syntax tree)

$$\dfrac{\dfrac{\text{some}^{\uparrow} : pr \xrightarrow{+} np^{+} \quad \text{dog}^{\uparrow} : pr}{\text{some dog}^{\uparrow} : pr \xrightarrow{+} t}> \qquad \dfrac{\dfrac{\text{chased}^{\downarrow} : e \xrightarrow{+} pr}{\text{chased}^{\uparrow} : np^{-} \xrightarrow{+} pr}\text{K} \quad \dfrac{\text{no}^{\uparrow} : pr \xrightarrow{-} np^{-} \quad \text{cat}^{\downarrow} : pr}{\text{no cat}^{\uparrow} : np^{-}}>}{\text{chased no cat}^{\uparrow} : pr}>}{\text{some dog chased no cat}^{\uparrow} : t}>$$

Polarized parse trees, and what they mean

Example (a polarized syntax tree)

$$
\dfrac{
\dfrac{\text{some}^{\uparrow} : pr \xrightarrow{+} np^{+} \quad \text{dog}^{\uparrow} : pr}{\text{some dog}^{\uparrow} : pr \xrightarrow{+} t}\, >
\qquad
\dfrac{
\dfrac{\text{chased}^{\downarrow} : e \xrightarrow{+} pr}{\text{chased}^{\uparrow} : np^{-} \xrightarrow{+} pr}\, \mathrm{K}
\quad
\dfrac{\text{no}^{\uparrow} : pr \xrightarrow{-} np^{-} \quad \text{cat}^{\downarrow} : pr}{\text{no cat}^{\uparrow} : np^{-}}\, >
}{\text{chased no cat}^{\uparrow} : pr}\, >
}{\text{some dog chased no cat}^{\uparrow} : t}\, >
$$

The rules actually have combinator constants which I elided before

Rules

$$
\dfrac{u^{d} : x \xrightarrow{m} y \quad v^{md} : x}{(uv)^{d} : y}\, >
\qquad
\dfrac{u^{md} : x}{(\mathrm{T}u)^{d} : (x \xrightarrow{m} y) \xrightarrow{+} y}\, \mathrm{T}
\qquad
\dfrac{u^{md} : u \xrightarrow{n} y \quad v^{md} : y \xrightarrow{n} z}{(\mathrm{B}uv)^{d} : (x \xrightarrow{mn} z)}\, \mathrm{B}
$$

$$
\dfrac{u^{md} : e \to b}{(r_{m}u)^{d} : np^{m} \xrightarrow{+} b}\, \mathrm{K}
\qquad
\dfrac{u^{d} : x \xrightarrow{m} y}{u^{d} : x \xrightarrow{\cdot} y}\, \mathrm{M}
\qquad
\dfrac{u^{=} : x \xrightarrow{m} y}{u^{d} : x \xrightarrow{m} y}\, \mathrm{W}
$$

Incidentally, I also polarized a few other combinators in addition to the CCG ones.
There almost certainly is a general algorithm to be had.

Example (What it means)

$$\dfrac{\dfrac{v^{\uparrow} : pr \xrightarrow{+} np^{+} \qquad w^{\uparrow} : pr}{vw^{\uparrow} : pr \xrightarrow{+} t}\,{\scriptstyle >} \qquad \dfrac{\dfrac{x^{\downarrow} : e \xrightarrow{+} pr}{r_{-}x^{\uparrow} : np^{-} \xrightarrow{+} pr}\,{\scriptstyle \mathrm{K}} \qquad \dfrac{y^{\uparrow} : pr \xrightarrow{-} np^{-} \qquad z^{\downarrow} : pr}{yz^{\uparrow} : np^{-}}\,{\scriptstyle >}}{(r_{-}x)(yz)^{\uparrow} : pr}\,{\scriptstyle >}}{\tau = (vw)((r_{-}x)(yz))^{\uparrow} : t}\,{\scriptstyle >}$$

The semantic term τ on the bottom is a combinator.

Soundness

The polarity arrows on the leaves mean that in every model,

$$[\![\tau]\!](v, w, x, y, z) : \mathbb{P}_{pr \xrightarrow{+} np^{+}} \times \mathbb{P}_{pr} \times (\mathbb{P}_{e \xrightarrow{+} pr})^{op} \times \mathbb{P}_{pr \xrightarrow{-} np^{-}} \times (\mathbb{P}_{pr})^{op} \quad \xrightarrow{+} \quad \mathbb{P}_{t}$$

At this point, my talk shifts from theory to practice.

My former PhD student Hai Hu designed and implemented
an arrow-tagging tool called ccg2Mono
based on the rules which we have seen,
and then he built an inference engine for NLI based on
monotonicity and natural logic called MonaLog.

`https://huhailinguist.github.io/`

What our program ccg2Mono does

I am cheating a little, so you can ask me on this

Input: a syntax tree a given to us by a CCG parser

$$\cfrac{\textsf{every} : \mathrm{NP/N} \qquad \cfrac{\textsf{cat} : \mathrm{N} \qquad \cfrac{\textsf{that} : (\mathrm{N\backslash N})/(\mathrm{S/NP}) \qquad \cfrac{\cfrac{\textsf{F} : \mathrm{NP}}{\textsf{F} : \mathrm{S/(S\backslash NP)}}\,\mathrm{T} \qquad \textsf{ch} : (\mathrm{S\backslash NP})/\mathrm{NP}}{\textsf{Fido chased} : \mathrm{S/NP}}\,\mathrm{B}}{\textsf{that Fido chased} : \mathrm{N\backslash N}}\,>}{\textsf{cat that Fido chased} : \mathrm{N}}\,<}{\cfrac{\textsf{every cat that Fido chased} : \mathrm{NP} \qquad \textsf{ran} : \mathrm{S\backslash NP}}{\textsf{every cat that Fido chased ran} : \mathrm{S}}\,<}\,>$$

Output: a polarized version, but note the new nodes and some changes

$$\cfrac{\textsf{F}^{\downarrow} : e \qquad \textsf{ch}^{\downarrow} : e \to (e \to t)}{\textsf{Fido chased}^{\downarrow} : e \to t}\,<$$

$$\cfrac{\textsf{that}^{\downarrow} : (e \to t) \xrightarrow{+} (pr \xrightarrow{+} pr) \qquad \textsf{Fido chased}^{\downarrow} : e \to t}{\textsf{that Fido chased}^{\downarrow} : pr \xrightarrow{+} pr}\,>$$

$$\cfrac{\textsf{cat}^{\downarrow} : pr \qquad \textsf{that Fido chased}^{\downarrow} : pr \xrightarrow{+} pr}{\textsf{cat that Fido chased}^{\downarrow} : pr}\,<$$

$$\cfrac{\textsf{every}^{\uparrow} : pr \xrightarrow{-} np^{+} \qquad \textsf{cat that Fido chased}^{\downarrow} : pr}{\textsf{every cat that Fido chased}^{\uparrow} : np^{+}}\,>$$

$$\cfrac{\textsf{ran}^{\uparrow} : e \xrightarrow{+} t}{\textsf{ran}^{\uparrow} : \mathrm{NP}^{+} \xrightarrow{+} t}\,\mathrm{K}$$

$$\cfrac{\textsf{every cat that Fido chased}^{\uparrow} : np^{+} \qquad \textsf{ran}^{\uparrow} : \mathrm{NP}^{+} \xrightarrow{+} t}{\textsf{every cat that Fido chased ran}^{\uparrow} : t}\,<$$

Occasionally the parser's input had to be “corrected”.

Output: a polarized version, but note the new nodes and some changes

$$
\dfrac{
  \dfrac{
    \text{every}^{\uparrow} : pr \xrightarrow{-} np^{+}
    \qquad
    \dfrac{
      \text{cat}^{\downarrow} : pr
      \qquad
      \dfrac{
        \text{that}^{\downarrow} : (e \to t) \xrightarrow{+} (pr \xrightarrow{+} pr)
        \qquad
        \dfrac{\text{F}^{\downarrow} : e \quad \text{ch}^{\downarrow} : e \to (e \to t)}{\text{Fido chased}^{\downarrow} : e \to t}<
      }{\text{that Fido chased}^{\downarrow} : pr \xrightarrow{+} pr}>
    }{\text{cat that Fido chased}^{\downarrow} : pr}<
  }{\text{every cat that Fido chased}^{\uparrow} : np^{+}}>
  \qquad
  \dfrac{\text{ran}^{\uparrow} : e \xrightarrow{+} t}{\text{ran}^{\uparrow} : \mathrm{NP}^{+} \xrightarrow{+} t}\,\textsc{k}
}{\text{every cat that Fido chased ran}^{\uparrow} : t}<
$$

By the way, if you don't like the (K) rule,
or the way the example lexicon treated verbs,
there are other analyses.

To me, we only have clear order intuitions about "low" types,
and so the type-raising treatment is appealing.

Examples of polarized sentences from MonaLog

system available at `https://github.com/huhailinguist/monalog`

Every$^{\uparrow}$ man$^{\downarrow}$ and$^{\uparrow}$ some$^{\uparrow}$ woman$^{\uparrow}$ sleeps$^{\uparrow}$

Every$^{\uparrow}$ man$^{\downarrow}$ and$^{\uparrow}$ no$^{\uparrow}$ woman$^{\downarrow}$ sleeps$^{=}$

If$^{\uparrow}$ some$^{\downarrow}$ man$^{\downarrow}$ walks$^{\downarrow}$, then$^{\uparrow}$ no$^{\uparrow}$ woman$^{\downarrow}$ runs$^{\downarrow}$

Every$^{\uparrow}$ man$^{\downarrow}$ does$^{\downarrow}$ n't$^{\uparrow}$ hit$^{\downarrow}$ every$^{\downarrow}$ dog$^{\uparrow}$

No$^{\uparrow}$ man$^{\downarrow}$ who$^{\downarrow}$ likes$^{\downarrow}$ every$^{\downarrow}$ dog$^{\uparrow}$ sleeps$^{\downarrow}$

Most$^{\uparrow}$ men$^{=}$ that$^{=}$ every$^{=}$ woman$^{=}$ hits$^{=}$ cried$^{\uparrow}$

Every$^{\uparrow}$ young$^{\downarrow}$ man$^{\downarrow}$ that$^{\uparrow}$ no$^{\uparrow}$ young$^{\downarrow}$ woman$^{\downarrow}$ hits$^{\uparrow}$ cried$^{\uparrow}$

A$^{\uparrow}$ special$^{\uparrow}$ report$^{\uparrow}$ found$^{\downarrow}$ no$^{\uparrow}$ incriminating$^{\downarrow}$ evidence$^{\downarrow}$

Toy example, but too hard for most people

Challenge

Assume: all skunks are mammals

Which of the following follows?

- All who love all who hate all skunks
  love all who hate all mammals.
- All who love all who hate some skunks
  love all who hate some skunks.

Solution

Clearly we have

- All who love all who hate all skunks
  love all who hate all skunks.

This is just All who are X are X

Toy example, but too hard for most people

Solution

ccg2Mono tells us

- All who love all who hate all skunks$^{\downarrow}$.
  love all who hate all skunks$^{\uparrow}$.

We assumed skunks $\leq$ mammals.
So we can move the second skunks up to mammals.

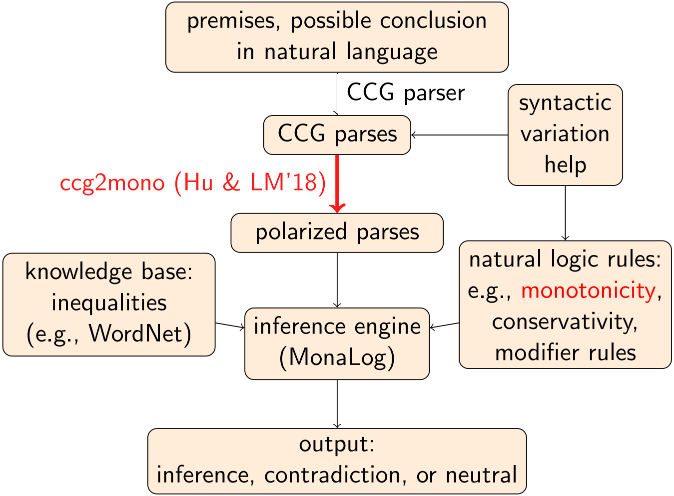
We conclude

- All who love all who hate all skunks$^{\downarrow}$.
  love all who hate all mammals$^{\uparrow}$.

- All who love all who hate some skunks$^{\uparrow}$.
  love all who hate some skunks$^{\downarrow}$.

So here we would make a different inference.

How arrow tagging fits in to our natural logic inference (NLI) system MonaLog



Results on the SICK dataset

system	P	R	acc.
majority baseline	–	–	56.36
Natural-logic-based: MonaLog‡			
MonaLog + pass2act	89.42	72.18	80.25†
MonaLog + $\exists$ transformation	89.43	71.53	79.11†
MonaLog + all rewrite help	83.75	70.66	77.19
MonaLog + all rewrite help	89.91	74.23	**81.66**†
Hybrid: MonaLog + BERT	83.09	85.46	85.38
Hybrid: MonaLog + BERT	85.65	87.33	**85.95**†
ML/DL-based systems			
BERT (base, uncased)	86.81	85.37	**86.74**
BERT (base, uncased)	84.62	84.27	85.00†
Yin & Schütze'17	–	–	**87.1**
Beltagy et al '16	–	–	85.1
Other logic-based systems			
Bjerva et al '14	93.6	60.6	81.6
Abzianidze'17	97.95	58.11	81.35
Martínez-Gomez et al '17	97.04	63.64	83.13
Yanaka et al '18	84.2	77.3	84.3

† = running on a corrected version of the SICK dataset.

‡ = P / R for MonaLog averaged across three labels.

Two talented undergraduate students

Zeming Chen and Qiyue Gao

built a polarity-tagging system using universal dependency parses rather than CCG parses.

+ More people can use it.

+ It performs better than ccg2Mono,
  partly because the parses are better,
  and partly because ccg2Mono misses some arrows,
  such as on attitude verbs: John refused to dance$^{\downarrow}$.

− This form of grammar doesn't have
  a connection to formal semantics,
  so one can't really prove soundness results
  the way we can with ccg2Mono.

Two talented undergraduate students

Zeming Chen and Qiyue Gao

built a polarity-tagging system using universal dependency parses rather than CCG parses.

It's called Udep2mono and you can get it at `https://github.com/eric11eca/Udep2Mono`.

Udep2Mono won a best paper award at IWCS 2021.

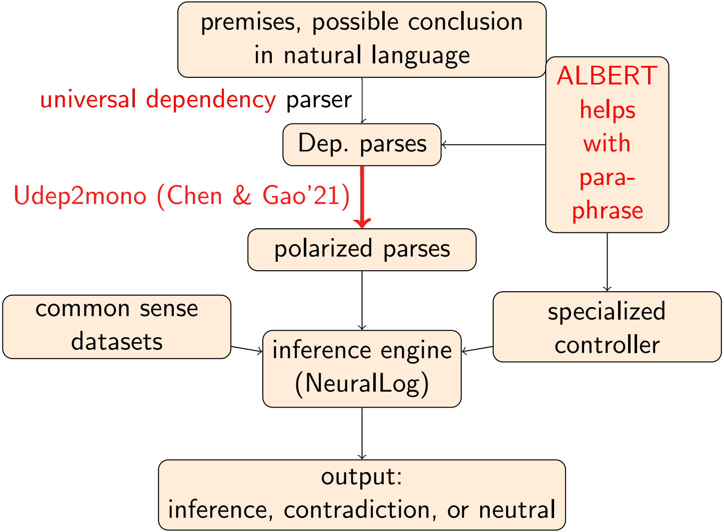
Some hard sentences

More dogs than cats sit.
Less than 5 people ran.
A dog who ate two rotten biscuits was sick for three days.
Every dog who likes most cats was chased by at least two of them.
Even if you are addicted to cigarettes you can smoke two a day.

Performance, roughly

NatLog (=MacCartney & Manning 2009) 70%
ccg2Mono 80%
Udep2Mono 97%

A hybrid system: NeuralLog



Model	**P**	**R**	**acc.**
ML/DL-based systems			
BERT (base, uncased)	86.8	85.4	86.7
?	–	–	87.1
?	–	–	85.1
Logic-based systems			
?	98.0	58.1	81.4
?	97.0	63.6	83.1
?	84.2	77.3	84.3
?	83.8	70.7	77.2
?+BERT	83.2	85.5	85.4
?	94.3	67.9	84.4
New System: NeuroLog (Chen, Gao, LM 2021)			
NeuralLog (full system)	88.0	**87.6**	**90.3**
– ALBERT-SV	68.9	79.3	71.4
– Monotonicity	74.5	75.1	74.7

What my collaborators and I have been doing

Thanks to

Hai Hu, Thomas Icard, Kyle Richardson, Zeming Chen, Qiyue Gao
Valeria de Paiva, Katerina Kalouli

- We extended monotonicity from vanilla CG to CCG.
- We have an order-enriched version of the typed lambda calculus.
- We have a running system that can polarize input sentences.
- We built a MonaLog and now NeuralLog that can solve a large NLI dataset.
- We can generate high-quality sentence pairs, helpful to a ML model.
- We have hybridized logic and machine learning and currently have the SOTA NLI system for inference on the SICK dataset.
- We have re-annotated SICK by hand and have a deep study of NLI annotation.

What is happening in this general area?

Check out next year's workshop
Natural Logic Meets Machine Learning