Teaching a Neural Network to Reason with Implicatives

Lauri Karttunen and Ignacio Cases
Stanford

CLASP, University of Gothenburg
May 29, 2019

An introduction to

Recursive Routing Networks: Learning to Compose Modules for Language Understanding

Ignacio Cases, Clemens Rosenbaum, Matthew Riemer, Atticus Geiger, Tim Klinger, Alex Tamkin, Olivia Li, Sandhini Agarwal, Joshua D. Greene, Dan Jurafsky, Christopher Potts, Lauri Karttunen

NAACL 2019, Minneapolis

Stanford Corpus of Implicatives

the order compelled him to appear as a witness	**entails**	he appeared as a witness
we have missed an opportunity to examine the art market today	**contradicts**	we have examined the art market today
Mr Odinga had not been forced to change his plans	**permits**	Mr Odinga had changed his plans

Table 2: Examples from SCI randomly chosen from the validation set. Each row contains a triplet formed by a premise (left column), a hypothesis (right column), and a label specifying one of the three possible relations (*entails*, *contradicts*, *permits*) holding between premise and hypothesis. The last row contains an example of a probabilistic implicative (see the main text).

92 constructions, 9671 triplets in the NAACL 2019 submission

Topics

What are implicatives?

Relations: entails, contradicts, permits (= neither entails nor contradicts)

Entailment vs. presupposition. Strawson entailment.

What are signatures?

Nested implicatives.

Two-way implicatives.

One-way implicatives and Invited Inferences

MacCartney relations: COVER

Phrasal implicatives

Recursive Routing Network model

Desiderata

Implicative: manage

Joan managed to solve the problem	Matrix clause
entails	Relation
Joan solved the problem	Complement clause
contradicts	
Joan did not solve the problem	
permits	= neither entails nor contradicts
The problem was not about mathematics	

Implicative: fail

Joan failed to solve the problem	Matrix clause
entails	Relation
Joan did not solve the problem	Complement clause
contradicts	
Joan solved the problem	
permits	= neither entails nor contradicts
The problem was about mathematics.	

No “Strawson” entailment 1

Joan solved the problem

does not entail

Joan managed to solve the problem

because *manage* has a presupposition:

It was difficult for Joan to solve the problem

The entailment only goes in one direction. The sentences in blue are not equivalent for us (as they are in MacCartney’s 2009 *NatLog* system).

A entails B just in case A satisfies all the presuppositions of B. That is what von Fintel 1999 calls “Strawson entailment.” We adopt this notion of entailment.

No “Strawson” entailment 2

Joan did not solve the problem

does not entail

Joan failed to solve the problem

because *fail* has a presupposition:

Joan tried to solve the problem or was expected to solve it

The entailment only goes in one direction. The sentences in blue are not equivalent (as they are in MacCartney’s 2009 *NatLog* system).

Signatures: pos|neg

The pos sign indicates the semantic relation of the matrix sentence to its complement in affirmative environments, the neg sign pertains to negative environments.

+ indicates entailment, − indicates contradiction, o stands for permits

manage is +|−

fail is −|+

promise is o|os

Nested implicatives

Implicatives can be nested:

Joan failed to manage to solve the problem

Joan managed to fail to solve the problem

both entail

Joan did not solve problem

but they have different presuppositions

What is the difference?

fail to manage vs. manage to fail

Theresa May failed to manage to deliver Brexit.

Tre Kronor managed to fail to win over the Finnish Lions.

Composition of signatures

manage o fail = manage to fail

+|- -|+ -|+

fail o manage = fail to manage

-|+ +|- -|+

promise o manage = promise to manage

o|o +|- o | o

manage o promise = manage to promise

+|- o|o o|o

fail o promise = fail to promise

-|+ o|o o|o

Composition of signatures

sig_1 o sig_2 = sig_2 if sig_1 is +|?

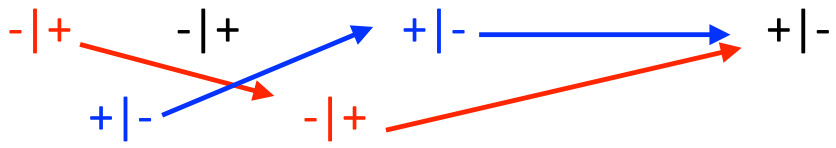
sig_1 o sig_2 = reverse(sig_2) if sig1 is -|?

sig_1 o sig_2 = o|o if sig1 is o|?

Composition of signatures is associative:

(sig_1 o sig_2) o sig_3 = sig_1 o (sig_2 o sig_3)

not o fail o manage = not fail to manage



The blue and the red path lead to the same result.

Two-way Implicatives

Two-way implicatives yield an entailment under both positive and negative polarity.

+|− verbs

manage, bother, dare, deign, remember (to), happen, turn out

−|+ verbs

fail, neglect, refuse, forget (to)

remember and *forget* are implicatives with *to* complementizer but factives with *that* as their complementizer.

I remembered/forgot that I locked the door

I didn't remember/forget that I locked the door

presuppose that I locked the door

It's not about *to* vs. *that*

turn out that and *turn out to*
are both implicative

It did not turn out that it was what I wanted to do.
It did not turn out to be what I wanted to do.

entail It was not what I wanted to do.

be bad that and *be bad to*
are both factive

It wasn't bad that we had one day of rain on our trip.
It wasn't bad to have one day of rain on our trip.

presuppose We had one day of rain or our trip.

One-way implicatives

There are four types of implicatives that yield an entailment only under one polarity. The entailments of able and force are polarity-preserving, refuse and hesitate reverse the polarity.

Ann was not able to speak.	⊏	Ann didn't speak.	o|−
Ann was forced to speak.	⊏	Ann spoke.	+|o
Ann refused to speak.	⊏	Ann didn't speak.	−|o
Ann didn't hesitate to speak.	⊏	Ann spoke.	o|+

With the other polarity there is no entailment but there may be a suggestion, an invited inference.

Invited inferences

In a neutral context where it has not been already mentioned or otherwise known what actually happened, all of the one-way implicatives are pushed towards being two-way implicatives unless the author explicitly indicates otherwise.

Ann was able to speak.	↝	Ann spoke.	(+)|−
Ann was not forced to speak.	↝	Ann didn't speak.	+|(−)
Ann did not refuse to speak.	↝	Ann spoke.	−|(+)
Ann hesitated to speak.	↝	Ann didn't speak.	(−)|+

This is a systematic effect although the strength of the invitation varies from one lexical item to another: very strong on able, weak on hesitate.

To explain this effect it is useful to look at MacCartney's NatLog system.

MacCartney Relations

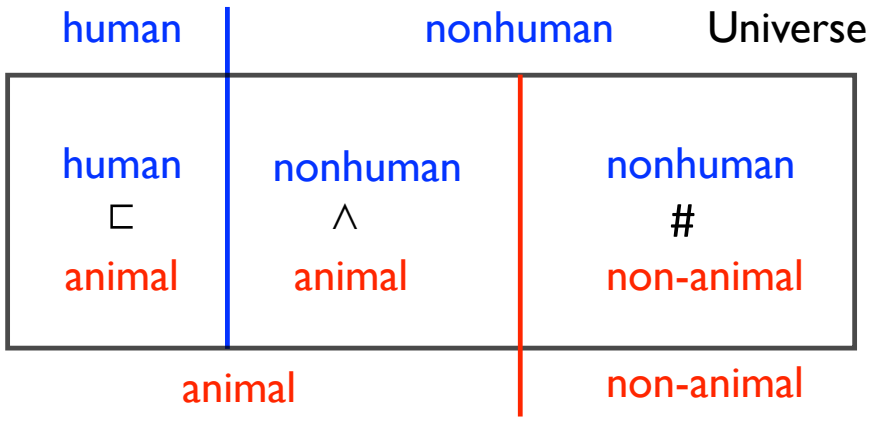
≡	equivalence	couch ≡ sofa
⊏	forward entailment	crow ⊏ bird
⊐	reverse entailment	European ⊐ French
^	negation	human ^ nonhuman
|	alternation	cat | dog
⌣	cover	animal ⌣ nonhuman
#	independence	hungry # hippo

Except for cover the relations are familiar. As MacCartney himself says, it is not immediately obvious what that relation could be useful for.

animal ⌣ nonhuman

$X \smile Y \equiv X \cap Y \neq \varnothing \wedge X \cup Y = U$

where U is the Universe of Discourse

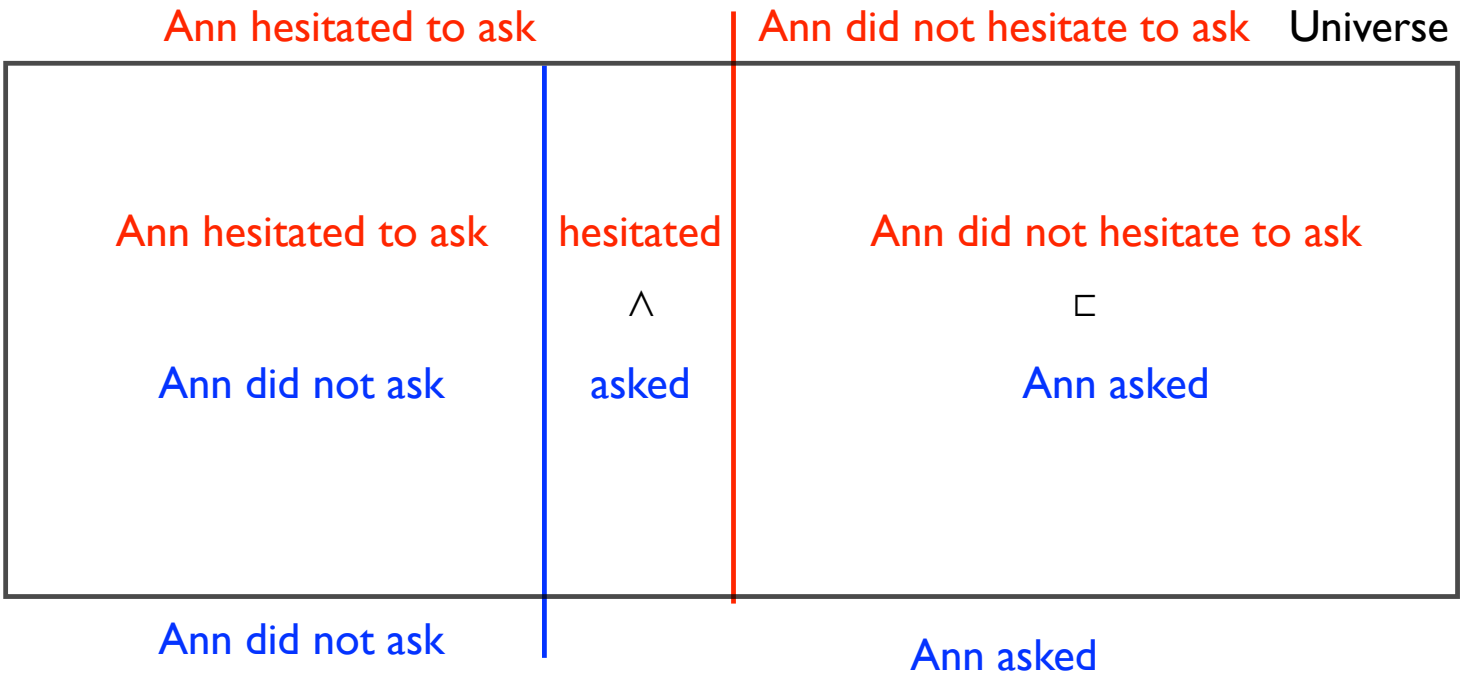


You are an animal!

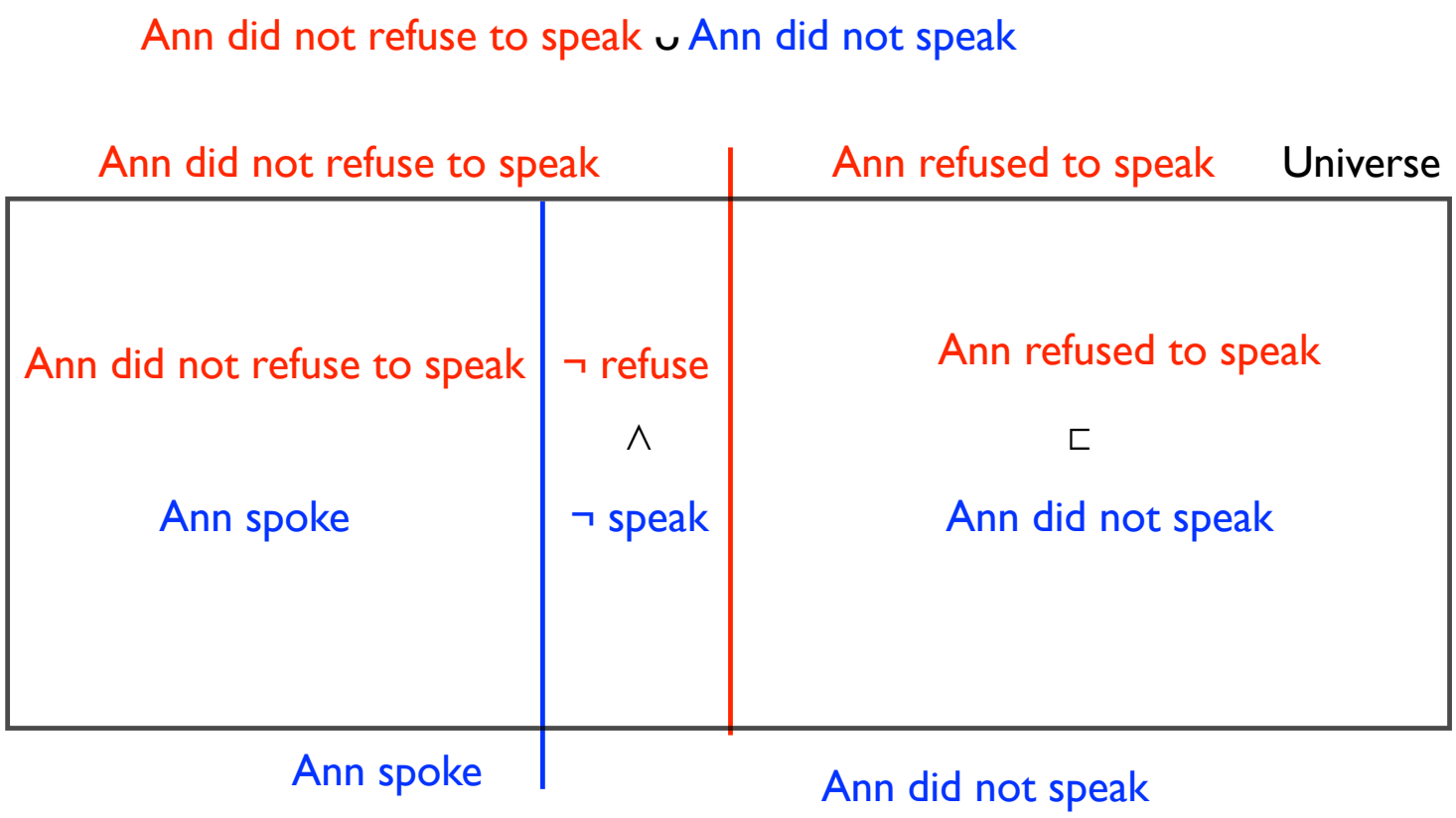
Why is this an insult?

One way implicatives: ○ | +

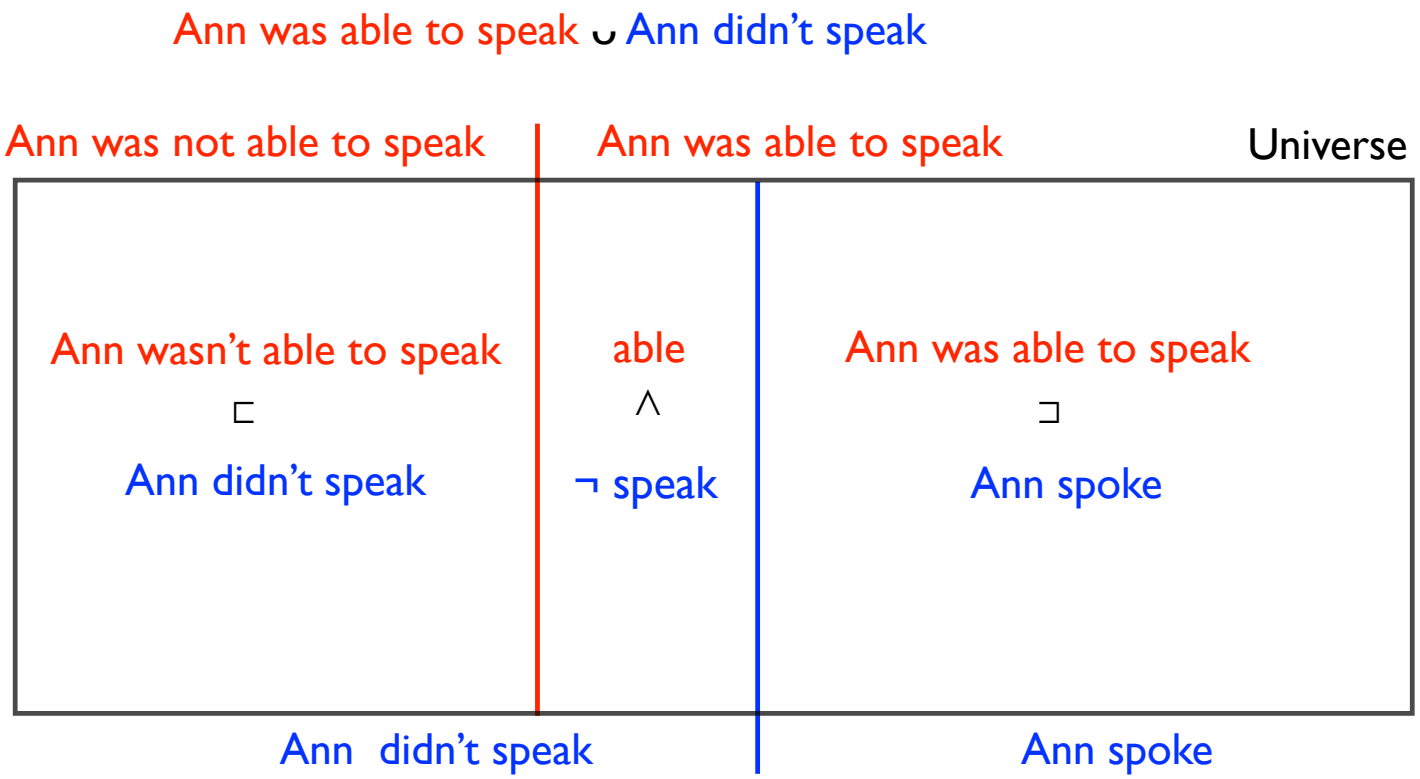
Ann hesitated to ask ∪ Ann asked



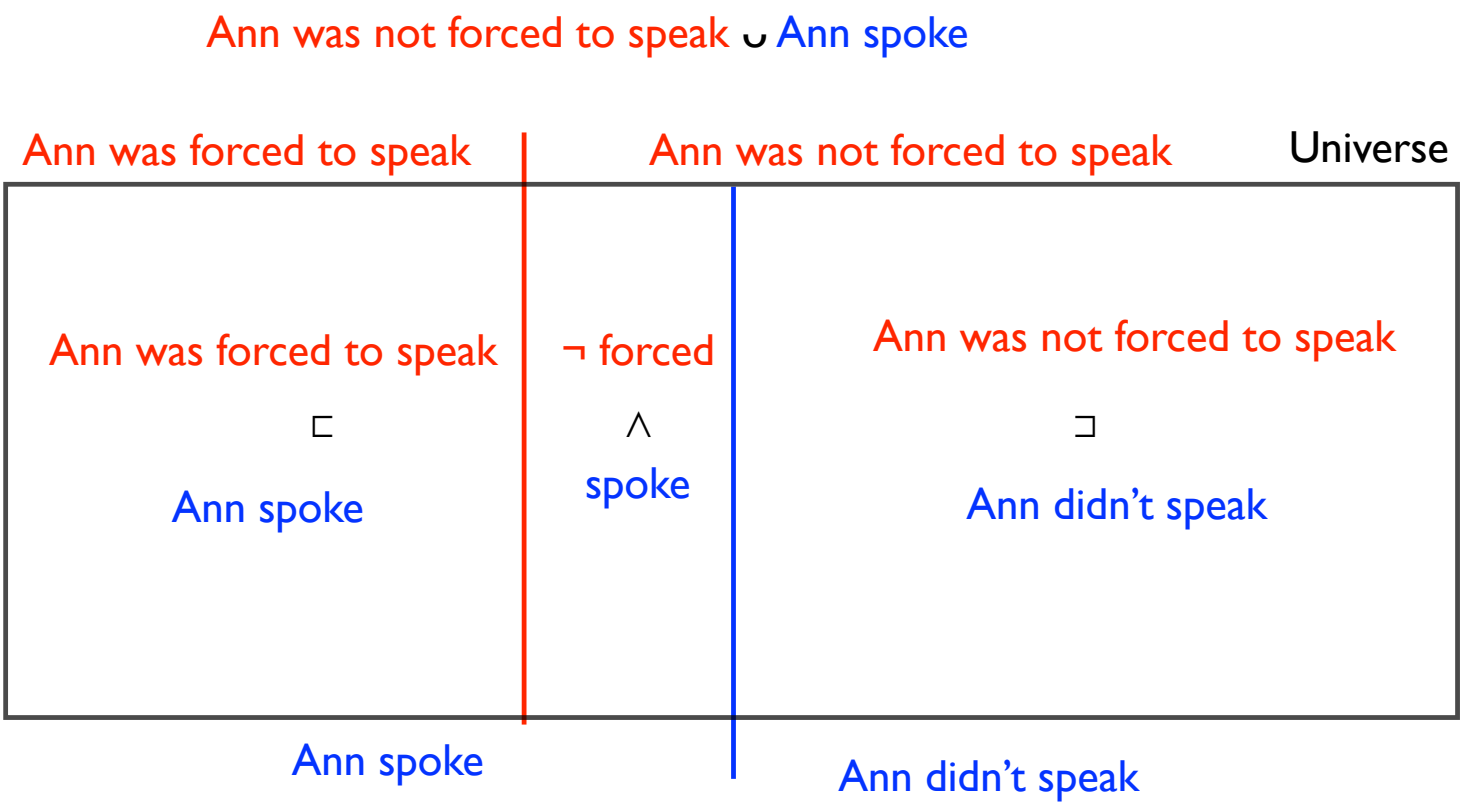
One way implicatives: –|o



One way implicatives: ○|–



One way implicatives: +|o



Explaining the invited inferences

As shown in the previous diagrams all the one-way implicatives are in a cover relation:

Ann was able to speak.	⌣	Ann did not speak.
Ann did not refuse to speak.	⌣	Ann did not speak.
Ann was not forced to speak.	⌣	Ann spoke.
Ann hesitated to speak.	⌣	Ann spoke.

In all cases the invited inference is the negation of the corresponding statement in the right, from the intersection of the two propositions.

Probabilistic Signatures

The likelihood of invited inferences varies greatly depending on the construction. They are very likely with *be able*, less likely with *prevent* and very unlikely with *hesitate*. To better match the actual usage we assign to most one-way implicatives a probabilistic signature. For example,

be able	.9|-
prevent	-|.7

Where .9|- instead of the logically correct o|- for *be able* means that in 90% of cases, we estimate that the author intends to communicate that not only was the protagonist able to something but that he actually did it. Similarly for *prevent*, we estimate that in 70% of cases the author wishes to convey that when the protagonist was not prevented from doing something she did it.

We did not give a probabilistic signature to the o|+ *hesitate*.

Contextual clues

In the case of probabilistic implicatives it is often possible to infer what the author has in mind, the annotation should be done case by case by a human instead of a blind algorithm. For example,

Carolina was able to score only one touchdown.

⇝ Carolina scored one touchdown.

This time we were not forced to swim naked in the sea.

⇝ We did not swim naked in the sea this time.

We were not forced to attend these meetings, every student was free to leave any time.

⇝ We attended these meetings voluntarily.

Phrasal two-way implicatives

+ | –

have the courage, wisdom

Julie had the chutzpah to ask the meter maid for a quarter.

I didn't have the courage to tell her that I loved her.

meet an obligation

We clearly fulfilled the obligation to pass a balanced budget.

Strausser hasn't met his responsibility to make improvements.

take the effort, asset, opportunity

She took the trouble to iron all the clothes.

I just didn't take the time to care for myself.

+ | –

use an asset, opportunity

I used the money to buy shoes and food.

Randy didn't use the opportunity to toot his own horn.

waste an asset

I wasted the money to buy a game that I cannot play.

I'm glad I didn't waste 90 minutes to see this film.

– | +

waste an opportunity

Mr. Spitzer wasted the opportunity to drive a harder bargain.

She didn't waste the chance to smile back at him.

fail an obligation

The Avatar failed his duty to bring peace to a broken world.

Orlando didn't neglect his duty to escort the dead.

Phrasal one-way implicatives

−|○

lack opportunity

She lost the chance to qualify for the final.

○|−

have ability

The defendant had no ability to pay the fine.

make effort

I have made no effort to check the accuracy of this blog.

○|+

show hesitation

She did not have any hesitation to don the role of a seductress.

Fonseka displayed no reluctance to carry out his orders.

VERB FAMILY	NOUN FAMILY	IMPLICATIVE SIGNATURE
HAVE	ABILITY | OPPORTUNITY	○|–
HAVE	COURAGE|WISDOM	+|–
LACK	ABILITY | OPPORTUNITY	–|○
MAKE	EFFORT	○|–
MEET	OBLIGATION	+|–
FAIL	OBLIGATION	–|+
SHOW	HESITATION	○|+
TAKE	ASSET | EFFORT	+|–
USE	ASSET | OPPORTUNITY	+|–
WASTE	ASSET	+|–
WASTE	OPPORTUNITY	–|+

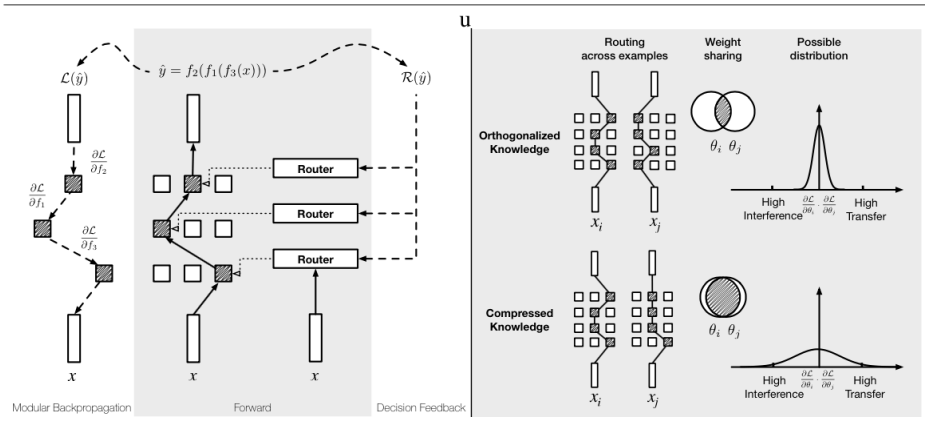
Verb families

FAIL	fail, neglect
HAVE	get, have, possess
LACK	discard, give up, lack, lose, miss, throw away
MAKE	do, make, undertake
MEET	acquit, do, fulfill, meet, perform (OBLIGATION)
SHOW	have, show, display
TAKE	grab, seize, snap, snatch, take
USE	expend, exploit, use, utilize
WASTE	drop, squander, waste

Noun families

ABILITY	ability, means, oomph, power
ASSET	asset, fortune, money, time
COURAGE	audacity, chutpah, courage, endurance, guts, impudence, nerve, stamina
EFFORT	attempt, effort, initiative, trouble
HESITATION	hesitation, qualms, reluctance, scrupels
OBLIGATION	duty, mission, obligation
OPPORTUNITY	chance, opportunity, occasion
WISDOM	expend, exploit, use, utilize

Recursive Routing Network



The model consists of multiple neural networks, a 3x3 collection in our case, and a router that picks a path of networks depending on the type of input it receives. This avoids the “catastrofic interference” we observed in an earlier, non-routed version on *take vow* o|o, *take chance* o|−, and *take time* +|−.

Desiderata I

1. Basics. The model should have a good performance on sentences containing implicative constructions it has seen in training, including sentences that are longer than the training examples, and containing lexical items not seen in training.

2. Generalization. The model should able to generalize in the way people do. For example, the SCI corpus includes the constructions *waste opportunity* and *waste chance*. If we remove all *waste chance* examples from training, will it get the right result when it encounters examples with *waste chance*?

3. Composition of signatures. The SCI corpus contains examples of nested implicatives. If we test the model with a nested construction it has seen before but with a different premise and hypothesis, will it give a correct response? Will it interpret correctly nested constructions it has not seen before composed of previously seen implicatives?

Desiderata II

4. Negation. Every statement contradicts its negation. Will the model learn this without specific training data?

5. **Symmetry of contradiction**. Whenever A contradicts B, B contradicts A as well, provided that A does not have presuppositions that B does not have. (Entailment for us is "Strawson entailment".)

6. Reflexivity and transitivity of entailment. Every statement entails itself. We do not have any training data of the type A entails A. Will the model discover that on its own? The semantics of nested implicatives is a special case of transitivity of entailment. If we break the nesting and construct pairs of triplets for testing the form A entails B, B entails C, will the model correctly conclude that A entails C?

7. Distant contradiction. If A entails C and B entails not C, will the model be able to conclude that A and B contradict each other?

References

Bowman, Samuel R., Gabor Angeli, Christopher Potts & Christopher D. Manning. 2015. A large annotated corpus for learning natural language inference. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 632–642. Association for Computational Linguistics.

von Fintel, Kai. 1999. NPI Licensing, Strawson Entailment, and Context Dependency. *Journal of Semantics* 16(2). 97–148.

Geis, Michael L. & Arnold M. Zwicky. 1971. On invited inferences. *Linguistic Inquiry* 2(4). 561–566. http://www.jstor.org/stable/4177664.

Grice, H. P. 1975. Logic and conversation. In Peter Cole & Jerry L. Morgan (eds.), *Syntax and Semantics 3: Speech Acts*, 41–58. New York: Academic Press.

Grice, Paul H. 1989. *Studies in the Way of Words*. Cambridge, MA: Harvard University.

Horn, Laurence R. 1985. Metalinguistic negation and pragmatic ambiguity. *Language* 61(1). 121–174.

Karttunen, Lauri. 1971. Implicative verbs. *Language* 47. 340–358. http://dx.doi.org/10.2307/412084.

Karttunen, Lauri. 2012. Simple and phrasal implicatives. In **SEM 2012*, 124–131. Montréal, Canada:

MacCartney, Bill & Christopher D. Manning. 2009. An extended model of natural logic. In *The 8th* International Conference on Computational Semantics (IWCS-8, 140–156. Tilburg, Netherlands: University of Tilburg.

Nairn, Rowan, Lauri Karttunen & Cleo Condoravdi. 2006. Computing relative polarity for textual inference. In Johan Bos & Alexander Koller (eds.), *Inference in Computational Semantics (ICoS-5)*, 67–76. Manchester, UK: University of Manchester. http://www.aclweb.org/anthology/ W06-39.

Williams, Adina, Nikita Nangia & Samuel Bowman. 2018. A broad-coverage challenge corpus for sentence understanding through inference. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, 1112–1122.