# Natural Language Reasoning with a Natural Theorem Prover

## Day 4: Natural Language Theorem Proving

Lasha Abzianidze

Utrecht University

33rd ESSLLI in Gaillimh, Éire
8-12 August 2022

# Where are we now

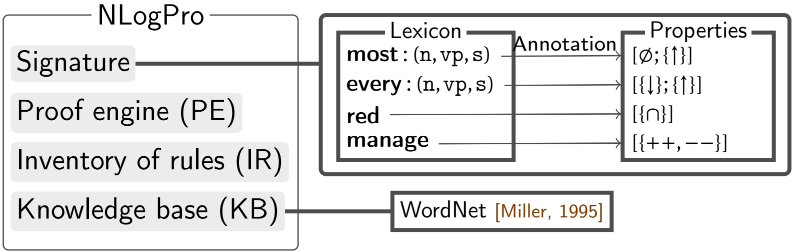
What we have done so far:

- Introduce Natural Tableau: a tableau system for natural logic, with more natural rules, with LLFs types with syntactic and semantci types
- Obtaining LLFs from CCG derivations of CCG parsers:simplifying, fixing and type-raising
- Rules that tackle erroneous PP-attachments (optional if the performance needs it)

What is today's plan:

- Describe a Natural Tableau-based theorem prover for natural language
- Describing the SICK and FraCaS NLI datasets
- Evaluation on FraCaS (on SICK will be tomorrow)
- Running the prover on google colab

# Natural logic theorem prover (NLogPro)



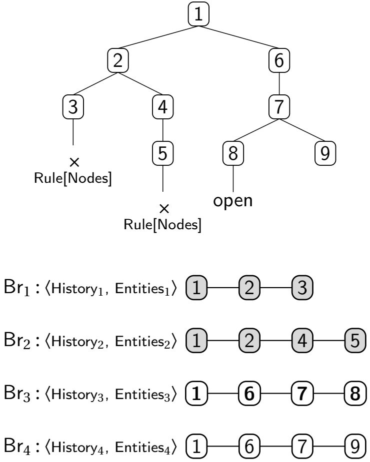

KB uses 4 relations from WordNet 3.0 (online version):

- derivation
- similarity
- hyponymy/hypernymy
- antonymy

⚠ No word sense disambiguation system is used.

# Two data structures

The proof engine builds both a tree and a list structures:



# Some derivable rules

Derivable rules are shortcuts for several rule applications.

$\exists^{n}_{\mathbb{F}}$

$$\frac{q_{\mathrm{n,vp,s}} N\, V : [\,] : \mathbb{F} \qquad N : [c_e] : \mathbb{T}}{V : [c] : \mathbb{F}}$$

$q \in \{\textbf{a}, \textbf{some}, \textbf{the}, \textbf{s}\}$

$\forall^{n}_{\mathbb{T}}$

$$\frac{q_{\mathrm{n,vp,s}} N\, V : [\,] : \mathbb{T} \qquad N : [c_e] : \mathbb{T}}{V : [c] : \mathbb{T}}$$

$q \in \{\textbf{every}, \textbf{the}\}$

$\mathrm{NO}^{n}_{\mathbb{T}}$

$$\frac{\textbf{no}_{\mathrm{n,vp,s}} N\, V : [\,] : \mathbb{T} \qquad N : [c_e] : \mathbb{T}}{V : [c] : \mathbb{F}}$$

$\exists^{v}_{\mathbb{F}}$

$$\frac{q_{\mathrm{n,vp,s}} N\, V : [\,] : \mathbb{F} \qquad V : [c_e] : \mathbb{T}}{N : [c] : \mathbb{F}}$$

$q \in \{\textbf{a}, \textbf{some}, \textbf{the}, \textbf{s}\}$

$\forall^{v}_{\mathbb{T}}$

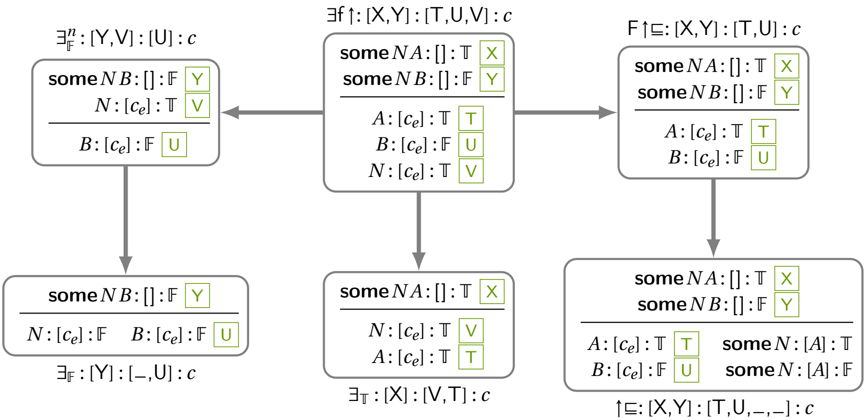
$$\frac{q_{\mathrm{n,vp,s}} N\, V : [\,] : \mathbb{T} \qquad V : [c_e] : \mathbb{F}}{N : [c] : \mathbb{F}}$$

$q \in \{\textbf{every}, \textbf{the}\}$

$\mathrm{NO}^{v}_{\mathbb{T}}$

$$\frac{\textbf{no}_{\mathrm{n,vp,s}} N\, V : [\,] : \mathbb{T} \qquad V : [c_e] : \mathbb{T}}{N : [c] : \mathbb{F}}$$

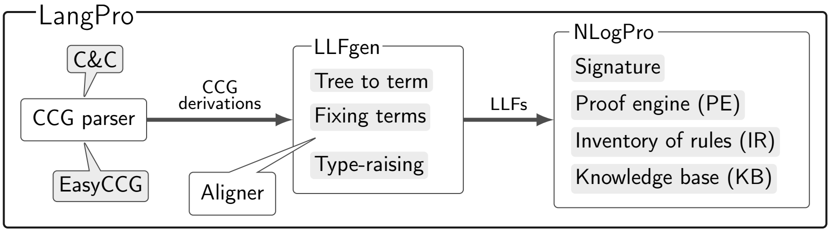
# Rule application subsumption




$$\exists^{n}_{\mathbb{F}} : [\mathsf{Y},\mathsf{V}] : [\mathsf{U}] : c \Rightarrow \exists_{\mathbb{F}} : [\mathsf{Y}] : [\_,\mathsf{U}] : c$$

# Natural language theorem prover (LangPro)

Chaining a CCG parser, the LLF generator and NLogPro results in a theorem prover for natural language.

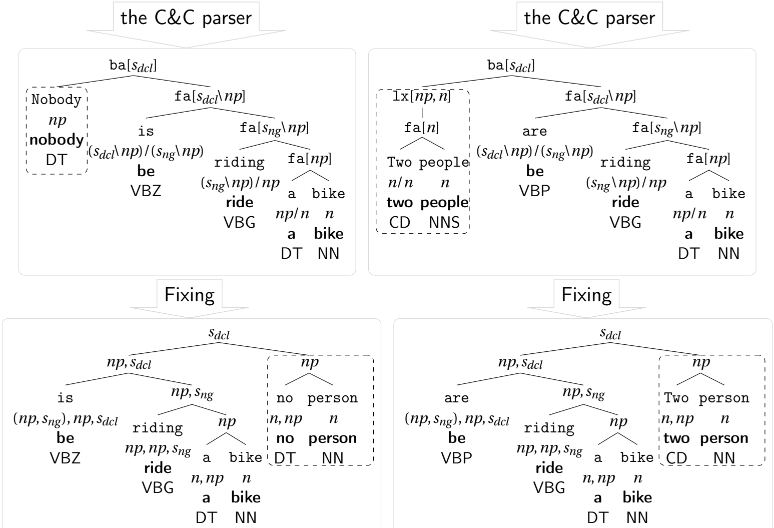



Online demo: http://naturallogic.pro/LangPro

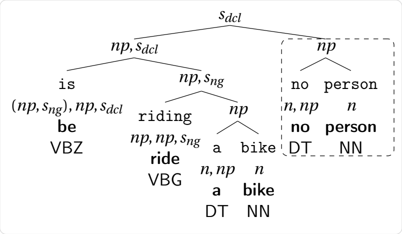
GitHub repo: https://github.com/kovvalsky/LangPro

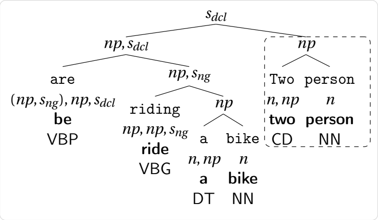
# LangPro in action

SICK-2865: `Nobody is riding a bike` $\overset{?}{\Rightarrow}$ `Two people are riding a bike`



# LangPro in action (2)




Type-raising

Type-raising

**no person** (**be** ($\lambda x.$ (**a bike**) ($\lambda y.$ **ride** $y$ $x$)))
a bike ($\lambda x.$ no person (be (ride $x$)))

**two person** (**be** ($\lambda x.$ (**a bike**) ($\lambda y.$ **ride** $y$ $x$)))
a bike ($\lambda x.$ two person (be (ride $x$)))

Proving by PE using IR & KB

intial nodes for entailment checking:
**no person** (**be** ($\lambda x.$ (**a bike**) ($\lambda y.$ **ride** $y$ $x$))) : [] : $\mathbb{T}$
**two person** (**be** ($\lambda x.$ (**a bike**) ($\lambda y.$ **ride** $y$ $x$))) : [] : $\mathbb{F}$

intial nodes for contradiction checking:
**no person** (**be** ($\lambda x.$ (**a bike**) ($\lambda y.$ **ride** $y$ $x$))) : [] : $\mathbb{T}$
**two person** (**be** ($\lambda x.$ (**a bike**) ($\lambda y.$ **ride** $y$ $x$))) : [] : $\mathbb{T}$

# LangPro in action (3)

1 **no person** (**be**($\lambda x.$ (**a bike**) ($\lambda y.$ **ride** $y\ x$))): $[\,] : \mathbb{T}$

2 **two person** (**be** ($\lambda x.$ (**a bike**) ($\lambda y.$ **ride** $y\ x$))): $[\,] : \mathbb{T}$

$\exists_{\mathbb{T}}[2]$

3 **person**: $[c] : \mathbb{T}$

4 **be**($\lambda x.$ (**a bike**) ($\lambda y.$ **ride** $y\ x$)): $[c] : \mathbb{T}$

$\mathrm{no}^{n}_{\mathbb{T}}[1,4]$

5 **person**: $[c] : \mathbb{F}$

6 $\times$

$$\frac{\textbf{no}\ A\ B : [\,] : \mathbb{T} \quad A : [c] : \mathbb{T}}{B : [c] : \mathbb{F}}\ \mathrm{no}^{n}_{\mathbb{T}}$$

$$\frac{N^{\mathrm{CD}}\ A\ B : [\,] : \mathbb{T}}{A : [c] : \mathbb{T} \quad B : [c] : \mathbb{T}}\ \exists_{\mathbb{T}}$$

# The SICK dataset

SICK [Marelli et al., 2014b] contains Sentences Involving Compositional Knowledge:

- 10K Text-Hypothesis pairs generated semi-automatically and annotated by humans with three labels: E, C, & N.
- Contains no encyclopedic knowledge, no named entities, relatively small vocabulary, less multiword expressions and no lengthy sentences ($\approx 9$ words per sentence).
- Contradictions (86%) rely too much on negative words and antonyms [Lai and Hockenmaier, 2014].
- A benchmark for the SemEval-14 RTE task [Marelli et al., 2014a]: Trial (5%), Train (45%), and test (50%).
- 84% of crowd workers' labels match the majority, i.e, gold labels.

# SICK construction

| | |
|---|---|
| **Original pair** | |
| **S0a:** *A sea turtle is hunting for fish* | **S0b:** *The turtle followed the fish* |
| **Normalized pair** | |
| **S1a:** *A sea turtle is hunting for fish* | **S1b:** *The turtle is following the fish* |
| **Expanded pair** | |
| Similar meaning | |
| **S2a:** *A sea turtle is hunting for food* | **S2b:** *The turtle is following the red fish* |
| Logically contradictory or at least highly contrasting meaning | |
| **S3a:** *A sea turtle is not hunting for fish* | **S3b:** *The turtle isn't following the fish* |
| Lexically similar but different meaning | |
| **S4a:** *A fish is hunting for a turtle in the sea* | **S4b:** *The fish is following the turtle* |

| Normalized sentence pairs | | Score | Label |
|---|---|---|---|
| **S1a:** *A sea turtle is hunting for fish* | **S2a:** *A sea turtle is hunting for food* | 4.5 | E |
| **S3a:** *A sea turtle is not hunting for fish* | **S1a:** *A sea turtle is hunting for fish* | 3.4 | C |
| **S4a:** *A fish is hunting for a turtle in the sea* | **S1a:** *A sea turtle is hunting for fish* | 3.9 | N |
| **S2b:** *The turtle is following the red fish* | **S1b:** *The turtle is following the fish* | 4.6 | E |
| **S1b:** *The turtle is following the fish* | **S3b:** *The turtle isn't following the fish* | 4 | C |
| **S1b:** *The turtle is following the fish* | **S4b:** *The fish is following the turtle* | 3.8 | C |
| **S1a:** *A sea turtle is hunting for fish* | **S2b:** *The turtle is following the red fish* | 4 | N |
| **S1a:** *A sea turtle is hunting for fish* | **S3b:** *The turtle isn't following the fish* | 3.2 | N |
| **S4b:** *The fish is following the turtle* | **S1a:** *A sea turtle is hunting for fish* | 3.2 | N |
| **S1b:** *The turtle is following the fish* | **S2a:** *A sea turtle is hunting for food* | 3.9 | N |
| **S1b:** *The turtle is following the fish* | **S3a:** *A sea turtle is not hunting for fish* | 3.4 | N |
| **S4a:** *A fish is hunting for a turtle in the sea* | **S1b:** *The turtle is following the fish* | 3.5 | N |
| **S1a:** *A sea turtle is hunting for fish* | **S1b:** *The turtle is following the fish* | 3.8 | N |

# SICK examples and stats

SICK-1241 GOLD: `neutral`

A woman is dancing and singing with other women

---

A woman is dancing and singing in the rain

SICK-341 GOLD: `contradiction`

There is no girl with a black bag on a crowded train

---

A girl with a black bag is on a crowded train

SICK-8381 GOLD: `entailment`

The young girl in blue is having fun on a slide

---

The young girl in blue is enjoying a slide

| Relatedness | neutral | contradiction | entailment | Total |
|---|---|---|---|---|
| [1,2) range | 10% | 0% | 0% | 10% (923) |
| [2,3) range | 13% | 1% | 0% | 14% (1373) |
| [3,4) range | 28% | 10% | 1% | 29% (3872) |
| [4,5] range | 7% | 3% | 27% | 37% (3672) |
| Total | 56.86% (5595) | 14.47% (1424) | 28.67% (2821) | 9840 |

# The FraCaS dataset

The FraCaS test suite [Cooper et al., 1996] was an early attempt to creating a semantic benchmark for NLP systems.

- Contains 346 problems, 45% of which are multi-premised.
- Covers GQs, plurals, anaphora, ellipsis, adjectives, comparatives, temporal reference, verbs and attitudes.
- Three-way annotated by the authors of the dataset.
- Contains some ambiguous sentences and a few erroneous problems.
- Requires almost no lexical or world knowledge

Later, the FraCaS question-answer pairs where converted into an NLI format [MacCartney and Manning, 2007]: online version

# FraCaS NLI problems

FraCaS-26 GOLD: `entailment`

Most Europeans are resident in Europe
All Europeans are people
All people who are resident in Europe can travel freely within Europe

---

Most Europeans can travel freely within Europe

FraCaS-61 GOLD: `undefined`

Both female commissioners used to be in business.

---

Both commissioners used to be in business.

FraCaS-171 GOLD: `entailment`

John wants to know how many men work part time.
And women.

---

John wants to know how many women work part time.

FraCaS-87 GOLD: `entailment`

Every representative and client was at the meeting.

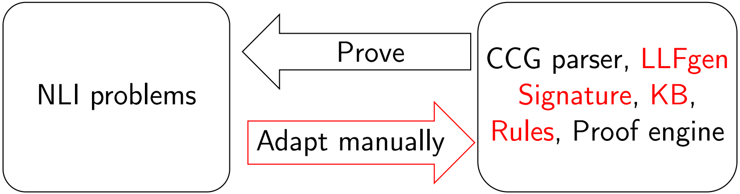
---

Every representative was at the meeting.

# Learning phase

The prover LangPro is (semi-automatically) trained on the NLI datasets [Abzianidze, 2016a].

- **Adaptation**:




  Used datasets: SICK-trial and FraCaS
- **Development**:
  Finding optimal values for certain parameters of the prover based on its performance on SICK-train.

**NB**: Only C&C parser is used in the learning phase in order to test LangPro for an unseen parser, EasyCCG, later.

# Adaptation: negative cases

We avoid fitting to the data and adopting unsound and non-general solutions.

The problems that were not solved during the adaptation:

- Sentence is not recognised as of category $S$ or failed to be parsed
- The error is analysis is too specific to fix:

| At | most | ten | commissioners | spend | time | at | home |
|---|---|---|---|---|---|---|---|
| $(S/S)/NP$ | $N/N$ | $N/N$ | $N$ | $(VP/PP)/NP$ | $N$ | $PP/NP$ | $N$ |

- Lexical relation is context dependent:

SICK-4505 GOLD: entailment

The doctors are healing a man

The doctor is helping the patient

SICK-384 GOLD: entailment

A white and tan dog is running through the tall and green grass

A white and tan dog is running through a field

# Adaptation: positive cases

The problems that were solved by upgrading one of the components of the prover:

- Treat **few** as ↓ in its 1st arg (*absolute* reading):

  FraCaS-76 GOLD: `entailment`

  Few committee members are from southern Europe
  
  ---
  
  Few female committee members are from southern Europe

- Introduce **fit** $\sqsubseteq$ **apply** and **food** $\sqsubseteq$ **meal**:

  SICK-4734 GOLD: `entailment`

  A man is fitting a silencer to a pistol
  
  ---
  
  A man is applying a silencer to a gun

  SICK-5110 GOLD: `entailment`

  A chef is preparing some food
  
  ---
  
  A chef is preparing a meal

# Development phase

Optimal values of the following parameters are searched:

- The number of word senses to consider at the same time;
- The upper bound for the number of rule applications;
- Whether to use a term aligner:
  - Weak aligner aligns everything except terms of type `np`:

    SICK-1022 GOLD: `contradiction`

    A woman is wearing sunglasses of large size and is holding newspapers in both hands

    There is no woman wearing sunglasses of large size and holding newspapers in both hands

    SICK-727 GOLD: `contradiction`

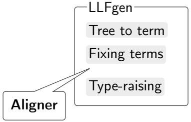
    The man in a grey t-shirt is sitting on a rock in front of the waterfall

    There is no man in a grey t-shirt sitting on a rock in front of the waterfall
  - Strong aligner aligns everything except terms of type `np` with ↓arg.

    SICK-423 GOLD: `contradiction`

    Two men are not holding fishing poles

    Two men are holding fishing poles
- Efficiency criterion of tableau rules.



# Efficiency criterion

Tableau rules have the following properties:

- Non-branching or branching (so called, $\alpha$ or $\beta$ rules);
- Semantic equivalence vs proper entailment;
- Consuming (so called, $\gamma$ rule) vs non-consuming;
- Producing (so called, $\delta$ rule) vs non-producing.

An example of an efficiency criterion:

$$EC = \langle \texttt{nonBr}, \texttt{semEqui}, \texttt{nonConsum}, \texttt{nonProd} \rangle$$

An efficiency vectors based on the $EC$ efficiency criterion:

- $V_{EC}(\wedge_{\mathbb{T}}) = 1111$
- $V_{EC}(\vee_{\mathbb{T}}) = 0111$
- $V_{EC}(\exists_{\mathbb{T}}) = 1110$
- $V_{EC}(\exists_{\mathbb{F}}) = 0001$

**What is the optimal efficiency criterion?**

# Greedy search for optimal parameters

| Acc% | Prec% | Rec% | Sense | Efficiency criterion | Aligner | RAL | Parser |
|---|---|---|---|---|---|---|---|
| 75.09 | 98.5 | 43.6 | 1 | [nonP,nonB,equi,nonC] | No | 200 | C&C |
| 76.42 | 98.3 | 46.8 | 1-5 | - | - | - | - |
| 76.89 | 97.8 | 48.1 | All | - | - | - | - |
| 78.44 | 97.9 | 51.7 | - | [equi,nonB,nonP,nonC] | - | - | - |
| 79.33 | 97.9 | 53.8 | - | - | Weak | - | - |
| 81.5 | 97.7 | 59.0 | - | - | Strong | - | - |
| 81.53 | 97.7 | 59.1 | - | - | Strong | 400 | - |
| 81.38 | 98.0 | 58.5 | - | - | Strong | 400 | EasyCCG |
| **82.6** | 97.7 | 61.6 | - | - | Strong | 400 | Both |

The results are given on the SICK-train problems.

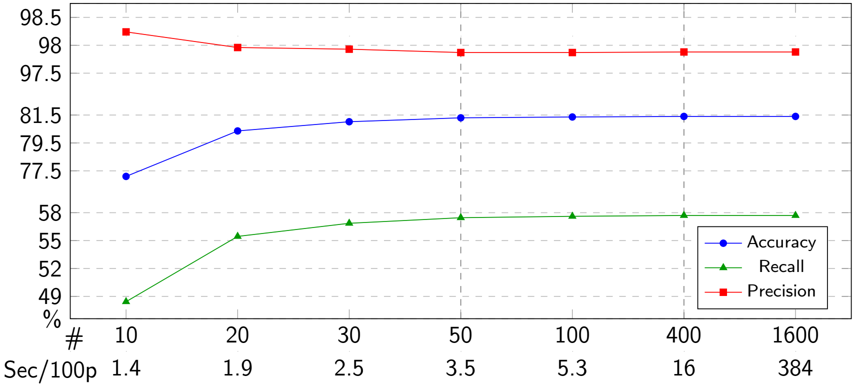
FraCaS-21 GOLD: entailment

The residents of member states have the right to live in Europe
All residents of member states are individuals
Every individual who has the right to live in Europe can travel freely within Europe

---

The residents of member states can travel freely within Europe

# Efficient and optimal rule application numbers




The results are given on the SICK-train problems.

# Solving FraCaS [Abzianidze, 2016b]

LangPro with C&C

| Gold\ccLP | yes | no | unk |
|---|---|---|---|
| yes | **51** | 0 | 19 + 4 |
| no | 1 | **14** | 2 |
| unk | 1 | 0 | **44 + 6** |

P = .97, R = .71, Acc = .81

+

LangPro with EasyCCG

| Gold\easyLP | yes | no | unk |
|---|---|---|---|
| yes | **52** | 0 | 22 |
| no | 1 | **12** | 4 |
| unk | 2 | 0 | **49** |

P = .96, R = .70, Acc = .80

=

LangPro

| Gold\LP | yes | no | unk |
|---|---|---|---|
| yes | **60** | 0 | 14 |
| no | 1 | **14** | 2 |
| unk | 2 | 0 | **49** |

P = **.96**, R = **.81**, Acc = **.87**

FraCaS-109 GOLD: `contradiction` LP: `entailment`

Just one accountant attended the meeting

Some accountants attended the meeting

# Related work (FraCaS)

[MacCartney and Manning, 2008] and [Angeli and Manning, 2014] employ a natural logic that is driven by sentence edits.

[Lewis and Steedman, 2013] employ Boxer-style [Bos et al., 2004] translation into FOL, Prover9 and distributional relation clustering.

[Mineshima et al., 2015, Haruta et al., 2020] also uses the Boxer-style translation but some HOGQs are treated as higher-order terms. Their inference system is implemented in the proof assistant Coq.

[Tian et al., 2014] and [Dong et al., 2014] uses abstract denotations obtained from DCS trees [Liang et al., 2011]: $\mathbf{man} \subset \pi_{\text{subj}}(\mathbf{read} \cap (W_{\text{subj}} \times \mathbf{book}_{\text{obj}}))$

[Bernardy and Chatzikyriakidis, 2017] uses Grammatical Framework and Coq.

[Hu et al., 2019] monotonicity calculus with trees obtain from CCG parsers.

[Kim et al., 2021] monotonicity reasoning with Unscoped Episodic Logical Forms.

# Comparison on FraCaS

| Sec (Sing/All) | Single-premised (Acc %) | | | | | | | | Overall (Acc %) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | BL | NL07,08 | LS | NL14 | T14a,b | M15 | K21 | **LP** | BL | LS13 | T14a,b | M15 | H20 | HM19 | $BC21^G$ | K21 | **LP** |
| 1 GQs (44/74) | 45 | 84 **98** | 70 | 95 | 80 93 | 82 | 73 | 93 | 50 | 62 | 80 95 | 78 | **97** | 88 | 93 | 70 | 95 |
| 2 Plur (24/33) | 58 | ~~42~~ **75** | - | ~~38~~ | - | 67 | - | **75** | 61 | - | - | 67 | - | - | 79 | - | **73** |
| 5 Adj (15/22) | 40 | 60 80 | - | **87** | - | **87** | - | **87** | 41 | - | - | 68 | **82** | - | 86 | - | 77 |
| 9 Att (9/13) | 67 | ~~56~~ 89 | - | ~~22~~ | - | 78 | - | **100** | 62 | - | - | 77 | **92** | - | 85 | - | **92** |
| 1,2,5,9 (92/142) | 50 | - **88** | - | - | - | 78 | - | **88** | 52 | - | - | 74 | - | - | 88 | - | **87** |

**BL** majority baseline, **NL07** [MacCartney and Manning, 2007], **NL08** [MacCartney and Manning, 2008], **NL14** [Angeli and Manning, 2014], **LS13** [Lewis and Steedman, 2013], **M15** [Mineshima et al., 2015], **T14a** [Tian et al., 2014], **T14b** [Dong et al., 2014], **HM19** [Hu et al., 2019], **H20** [Haruta et al., 2020], **K21** [Kim et al., 2021], and **BC21** [Bernardy and Chatzikyriakidis, 2021] (with gold trees)

# Conclusion

- The theorem prover for natural logic;
- The theorem prover for natural language is a pipeline:
  CCG parser + LLFgen + natural logic prover + WordNet;
- Play with it: http://naturallogic.pro
- Clone or fork it: https://github.com/kovvalsky/LangPro

# References I


Abzianidze, L. (2016a). *A natural proof system for natural language*. PhD thesis, Tilburg University.

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

Angeli, G. and Manning, C. D. (2014). Naturalli: Natural logic inference for common sense reasoning. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Bernardy, J.-P. and Chatzikyriakidis, S. (2017). A type-theoretical system for the FraCaS test suite: Grammatical framework meets coq. In *IWCS 2017 - 12th International Conference on Computational Semantics - Long papers*.

Bernardy, J.-P. and Chatzikyriakidis, S. (2021). Applied temporal analysis: A complete run of the FraCaS test suite. In *Proceedings of the 14th International Conference on Computational Semantics (IWCS)*, pages 11–20, Groningen, The Netherlands (online). Association for Computational Linguistics.

Bos, J., Clark, S., Steedman, M., Curran, J. R., and Hockenmaier, J. (2004). Wide-coverage semantic representations from a ccg parser. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING '04)*, pages 1240–1246, Geneva, Switzerland.

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.

# References II


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 Computation*, pages 585–594, Phuket,Thailand. Department of Linguistics, Chulalongkorn University.

Haruta, I., Mineshima, K., and Bekki, D. (2020). Combining event semantics and degree semantics for natural language inference. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 1758–1764, Barcelona, Spain (Online). International Committee on Computational Linguistics.

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.

Kim, G., Juvekar, M., Ekmekciu, J., Duong, V., and Schubert, L. (2021). A (mostly) symbolic system for monotonic inference with unscoped episodic logical forms. In *Proceedings of the 1st and 2nd Workshops on Natural Logic Meets Machine Learning (NALOMA)*, pages 71–80, Groningen, the Netherlands (online). Association for Computational Linguistics.

Lai, A. and Hockenmaier, J. (2014). Illinois-lh: A denotational and distributional approach to semantics. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 329–334, Dublin, Ireland. Association for Computational Linguistics and Dublin City University.

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

# References III


Liang, P., Jordan, M. I., and Klein, D. (2011). Learning dependency-based compositional semantics. In *Association for Computational Linguistics (ACL)*, pages 590–599.

MacCartney, B. and Manning, C. D. (2007). Natural logic for textual inference. In *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, RTE '07, pages 193–200, Stroudsburg, PA, USA. Association for Computational Linguistics.

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.

Marelli, M., Menini, S., Baroni, M., Bentivogli, L., Bernardi, R., and Zamparelli, R. (2014a). Semeval-2014 task 1: Evaluation of compositional distributional semantic models on full sentences through semantic relatedness and textual entailment. In *Proceedings of SemEval 2014 (International Workshop on Semantic Evaluation)*, pages 1–8, East Stroudsburg PA. ACL.

Marelli, M., Menini, S., Baroni, M., Bentivogli, L., Bernardi, R., and Zamparelli, R. (2014b). A sick cure for the evaluation of compositional distributional semantic models. In Calzolari, N., Choukri, K., Declerck, T., Loftsson, H., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., and Piperidis, S., editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland. European Language Resources Association (ELRA).

Miller, G. A. (1995). Wordnet: A lexical database for english. *Communications of the ACM*, 38(11):39–41.

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, Lisbon, Portugal. Association for Computational Linguistics.

# References IV

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)*, pages 79–89, Baltimore, Maryland. Association for Computational Linguistics.