A Natural Proof System for Natural Language

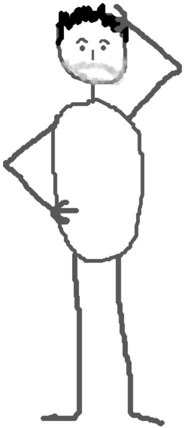
Lecture 1: Natural Language Inference & Tableau Method

Lasha Abzianidze

university of groningen

TbiLLAI 2019 in Tbilisi, Georgia

About me



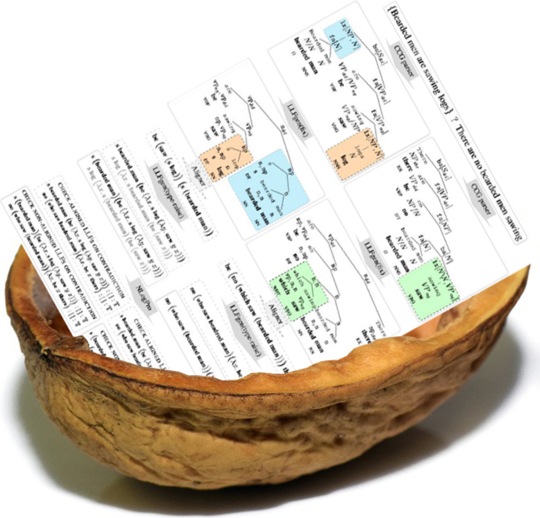
About the course

The following research questions are at the heart of the course:

- How are the meanings of natural language sentences related to each other?
- How to systematically reason with natural language sentences?
- How to get an explainable reasoning system?
- Can we use expressive but at the same time *friendly* meaning representations?

There will be a lot of *natural* trees

Course in a nutshell



Topics per day

Mon Natural Language Inference & Tableau Method

Tue 🛏

Wed 🛏

Thu 🚌

Fri Natural Tableau System

Sat Natural Language Inference with Natural Theorem Prover

Course web page: naturallogic.pro/Teaching/tbillai19/

Natural Language Understanding

Natural language understanding is one of the main problems of Artificial Intelligence and Natural Language Processing (NLP).

How to test whether a machine/program understands a natural language?

estimate semantic competence in a natural language

Given two sentences $S1$ and $S2$, detect whether $S1$ entails $S2$.

$S1$: A brown cat is lying on a yellow mat

$S2$: There is an animal on a yellow-colored mat

Answer: Yes

$S1$: The cat is lying on a mat

$S2$: The cat is rolling on a yellow mat

Answer: No

Recognizing Textual Entailment (2005-2013)

The task of Recognizing Textual Entailment (RTE) was introduced by [Dagan et al., 2005]:

- Textual entailment is defined as a *directional relationship* between pairs of text expressions, denoted by T (the entailing "Text") and H (the entailed "Hypothesis"). We say that T entails H if humans reading T would typically infer that H is most likely true.
- An RTE task: given two texts, T (text) and H (hypothesis), detect textual entailment from T to H.

RTE problems

RTE2 GOLD: `non-entailment`

Drew Walker, NHS Tayside's public health director, said: "It is important to stress that this is not a confirmed case of rabies."

A case of rabies was confirmed

RTE2 GOLD: `entailment`

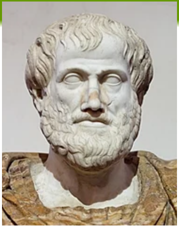
About two weeks before the trial started, I was in Shapiro's office in Century City

---

Shapiro works in Century City

Taken from [Dagan et al., 2013]

Long before RTE: Aristotle's syllogisms



Aristotle's syllogisms (4th century BC):

- 256 RTE problems
- Text consists of two sentences
- 24(!) of the problems are entailment

OAO-3 GOLD: `entailment`

Some cats have no tails
All cats are mammals

Some mammals have no tails

IAE-3 GOLD: `contradiction`

Some dogs have spots
All dogs are mammals

No mammals have spots

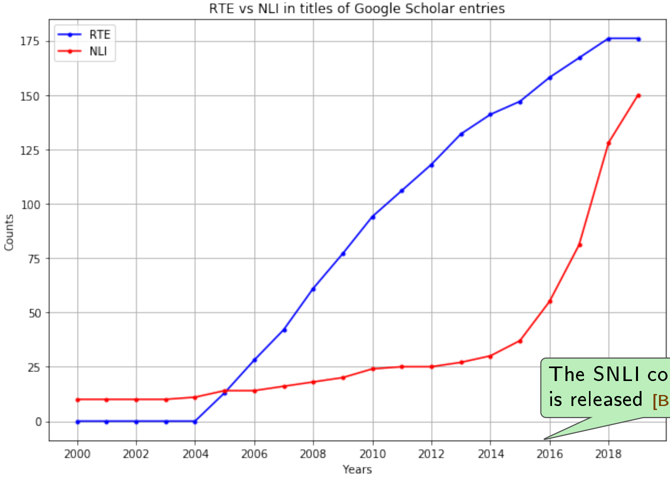
IAA-1 GOLD: `neutral`

Some vehicles are electric
All cars are vehicles

All cars are electric

After RTE: Natural Language Inference

Natural Language Inference is a recent term for Recognizing Textual Entailment.



The SNLI corpus for training ANNs is released [Bowman et al., 2015]

Modern NLI

The Stanford NLI (SNLI) corpus [Bowman et al., 2015]:

- Large corpus: 570K premise-hypothesis pairs
- Tackling entity & event co-reference by grounding in images
- Premises are image captions and hypotheses are generated by crowd workers
- Premise-hypothesis pairs are annotated by 5 crowd workers with 3 labels

Examples from SNLI*

SNLI-3581451227.jpg#4r1c GOLD: contradiction[1c]

A little girl and boy after a wedding in a field

the sail boat sank in the ocean

SNLI-475816542.jpg#2r1c GOLD: contradiction[3c2n]

A black and a brown dog are running toward the camera.

A black and a brown cat are running toward the camera.

SNLI-4837051771.jpg#2r1n GOLD: neutral[2e3n]

A small ice cream stand with two people standing near it.

Two people in line to buy icecream.

*Arguable examples

Critical look at SNLI

In the test part of SNLI, 12.2% of problems get 2vs3 annotations.

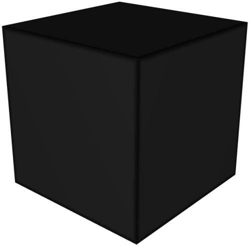
Annotation artifacts inflate systems' performance [Poliak et al., 2018, Gururangan et al., 2018]:

- Hypothesis only baselines score strikingly high wrt the majority class baseline: 69.2% vs 33.8% accuracy
- *animal*, *outdoors*, and *person* often in `entailment` hypotheses
- *tall*, *sad*, and *first* often in `neutral` hypotheses
- *cat*, *sleeping*, and *no* often in `contradiction` hypotheses

Smart black boxes

Author performance on test-SNLI (91.4%) is already suppressed by a deep neural network-based system (91.6%) [Liu et al., 2019]

☞SNLI leaderboard



What does this mean?

Specially dedicated workshops:
BlackboxNLP 2018 and 2019

Harder challenges aka *Task-Independent Sentence Understanding*:

☞GLUE leaderboard

☞SuperGLUE leaderboard

Explainable reasoning

Explainable reasoning is a feature associated with a white box systems:

- Explain entailment by providing some sort of proof or argument
- Explain contradiction by highlighting the incompatible cases
- Explain neutral relation by providing counterexamples for entailment and contradiction

e-SNLI – NLI with natural language explanations [Camburu et al., 2018]:

- Crowd workers gave explanations for labelled problems;
- Relevant words were highlighted;
- Systems need to predict a label and an explanation;

Examples from e-SNLI

SNLI-3581451227.jpg#4r1c GOLD: contradiction[1c]

A little girl and boy after a wedding in a field

the sail boat sank in the ocean

A girl and boy are people, not a thing, as a sail boat is. You cannot be in a field and in the ocean at the same time

SNLI-475816542.jpg#2r1c GOLD: contradiction[3c2e]

A black and a brown dog are running toward the camera.

A black and a brown cat are running toward the camera.

They refer to a dog, not a cat

The animal is either a cat or a dog

A dog cannot be a cat

Examples from e-SNLI (II)

SNLI-4837051771.jpg#2r1n   GOLD: neutral[2e3n]

A small ice cream stand with two people standing near it.

Two people in line to buy icecream.

Being near a stand doesn't mean you have to buy anything

Just because two people are standing near an ice cream stand, doesn't mean they are in line to buy ice cream

People who are standing near an ice cream stand are not always in line to buy ice cream

Shortcomings of NLI systems

- Few NLI systems are able to reason over multiple premises.
- Most NLI systems do not use logic-based reasoning: poor at processing Booleans (e.g., *or, not*) and quantifiers (e.g., *every, no*).

  P1: Most boxers have been knocked out
  P2: All boxers are athletes
  P3: All athletes who has been knocked out has a broken nose
  C: Most boxers have a broken nose

- SOTA NLI systems are not explanatory (though pretty good).
- Most RTE systems can be fooled easily (i.e. not having high precision)

  SICK-1745 GOLD:: `neutral`
  T: A man is pushing the buttons of a microwave
  H: A man is being pushed toward the buttons of a microwave

Shortcomings of logic-based NLI systems

- Their logic is often not expressive enough to model some aspects of linguistics semantics: higher-order terms like generalized quantifiers (e.g., *few*, *most*) and subsective modifiers e.g. *competent*, *slowly*.
- Translation of linguistic semantics into formal logic is usually a complex and immense problem, e.g., NL text into first-order logic [Bos, 2008].
- After the translation, information about constituency and syntax is not available in a formal language while the information is often crucial for *shallow* reasoning, e.g., monotonicity reasoning.

  SICK-8145 GOLD: entailment

  T: A woman in blue has a yellow ball in the mitt

  H: A woman in blue has a yellow ball in the hand

How logic can be natural?

Natural logic is a hypothetical logic which is built in natural language and represents its integral part.

It is a theory about "the regularities governing the notion of a valid argument for reasoning in natural language" [Lakoff, 1970].

"Natural logic is a somewhat loose [...] term for [...] attempts [...] at describing basic patterns of human reasoning directly in natural language without the intermediate of some formal system" [van Benthem, 2008].

Natural logic is "the study of inference in natural language, done as close as possible to the surface forms" [Moss, 2010b].

Monotonicity reasoning

The most popular and success story of natural logic is monotonicity reasoning.

Monotonicitity reasoning is about replacing phrases in a premise in such a way that the obtained sentences are entailment of the premise.

GOLD: `entailment`

P: Every man who consumed alcohol devoured most snacks

H: Every young man who drank beer ate some snacks

GOLD: `entailment`

P: $3 \times [s_3(x) = x + 3](2) \;\leq\; [p_3(x) = x^3]\big([m_4(x) = x \pmod 4](7)\big)$

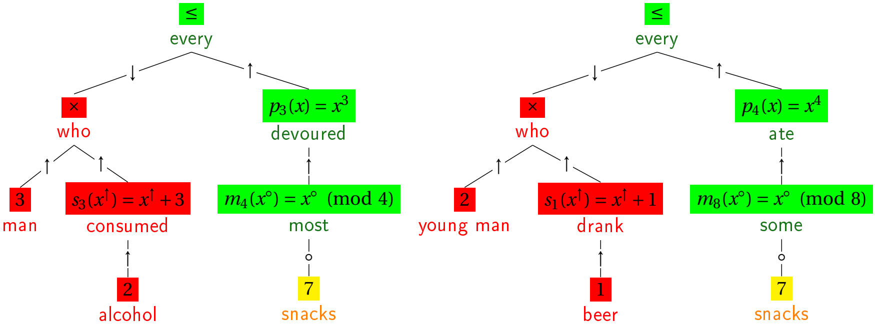
H: $2 \times [s_1(x) = x + 1](1) \;\leq\; [p_4(x) = x^4]\big([m_8(x) = x \pmod 8](7)\big)$

Monotonicity reasoning in action

GOLD: `entailment`

P: $3 \times [s_3(x) = x+3](2) \;\leq\; [p_3(x) = x^3]\big([m_4(x) = x \pmod 4](7)\big)$

H: $2 \times [s_1(x) = x+1](1) \;\leq\; [p_4(x) = x^4]\big([m_8(x) = x \pmod 8](7)\big)$



GOLD: `entailment`

P: Every man who consumed alcohol devoured most snacks

H: Every young man who drank beer ate some snacks

Related work

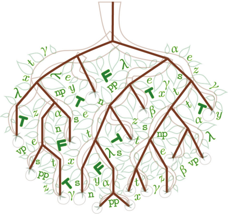
Other works on monotonicity reasoning and natural logic:

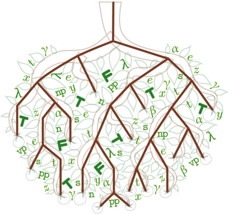
- First study of monotonicity reasoning as a formal calculus [Van Benthem, 1986, van Benthem, 1987, Sánchez-Valencia, 1991]
- Moving from syllogistic logics towards natural logic [Moss, 2010a]
- A tableau proof system for a fragment of natural logic [Muskens, 2010]
- Formal system for extended monotonicity reasoning [MacCartney and Manning, 2008, Icard, 2012, Icard and Moss, 2014]

Working systems:

- Monotonicity-based inference system for a fragment of English, operating on categorical grammar derivation trees [Fyodorov et al., 2003, Zamansky et al., 2006]
- Implementation of syllogistic logic with monotonicity [Eijck, 2005]
- Two implementations of extended syllogistic logics [Hemann et al., 2015]
- Natural language inference using polarity-marked parse trees [Hu et al., 2019]

Semantic Tableau Method





Logic & proof systems

Logic consists of four components:

- Intuitive non-formal motivation
- Syntax of formulas: well-formed formulas vs ill-formed ones
- Semantics associated with the formulas
- Some type of proof calculus

A proof calculus/system:

- employed to systematically capture valid formulas and arguments
- is a syntactic game: there are legal and illegal moves
- comes in several flavours
- is usually a sound and complete

Semantic tableau method

A semantic tableau method [Beth, 1955, Hintikka, 1955] is a proof procedure for formal logics that checks formulas with truth constraints:

Input: A set of signed formulas
$P_1 : \mathbb{T}, \ldots, P_m : \mathbb{T},\ Q_1 : \mathbb{F}, \ldots, Q_n : \mathbb{F}$

Output: some or no model satisfying the truth constraints on the formulas
☞ A model search problem

Prove or refute

Whenever it rains, the roof leaks

How to verify truth of this statement?

Show that:

- In every situation it is true
  Check every situation when it rains and show the roof leaking
- In some situation it is not true
  Find some situation when it rains and the roof isn't leaking

Approval route

Refutation route

Proving by failing to refute

A tableau method tries to refute statement in order to prove it:

1. Given $P_1, ..., P_m \vDash Q$ to prove
2. Try to refute $P_1, ..., P_m \vDash Q$
   1. Build the counterexample: $P_1 : \mathbb{T}, \ldots, P_m : \mathbb{T},\ Q : \mathbb{F}$
   2. Try to satisfy the counterexample
3. If refutation succeeded, $P_1, ..., P_m \vDash Q$ is disproved
4. Otherwise $P_1, ..., P_m \vDash Q$ is proved

Propositional tableau method (signed version)

Prove: $P \wedge Q \vDash Q \wedge \neg P$

Counterexample: $P \wedge Q : \mathbb{T},\ Q \wedge \neg P : \mathbb{F}$

Propositional tableau rules:

$\wedge_{\mathbb{T}}$: $X \wedge Y : \mathbb{T}$ ⟹ $X : \mathbb{T}$, $Y : \mathbb{T}$

$\wedge_{\mathbb{F}}$: $X \wedge Y : \mathbb{F}$ ⟹ $X : \mathbb{F}$ | $Y : \mathbb{F}$

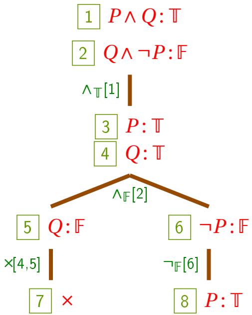
$\vee_{\mathbb{T}}$: $X \vee Y : \mathbb{T}$ ⟹ $X : \mathbb{T}$ | $Y : \mathbb{T}$

$\vee_{\mathbb{F}}$: $X \vee Y : \mathbb{F}$ ⟹ $X : \mathbb{F}$, $Y : \mathbb{F}$

$\neg_{\mathbb{F}}$: $\neg X : \mathbb{F}$ ⟹ $X : \mathbb{T}$

$\neg_{\mathbb{T}}$: $\neg X : \mathbb{T}$ ⟹ $X : \mathbb{F}$

$\times$: $X : \mathbb{T}$, $X : \mathbb{F}$ ⟹ $\times$



A situation supporting a counterexample: $P : \mathbb{T}, Q : \mathbb{T}$

Closed tableau

Prove: $\neg(P \wedge Q) \vDash \neg P \vee \neg Q$ Proved!

Counterexample: $\neg(P \wedge Q) : \mathbb{T},\ \neg P \vee \neg Q : \mathbb{F}$

Propositional tableau rules:

$\wedge_{\mathbb{T}}$
$X \wedge Y : \mathbb{T}$
$X : \mathbb{T}$ $Y : \mathbb{T}$

$\wedge_{\mathbb{F}}$	
$X \wedge Y : \mathbb{F}$	
$X : \mathbb{F}$	$Y : \mathbb{F}$

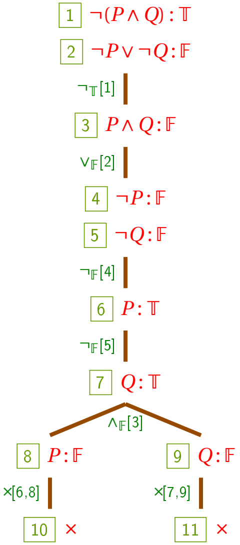
$\vee_{\mathbb{T}}$	
$X \vee Y : \mathbb{T}$	
$X : \mathbb{T}$	$Y : \mathbb{T}$

$\vee_{\mathbb{F}}$
$X \vee Y : \mathbb{F}$
$X : \mathbb{F}$ $Y : \mathbb{F}$

$\neg_{\mathbb{F}}$
$\neg X : \mathbb{F}$
$X : \mathbb{T}$

$\neg_{\mathbb{T}}$
$\neg X : \mathbb{T}$
$X : \mathbb{F}$

$\times$
$X : \mathbb{T}$ $X : \mathbb{F}$
$\times$



Different proof strategy

Prove: $\neg(P \wedge Q) \vDash \neg P \vee \neg Q$ Prover!

Counterexample: $\neg(P \wedge Q) : \mathbb{T}$, $\neg P \vee \neg Q : \mathbb{F}$

Propositional tableau rules:

$\wedge_{\mathbb{T}}$: $X \wedge Y : \mathbb{T}$ / $X : \mathbb{T}$, $Y : \mathbb{T}$

$\wedge_{\mathbb{F}}$: $X \wedge Y : \mathbb{F}$ / $X : \mathbb{F}$ | $Y : \mathbb{F}$

$\vee_{\mathbb{T}}$: $X \vee Y : \mathbb{T}$ / $X : \mathbb{T}$ | $Y : \mathbb{T}$

$\vee_{\mathbb{F}}$: $X \vee Y : \mathbb{F}$ / $X : \mathbb{F}$, $Y : \mathbb{F}$

$\neg_{\mathbb{F}}$: $\neg X : \mathbb{F}$ / $X : \mathbb{T}$

$\neg_{\mathbb{T}}$: $\neg X : \mathbb{T}$ / $X : \mathbb{F}$

$\times$: $X : \mathbb{T}$, $X : \mathbb{F}$ / $\times$

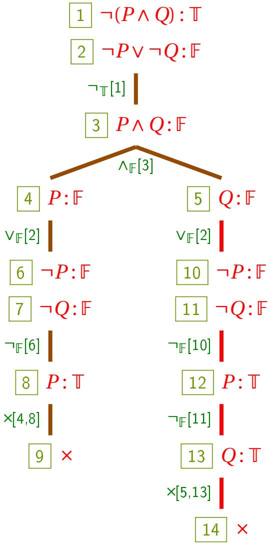


Tableau exercise

1 $P \vee (Q \wedge \neg R) : \mathbb{T}$

2 $(P \vee \neg R) \wedge (\neg Q \vee Q) : \mathbb{F}$

- 3 $P \vee \neg R : \mathbb{F}$
  - 5 $P : \mathbb{F}$
  - 6 $\neg R : \mathbb{F}$
    - 7 $P : \mathbb{T}$
      - 9 $\times$
    - 8 $Q \wedge \neg R : \mathbb{T}$
      - 10 $Q : \mathbb{T}$
      - 11 $\neg R : \mathbb{T}$
        - 12 $\times$
- 4 $\neg Q \vee Q : \mathbb{F}$
  - 13 $\neg Q : \mathbb{F}$
  - 14 $Q : \mathbb{F}$
    - 15 $Q : \mathbb{T}$
      - 16 $\times$

Quiz

1. If propositional formula ϕ is built up from n Boolean connectives, at most how many rule applications will be applicable to the tableau started with $\phi : \mathbb{T}$?
2. ...started with $\phi : \mathbb{F}$?
3. Can you think of tableau rules for $\rightarrow_{\mathbb{T}}$ and $\rightarrow_{\mathbb{F}}$?

Rules for quantifiers

Rules for $\exists$:

$\exists_{\mathbb{T}}$
$\exists x.\phi : \mathbb{T}$
$\phi[x/c] : \mathbb{T}$
c is fresh

$\exists^c_{\mathbb{F}}$
$\exists x.\phi : \mathbb{F}$
$\phi[x/c] : \mathbb{F}$
c is old

Rules for $\forall$:

$\forall_{\mathbb{F}}$
$\forall x.\phi : \mathbb{F}$
$\phi[x/c] : \mathbb{F}$
c is fresh

$\forall^c_{\mathbb{T}}$
$\forall x.\phi : \mathbb{T}$
$\phi[x/c] : \mathbb{T}$
c is old

⚠ Dangerous zone!

$\boxed{1}$ $\forall x.\exists y.\mathtt{love}(x,y) : \mathbb{T}$

$\boxed{2}$ $\forall z.\mathtt{love}(z,z) : \mathbb{F}$

$\forall_{\mathbb{F}}[2]$

$\boxed{3}$ $\mathtt{love}(c,c) : \mathbb{F}$

$\forall^c_{\mathbb{T}}[1]$

$\boxed{4}$ $\exists y.\mathtt{love}(c,y) : \mathbb{T}$

$\exists^c_{\mathbb{T}}[4]$

$\boxed{5}$ $\mathtt{love}(c,d) : \mathbb{T}$

$\forall^d_{\mathbb{T}}[1]$

$\boxed{6}$ $\exists y.\mathtt{love}(d,y) : \mathbb{T}$

$\vdots$

Non-empty domain

Rules for $\exists$:

$\exists_{\mathbb{T}}$
$\exists x.\phi : \mathbb{T}$
$\phi[x/c] : \mathbb{T}$
c is fresh

$\exists^{c}_{\mathbb{F}}$
$\exists x.\phi : \mathbb{F}$
$\phi[x/c] : \mathbb{F}$
c is old

Rules for $\forall$:

$\forall_{\mathbb{F}}$
$\forall x.\phi : \mathbb{F}$
$\phi[x/c] : \mathbb{F}$
c is fresh

$\forall^{c}_{\mathbb{T}}$
$\forall x.\phi : \mathbb{T}$
$\phi[x/c] : \mathbb{T}$
c is old

$\boxed{1}$ $\forall x.\big(\mathtt{run}(x) \wedge \neg\mathtt{run}(x)\big) : \mathbb{T}$

Non-empty domain constraint:
you can always have an entity

Conclusion: Natural Language Inference

The RTE/NLI task can be seen "as the best way of testing an NLP system's semantic capacity" [Cooper et al., 1996].

The NLI task is popular: many benchmarks and datasets

NLI systems comes with many flavours but we focus on logic-based ones

Monotonicity reasoning, the signature of natural logic

- Polarity marking
- String edit and word replacement reasoning

Conclusion: Tableau Method

- A semantic tableau method
  "today [it is] one of the most popular, since it appears to bring together the proof-theoretical and the semantical approaches to the presentation of a logical system and is also very intuitive. In many universities it is the style first taught to students." [D'Agostino et al., 1999].

- Propositional tableau system: when applying a rule to a tableau entry, remember to do so for each branch it sits on.

- Dangerous zone: First-order logic tableau might not terminate

References I

Beth, E. W. (1955). Semantic Entailment and Formal Derivability. *Koninklijke Nederlandse Akademie van Wentenschappen, Proceedings of the Section of Sciences*, 18:309–342.

Bos, J. (2008). Wide-coverage semantic analysis with boxer. In Bos, J. and Delmonte, R., editors, *Semantics in Text Processing. STEP 2008 Conference Proceedings*, Research in Computational Semantics, pages 277–286. College Publications.

Bowman, S. R., Angeli, G., Potts, C., and Manning, C. D. (2015). A large annotated corpus for learning natural language inference. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 632–642, Lisbon, Portugal. Association for Computational Linguistics.

Camburu, O.-M., Rocktäschel, T., Lukasiewicz, T., and Blunsom, P. (2018). e-snli: Natural language inference with natural language explanations. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R., editors, *Advances in Neural Information Processing Systems 31*, pages 9539–9549. Curran Associates, Inc.

Cooper, R., Crouch, D., Eijck, J. V., Fox, C., Genabith, J. V., Jaspars, J., Kamp, H., Milward, D., Pinkal, M., Poesio, M., Pulman, S., Briscoe, T., Maier, H., and Konrad, K. (1996). *FraCaS: A Framework for Computational Semantics*. Deliverable D16.

Dagan, I., Glickman, O., and Magnini, B. (2005). The pascal recognising textual entailment challenge. In *Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment*.

Dagan, I., Roth, D., Sammons, M., and Zanzotto, F. M. (2013). *Recognizing Textual Entailment: Models and Applications*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers.

References II

D'Agostino, M., Gabbay, D. M., Hähnle, R., and Posegga, J., editors (1999). *Handbook of Tableau Methods*. Springer.

Eijck, J. V. (2005). Syllogistics = monotonicity + symmetry + existential import.

Fyodorov, Y., Winter, Y., and Francez, N. (2003). Order-based inference in natural logic. *Logic Journal of the IGPL*, 11(4):385–416.

Gururangan, S., Swayamdipta, S., Levy, O., Schwartz, R., Bowman, S., and Smith, N. A. (2018). Annotation artifacts in natural language inference data. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 107–112, New Orleans, Louisiana. Association for Computational Linguistics.

Hemann, J., Swords, C., and Moss, L. (2015). Two advances in the implementations of extended syllogistic logics. In Balduccini1, M., Mileo, A., Ovchinnikova, E., Russo, A., and Schüuller, P., editors, *Joint Proceedings of the 2nd Workshop on Natural Language Processing and Automated Reasoning, and the 2nd International Workshop on Learning and Nonmonotonic Reasoning at LPNMR 2015*, pages 1–14.

Hintikka, J. (1955). *Two Papers on Symbolic Logic: Form and Content in Quantification Theory and Reductions in the Theory of Types*. Number 8 in Acta philosophica Fennica. Societas Philosophica.

Hu, H., Chen, Q., and Moss, L. (2019). Natural language inference with monotonicity. In *Proceedings of the 13th International Conference on Computational Semantics - Short Papers*, pages 8–15, Gothenburg, Sweden. Association for Computational Linguistics.

Icard, T. F. (2012). Inclusion and exclusion in natural language. *Studia Logica*, 100(4):705–725.

References III

Icard, T. F. and Moss, L. S. (2014). Recent progress on monotonicity. *Linguistic Issues in Language Technology*, 9.

Lakoff, G. (1970). Linguistics and natural logic. In Davidson, D. and Harman, G., editors, *Semantics of Natural Language*, volume 40 of *Synthese Library*, pages 545–665. Springer Netherlands.

Liu, X., He, P., Chen, W., and Gao, J. (2019). Multi-task deep neural networks for natural language understanding. *arXiv preprint arXiv:1901.11504*.

MacCartney, B. and Manning, C. D. (2008). Modeling semantic containment and exclusion in natural language inference. In Scott, D. and Uszkoreit, H., editors, *COLING*, pages 521–528.

Moss, L. S. (2010a). Logics for natural language inference. Expanded version of lecture notes from a course at ESSLLI 2010.

Moss, L. S. (2010b). Natural logic and semantics. In Aloni, M., Bastiaanse, H., de Jager, T., and Schulz, K., editors, *Logic, Language and Meaning: 17th Amsterdam Colloquium, Amsterdam, The Netherlands, December 16-18, 2009, Revised Selected Papers*, pages 84–93. Springer Berlin Heidelberg, Berlin, Heidelberg.

Muskens, R. (2010). An analytic tableau system for natural logic. In Aloni, M., Bastiaanse, H., de Jager, T., and Schulz, K., editors, *Logic, Language and Meaning*, volume 6042 of *Lecture Notes in Computer Science*, pages 104–113. Springer Berlin Heidelberg.

References IV

Poliak, A., Naradowsky, J., Haldar, A., Rudinger, R., and Van Durme, B. (2018). Hypothesis only baselines in natural language inference. In *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics*, pages 180–191, New Orleans, Louisiana. Association for Computational Linguistics.

Sánchez-Valencia, V. (1991). Categorial grammar and natural reasoning. ILTI Publication Series for Logic, Semantics, and Philosophy of Language LP-91-08, University of Amsterdam.

Van Benthem, J. (1986). *Essays in Logical Semantics*, volume 29 of *Studies in Linguistics and Philosophy*. Springer Netherlands.

van Benthem, J. (1987). Meaning: Interpretation and inference. *Synthese*, 73(3):451–470.

van Benthem, J. (2008). A brief history of natural logic. In *Technical Report PP-2008-05*. Institute for Logic, Language & Computation.

Zamansky, A., Francez, N., and Winter, Y. (2006). A 'natural logic' inference system using the lambek calculus. *Journal of Logic, Language and Information*, 15(3):273–295.