Monotonicity in Natural Language Inference: An Update on Theory and Practice

Larry Moss

CLASP Seminar

May 29, 2019

The goal of this project is to solve inference problems in natural language such as the following:

entail, *contradict* or *neural*?

P: A flute is being played by a girl
H: *There is no woman playing a flute*

entail, *contradict* or *neural*?

P1: Most Europeans are resident in Europe
P2: All Europeans are people
P3: All people who are resident in Europe can travel freely within Europe
H: *Most Europeans can travel freely within Europe*

Often referred to as **Natural Language Inference** (NLI) and in the recent past as **Recognizing Textual Entailment** (RTE).

Goal

The goal is to build tools that can help with several automatic inference tasks such as the FraCas textual inference problem set.

fracas-013 answer: yes

P1 Both leading tenors are excellent.
P2 Leading tenors who are excellent are indispensable.
Q Are both leading tenors indispensable?
H Both leading tenors are indispensable.

fracas-014 answer: no

P1 Neither leading tenor comes cheap.
P2 One of the leading tenors is Pavarotti.
Q Is Pavarotti a leading tenor who comes cheap?
H Pavarotti is a leading tenor who comes cheap.

The dominant approach: Machine Learning

system	P	R	acc.
majority baseline	–	–	56.36
Natural-logic-based: MonaLog[‡] (this work)			
MonaLog + pass2act	89.42	72.18	80.25[†]
MonaLog + existential trans.	89.43	71.53	79.11[†]
MonaLog + all	83.75	70.66	77.19
MonaLog + all	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 and Schütze (2017)	–	–	**87.1**
Beltagy et al. (2016)	–	–	85.1
Logic-based systems			
Bjerva et al. (2014)	93.6	60.6	81.6
Abzianidze (2015)	97.95	58.11	81.35
Martínez-Gómez et al. (2017)	97.04	63.64	83.13
Yanaka et al. (2018)	84.2	77.3	84.3

The dominant approach: Machine Learning

Logic-based approaches

- Tableau (Abzianidze, following Muskens)
- Translation to a richer logical form,
  then call a theorem prover
  (Yanaka, also Bekki, Mineshima, etc.)
- Natural Logic: monotonicity calculus + special rules
  (Hu, Icard, M, Tune)

This is an entry for a United States National Science Foundation contest on mathematics outreach for the general public.

Monotonicity: review from the video

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

$$\frac{v^{\downarrow} \quad w^{\downarrow} \quad z^{\uparrow}}{(z - (v + w))^{\uparrow}} \tag{1}$$

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

Another way to say (1):
$v, w, z \mapsto z - (v + w)$ is an increasing function

$$\mathbb{R}^{op} \times \mathbb{R}^{op} \times \mathbb{R} \to \mathbb{R}$$

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

$$\frac{v^{\downarrow} \quad w^{=} \quad z^{\uparrow}}{(z - (v + w))^{\uparrow}} \tag{2}$$

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
$\| \; \| : \mathbb{R} \to \mathbb{R}$	is neither

And we can write (2) as
$v, w, z \mapsto z - (v + w)$ is an increasing function

$$\mathbb{R}^{op} \times \mathbb{R}^{\flat} \times \mathbb{R} \to \mathbb{R}$$

(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
siamese $\leq$ cat	hit vigorously $\leq$ hit	two $\leq$ three
bird $\leq$ scooter	hit lightly $\leq$ hit	three $\leq$ four

The goals of our work

The project that Hai Hu and I are engaged in aims to understand the polarizations $\uparrow$, $\downarrow$, and $=$ both in theoretical and practical ways.

Theoretical contribution

A system that can account for the polarization of many more English sentences that previously.

A much more solid understanding of all the math.

Practical contribution

A system that accepts input from an off-the-shelf parser for Combinatory Categorial Grammar (CCG) and returns the polarization of the semantic function determined by the parse.

An algorithm for inference that uses the "arrow information".

Experience with machine learners.

I assume that you have seen categorial grammar

$$\frac{\text{Dana: NP} \qquad \dfrac{\text{praised: } (\mathrm{S}\backslash\mathrm{NP})/\mathrm{NP} \qquad \text{Kim: NP}}{\text{praised Kim: } \mathrm{S}\backslash\mathrm{NP}}}{\text{Dana praised Kim: S}}$$

From our lexicon

(Dana, NP)
(Kim, NP)
(praised, $(\mathrm{S}\backslash\mathrm{NP})/\mathrm{NP}$)

The leaves must match the categories in the lexicon,
and going down we use directed cancellation.

A key point is that CG reconstructs traditional categories
like verb phrase
as complex categories: $\mathrm{S}\backslash\mathrm{NP}$

We take a single base type, R.

Another lexicon

plus : (R/R)/R	minus : (R/R)/R
times : (R/R)/R	div2 : (R/R)/R
v : R	w : R
x : R	y : R
z : R	
1 : R	2 : R

We get terms in Polish notation

$$\dfrac{\dfrac{\mathsf{minus} : \mathsf{(R/R)/R} \quad \mathsf{z} : \mathsf{R}}{\mathsf{minus\ z} : \mathsf{R/R}} \qquad \dfrac{\dfrac{\mathsf{plus} : \mathsf{(R/R)/R} \quad \mathsf{v} : \mathsf{R}}{\mathsf{plus\ v} : \mathsf{R/R}} \qquad \mathsf{w} : \mathsf{R}}{\mathsf{plus\ v\ w} : \mathsf{R}}}{\mathsf{minus\ z\ plus\ v\ w} : \mathsf{R}}$$

We think of the tree as justifying the fact that

$$z - (v + w)$$

is a term of syntactic category R,
based on the assumptions at the leaves.

The semantics will use
higher-order (one-place) functions on the real numbers.

We take sets for our semantic domains,
using function sets for the two slashes:

$$\begin{array}{lcl} D_{\mathsf{R}} & = & \mathbb{R} \\ D_{X\backslash Y} & = & D_X \to D_Y \\ D_{Y/X} & = & D_X \to D_Y \end{array}$$

Then automatically,

$$D_{\mathsf{R}/\mathsf{R}} = \mathbb{R} \to \mathbb{R}.$$

And

$$D_{(\mathsf{R}/\mathsf{R})/\mathsf{R}} = \mathbb{R} \to (\mathbb{R} \to \mathbb{R}).$$

As one particular model, we take

$$
\begin{aligned}
[\![\text{v}]\!] &= 4\\
[\![\text{w}]\!] &= 2\\
[\![\text{x}]\!] &= 65\\
[\![\text{y}]\!] &= -3\\
[\![\text{z}]\!] &= 0\\
\\
[\![\text{1}]\!] &= 1\\
[\![\text{2}]\!] &= 2\\
[\![\text{plus}]\!](a)(b) &= a + b\\
[\![\text{minus}]\!](a)(b) &= a - b\\
[\![\text{times}]\!](a)(b) &= a \cdot b\\
\\
[\![\text{div2}]\!](a)(b) &= 2^{a \div b}
\end{aligned}
$$

A lot of these choices are "standard";
it would not be sensible to do it differently.

The semantics works by function application

$$\dfrac{\dfrac{\text{minus} : (R/R)/R \quad \text{z} : R}{\text{minus z} : R/R} \quad \dfrac{\dfrac{\text{plus} : (R/R)/R \quad \text{v} : R}{\text{plus v} : R/R} \quad \text{w} : R}{\text{plus v w} : R}}{\text{minus z plus v w} : R}$$

The semantics is

$$\begin{aligned}
[\![\text{minus z plus v w}]\!] &= [\![\text{minus}]\!]([\![\text{z}]\!])([\![\text{plus}]\!]([\![\text{v}]\!])([\![\text{w}]\!])) \\
&= [\![\text{minus}]\!]([\![\text{z}]\!])([\![\text{v}]\!] + [\![\text{w}]\!]) \\
&= [\![\text{z}]\!] - ([\![\text{v}]\!] + [\![\text{w}]\!]) \\
&= 0 - (4 + 2) \\
&= -6
\end{aligned}$$

Consider

$$f(v^{\uparrow}, w^{\uparrow}, x^{\uparrow}, y^{\downarrow}, z^{\downarrow}) \quad = \quad \frac{x - y}{2^{z-(v+w)}}.$$

To fit it all on the screen, let's drop the types:

$$\dfrac{\dfrac{\text{div2} \qquad \dfrac{\dfrac{\text{minus} \quad \text{x}}{\text{minus x}} \quad \text{y}}{\text{minus x y}}}{\text{div2 minus x y}} \qquad \dfrac{\dfrac{\text{minus} \quad \text{z}}{\text{minus z}} \qquad \dfrac{\dfrac{\text{plus} \quad \text{v}}{\text{plus v}} \quad \text{w}}{\text{plus v w}}}{\text{minus z plus v w}}}{\text{div2 minus x y minus z plus v w}}$$

div2(t)(u) is supposed to mean $2^{t \div u}$.

Can we determine the polarities of the variables from the tree?

Go from the root to the leaves, marking

green for $\uparrow$ red for $\downarrow$

Flip colors on the right branches of nodes marked *div2* and *minus*.

Can we determine the polarities of the variables from the tree?

Go from the root to the leaves, marking

green for ↑ red for ↓

Flip colors on the right branches of nodes marked *div2* and *minus*.

$$\frac{\dfrac{\mathit{div2} \qquad \dfrac{\dfrac{\mathit{minus} \quad x}{\mathit{minus}\ x} \quad y}{\mathit{minus}\ x\ y}}{\mathit{div2}\ \mathit{minus}\ x\ y} \qquad \dfrac{\dfrac{\mathit{minus} \quad z}{\mathit{minus}\ z} \qquad \dfrac{\dfrac{\mathit{plus} \quad v}{\mathit{plus}\ v} \quad w}{\mathit{plus}\ v\ w}}{\mathit{minus}\ z\ \mathit{plus}\ v\ w}}{\mathit{div2}\ \mathit{minus}\ x\ y\ \mathit{minus}\ z\ \mathit{plus}\ v\ w}$$

Can we determine the polarities of the variables from the tree?

Go from the root to the leaves, marking

green for ↑ red for ↓

Flip colors on the right branches of nodes marked *div2* and *minus*.

$$\dfrac{\dfrac{\mathit{div2} \quad \dfrac{\dfrac{\mathit{minus} \quad x}{\mathit{minus}\ x} \quad y}{\mathit{minus}\ x\ y}}{\color{green}{\mathit{div2}\ \mathit{minus}\ x\ y}} \quad \dfrac{\dfrac{\mathit{minus} \quad z}{\mathit{minus}\ z} \quad \dfrac{\dfrac{\mathit{plus} \quad v}{\mathit{plus}\ v} \quad w}{\mathit{plus}\ v\ w}}{\color{red}{\mathit{minus}\ z\ \mathit{plus}\ v\ w}}}{\color{green}{\mathit{div2}\ \mathit{minus}\ x\ y\ \mathit{minus}\ z\ \mathit{plus}\ v\ w}}$$

Can we determine the polarities of the variables from the tree?

Go from the root to the leaves, marking

green for ↑ red for ↓

Flip colors on the right branches of nodes marked *div2* and *minus*.

minus *x* *plus* *v*

minus x *y* *minus* *z* *plus v* *w*

div2 *minus x y* *minus z* *plus v w*

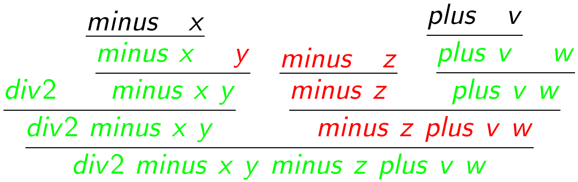
div2 minus x y *minus z plus v w*

div2 minus x y minus z plus v w

Go from the root to the leaves, marking

green for ↑ red for ↓

Flip colors on the right branches of nodes marked *div2* and *minus*.



Can we determine the polarities of the variables from the tree?

Go from the root to the leaves, marking

$\color{green}{\text{green}}$ for $\uparrow$ $\qquad$ $\color{red}{\text{red}}$ for $\downarrow$

$$\color{green}{\dfrac{\dfrac{\mathit{div2} \qquad \dfrac{\dfrac{\mathit{minus} \quad x}{\mathit{minus}\ x} \quad y}{\mathit{minus}\ x\ y}}{\mathit{div2}\ \mathit{minus}\ x\ y} \qquad {\color{red}{\dfrac{\dfrac{\mathit{minus} \quad z}{\mathit{minus}\ z} \qquad {\color{green}{\dfrac{\dfrac{\mathit{plus} \quad v}{\mathit{plus}\ v} \quad w}{\mathit{plus}\ v\ w}}}}{\mathit{minus}\ z\ \mathit{plus}\ v\ w}}}}{\mathit{div2}\ \mathit{minus}\ x\ y\ \mathit{minus}\ z\ \mathit{plus}\ v\ w}}$$

This agrees with what we saw before:

$$f(v^{\color{red}\uparrow}, w^{\color{red}\uparrow}, x^{\color{red}\uparrow}, y^{\color{red}\downarrow}, z^{\color{red}\downarrow}) \quad = \quad \frac{x - y}{2^{z-(v+w)}}$$

Historical influences on this project

CG

Husserl, Frege, Lesniewski (antecedents)
Ajdukiewicz, Bar Hillel ("vanilla" CG)
Lambek, Steedman (extra rules)
van Benthem (syntax-semantics interface)

NL semantics and proof theory, especially related to monotonicity

Leibniz, Sommers 1982 (antecedents)
Montague 1973 (semantics), Fitch 1973 (rules)
Keenan, van Benthem 1986, Sánchez Valencia 1991
Dowty 1994 (internalization)

Inference in computational linguistics

Nairn, Condoravdi, and Karttunen 2006
MacCartney and Manning 2009

Historical influences on this project

- van Benthem 1986, 1991: combine vanilla CG with inference
- Nairn, Condoravdi, and Karttunen 2006:
  something similar (!),
  but not noticed as such,
  not using CG, and not aimed at the same issues

- Steedman: CCG, a working system

- Dowty 1994: internalization of inferential features
  in the type system

- MacCartney and Manning 2009: get something to work.

The problem with Ajdukiewicz/Bar-Hillel CG

The problem is that this form of grammar cannot work out in practice.

I was looking for a form of grammar which has

- a syntax-semantics interface using functions
- can parse a wider class of sentences
- can even work with real text

I settled on CCG.
The important thing is the new rules type raising and composition.

Rules of CCG

general rules of CCG (a few missing)

$$\frac{Y \quad X\backslash Y}{X}\,< \qquad \frac{X/Y \quad Y}{X}\,> \qquad \frac{Y}{X/(X\backslash Y)}\,\mathrm{T}$$

$$\frac{X}{Y\backslash(Y/X)}\,\mathrm{T} \qquad \frac{X/Y \quad Y/Z}{X/Z}\,\mathrm{B} \qquad \frac{Y\backslash Z \quad X\backslash Y}{X\backslash Z}\,\mathrm{B}$$

A tiny lexicon

word	category	word	category
every	NP/N	Fido	NP
cat	N	chased	$(\mathrm{S}\backslash\mathrm{NP})/\mathrm{NP}$
that	$(\mathrm{N}\backslash\mathrm{N})/(\mathrm{S}/\mathrm{NP})$	ran	$\mathrm{S}\backslash\mathrm{NP}$

$$\frac{\text{F} : \mathrm{NP}}{\text{F} : \mathrm{S}/(\mathrm{S}\backslash\mathrm{NP})}\,\mathrm{T}$$

$$\frac{\text{F} : \mathrm{S}/(\mathrm{S}\backslash\mathrm{NP}) \quad \text{ch} : (\mathrm{S}\backslash\mathrm{NP})/\mathrm{NP}}{\text{Fido chased} : \mathrm{S}/\mathrm{NP}}\,\mathrm{B}$$

$$\frac{\text{that} : (\mathrm{N}\backslash\mathrm{N})/(\mathrm{S}/\mathrm{NP}) \quad \text{Fido chased} : \mathrm{S}/\mathrm{NP}}{\text{that Fido chased} : \mathrm{N}\backslash\mathrm{N}}\,>$$

$$\frac{\text{cat} : \mathrm{N} \quad \text{that Fido chased} : \mathrm{N}\backslash\mathrm{N}}{\text{cat that Fido chased} : \mathrm{N}}\,<$$

$$\frac{\text{every} : \mathrm{NP}/\mathrm{N} \quad \text{cat that Fido chased} : \mathrm{N}}{\text{every cat that Fido chased} : \mathrm{NP}}\,>$$

$$\frac{\text{every cat that Fido chased} : \mathrm{NP} \quad \text{ran} : \mathrm{S}\backslash\mathrm{NP}}{\text{every cat that Fido chased ran} : \mathrm{S}}\,<$$

General theory: preorders and monotone functions

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 set of truth values $\mathbb{2} = \{\mathsf{T}, \mathsf{F}\}$ is a preorder, with $\mathsf{F} \leq \mathsf{T}$.

The set of real numbers $\mathbb{R}$ is a preorder, with the usual $\leq$.

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 $\mathbb{P}^X$ is the pointwise order:

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

Three more constructions of preorders

Definition

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

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

Definition

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

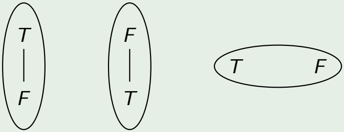
$$p \leq q \text{ in } \mathbb{P}^{\flat} \quad \text{iff} \quad p = q$$

Definition

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

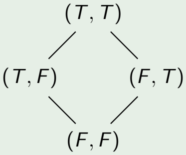
$$(p, q) \leq (p', q') \text{ in } \mathbb{P} \times \mathbb{Q} \quad \text{iff} \quad p \leq p' \text{ in } \mathbb{P}, \text{ and } q \leq q' \text{ in } \mathbb{Q}$$

Example

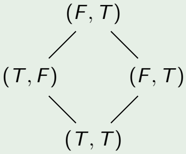


2 $\quad$ 2^{op} $\quad$ $2^{\flat}$

Example

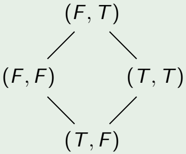


$\mathbb{2} \times \mathbb{2}$

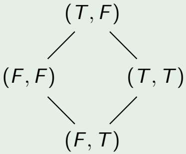


$(\mathbb{2} \times \mathbb{2})^{op}$

Example



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



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

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

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

For example, $\neg$ is antitone on $\mathbb{2}$.

So we have $\neg : \mathbb{2} \xrightarrow{-} \mathbb{2}$.

Monotone and Antitone functions

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

$f : \mathbb{P} \xrightarrow{\cdot} \mathbb{Q}$

For a "random" function f,
we write $f : \mathbb{P} \xrightarrow{\cdot} \mathbb{Q}$.

So this means "in general, neither monotone nor antitone."

Preorder enrichment of grammar

To derive the $\uparrow$ and $\downarrow$ polarities, we need to change the entire architecture of CG,
and indeed to change everything about the semantics,
going from sets to preorders.

For example, standard CG has function types $X \to Y$.

In the preorder enrichment, we have

- $X \xrightarrow{+} Y$ (monotone functions)
- $X \xrightarrow{-} Y$ (antitone functions)
- $X \xrightarrow{\cdot} Y$ (all functions)

To derive the $\uparrow$ and $\downarrow$ polarities, we need to change the entire architecture of CG,
and indeed to change everything about the semantics,
going from sets to preorders.

For example, standard CG has function types $X \to Y$.

In the preorder enrichment, we have

- $X \xrightarrow{+} Y$ (monotone functions)
- $X \xrightarrow{-} Y$ (antitone functions)
- $X \xrightarrow{\cdot} Y$ (all functions)

We start with

$$\begin{array}{lcl} \mathbb{P}_e & = & \text{the flat order on some set} \\ \mathbb{P}_t & = & 2 \\ \mathbb{P}_{num} & = & \mathbb{N} \end{array}$$

A lexicon

By flatness, $e \to t$ is the same as $e \stackrel{+}{\to} t$ and $e \stackrel{-}{\to} t$

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

The CCG rules have many different polarized and marked versions

function application in all flavors of CG

$$\frac{a : X/Y \quad b : Y}{ab : X} >$$

This splits into many versions

$$\frac{a^{\uparrow} : x \xrightarrow{+} y \quad b : x^{\uparrow}}{(ab)^{\uparrow} : y} > \qquad \frac{a^{\downarrow} : x \xrightarrow{-} y \quad b : x^{\uparrow}}{(ab)^{\downarrow} : y} >$$

$$\frac{a^{\downarrow} : x \xrightarrow{+} y \quad b : x^{\downarrow}}{(ab)^{\downarrow} : y} > \qquad \frac{a^{\uparrow} : x \xrightarrow{-} y \quad b : x^{\downarrow}}{(ab)^{\uparrow} : y} >$$

There are yet more versions when we use the polarity $=$.

Notation to summarize these facts

Pattern

$$\frac{f^{d} : x \xrightarrow{m} y \quad g : x^{md}}{f(g)^{d} : y} >$$

We combine markings and polarities as in the table below:

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

We take a single base type r, and as our lexicon we take

plus : $(r/r)/r$	minus : $(r/r)/r$
times : $(r/r)/r$	div2 : $(r/r)/r$
v : r	w : r
x : r	y : r
z : r	
1 : r	2 : r

Polish notation for $z - (v + w)$

$$\dfrac{\dfrac{\text{minus} : (r/r)/r \quad \text{z} : r}{\text{minus z} : r/r} \qquad \dfrac{\dfrac{\text{plus} : (r/r)/r \quad \text{v} : r}{\text{plus v} : r/r} \quad \text{w} : r}{\text{plus v w} : r}}{\text{minus z plus v w} : r}$$

Note that we have polarity facts:

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

What we want to do is to illustrate our algorithm on this relatively simple example.

What we need to do in order to polarize the tree

We want to use

$$\frac{f^{d} : x \stackrel{m}{\rightarrow} y \quad g : x^{md}}{f(g)^{d} : y} >$$

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

Input

$$\frac{\dfrac{\text{minus} : r \rightarrow (r \rightarrow r) \quad \text{z} : r}{\text{minus z} : r \rightarrow r} \quad \dfrac{\dfrac{\text{plus} : r \rightarrow (r \rightarrow r) \quad \text{v} : r}{\text{plus v} : r \rightarrow r} \quad \text{w} : r}{\text{plus v w} : r}}{\text{minus z plus v w} : r}$$

$\text{plus}^{\downarrow} : r \stackrel{+}{\rightarrow} (r \stackrel{+}{\rightarrow} r)$

$\text{minus}^{\uparrow} : r \stackrel{+}{\rightarrow} (r \stackrel{-}{\rightarrow} r)$

Expected output

$$\frac{\dfrac{\text{minus}^{\uparrow} : r \stackrel{+}{\rightarrow} (r \stackrel{-}{\rightarrow} r) \quad \text{z}^{\uparrow} : r}{\text{minus z}^{\uparrow} : r \stackrel{-}{\rightarrow} r} \quad \dfrac{\dfrac{\text{plus}^{\downarrow} : r \stackrel{+}{\rightarrow} (r \stackrel{+}{\rightarrow} r) \quad \text{v}^{\downarrow} : r}{\text{plus v}^{\downarrow} : r \stackrel{+}{\rightarrow} r} \quad \text{w}^{\downarrow} : r}{\text{plus v w}^{\downarrow} : r}}{\text{minus z plus v w}^{\uparrow} : r}$$

We want to use

$$\frac{f^{d} : x \xrightarrow{m} y \quad g : x^{md}}{fg^{d} : y} >$$

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

Input

$$\frac{\dfrac{\text{minus} : r \to (r \to r) \quad \text{z} : r}{\text{minus z} : r \to r} \quad \dfrac{\dfrac{\text{plus} : r \to (r \to r) \quad \text{v} : r}{\text{plus v} : r \to r} \quad \text{w} : r}{\text{plus v w} : r}}{\text{minus z plus v w} : r}$$

$\text{plus}^{\downarrow} : r \xrightarrow{+} (r \xrightarrow{+} r)$

$\text{minus}^{\uparrow} : r \xrightarrow{+} (r \xrightarrow{-} r)$

More on how it's done

We want to use

$$\dfrac{f^{d} : x \xrightarrow{m} y \quad g : x^{md}}{fg^{d} : y} >$$

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

Input

$$\dfrac{\dfrac{\text{minus} : r \to (r \to r) \quad \text{z} : r}{\text{minus z} : r \to r} \quad \dfrac{\dfrac{\text{plus} : r \to (r \to r) \quad \text{v} : r}{\text{plus v} : r \to r} \quad \text{w} : r}{\text{plus v w} : r}}{\text{minus z plus v w} : r}$$

$\text{plus}^{\downarrow} : r \xrightarrow{+} (r \xrightarrow{+} r)$

$\text{minus}^{\uparrow} : r \xrightarrow{+} (r \xrightarrow{-} r)$

Extra requirement

The bottom of the tree should get $\uparrow$

We want to use

$$\frac{f^{d} : x \xrightarrow{m} y \quad g : x^{md}}{fg^{d} : y} >$$

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

Starting

$$\dfrac{\dfrac{\text{minus} : r \xrightarrow{+} (r \xrightarrow{-} r) \quad \text{z} : r}{\text{minus z} : r \rightarrow r} \quad \dfrac{\dfrac{\text{plus} : r \xrightarrow{+} (r \xrightarrow{+} r) \quad \text{v} : r}{\text{plus v} : r \rightarrow r} \quad \text{w} : r}{\text{plus v w} : r}}{\text{minus z plus v w}^{\uparrow} : r}$$

We want to use

$$\frac{f^{d} : x \xrightarrow{m} y \quad g : x^{md}}{fg^{d} : y} >$$

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

Let's use variables for the missing polarities

$$\dfrac{\dfrac{\text{minus}^{d} : r \xrightarrow{+} (r \xrightarrow{-} r) \quad \text{z} : r}{\text{minus z} : r \to r} \qquad \dfrac{\dfrac{\text{plus}^{e} : r \xrightarrow{+} (r \xrightarrow{+} r) \quad \text{v} : r}{\text{plus v} : r \to r} \quad \text{w} : r}{\text{plus v w} : r}}{\text{minus z plus v w}^{\uparrow} : r}$$

We want to use

$$\frac{f^{d} : x \xrightarrow{m} y \quad g : x^{md}}{fg^{d} : y} > $$

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

Then we fill in the rest based on this and the overall pattern

$$\dfrac{\dfrac{\text{minus}^{d} : r \xrightarrow{+} (r \xrightarrow{-} r) \quad \text{z}^{+d} : r}{\text{minus z}^{d} : r \xrightarrow{-} r} \quad \dfrac{\dfrac{\text{plus}^{e} : r \xrightarrow{+} (r \xrightarrow{+} r) \quad \text{v}^{+e} : r}{\text{plus v}^{e} : r \xrightarrow{+} r} \quad \text{w} : r}{\text{plus v w}^{e} : r}}{\text{minus z plus v w}^{\uparrow} : r}$$

We have constraints at the bottom: $\uparrow = d$, and $-d = e$.

More on how it's done

We want to use

$$\frac{f^{d} : x \xrightarrow{m} y \quad g : x^{md}}{fg^{d} : y} >$$

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

Then we fill in the rest based on this and the overall pattern

$$\frac{\dfrac{\text{minus}^{d} : r \xrightarrow{+} (r \xrightarrow{-} r) \quad \text{z}^{+d} : r}{\text{minus z}^{d} : r \xrightarrow{-} r} \quad \dfrac{\dfrac{\text{plus}^{e} : r \xrightarrow{+} (r \xrightarrow{+} r) \quad \text{v}^{+e} : r}{\text{plus v}^{e} : r \xrightarrow{+} r} \quad \text{w} : r}{\text{plus v w}^{e} : r}}{\text{minus z plus v w}^{\uparrow} : r}$$

We have constraints at the bottom: $\uparrow = d$, and $-d = e$.

So we solve this to get the desired output

$d = \uparrow \qquad e = \downarrow \qquad \text{v}^{\downarrow} \qquad \text{w}^{\downarrow} \qquad \text{z}^{\uparrow}$

general rules of CCG (a few missing)

$$\frac{Y \quad X\backslash Y}{X}\ <$$

$$\frac{X/Y \quad Y}{X}\ >$$

$$\frac{Y}{X/(X\backslash Y)}\ \mathrm{T}$$

$$\frac{X}{Y\backslash (Y/X)}\ \mathrm{T}$$

$$\frac{X/Y \quad Y/Z}{X/Z}\ \mathrm{B}$$

$$\frac{Y\backslash Z \quad X\backslash Y}{X\backslash Z}\ \mathrm{B}$$

We next want to see the preordered version of (B).

If $f : \mathbb{P} \to \mathbb{Q}$ is monotone and $g : \mathbb{Q} \to \mathbb{R}$ is monotone, then $g \circ f$ is monotone.

If $f : \mathbb{P} \to \mathbb{Q}$ is monotone and $g : \mathbb{Q} \to \mathbb{R}$ is antitone, then $g \circ f$ is antitone.

If $f : \mathbb{P} \to \mathbb{Q}$ is antitone and $g : \mathbb{Q} \to \mathbb{R}$ is monotone, then $g \circ f$ is antitone.

If $f : \mathbb{P} \to \mathbb{Q}$ is antitone and $g : \mathbb{Q} \to \mathbb{R}$ is antitone, then $g \circ f$ is monotone.

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 $\xrightarrow{\cdot}$.

Interim summary

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

We introduce 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 $f_1(p_1) = f_2(p_2)$.

the (B) rule

$$\frac{f^{d} : x \xrightarrow{m} y \quad g^{md} : y \xrightarrow{n} z}{(g \circ f)^{d} : x \xrightarrow{mn} z}\ \mathrm{B}$$

the (T) rule

$$\frac{f^{md} : x}{(\lambda g.g(f))^{d} : (x \xrightarrow{m} y) \xrightarrow{+} y}\ \mathrm{T}$$

Note that the last marking is $+$.

From the language of monotonicity to the monotonicty of language

The syntactic types start with $\textsc{s}$, $\textsc{n}$, and $\textsc{np}$,
just as in CG.
(To handle numbers, we also add $\textsc{num}$.)
For the semantic types, we start with base types, e, t, and num.

We then form complex types:

- If x and y are types, so are $x \xrightarrow{+} y$, $x \xrightarrow{-} y$, and $x \xrightarrow{\cdot} y$.

Abbreviations

(et) abbreviates $e \xrightarrow{\cdot} t$.

np^{+} abbreviates $et \xrightarrow{+} t$.

np^{-} abbreviates $et \xrightarrow{-} t$.

np abbreviates $et \xrightarrow{\cdot} t$.

For the determiners, our lexicon uses order-enriched types

word	type
every	$\mathrm{N} \xrightarrow{-} \mathrm{NP}^{+}$
some	$\mathrm{N} \xrightarrow{+} \mathrm{NP}^{+}$

word	type
no	$\mathrm{N} \xrightarrow{-} \mathrm{NP}^{-}$
most	$\mathrm{N} \xrightarrow{\cdot} \mathrm{NP}^{+}$

These are basically the internalized types first considered by Dowty.

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

The syntax tree is given to us by the parser:

$$\dfrac{\text{F} : \mathrm{NP}}{\text{F} : \mathrm{S}/(\mathrm{S}\backslash\mathrm{NP})}\ {\scriptstyle \mathrm{T}}$$

$$\dfrac{\text{F} : \mathrm{S}/(\mathrm{S}\backslash\mathrm{NP}) \qquad \text{ch} : (\mathrm{S}\backslash\mathrm{NP})/\mathrm{NP}}{\text{Fido chased} : \mathrm{S}/\mathrm{NP}}\ {\scriptstyle \mathrm{B}}$$

$$\dfrac{\text{that} : (\mathrm{N}\backslash\mathrm{N})/(\mathrm{S}/\mathrm{NP}) \qquad \text{Fido chased} : \mathrm{S}/\mathrm{NP}}{\text{that Fido chased} : \mathrm{N}\backslash\mathrm{N}}\ {\scriptstyle >}$$

$$\dfrac{\text{cat} : \mathrm{N} \qquad \text{that Fido chased} : \mathrm{N}\backslash\mathrm{N}}{\text{cat that Fido chased} : \mathrm{N}}\ {\scriptstyle <}$$

$$\dfrac{\text{every} : \mathrm{NP}/\mathrm{N} \qquad \text{cat that Fido chased} : \mathrm{N}}{\text{every cat that Fido chased} : \mathrm{NP}}\ {\scriptstyle >}$$

$$\dfrac{\text{every cat that Fido chased} : \mathrm{NP} \qquad \text{ran} : \mathrm{S}\backslash\mathrm{NP}}{\text{every cat that Fido chased ran} : \mathrm{S}}\ {\scriptstyle <}$$

This tree has a semantics which is suggested below:

$$\dfrac{\text{F} : \mathrm{NP}}{\text{F} : (\mathrm{NP} \to \mathrm{S}) \to \mathrm{S}}\ {\scriptstyle \mathrm{T}}$$

$$\dfrac{\text{F} : (\mathrm{NP} \to \mathrm{S}) \to \mathrm{S} \qquad \text{ch} : \mathrm{NP} \to (\mathrm{NP} \to \mathrm{S})}{\text{Fido ch} : \mathrm{NP} \to \mathrm{S}}\ {\scriptstyle \mathrm{B}}$$

$$\dfrac{\text{that} : (\mathrm{NP} \to \mathrm{S}) \to (\mathrm{N} \to \mathrm{N}) \qquad \text{Fido ch} : \mathrm{NP} \to \mathrm{S}}{\text{that Fido chased} : \mathrm{N} \to \mathrm{N}}\ {\scriptstyle >}$$

$$\dfrac{\text{cat} : \mathrm{N} \qquad \text{that Fido chased} : \mathrm{N} \to \mathrm{N}}{\text{cat that Fido chased} : \mathrm{N}}\ {\scriptstyle <}$$

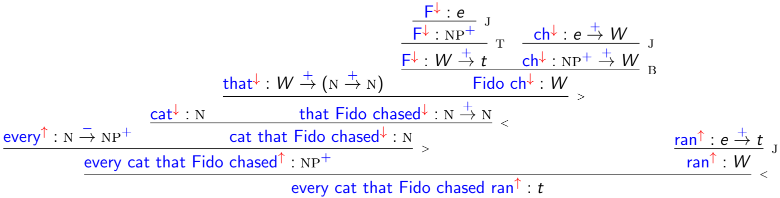
$$\dfrac{\text{every} : \mathrm{N} \to \mathrm{NP} \qquad \text{cat that Fido chased} : \mathrm{N}}{\text{every cat that Fido chased} : \mathrm{NP}}\ {\scriptstyle >}$$

$$\dfrac{\text{every cat that Fido chased} : \mathrm{NP} \qquad \text{ran} : \mathrm{NP}}{\text{every cat that Fido chased ran} : \mathrm{S}}$$

"The structure of every sentence is a lesson in logic."

John Stuart Mill (1867)

Saving on notation by writing W for $\text{NP}^{+} \xrightarrow{+} t$:



The arrows *could* be determined just by parsing from our rules,
but since we want to use the parse given to us by a parser,
we aim for an algorithm that polarizes an upolarized CCG tree.

I am omitting discussion of our actual algorithm.
You could take it to be constraint satisfaction,
but it's possible to be much more direct.

$$\dfrac{\dfrac{\dfrac{\textsf{ch}^{=} : e \xrightarrow{+} pr}{\textsf{ch}^{=} : np \xrightarrow{+} pr}\,\textsc{k}}{\textsf{ch}^{\uparrow} : np \xrightarrow{+} pr}\,\textsc{w}}{}$$

$$\dfrac{\textsf{some cat}^{\uparrow} : np^{+}}{\textsf{some cat}^{\uparrow} : np}\,\textsc{m} \qquad \dfrac{\textsf{no arm}^{\downarrow} : np^{-}}{\textsf{no arm}^{\downarrow} : np}\,\textsc{m}$$

$$\dfrac{\textsf{and} : np \xrightarrow{+} (np \xrightarrow{+} np) \qquad \textsf{no arm}^{\downarrow} : np}{\textsf{and no arm} : np \xrightarrow{+} np}\,>$$

$$\dfrac{\textsf{some cat}^{\uparrow} : np \qquad \textsf{and no arm} : np \xrightarrow{+} np}{\textsf{some cat and no armadillo}^{\uparrow} : np}\,<$$

$$\dfrac{\textsf{ch}^{\uparrow} : np \xrightarrow{+} pr \qquad \textsf{some cat and no armadillo}^{\uparrow} : np}{\textsf{chased some cat and no armadillo}^{\uparrow} : pr}\,>$$

$$\dfrac{\textsf{F}^{\uparrow} : e \qquad \textsf{chased some cat and no armadillo}^{\uparrow} : pr}{\textsf{Fido chased some cat and no armadillo}^{\uparrow} : t}\,<$$

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

We start with preorders for the base types:

$$\begin{array}{lcl} \mathbb{P}_e & = & \text{the flat preorder} \\ & & \text{on an arbitrary set } X \\ \mathbb{P}_t & = & 2 \\ \mathbb{P}_{num} & = & \mathbb{N} \end{array}$$

Each type x gives us a preorder $\mathbb{P}_x$ using the following rules

$$\begin{array}{lcl} \mathbb{P}_{x \xrightarrow{+} y} & = & \mathbb{P}_x \xrightarrow{+} \mathbb{P}_y \\ \mathbb{P}_{x \xrightarrow{-} y} & = & \mathbb{P}_x \xrightarrow{-} \mathbb{P}_y \\ \mathbb{P}_{x \xrightarrow{\cdot} y} & = & \mathbb{P}_x \xrightarrow{\cdot} \mathbb{P}_y \end{array}$$

A lexicon

item	category	semantic type
Fido, Felix	NP	e
cat, dog	N	$n = pr$
swim, run	IV = S\NP	pr
chase, see, hit, kiss	TV = IV/NP	$e \to pr$
every	DET = NP/N	$pr \xrightarrow{-} np^{+}$
some	NP/N	$pr \xrightarrow{+} np^{+}$
no	NP/N	$pr \xrightarrow{-} np^{-}$
most	NP/N	$pr \xrightarrow{\cdot} np^{+}$
who	(N\N)/(S/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	X/(X\X)	$x \xrightarrow{+} (x \xrightarrow{+} x)$
one, two, three	NUM	num
more than	DET/NUM	$num \xrightarrow{-} (pr \xrightarrow{+} np^{+})$
less than	DET/NUM	$num \xrightarrow{+} (pr \xrightarrow{-} np^{-})$
if ... then ...	(S\S)/S	$t \xrightarrow{-} (t \xrightarrow{+} t)$

A lexicon

item	category	type
Fido, Felix	NP	e
cat, dog	N	$n = pr$
swim, run	$\mathrm{IV} = \mathrm{S}\backslash\mathrm{NP}$	pr
chase, see, hit, kiss	$\mathrm{TV} = \mathrm{IV}/\mathrm{NP}$	$e \rightarrow pr$

For these content words, a model has an interpretation that can be any element of the listed semantic type:

$$\begin{aligned}
[\![Fido]\!], [\![Felix]\!], \ldots &\in \mathbb{P}_e \\
[\![cat]\!], [\![dog]\!], \ldots &\in \mathbb{P}_{pr} \\
[\![swim]\!], [\![run]\!], \ldots &\in \mathbb{P}_{pr} \\
[\![chase]\!], [\![see]\!], \ldots &\in \mathbb{P}_{e \rightarrow et}
\end{aligned}$$

Models: the function words have standard values

Three new facts, for x a Boolean category

$$\frac{f^{=} : e \to x}{(f_\star)^{=} : \mathrm{NP} \xrightarrow{+} x}\ \mathrm{I} \qquad \frac{f^{d} : e \to x}{(f_\star^{+})^{d} : \mathrm{NP}^{+} \xrightarrow{+} x}\ \mathrm{J} \qquad \frac{f^{d} : e \to x}{(f_\star^{-})^{\mathit{flip}\, d} : \mathrm{NP}^{-} \xrightarrow{+} x}\ \mathrm{K}$$

The last is the most subtle rule of the system.

It is related to the our type for transitive verbs:

$$e \to (e \to t)$$

This is a departure from what one would expect from CG:

$$\mathsf{np}^{+} \xrightarrow{+} (\mathsf{np}^{+} \xrightarrow{+} t)$$

or perhaps

$$np \xrightarrow{+} (np \xrightarrow{+} t)$$

Rules again, but with an explanation

Rules

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

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

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

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

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

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

The $>$ in the application rule is function application.

The T in the type-raising rule is the Montague lift.

The B in the type-raising rule is function composition, backwards.

The r_m in the K rule is from our refinement of the Justification Theorem.

In the M rule, we have a trivial inclusion.

The W rule is trivial.

What our polarized trees mean, on a semantic level

Example (a polarized syntax tree)

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

Example (Abstract the words and move from syntax to semantics)

$$\dfrac{\dfrac{v^{\uparrow} : pr \overset{+}{\to} np^{+} \quad w^{\uparrow} : pr}{vw^{\uparrow} : pr \overset{+}{\to} t}\,> \qquad \dfrac{\dfrac{x^{\downarrow} : e \overset{+}{\to} pr}{r\_x^{\uparrow} : np^{-} \overset{+}{\to} pr}\,\mathrm{K} \quad \dfrac{y^{\uparrow} : pr \overset{-}{\to} np^{-} \quad z^{\downarrow} : pr}{yz^{\uparrow} : np^{-}}\,>}{(r\_x)(yz)^{\uparrow} : pr}\,>}{\tau = (vw)((r\_x)(yz))^{\uparrow} : t}\,>$$

Note that the semantic term τ on the bottom is a combinator term.

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

$$[\![\tau]\!] : \mathbb{P}_{pr \overset{+}{\to} np^{+}} \times \mathbb{P}_{pr} \times (\mathbb{P}_{e \overset{+}{\to} pr})^{op} \times \mathbb{P}_{pr \overset{-}{\to} np^{-}} \times (\mathbb{P}_{pr})^{op} \quad \overset{+}{\to} \quad \mathbb{P}_{t}$$

What we are given

We are given syntax trees like the following:

Example

$$\dfrac{\dfrac{\text{some} : \mathrm{NP}/\mathrm{N} \quad \text{dog} : \mathrm{N}}{\text{some dog} : \mathrm{NP}}> \qquad \dfrac{\text{chased} : (\mathrm{S}\backslash\mathrm{NP})/\mathrm{NP} \quad \dfrac{\text{no} : \mathrm{NP}/\mathrm{N} \quad \text{cat} : \mathrm{N}}{\text{no cat} : \mathrm{NP}}>}{\text{chased no cat} : \mathrm{S}\backslash\mathrm{NP}}>}{\text{some dog chased no cat} : \mathrm{S}}<$$

The standard semantics is given by a translation to (unmarked) (e, t)-types:

$$\begin{array}{lcl}
\mathrm{N} & \mapsto & et \\
\mathrm{NP} & \mapsto & (et \to t) \\
\mathrm{DET} = \mathrm{NP}/\mathrm{N} & \mapsto & (et) \to ((et) \to t) \\
\hline
\mathrm{X}/\mathsf{Y} & \mapsto & \mathsf{Y}^{tr} \to \mathrm{X}^{tr} \\
\mathsf{Y}\backslash\mathrm{X} & \mapsto & \mathsf{Y}^{tr} \to \mathrm{X}^{tr}
\end{array}$$

Again, we are given syntax trees like the following:

Example

$$\dfrac{\dfrac{\text{some} : \mathrm{NP}/\mathrm{N} \quad \text{dog} : \mathrm{N}}{\text{some dog} : \mathrm{NP}}> \qquad \dfrac{\text{chased} : (\mathrm{S}\backslash\mathrm{NP})/\mathrm{NP} \quad \dfrac{\text{no} : \mathrm{NP}/\mathrm{N} \quad \text{cat} : \mathrm{N}}{\text{no cat} : \mathrm{NP}}>}{\text{chased no cat} : \mathrm{S}\backslash\mathrm{NP}}>}{\text{some dog chased no cat} : \mathrm{S}}<$$

But the word order information plays no role in the semantics, so we rather think of it as

Example (Now think of functions)

$$\dfrac{\dfrac{\text{some} : \mathrm{N} \to \mathrm{NP} \quad \text{dog} : \mathrm{N}}{\text{some dog} : \mathrm{NP}}> \qquad \dfrac{\text{chased} : \mathrm{NP} \to (\mathrm{NP} \to \mathrm{S}) \quad \dfrac{\text{no} : \mathrm{N} \to \mathrm{NP} \quad \text{cat} : \mathrm{N}}{\text{no cat} : \mathrm{NP}}>}{\text{chased no cat} : \mathrm{NP} \to \mathrm{S}}>}{\text{some dog chased no cat} : \mathrm{S}}<$$

Example (Desired output)

$$\dfrac{\dfrac{\text{some}^{\uparrow} : \mathrm{N} \xrightarrow{+} \mathrm{NP}^{+} \quad \text{dog}^{\uparrow} : \mathrm{N}}{\text{some dog}^{\uparrow} : \mathrm{NP}^{+}}> \qquad \dfrac{\text{ch}^{\uparrow} : \mathrm{NP}^{-} \xrightarrow{+} (\mathrm{NP}^{+} \xrightarrow{+} \mathrm{S}) \quad \dfrac{\text{no}^{\uparrow} : \mathrm{N} \xrightarrow{-} \mathrm{NP}^{-} \quad \text{cat}^{\downarrow} : \mathrm{N}}{\text{no cat}^{\uparrow} : \mathrm{NP}^{-}}>}{\text{chased no cat}^{\uparrow} : \mathrm{NP}^{+} \xrightarrow{+} \mathrm{S}}>}{\text{some dog chased no cat}^{\uparrow} : \mathrm{S}}<$$

We want to make choices on all of the verb types, markings, and polarities

Example (Now think of functions)

$$\dfrac{\dfrac{\textsf{some} : \mathrm{N} \to \mathrm{NP} \quad \textsf{dog} : \mathrm{N}}{\textsf{some dog} : \mathrm{NP}}> \qquad \dfrac{\textsf{chased} : \mathrm{NP} \to (\mathrm{NP} \to \mathrm{S}) \quad \dfrac{\textsf{no} : \mathrm{N} \to \mathrm{NP} \quad \textsf{cat} : \mathrm{N}}{\textsf{no cat} : \mathrm{NP}}>}{\textsf{chased no cat} : \mathrm{NP} \to \mathrm{S}}>}{\textsf{some dog chased no cat} : \mathrm{S}}<$$

Example (Desired output)

$$\dfrac{\dfrac{\textsf{some}^{\uparrow} : \mathrm{N} \xrightarrow{+} \mathrm{NP}^{+} \quad \textsf{dog}^{\uparrow} : \mathrm{N}}{\textsf{some dog}^{\uparrow} : \mathrm{NP}^{+}}> \qquad \dfrac{\textsf{ch}^{\uparrow} : \mathrm{NP}^{-} \xrightarrow{+} (\mathrm{NP}^{+} \xrightarrow{+} \mathrm{S}) \quad \dfrac{\textsf{no}^{\uparrow} : \mathrm{N} \xrightarrow{-} \mathrm{NP}^{-} \quad \textsf{cat}^{\downarrow} : \mathrm{N}}{\textsf{no cat}^{\uparrow} : \mathrm{NP}^{-}}>}{\textsf{chased no cat}^{\uparrow} : \mathrm{NP}^{+} \xrightarrow{+} \mathrm{S}}>}{\textsf{some dog chased no cat}^{\uparrow} : \mathrm{S}}<$$

Our same Soundness Theorem before would tell us that

- If the leaves of the tree belong to the semantic spaces associated with their categories
- and If the tree matches our rule set at every non-leaf node
- and if the root of the tree has category S and polarization $\uparrow$

Then the polarity arrows on the leaves are correct semantic statements in every model.

Examples of polarized sentences from our system

No$^{\uparrow}$ man$^{\downarrow}$ walks$^{\downarrow}$
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$^{\downarrow}$ that$^{\downarrow}$ 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$^{\downarrow}$

Monotonicity + Natural Logic at work: the FRaCaS dataset

entail, *contradict* or *neural*?

P: A schoolgirl with a black bag is on a crowded train
H: No schoolgirl is on a crowded train

entail, *contradict* or *neural*?

P: A schoolgirl with a black bag is on a crowded train
H: A girl is on a train

Monotonicity + Natural Logic at work: the FRaCaS dataset

system	MM08	AM14	LS13	T14	D14	M15	A16	ours
multi-premise?	N	N	Y	Y	Y	Y	Y	Y
# problems	44	44	74	74	74	74	74	74
Acc. (%)	97.73	95	62	80	95	77	93	88

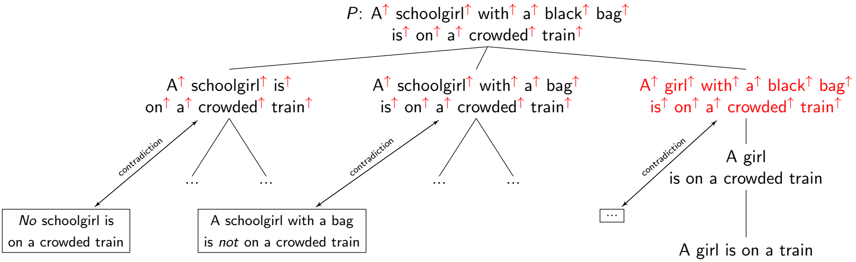
MM08: MacCartney and Manning (2008).
AM14: Angeli and Manning (2014).
LS13: Lewis and Steedman (2013).
T14: Tian et al. (2014). D14: Dong et al. (2014).
M15: Mineshima et al. (2015).
A16: Abzianidze (2016).
ours: Joint work with Hai Hu and Qi Chen (2019)

Monotonicity + Natural Logic at work: the FRaCaS dataset

Truth / Pred	E	U	C
Entail	29	7	0
Unknown	0	33	0
Contradict	0	2	3

Confusion matrix of our system.
Our system achieves 100% precision and comparable accuracy with others.

How the algorithm works, roughly



The MonaLog Inference algorithm

Input: sentence pair: *every man walks ?? every young man walks*

Step 1: get polarized sentences

Step 2: extract all adjs, nouns, adverbs, verbs, RC,
and add to **knowledge base** $\mathcal{K}$ the following:

adj n $\leq$ *n*,
n pp $\leq$ *n*,
n RC $\leq$ *n*.

E.g., *small dog* $\leq$ *dog*, *dog from France* $\leq$ *dog*, *dog that barks* $\leq$ *dog*.
v a $\leq$ *v*. E.g., *walk fast* $\leq$ *walk*.

Help yourself to WordNet information:
poodle $\leq$ *dog*, *dog* $|$ *cat*, *big* $\perp$ *small*

If we say

$$\text{Tricia isa doctor}$$

then we would add to our knowledge base

$$\text{every doctor} \leq \text{Tricia}$$

and also

$$\text{Tricia } \leq \text{some doctor}$$

- FraCaS: 346 problems
- See our paper at IWCS (Hu et al., 2019)

An example where monotonicity is not enough

P1: Most Europeans are resident in Europe
P2: All Europeans are people
P3: All people who are resident in Europe can travel freely within Europe

H: *Most Europeans can travel freely within Europe*

$$\frac{\text{Det } x\ y \qquad \text{All } x\ z}{\text{Det } x\ (y \wedge z)}\ \mathrm{DET}$$

Monotonicity + Natural Logic at work: how it works on the SICK dataset

- SICK (Sentences Involving Compositional Knowledge)
- 10,000 English sentence pairs, generated from image, video descriptions, annotated by Turkers.

id	premise	hypothesis	orig. label	corr. label
219	There is no girl in white dancing	A girl in white is dancing	C	C
294	Two girls are lying on the ground	Two girls are sitting on the ground	N	C
743	A couple who have just got married are walking down the isle	The bride and the groom are leaving after the wedding	E	N
1645	A girl is on a jumping car	One girl is jumping on the car	E	N
1981	A truck is quickly going down a hill	A truck is quickly going up a hill	N	C
8399	A man is playing guitar next to a drummer	A guitar is being played by a man next to a drummer	E	n.a.

Table: Examples from SICK Marelli et al. (2014) and corrected SICK Kalouli et al. (2018) w/ their syntactic variations.

Experiment 1: Solve SICK, using MonaLog (+ BERT)

- MonaLog:

1. Syntactic transformations:

a) pass2act; b) there be no N doing sth. → No N is doing sth.

2. Generate entailments and contradictions from *premise*.
3. If *hypothesis* in E/C, then return E/C, else return Neutral.

- MonaLog + BERT:

If MonaLog returns E/C, then use MonaLog, else use BERT.

Results: Experiment 1

system	P	R	acc.
majority baseline	–	–	56.36
Natural-logic-based: MonaLog‡ (this work)			
MonaLog + pass2act	89.42	72.18	80.25†
MonaLog + existential trans.	89.43	71.53	79.11†
MonaLog + all	83.75	70.66	77.19
MonaLog + all	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 and Schütze (2017)	–	–	**87.1**
Beltagy et al. (2016)	–	–	85.1
Logic-based systems			
Bjerva et al. (2014)	93.6	60.6	81.6
Abzianidze (2015)	97.95	58.11	81.35
Martínez-Gómez et al. (2017)	97.04	63.64	83.13
Yanaka et al. (2018)	84.2	77.3	84.3

Performance on the SICK test set.

† = corrected SICK.

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

Results involving BERT are averaged across six runs; same for later experiments.

id	premise	hypothesis	SICK	corrected SICK	MonaLog
359	There is no dog chasing another or holding a stick in its mouth	Two dogs are running and carrying an object in their mouths	N	n.a.	C
912	A woman is being kissed by a man	A lady is being kissed by a man	E	N	E
1402	A man is crying	A man is screaming	N	n.a.	E
1760	A flute is being played by a girl	There is no woman playing a flute	N	n.a.	C
2897	The man is lifting weights	The man is lowering barbells	N	n.a.	E
2922	A herd of caribous is not crossing a road	A herd of deer is crossing a street	N	n.a.	C
3403	A man is folding a tortilla	A man is unfolding a tortilla	N	n.a.	C
4333	A woman is picking a can	A woman is taking a can	E	N	E
5138	A man is doing a card trick	A man is doing a magic trick	N	n.a.	E
5268	Somebody is folding a piece of paper	A person is folding a piece of paper	E	C	E
5793	A man is cutting a fish	A woman is slicing a fish	N	n.a.	C

Examples of incorrect answers by MonaLog;

Experiment 2: generate new pairs, add to training data. Then repeat Experiment 1

1. Pair the generated entailments/contradictions with the input premise.
2. Add newly generated pairs to SICK.train.

Repeat Experixment 1.

training data	# E	# N	# C	acc.
SICK.train: baseline	1.2k	2.5k	0.7k	85.00
1/4 gen. + SICK.train	8k	2.5k	4k	85.30
1/2 gen. + SICK.train	15k	2.5k	7k	85.81
all gen. + SICK.train	30k	2.5k	14k	86.51
E, C prob. threshold = 0.95	30k	2.5k	14k	86.71
Hybrid baseline	1.2k	2.5k	0.7k	85.95
Hybrid + all gen.	30k	2.5k	14k	87.16
Hybrid + all gen. + threshold	30k	2.5k	14k	**87.49**

Table: Results of BERT trained on MonaLog-generated entailments and contradictions plus SICK.train (using the corrected SICK set).

Results: Exp 2 generated sentence pairs

label	premise	hypothesis	comm.
E	A woman be not cooking something	A person be not cooking something	correct
E	A man be talk to a woman who be seat beside he and be drive a car	A man be talk	correct
E	A south African plane be not fly in a blue sky	A south African plane be not fly in a very blue sky in a blue sky	unnat.
C	No panda be climb	Some panda be climb	correct
C	A man on stage be sing into a microphone	A man be not sing into a microphone	correct
C	No man rapidly be chop some mushroom with a knife	Some man rapidly be chop some mushroom with a knife with a knife	unnat.
E	Few$^{\uparrow}$ people$^{\downarrow}$ be$^{\downarrow}$ eat$^{\downarrow}$ at$^{\downarrow}$ red$^{\downarrow}$ table$^{\downarrow}$ in$^{\downarrow}$ a$^{\downarrow}$ restaurant$^{\downarrow}$ without$^{\downarrow}$ light$^{\uparrow}$	Few$^{\uparrow}$ large$^{\downarrow}$ people$^{\downarrow}$ be$^{\downarrow}$ eat$^{\downarrow}$ at$^{\downarrow}$ red$^{\downarrow}$ table$^{\downarrow}$ in$^{\downarrow}$ a$^{\downarrow}$ Asian$^{\downarrow}$ restaurant$^{\downarrow}$ without$^{\downarrow}$ light$^{\uparrow}$	correct

Table: Sentence pairs generated by MonaLog, lemmatized.

Let's stump BERT: work in progress

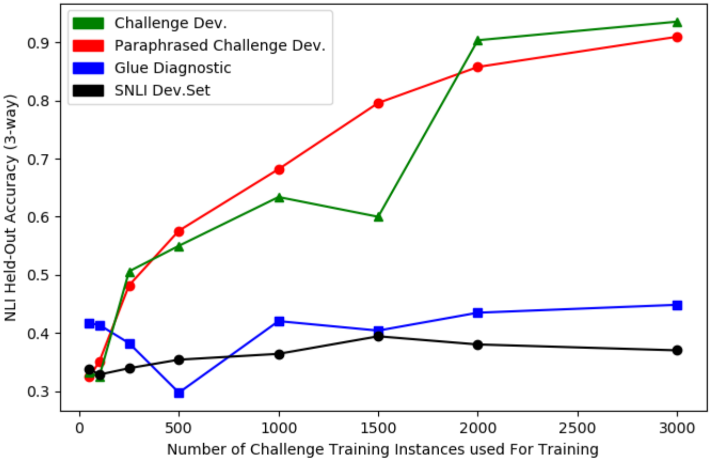
11 all black mammals saw exactly 5 stallions who danced
a black dogs saw exactly 6 stallions who danced
CONTRADICTION

67 every brown mammal did not touch some but not all stallions without faint odor
a brown or black mammal did touched some but not all stallions without faint odor
CONTRADICTION

93 at least 6 old poodles who were not sad did not stare at the table
at most 3 old dogs who were not sad did not stare at the table
CONTRADICTION

19 some bulldog who touched every object was sad
a bulldog who touched each mailbox was sad
ENTAILMENT

Lots of results so far but no clear conclusions



A parting point on the three logic-based approaches to NLI

	pros	cons
Machine Learning-based	wide coverage	no idea what is going on
Logic-based	high precision	translation to logic form: hard!
NatLog-based	high precision	cannot handle much syntactic variation

And ours generates training data for free.

Summary

- We entended monotonicity from vanilla CG to CCG.
- We have a running system that can polarize input sentences.
- We built a Natural-Logic-based system that can solve a large NLI dataset.
- Our system can generate high-quality sentence pairs, helpful to a ML model.
- Remaining problems:
  - string-comparison still too brittle;
  - it is hard to generate neutral pairs;
  - contradiction is sometimes hard to define.

Thomas Icard
monotonicity/polarity versions of the typed lambda calculus
MonaLog inference engine.

Will Tune
2017 IU Math PhD on properties of the monotonicity/polarity typed lambda calculus

Hai Hu, current Computational Linguistics PhD student at IU
implementation of the polarization algorithm, MonaLog inference

Qi Chen, Atreyee Mukherjee, Sandra Kübler
Work on SICK and ML in general

Kyle Richardson
latest work on stumping BERT

Abzianidze, L. (2015). A tableau prover for natural logic and language. In *EMNLP*, pages 2492–2502.

Abzianidze, L. (2016). Natural solution to fracas entailment problems. In *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*, pages 64–74.

Angeli, G. and Manning, C. (2014). NaturalLI: Natural logic inference for common sense reasoning. In *EMNLP*, pages 534–545.

Beltagy, I., Roller, S., Cheng, P., Erk, K., and Mooney, R. J. (2016). Representing meaning with a combination of logical and distributional models. *Computational Linguistics*, 42(4):763–808.

Bjerva, J., Bos, J., Van der Goot, R., and Nissim, M. (2014). The meaning factory: Formal semantics for recognizing textual entailment and determining semantic similarity. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 642–646.

Dong, Y., Tian, R., and Miyao, Y. (2014). Encoding generalized quantifiers in dependency-based compositional semantics. In

Proceedings of the 28th Pacific Asia Conference on Language, Information and Computing.

Hu, H., Chen, Q., and Moss, L. S. (2019). Natural language inference with monotonicity. In *Proceedings of the 13th International Conference on Computational Semantics (IWCS)*.

Kalouli, A.-L., Real, L., and de Paiva, V. (2018). Wordnet for easy textual inferences. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC)*, Miyazaki, Japan.

Lewis, M. and Steedman, M. (2013). Combined distributional and logical semantics. *Transactions of the Association of Computational Linguistics*, 1:179–192.

MacCartney, B. and Manning, C. D. (2008). Modeling semantic containment and exclusion in natural language inference. In *Proceedings of COLING*, pages 521–528. Association for Computational Linguistics.

Marelli, M., Menini, S., Baroni, M., Bentivogli, L., Bernardi, R., and Zamparelli, R. (2014). A SICK cure for the evaluation of compositional distributional semantic models.

Martínez-Gómez, P., Mineshima, K., Miyao, Y., and Bekki, D. (2017). On-demand injection of lexical knowledge for recognising textual entailment. In *Proceedings of the 15th Conference of the European Chapter of the ACL*, pages 710–720.

Mineshima, K., Martínez-Gómez, P., Miyao, Y., and Bekki, D. (2015). Higher-order logical inference with compositional semantics. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2055–2061.

Tian, R., Miyao, Y., and Matsuzaki, T. (2014). Logical inference on dependency-based compositional semantics. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 79–89.

Yanaka, H., Mineshima, K., Martínez-Gómez, P., and Bekki, D. (2018). Acquisition of phrase correspondences using natural deduction proofs. In *Proceedings of the 2018 Conference of the North American Chapter of the ACL: Human Language Technologies, (NAACL-HLT)*, pages 756–766, New Orleans, LA.

Yin, W. and Schütze, H. (2017). Task-specific attentive pooling of phrase alignments contributes to sentence matching. In

Proceedings of the 15th Conference of the European Chapter of the ACL, pages 699–709.