A Natural Proof System for Natural Language

Lecture 3: Natural language inference with a natural theorem prover

Lasha Abzianidze

university of groningen

TbiLLAI 2019 in Tbilisi, Georgia

Previous lecture

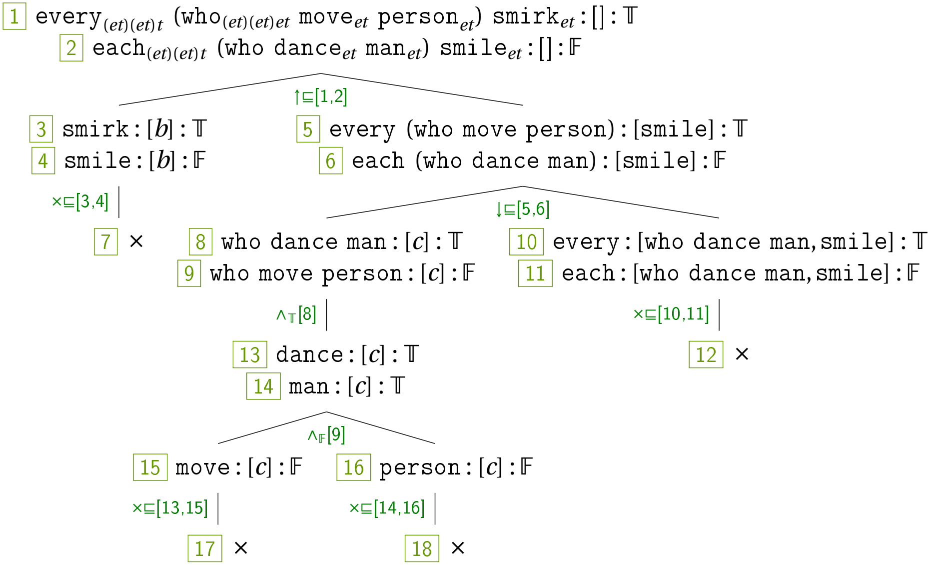
Lambda Logical Forms (LLFs):

- Simply typed λ-terms with e and t basic types;
- No logical constants, only lexical terms and λ-abstraction;
- Similar to syntactic trees of natural language expressions.

Natural tableau system:

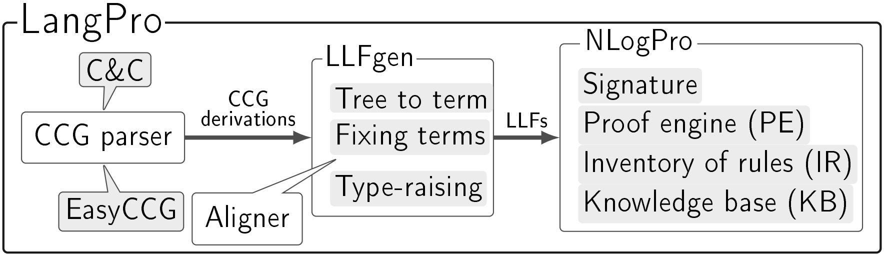
- Tableau entries: $\underbrace{\texttt{LLF:argumentList}}_{\text{Binary format of a term}}\texttt{:truthSign}$
- Monotonicity reasoning achieved with special monotonicity rules

Closed natural tableau



Today's lecture

A natural tableau theorem prover:



Solving natural language inference problems with the prover

LLFs and Categorial Grammar

LLFs are similar to formal derivations studied in Categorial Grammars (CGs) [Ajdukiewicz, 1935, Hillel, 1953].

- CGs treat each lexical item as a function;
- Categorial Type-Transparency principle links syntactic types to semantic ones.

Combinatory Categorial Grammar (CCG) [Steedman, 2000] is the only CG that is scaled up for wide-coverage text processing:

- CCG is well-studied from linguistic perspectives;
- There exists robust and accurate wide-coverage parsers for CCG, e.g., C&C parser [Clark and Curran, 2007] and EasyCCG

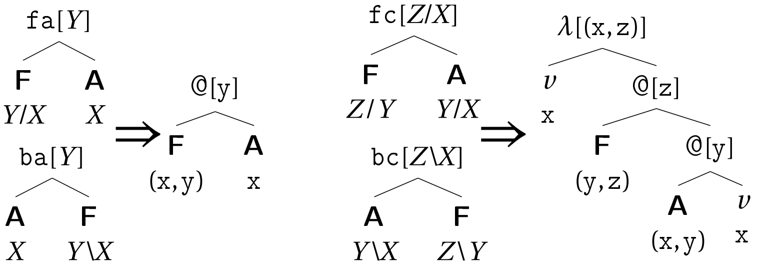
From CCG Trees to LLFs

Producing an LLF from a CCG derivation consists of several steps:

CCG Tree → **CCG Term** → **Fixed CCG Term** → **LLFs** ⇢ **FOL** / **DRT**

CCG Tree → CCG term

CCG Tree → **CCG Term** → **Fixed CCG Term** → **LLFs** ⇢ **FOL** / **DRT**



CCG term → fixed CCG term

CCG Tree → **CCG Term** → **Fixed CCG Term** → **LLFs** ⇢ FOL / DRT

$$[\mathtt{Dow}^{\mathrm{PER}}_{\mathrm{n,n}}\ \mathtt{Jones}^{\mathrm{PER}}_{\mathrm{n}}]_{\mathrm{np}} \rightsquigarrow \mathtt{Dow\_Jones}_{\mathrm{np}}$$

$$\mathtt{nobody}_{\mathrm{np}} \rightsquigarrow \mathtt{no}_{\mathrm{n,np}}\ \mathtt{person}_{\mathrm{n}}$$

$$[\mathtt{ice}_{\mathrm{n}}]_{\mathrm{np}} \rightsquigarrow \mathtt{a}_{\mathrm{n,np}}\ \mathtt{ice}_{\mathrm{n}}$$

$$[\mathtt{two}_{\mathrm{n,n}}\ \mathtt{dogs}_{\mathrm{n}}]_{\mathrm{np}} \rightsquigarrow \mathtt{two}_{\mathrm{n,np}}\ \mathtt{dogs}_{\mathrm{n}}$$

$$[\mathtt{working}_{\mathrm{np,s}}]_{\mathrm{n,n}} \rightsquigarrow \mathtt{who}_{\mathrm{(np,s),n,n}}\ \mathtt{working}$$

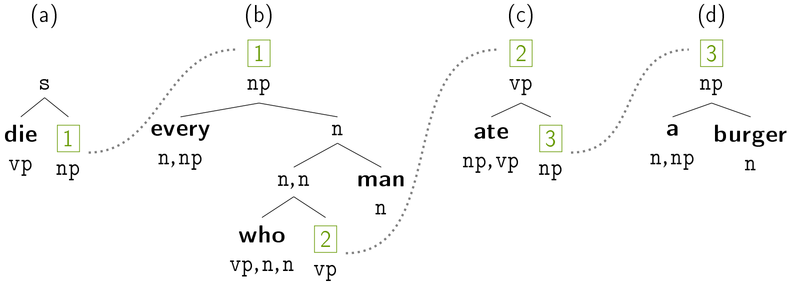
fixed CCG term → LLFs

Every man who ate a burger died

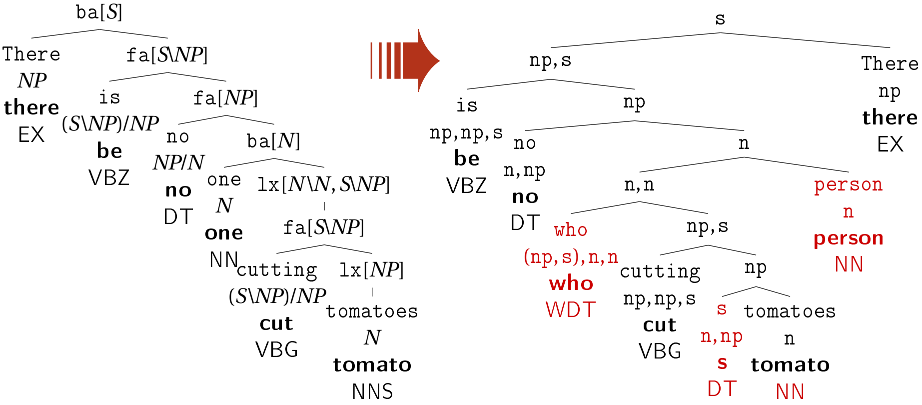
$\mathbf{die}_{\mathrm{vp}}\ (\mathbf{every}_{\mathrm{n,np}}(\mathbf{who}_{\mathrm{vp,n,n}}\ (\mathbf{eat}_{\mathrm{np,vp}}\ (\mathbf{a}_{\mathrm{n,np}}\ \mathbf{burger}_{\mathrm{n}}))\ \mathbf{man}_{\mathrm{n}}))\ \rightsquigarrow$

$\mathbf{EVERY}_{\mathrm{n,vp,s}}(\mathbf{who}\ (\lambda x.\mathbf{A}_{\mathrm{n,vp,s}}\ \mathbf{burger}\ (\lambda y.\mathbf{eat}\ y_{\mathrm{np}}\ x_{\mathrm{np}}))\mathbf{man})\mathbf{die}$

$\mathbf{A}_{\mathrm{n,vp,s}}\ \mathbf{burger}(\lambda y.\mathbf{EVERY}_{\mathrm{n,vp,s}}(\mathbf{who}\ (\lambda x.\mathbf{eat}\ y_{\mathrm{np}}\ x_{\mathrm{np}})\mathbf{man})\mathbf{die})$



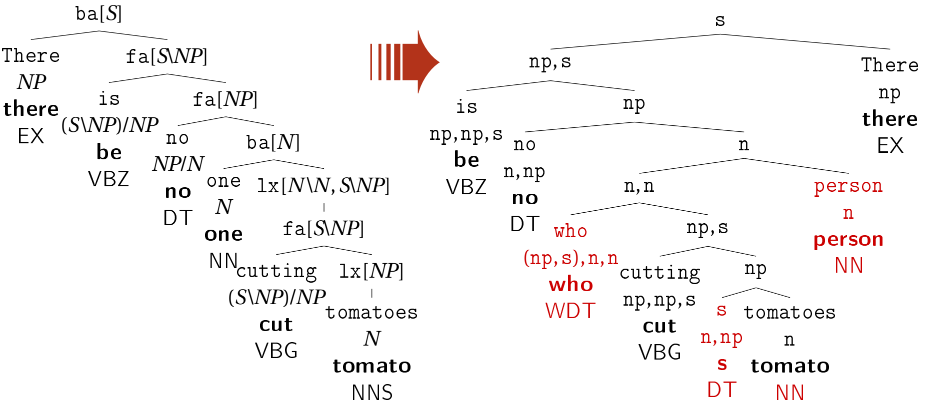
LLFgen (example)



There is no one cutting tomatoes ⇝

be(no(who(cut(s tomato))person))there

LLFgen (example)



be(**no**(**who**(**cut**(**s tomato**))**person**))**there** $\rightsquigarrow$

no (**who** ($\lambda x.$ **s tomato** ($\lambda y.$ **cut** $y\,x$)) **person**) ($\lambda z.$**be** z **there**)

s tomato ($\lambda y.$**no** (**who** (**cut** y) **person**) ($\lambda z.$**be** z **there**))

Uninformative *et*-based types

Semantic types based on e and t are uninformative from a syntactic point of view:

$h_{et} : [c_e]$ — $\texttt{cat}_{et} : [c_e]$ or $\texttt{sleep}_{et} : [c_e]$

$A_{(et)et}\ B_{et} : [c_e]$ — $\texttt{little}_{(et)et}\ \texttt{bird}_{et} : [c_e]$ or $\texttt{high}_{(et)et}\ \texttt{fly}_{et} : [c_e]$

$a_{(et)et}(b_{eet} c_e) : [c_e]$ — $\texttt{quietly}_{(et)et}\ (\texttt{follow}_{eet}\ \texttt{john}_e) : [c_e]$ or $\texttt{wife}_{(et)et}\ (\texttt{of}_{eet}\ \texttt{john}_e) : [c_e]$

Extending the type system

Add syntactic types to semantic ones:

$$\{e, t\} + \{\mathtt{np}, \mathtt{s}, \mathtt{n}, \mathtt{pp}\}$$

A partial order subtyping relation (<:) serves as an interface between syntactic and semantic types:

- $\mathtt{s} <: t$
- $e <: \mathtt{np}$
- $\mathtt{n} <: et$
- $\mathtt{pp} <: et$
- $(\alpha_1, \alpha_2) <: (\beta_1, \beta_2)$ iff $\beta_1 <: \alpha_1$ and $\alpha_2 <: \beta_2$

Syntactic terms

An additional typing rule:

if $A : \alpha$ and $\alpha <: \beta$, then $A : \beta$ too.

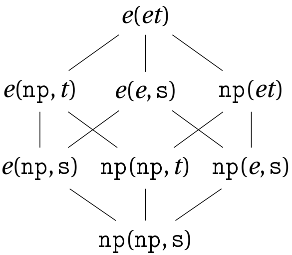
Terms of multiple types:

- $\mathbf{cat}_{\mathrm{n}}$ is of type et
- $\mathbf{red}_{\mathrm{n,n}}$ is of type (n, et) ~~and (et, et)~~
- $\mathbf{see}_{\mathrm{np,np,s}}$ is of type $\mathrm{np}(\mathrm{np}, t)$, eet, ...

Syntactic and semantic terms together:

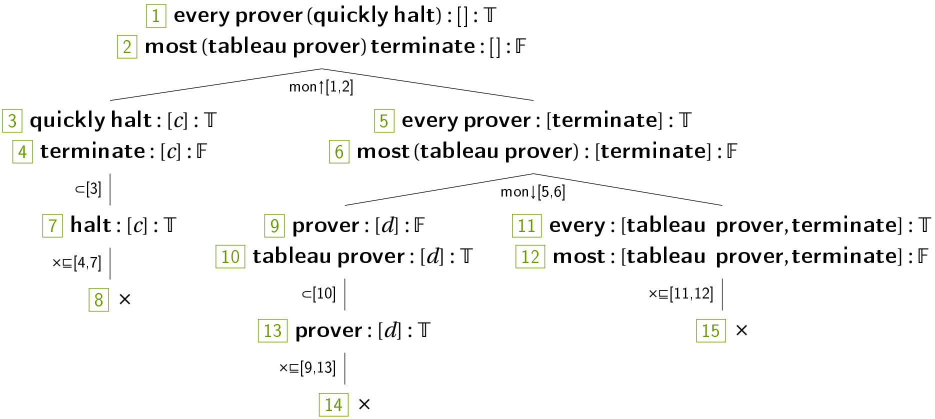
$\mathbf{cat}_{\mathrm{n}}\, c_e$, $\mathbf{love}_{\mathrm{np,np,s}} \mathbf{john}_{\mathrm{np}}\, c_e$, $\mathbf{on}_{\mathrm{pp}}\, d_e$

Types of $\mathbf{see}_{\mathrm{np,np,s}}$



No much changes in tableau proofs

Only the **style** of the terms is changed

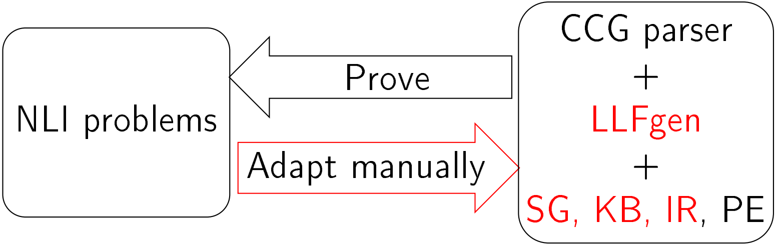


Linguistic Rules

Linguistic rules, in contrast to the algebraic rules, account for a certain syntactic constructions.

We will also include the rules that *remedy* errors introduced in the syntactic derivation trees.

Most of the linguistic rules are collected in a data-driven fashion:



Rules for prepositions

PP@N$_{\mathbb{T}}$

$$p^{\mathtt{IN}}_{\mathrm{np,n,n}}\, d\, N : [c] : \mathbb{T}$$

$$N : [c] : \mathbb{T}$$
$$p_{\mathrm{np,pp}} : [d, c] : \mathbb{T}$$

$\textbf{with}_{\mathrm{np,n,n}}\ g_e\ \textbf{bicycle}_{\mathrm{n}} : [c_e] : \mathbb{T}$

$\textbf{bicycle} : [c] : \mathbb{T}$

$\textbf{with}_{\mathrm{np,pp}} : [g, c] : \mathbb{T}$

PP@N$_{\mathbb{F}}$

$$p^{\mathtt{IN}}_{\mathrm{np,n,n}}\, d\, N : [c] : \mathbb{F}$$

$N : [c] : \mathbb{F}$ | $p_{\mathrm{np,pp}} : [g, c] : \mathbb{F}$

$\textbf{with}_{\mathrm{np,n,n}}\ g_e\ \textbf{bicycle}_{\mathrm{n}} : [c_e] : \mathbb{F}$

$\textbf{bicycle} : [c] : \mathbb{F}$ | $\textbf{with}_{\mathrm{np,pp}} : [g, c] : \mathbb{F}$

Treating PP attachment

SICK-9069 GOLD: entailment BY: C&C

Two boys are [[laying$_{VP}$ [in the ocean]$_{VP\backslash VP}$] [close to the beach]$_{VP\backslash VP}$]

Two boys are [[laying$_{VP/PP}$ [in the water]$_{PP}$] [close to the beach]$_{VP\backslash VP}$]

V@PP

$$V_{\mathrm{pp},\alpha}\,(p^{\mathrm{IN}}_{\mathrm{np,pp}}\,D) : [\vec{C}] : \mathbb{X}$$

$$p^{\mathrm{IN}}_{\mathrm{np},\alpha,\alpha}\,D\,V_\alpha : [\vec{C}] : \mathbb{X}$$

$$\alpha = (\mathrm{np}^*, \mathrm{vp})$$

$$\mathbf{lie}_{\mathrm{pp,vp}}\ (\mathbf{in}_{\mathrm{np,pp}}\ o_e) : [c] : \mathbb{F}$$

$$\mathbf{in}_{\mathrm{np,vp,vp}}\ o_e\ \mathbf{lie}_{\mathrm{vp}} : [c] : \mathbb{F}$$

Treating PP attachment (II)

SICK-340 GOLD: entailment BY: C&C

[schoolgirl$_{N/PP}$ [with a black bag]$_{PP}$] is on a crowded train

[girl$_N$ [with a black bag]$_{N\backslash N}$] is on a crowded train

N@PP$_{\mathbb{T}}$

$N_{\text{pp,n}}\ P_{\text{pp}} : [c_e] : \mathbb{T}$

$N_{\text{n}} : [c] : \mathbb{T}$

$P : [c] : \mathbb{T}$

N@PP$_{\mathbb{F}}$

$N_{\text{pp,n}}\ P_{\text{pp}} : [c_e] : \mathbb{F}$

$N_{\text{n}} : [c] : \mathbb{F}$ | $P : [c] : \mathbb{F}$

1 **schoolgirl**$_{\text{pp,n}}$ (**with**$_{\text{np,pp}}$ b_e) : $[g_e]$: $\mathbb{T}$

2 **with**$_{\text{np,n,n}}$ b **girl**$_{\text{n}}$: $[g]$: $\mathbb{F}$

N@PP$_{\mathbb{T}}$[1]

3 **schoolgirl**$_{\text{n}}$: $[g]$: $\mathbb{T}$

4 **with**$_{\text{np,pp}}$: $[b, g]$: $\mathbb{T}$

N@PP$_{\mathbb{F}}$[2]

5 **girl** : $[g]$: $\mathbb{F}$ | 6 **with**$_{\text{np,pp}}$: $[b, g]$: $\mathbb{F}$

$\times\sqsubseteq$[3,5] | $\times\sqsubseteq$[4,6]

7 $\times$ | 8 $\times$

Expletive *there*

[1] $\textbf{some}_{\mathrm{n,vp,s}}\ \textbf{dog}_{\mathrm{n}}\ (\textbf{be}_{\mathrm{vp,vp}}\ \textbf{run}_{\mathrm{vp}}) : [\,] : \mathbb{T}$

[2] $\textbf{some}_{\mathrm{n,vp,s}}\ (\textbf{that}_{\mathrm{vp,n,n}}\ \textbf{move}_{\mathrm{vp}}\ \textbf{animal}_{\mathrm{n}})(\lambda x.\,\textbf{be}_{\mathrm{np,vp}}\ x\ \textbf{there}_{\mathrm{np}_{thr}}) : [\,] : \mathbb{F}$

$\exists_{\mathbb{T}}[1]$

[3] $\textbf{dog} : [c_e] : \mathbb{T}$

[4] $\textbf{berun} : [c_e] : \mathbb{T}$

aux[4]

[5] $\textbf{run} : [c] : \mathbb{T}$

$\lambda{<}, \exists_{\mathbb{F}}[2]$

[6] $\textbf{that move animal} : [c] : \mathbb{F}$ | [7] $\textbf{be}\ c\ \textbf{there} : [\,] : \mathbb{F}$

$\wedge_{\mathbb{F}}[6]$ | A>, A>[7]

[11] $\textbf{move} : [c] : \mathbb{F}$ | [12] $\textbf{animal} : [c] : \mathbb{F}$ | [9] $\textbf{be} : [c, \textbf{there}] : \mathbb{F}$

$\times\sqsubseteq[5,11]$ | $\times\sqsubseteq[3,12]$ | ×THR[9]

[13] × | [14] × | [10] ×

×THR

$\textbf{be}_{\mathrm{np,vp}} : [C, D] : \mathbb{F}$

×

$\textbf{there}_{\mathrm{np}_{thr}} \in \{C, D\}$

Other closure rules

Open compound nouns:

$$\frac{N_{\mathrm{n}} : [d] : \mathbb{T} \quad H_{\mathrm{pp,n}}(prp\ d) : [c] : \overline{\mathbb{X}} \quad A_{\mathrm{n,n}}\ H_{\mathrm{n}} : [c] : \mathbb{X}}{\times}\ (\times\mathrm{CPN})$$

$N \approx A$ or $N \approx_d A$

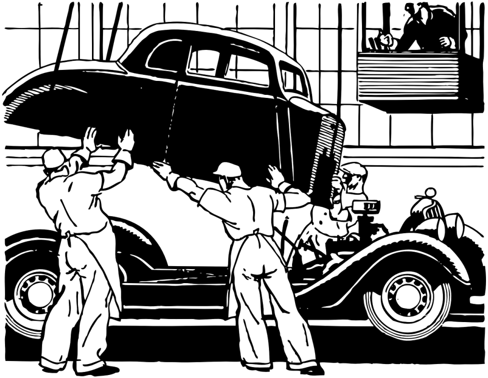
$$\frac{\textbf{protection}_{\mathrm{n}} : [d_e] : \mathbb{T} \quad \textbf{gear}_{\mathrm{pp,n}}(\textbf{for}_{\mathrm{np,pp}}\ d_e) : [c_e] : \mathbb{F} \quad \textbf{protective}_{\mathrm{n,n}}\ \textbf{gear}_{\mathrm{n}} : [c_e] : \mathbb{T}}{\times}\ (\times\mathrm{CPN}^*)$$

Light verb constructions:

$$\frac{l_{\vec{\alpha},\mathrm{vp}} : [c, \vec{D}] : \mathbb{X} \quad u_{\mathrm{n}} : [c] : \mathbb{T} \quad v_{\vec{\alpha},\mathrm{s}} : [D] : \overline{\mathbb{X}}}{\times}\ (\times\mathrm{LVC})$$

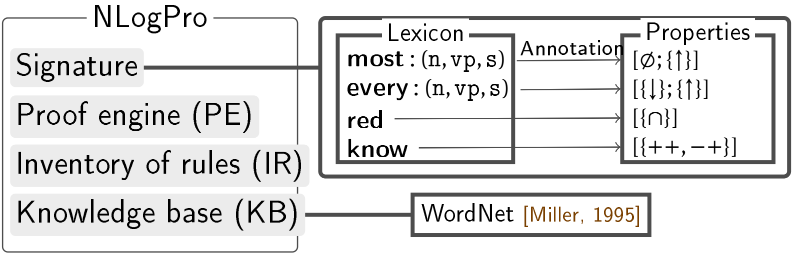
$l \in \{$**do**, **get**, **give**, **have**, **make**, **take**$\}$, $\vec{\alpha}$ is formed by np and pp, and $u \approx_d v$

$$\frac{\textbf{do}_{\mathrm{np,vp}} : [d_e, h_e] : \mathbb{T} \quad \textbf{dance}_{\mathrm{n}} : [d_e] : \mathbb{T} \quad \textbf{dance}_{\mathrm{vp}} : [h_e] : \mathbb{F}}{\times}\ (\times\mathrm{LVC}^*)$$



Building a natural theorem prover

Natural logic theorem prover (NLogPro)



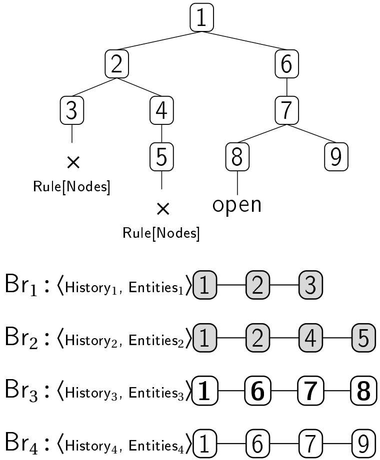
KB uses 4 relations from WordNet 3.0:

- derivation
- similarity
- hyponymy/hypernymy
- antonymy

⚠ No word sense disambiguation system is used.

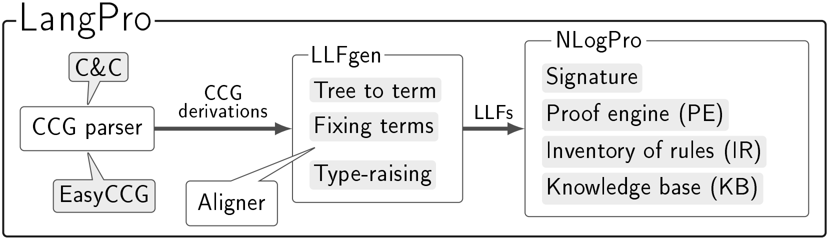
Two data structures

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



Natural language theorem prover

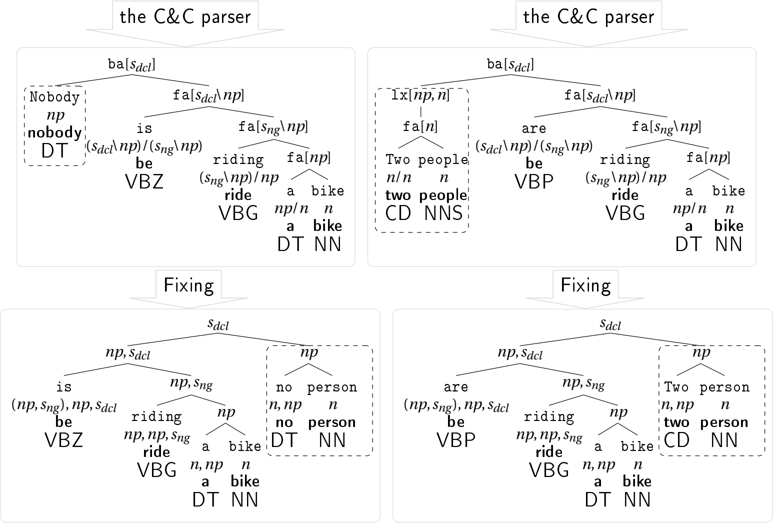
Chaining a CCG parser, the LLF generator and NLogPro results in a theorem prover for natural language.



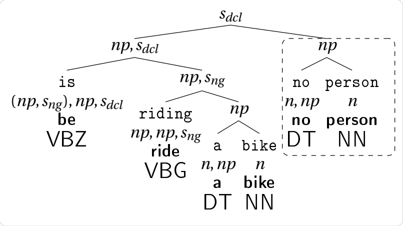
Online demo at: http://naturallogic.pro
git clone: https://github.com/kovvalsky/LangPro

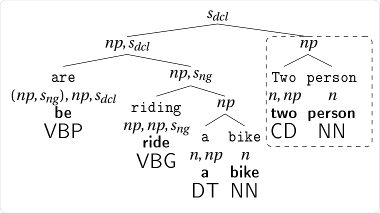
LangPro in action

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)): $[c] : \mathbb{T}$

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

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

6 ×

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

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

The SICK dataset

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

- 10K Text-Hypothesis pairs 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 ($\tilde{9}$ words per sentence).
- SemEval-14 RTE benchmark [Marelli et al., 2014a]
- 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

The FraCaS dataset

The FraCaS test suite [Cooper et al., 1996]:

- Contains 346 problems (45% multi-premised)
- Covers GQs, plurals, anaphora, ellipsis, adjectives, comparatives, temporal reference, verbs and attitudes.
- Three-way annotated by the authors of the dataset.
- Requires almost no lexical or world knowledge

The NLI dataset derived from FraCaS [MacCartney and Manning, 2007].

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`

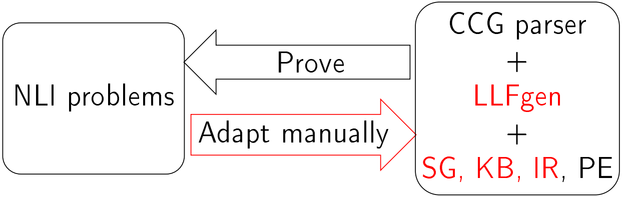
John wants to know how many men work part time.
And women.

John wants to know how many women work part time.

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

Adaptation

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** ⊑ **apply** and **food** ⊑ **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 are searched for:

- The number of word senses to consider;
- The max number of rule application limit (RAL);
- Whether to use a term aligner:
  - Weak aligner aligns everything but terms of type `np`:

    SICK-727   GOLD: `contradiction`

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

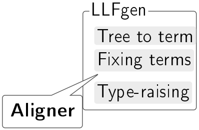
    There is no man in a grey t-shirt sitting on a rock in front of the waterfall

  - Strong aligner aligns everything but terms of type terms of type `np` with ↓arg.

    SICK-423   GOLD: `contradiction`

    Two men are not holding fishing poles

    Two men are holding fishing poles

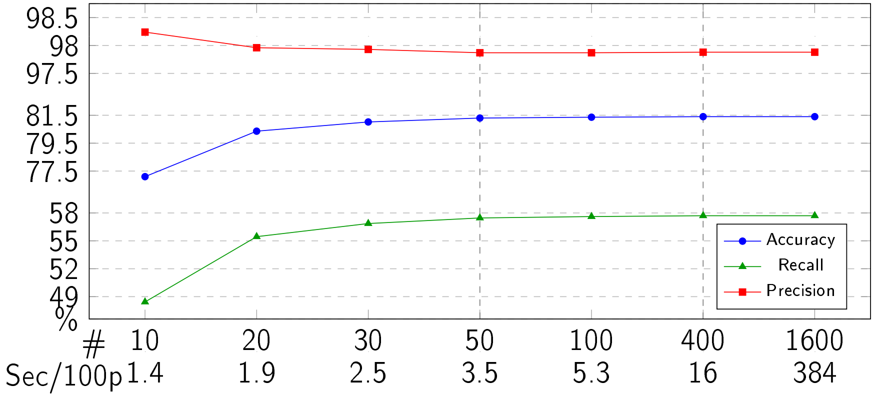


Greedy search for optimal parameters

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.

Efficient and optimal RAL



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] 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]:

$$\mathbf{man} \subset \pi_{\mathsf{subj}}\big(\mathbf{read} \cap (W_{\mathsf{subj}} \times \mathbf{book}_{\mathsf{obj}})\big)$$

[Bernardy and Chatzikyriakidis, 2017] uses Grammatical Framework and Coq. They use gold standard GF trees.

Comparison on FraCaS

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

NL07 [MacCartney and Manning, 2007], **NL08** [MacCartney and Manning, 2008], **NLI** [Angeli and Manning, 2014], **LS** [Lewis and Steedman, 2013],
M15 [Mineshima et al., 2015], **T14a** [Tian et al., 2014] and **T14b** [Dong et al., 2014]

Advantages of our approach over the related ones include:

- Reasoning (with the semantic tableau) over multiple-premises;
- Logical forms close to surface forms;
- Underlying expressive high-order logic.

Curing SICK [Abzianidze, 2015]

Gold SICK-test \ LangPro	Ent	Cont	Neut
Entailment	**805**	0	609
Contradiction	2	**482**	236
Neutral	26	7	**2760**

P=**97.4%**, R=**60.3%**, Acc=**82.14%**

Mainly the usage of WordNet and noisy gold labels are blamed for false proofs.

ID	G/LP	Premise	Conclusion
1405	N/E	A **prawn** is being cut by a woman	A woman is cutting **shrimps**
4443	N/E	A man is singing to a **girl**	A man is singing to a **woman**
2870	N/C	Two people are riding a **motorcycle**	Nobody is riding a **bike**
8913	N/C	A couple is not looking at a map	A couple is looking at a map
363	C/C	P: A soccer ball is not rolling into a goal net C: A soccer ball is rolling into a goal net	

False neutrals

Reason for false neutrals are knowledge sparsity (ca 50%), a lack of rules (ca 25%), wrong labels and parsing mistakes.

ID	G/LP	Premise	Conclusion
4974	E/N	Someone is holding a hedgehog	Someone is holding a small animal
6258	E/N	P: A policeman is sitting on a **motorcycle** C: The cop is sitting on a police **bike**	
4553	E/N	P: A man is emptying a container made of plastic C: A man is emptying a plastic container	
4720	E/N	A monkey is practicing martial arts	A chimp is practicing martial arts
6447	C/N	P: [A small boy [in a yellow shirt]] is laughing on the beach C: There is no small boy [in a yellow shirt [laughing on the beach]]	

Comparison on SICK

SemEval-14 systems	Prec%	Rec%	Acc%	(+LP)	NWS%
Baseline (majority)	-	-	56.69		39.7
Illinois-LH	81.56	**81.87**	**84.57**	(+0.65)	72.8
ECNU	84.37	74.37	83.64	(+1.77)	72.7
UNAL-NLP	81.99	76.80	83.05	(+1.48)	71.2
SemantiKLUE	85.40	69.63	82.32	(+2.84)	71.5
The Meaning Factory	93.63	60.64	81.59	(+2.78)	73.0
UTexas (Prob-FOL)	**97.87**	38.71	73.23	(+9.44)	62.5
LangPro	**97.35**	60.31	82.14		**74.8**

"Hard" problems

The problems from SICK-test that were proved correctly by both ccLangPro and easyLangPro but failed by all the top five systems at the SemEval-14 task.

ID	G	Text	Hypothesis
247	C	T: The woman is not wearing glasses or a headdress	H: A woman is wearing an Egyptian headdress
406	E	T: A group of scouts are hiking through the grass	H: People are walking
2895	C	The man isn't lifting weights	The man is lifting barbells
3527	E	T: A person is jotting something with a pencil	H: A person is writing
3570	C	The piece of paper is not being cut	Paper is being cut with scissors
3608	N	T: A monkey is riding a bike	H: A bike is being ridden over a monkey
3806	E	A man in a hat is playing a harp	A man is playing an instrument
4479	E	The boy is playing the piano	The boy is playing a musical instrument

Conclusion

Natural Tableau is a wide-coverage but still logic-based reasoning system inspired by Natural Logic.

It represents a proof-theoretic approach to NLI.

Natural tableau was successfully scaled up for the NLI task:
CCG parser + LLFgen + theorem prover

Pros and cons of Natural Tableau:

- ✔ Employs higher-order logic to model linguistic semantics;
- ✔ Allows deep logical and shallow (e.g. monotonicity) reasoning;
- ✔ Getting logical form is similar to syntactic parsing;
- ✘ Heavily hinges on CCG parsing;
- ✔ Proofs are highly reliable ($\leq 3\%$ false proofs);
- ✘ Suffers from multi-sense words;
- ✘ No fully automated learning from data yet;
- ✔ Its decision procedure is transparent and explanatory;

Future work

There are really many directions for future work:

- Explore different types of RTE data, e.g., the newswire or human generated data [Bowman et al., 2015];
- Incorporate more knowledge in KB, e.g., paraphrase database [Ganitkevitch et al., 2013].
- Model different phenomena: comparatives, anaphora, cardinals, etc.
- Pairing with distributional semantics: $R(w_1, w_2, r)$ and weighted closure branches;
- Acquisition of lexical knowledge: abductive reasoning;
- Generate LLFs from Universal Dependency trees

  + the Universal Semantic Tagging [Abzianidze and Bos, 2017]

  $\rightarrow^?$ Multilingual Natural Tableau

References I

Abzianidze, L. (2015). A tableau prover for natural logic and language. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2492–2502, Lisbon, Portugal. Association for Computational Linguistics.

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.

Abzianidze, L. and Bos, J. (2017). Towards universal semantic tagging. In *IWCS 2017 — 12th International Conference on Computational Semantics — Short papers*.

Ajdukiewicz, K. (1935). Die syntaktische Konnexität. *Studia Philosophica*, 1:1–27. English translation "Syntactic Connexion" by H. Weber in McCall, S. (Ed.) *Polish Logic*, pp. 207–231, Oxford University Press, Oxford, 1967.

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*.

References II

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.

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.

Clark, S. and Curran, J. R. (2007). Wide-coverage efficient statistical parsing with ccg and log-linear models. *Computational Linguistics*, 33.

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.

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.

Ganitkevitch, J., Van Durme, B., and Callison-Burch, C. (2013). PPDB: The paraphrase database. In *Proceedings of NAACL-HLT*, pages 758–764, Atlanta, Georgia. Association for Computational Linguistics.

Hillel, Y. B. (1953). A Quasi-Arithmetical Notation for Syntactic Description. *Language*, 29(1):47–58.

References III

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

Lewis, M. and Steedman, M. (2014). A* CCG parsing with a supertag-factored model. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 990–1000, Doha, Qatar. Association for Computational Linguistics.

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.

References IV

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.

Steedman, M. (2000). *The Syntactic Process*. MIT Press, Cambridge, MA, USA.

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.