the unicorn of constant-time parsing

Konstantinos Kogkalidis
Utrecht Institute of Linguistics OTS, Utrecht University

End-to-End Compositional Models of Vector-Based Semantics
ESSLLI, August 2022, Galway

A composition calculus for vector-based semantic modelling with a localization for Dutch

Overview

- Grammar
- Supertagging
- Parsing
- Unicorns

A type grammar for the 21st century

$\mathrm{ILL}_{\multimap}$ plus $\diamondsuit, \Box$ modalities for *dependency domain demarcation*.

Types inductively defined by:

$$\mathbb{T} := A \mid T \multimap T \mid \diamondsuit^{d} T \mid \Box^{d} T \qquad A \in \mathbb{A},\, T \in \mathbb{T}$$

- $\multimap$ – linear function builder
- $\diamondsuit$ – reserved for "necessary arguments", i.e. complements
- $\Box$ – reserved for "optional functors", i.e. adjuncts

..and its term calculus

$$\frac{}{\mathsf{c} : T \vdash \mathsf{c} : T}\ Lex \qquad \frac{\Gamma \vdash \mathsf{s} : T_1 \multimap T_2 \quad \Delta \vdash \mathsf{t} : T_1}{\Gamma, \Delta \vdash \mathsf{s}\ \mathsf{t} : T_2} \multimap E$$

$$\frac{\Gamma \vdash \mathsf{t} : T}{\langle \Gamma \rangle^d \vdash \triangle^d\, \mathsf{t} : \Diamond^d T} \Diamond^d I \qquad \frac{\Gamma \vdash \mathsf{s} : \Box^d T}{\langle \Gamma \rangle^d \vdash \blacktriangledown^d \mathsf{s} : T} \Box^d E$$

$$\frac{}{\mathrm{x} : T \vdash \mathrm{x} : T}\ Ax \qquad \frac{\Gamma, \mathrm{x} : T_1 \vdash \mathsf{s} : T_2}{\Gamma \vdash \lambda \mathrm{x}.\mathsf{s} : T_1 \multimap T_2} \multimap I$$

$$\frac{\langle \Gamma \rangle^d \vdash \mathsf{s} : T}{\Gamma \vdash \blacktriangle^d \mathsf{s} : \Box^d T} \Box^d I \qquad \frac{\Gamma[\langle \mathrm{x} : T_1 \rangle^d] \vdash \mathsf{t} : T_2 \quad \Delta \vdash \mathsf{s} : \Diamond^d T_1}{\Gamma[\Delta] \vdash \mathsf{t}[\mathrm{x} \mapsto \triangledown^d \mathsf{s}] : T_2} \Diamond^d E$$

$$\frac{\langle \Gamma, \langle \mathrm{x} : T_1 \rangle^{\mathsf{X}}, \Delta \rangle^d \vdash \mathsf{s} : T_2}{\langle \Gamma, \Delta \rangle^d, \langle \mathrm{x} : T_1 \rangle^{\mathsf{X}} \vdash \mathsf{s} : T_2} \mathsf{X}$$

..and its term calculus

$$\frac{}{\mathtt{c} : T \vdash \mathtt{c} : T}\ Lex \qquad \frac{\Gamma \vdash \mathtt{s} : T_1 \multimap T_2 \quad \Delta \vdash \mathtt{t} : T_1}{\Gamma, \Delta \vdash \mathtt{s}\ \mathtt{t} : T_2}\ \multimap E$$

$$\frac{\Gamma \vdash \mathtt{t} : T}{\langle \Gamma \rangle^d \vdash \triangle^d\, \mathtt{t} : \Diamond^d T}\ \Diamond^d I \qquad \frac{\Gamma \vdash \mathtt{s} : \Box^d T}{\langle \Gamma \rangle^d \vdash \blacktriangledown^d \mathtt{s} : T}\ \Box^d E$$

$$\frac{}{\mathrm{x} : T \vdash \mathrm{x} : T}\ Ax \qquad \frac{\Gamma, \mathrm{x} : T_1 \vdash \mathtt{s} : T_2}{\Gamma \vdash \lambda \mathrm{x}.\mathtt{s} : T_1 \multimap T_2}\ \multimap I$$

$$\frac{\langle \Gamma \rangle^d \vdash \mathtt{s} : T}{\Gamma \vdash \blacktriangle^d \mathtt{s} : \Box^d T}\ \Box^d I \qquad \frac{\Gamma[\langle \mathrm{x} : T_1 \rangle^d] \vdash \mathtt{t} : T_2 \quad \Delta \vdash \mathtt{s} : \Diamond^d T_1}{\Gamma[\Delta] \vdash \mathtt{t}[\mathrm{x} \mapsto \triangledown^d \mathtt{s}] : T_2}\ \Diamond^d E$$

$$\frac{\langle \Gamma, \langle \mathrm{x} : T_1 \rangle^{\mathsf{X}}, \Delta \rangle^d \vdash \mathtt{s} : T_2}{\langle \Gamma, \Delta \rangle^d, \langle \mathrm{x} : T_1 \rangle^{\mathsf{X}} \vdash \mathtt{s} : T_2}\ \mathsf{X}$$

..and its term calculus

$$\frac{}{\mathtt{c} : T \vdash \mathtt{c} : T}\ Lex \qquad \frac{\Gamma \vdash \mathtt{s} : T_1 \multimap T_2 \quad \Delta \vdash \mathtt{t} : T_1}{\Gamma, \Delta \vdash \mathtt{s}\ \mathtt{t} : T_2} \multimap E$$

$$\frac{\Gamma \vdash \mathtt{t} : T}{\langle \Gamma \rangle^{d} \vdash \vartriangle^{d} \mathtt{t} : \Diamond^{d} T}\ \Diamond^{d} I \qquad \frac{\Gamma \vdash \mathtt{s} : \Box^{d} T}{\langle \Gamma \rangle^{d} \vdash \blacktriangledown^{d} \mathtt{s} : T}\ \Box^{d} E$$

$$\frac{}{\mathtt{x} : T \vdash \mathtt{x} : T}\ Ax \qquad \frac{\Gamma, \mathtt{x} : T_1 \vdash \mathtt{s} : T_2}{\Gamma \vdash \lambda \mathtt{x}.\mathtt{s} : T_1 \multimap T_2} \multimap I$$

$$\frac{\langle \Gamma \rangle^{d} \vdash \mathtt{s} : T}{\Gamma \vdash \blacktriangle^{d} \mathtt{s} : \Box^{d} T}\ \Box^{d} I \qquad \frac{\Gamma[\langle \mathtt{x} : T_1 \rangle^{d}] \vdash \mathtt{t} : T_2 \quad \Delta \vdash \mathtt{s} : \Diamond^{d} T_1}{\Gamma[\Delta] \vdash \mathtt{t}[\mathtt{x} \mapsto \triangledown^{d} \mathtt{s}] : T_2}\ \Diamond^{d} E$$

$$\frac{\langle \Gamma, \langle \mathtt{x} : T_1 \rangle^{\mathsf{X}}, \Delta \rangle^{d} \vdash \mathtt{s} : T_2}{\langle \Gamma, \Delta \rangle^{d}, \langle \mathtt{x} : T_1 \rangle^{\mathsf{X}} \vdash \mathtt{s} : T_2}\ \mathsf{X}$$

..and its term calculus

$$\frac{}{\mathtt{c} : T \vdash \mathtt{c} : T}\ Lex \qquad \frac{\Gamma \vdash \mathtt{s} : T_1 \multimap T_2 \quad \Delta \vdash \mathtt{t} : T_1}{\Gamma, \Delta \vdash \mathtt{s}\ \mathtt{t} : T_2} \multimap E$$

$$\frac{\Gamma \vdash \mathtt{t} : T}{\langle \Gamma \rangle^d \vdash \vartriangle^d \mathtt{t} : \Diamond^d T}\ \Diamond^d I \qquad \frac{\Gamma \vdash \mathtt{s} : \Box^d T}{\langle \Gamma \rangle^d \vdash \blacktriangledown^d \mathtt{s} : T}\ \Box^d E$$

$$\frac{}{\mathrm{x} : T \vdash \mathrm{x} : T}\ Ax \qquad \frac{\Gamma, \mathrm{x} : T_1 \vdash \mathtt{s} : T_2}{\Gamma \vdash \lambda \mathrm{x}.\mathtt{s} : T_1 \multimap T_2} \multimap I$$

$$\frac{\langle \Gamma \rangle^d \vdash \mathtt{s} : T}{\Gamma \vdash \blacktriangle^d \mathtt{s} : \Box^d T}\ \Box^d I \qquad \frac{\Gamma[\langle \mathrm{x} : T_1 \rangle^d] \vdash \mathtt{t} : T_2 \quad \Delta \vdash \mathtt{s} : \Diamond^d T_1}{\Gamma[\Delta] \vdash \mathtt{t}[\mathrm{x} \mapsto \triangledown^d \mathtt{s}] : T_2}\ \Diamond^d E$$

$$\frac{\langle \Gamma, \langle \mathrm{x} : T_1 \rangle^{\mathsf{X}}, \Delta \rangle^d \vdash \mathtt{s} : T_2}{\langle \Gamma, \Delta \rangle^d, \langle \mathrm{x} : T_1 \rangle^{\mathsf{X}} \vdash \mathtt{s} : T_2}\ \mathsf{X}$$

..and its term calculus

$$\frac{}{\mathtt{c} : T \vdash \mathtt{c} : T}\ Lex \qquad \frac{\Gamma \vdash \mathtt{s} : T_1 \multimap T_2 \quad \Delta \vdash \mathtt{t} : T_1}{\Gamma, \Delta \vdash \mathtt{s}\ \mathtt{t} : T_2} \multimap E$$

$$\frac{\Gamma \vdash \mathtt{t} : T}{\langle \Gamma \rangle^d \vdash \triangle^d\ \mathtt{t} : \Diamond^d T}\ \Diamond^d I \qquad \frac{\Gamma \vdash \mathtt{s} : \Box^d T}{\langle \Gamma \rangle^d \vdash \blacktriangledown^d \mathtt{s} : T}\ \Box^d E$$

$$\frac{}{\mathtt{x} : T \vdash \mathtt{x} : T}\ Ax \qquad \frac{\Gamma, \mathtt{x} : T_1 \vdash \mathtt{s} : T_2}{\Gamma \vdash \lambda \mathtt{x}.\mathtt{s} : T_1 \multimap T_2} \multimap I$$

$$\frac{\langle \Gamma \rangle^d \vdash \mathtt{s} : T}{\Gamma \vdash \blacktriangle^d \mathtt{s} : \Box^d T}\ \Box^d I \qquad \frac{\Gamma[\langle \mathtt{x} : T_1 \rangle^d] \vdash \mathtt{t} : T_2 \qquad \Delta \vdash \mathtt{s} : \Diamond^d T_1}{\Gamma[\Delta] \vdash \mathtt{t}[\mathtt{x} \mapsto \triangledown^d \mathtt{s}] : T_2}\ \Diamond^d E$$

$$\frac{\langle \Gamma, \langle \mathtt{x} : T_1 \rangle^{\mathsf{X}}, \Delta \rangle^d \vdash \mathtt{s} : T_2}{\langle \Gamma, \Delta \rangle^d, \langle \mathtt{x} : T_1 \rangle^{\mathsf{X}} \vdash \mathtt{s} : T_2}\ \mathsf{X}$$

..and its term calculus

$$\frac{}{\mathtt{c} : T \vdash \mathtt{c} : T}\ Lex \qquad \frac{\Gamma \vdash \mathtt{s} : T_1 \multimap T_2 \quad \Delta \vdash \mathtt{t} : T_1}{\Gamma, \Delta \vdash \mathtt{s}\ \mathtt{t} : T_2} \multimap E$$

$$\frac{\Gamma \vdash \mathtt{t} : T}{\langle \Gamma \rangle^d \vdash \triangle^d\, \mathtt{t} : \Diamond^d T}\ \Diamond^d I \qquad \frac{\Gamma \vdash \mathtt{s} : \Box^d T}{\langle \Gamma \rangle^d \vdash \blacktriangledown^d \mathtt{s} : T}\ \Box^d E$$

$$\frac{}{\mathtt{x} : T \vdash \mathtt{x} : T}\ Ax \qquad \frac{\Gamma, \mathtt{x} : T_1 \vdash \mathtt{s} : T_2}{\Gamma \vdash \lambda \mathtt{x}.\mathtt{s} : T_1 \multimap T_2} \multimap I$$

$$\frac{\langle \Gamma \rangle^d \vdash \mathtt{s} : T}{\Gamma \vdash \blacktriangle^d \mathtt{s} : \Box^d T}\ \Box^d I \qquad \frac{\Gamma[\langle \mathtt{x} : T_1 \rangle^d] \vdash \mathtt{t} : T_2 \quad \Delta \vdash \mathtt{s} : \Diamond^d T_1}{\Gamma[\Delta] \vdash \mathtt{t}[\mathtt{x} \mapsto \triangledown^d \mathtt{s}] : T_2}\ \Diamond^d E$$

$$\frac{\langle \Gamma, \langle \mathtt{x} : T_1 \rangle^{\mathsf{X}}, \Delta \rangle^d \vdash \mathtt{s} : T_2}{\langle \Gamma, \Delta \rangle^d, \langle \mathtt{x} : T_1 \rangle^{\mathsf{X}} \vdash \mathtt{s} : T_2}\ \mathsf{X}$$

Example

example too big, send help

Now what?

The standard categorial pipeline:

- read a sentence
- assign a type to each word
- perform (a) phrasal composition
- ???
- profit

Now what?

The standard categorial pipeline:

- read a sentence
- assign a type to each word
- perform (a) phrasal composition
- ???
- profit

(pre-)history

$p(t_1, \ldots t_n \mid w_1, \ldots w_n) \approx$

- $\prod_i^n (t_i \mid w_i)$
  co-occurrence-based statistical models (90s)
- $\prod_i^n (t_i \mid w_{i-\kappa} \ldots w_{i+\kappa})$
  window-based n-gram models (00s), ffns (early 10s)
- $\prod_i^n (t_i \mid w_1, \ldots w_n)$
  sequence encoders (mid 10s)
- $\prod_i^n (t_i \mid t_1, \ldots t_{i-1}, w_1, \ldots w_n)$
  seq2seq (late 10s)

$\square^{det}(n \multimap np)$	n	$*$	np	$\Diamond^{obj} pron \multimap \Diamond^{su} np \multimap ssub$
the	role	that	types	play

$* := \Diamond^{relcl}(\Diamond^{obj1} pron \multimap ssub) \multimap \square^{mod}(np \multimap np)$

(pre-)history

$p(t_1, \ldots t_n \mid w_1, \ldots w_n) \approx$

- $\prod_i^n (t_i \mid w_i)$
  co-occurrence-based statistical models (90s)
- $\prod_i^n (t_i \mid w_{i-\kappa} \ldots w_{i+\kappa})$
  window-based n-gram models (00s), ffns (early 10s)
- $\prod_i^n (t_i \mid w_1, \ldots w_n)$
  sequence encoders (mid 10s)
- $\prod_i^n (t_i \mid t_1, \ldots t_{i-1}, w_1, \ldots w_n)$
  seq2seq (late 10s)

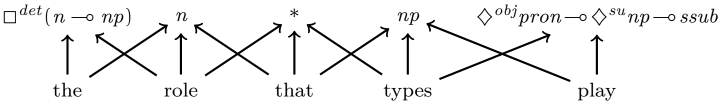
$\square^{det}(n \multimap np)$	n	$*$	np	$\diamond^{obj} pron \multimap \diamond^{su} np \multimap ssub$
↑	↑	↑	↑	↑
the	role	that	types	play

$* := \diamond^{relcl}(\diamond^{obj1} pron \multimap ssub) \multimap \square^{mod}(np \multimap np)$

(pre-)history

$p(t_1, \ldots t_n \mid w_1, \ldots w_n) \approx$

- $\prod_i^n (t_i \mid w_i)$
  co-occurrence-based statistical models (90s)
- $\prod_i^n (t_i \mid w_{i-\kappa} \ldots w_{i+\kappa})$
  window-based n-gram models (00s), ffns (early 10s)
- $\prod_i^n (t_i \mid w_1, \ldots w_n)$
  sequence encoders (mid 10s)
- $\prod_i^n (t_i \mid t_1, \ldots t_{i-1}, w_1, \ldots w_n)$
  seq2seq (late 10s)



$* := \Diamond^{relcl}(\Diamond^{obj1} pron \multimap ssub) \multimap \Box^{mod}(np \multimap np)$

(pre-)history

$p(t_1, \ldots t_n \mid w_1, \ldots w_n) \approx$

- $\prod_i^n (t_i \mid w_i)$
  co-occurrence-based statistical models (90s)
- $\prod_i^n (t_i \mid w_{i-\kappa} \ldots w_{i+\kappa})$
  window-based n-gram models (00s), ffns (early 10s)
- $\prod_i^n (t_i \mid w_1, \ldots w_n)$
  sequence encoders (mid 10s)
- $\prod_i^n (t_i \mid t_1, \ldots t_{i-1}, w_1, \ldots w_n)$
  seq2seq (late 10s)

$\Box^{det}(n \multimap np)$ $\quad n \quad * \quad np \quad \Diamond^{obj} pron \multimap \Diamond^{su} np \multimap ssub$

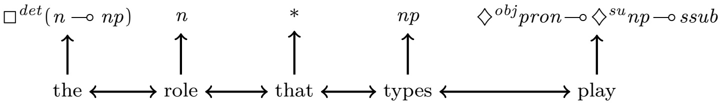
the ⟷ role ⟷ that ⟷ types ⟷ play

$* := \Diamond^{relcl}(\Diamond^{obj1} pron \multimap ssub) \multimap \Box^{mod}(np \multimap np)$

(pre-)history

$p(t_1, \ldots t_n \mid w_1, \ldots w_n) \approx$

- $\prod_i^n (t_i \mid w_i)$
  co-occurrence-based statistical models (90s)
- $\prod_i^n (t_i \mid w_{i-\kappa} \ldots w_{i+\kappa})$
  window-based n-gram models (00s), ffns (early 10s)
- $\prod_i^n (t_i \mid w_1, \ldots w_n)$
  sequence encoders (mid 10s)
- $\prod_i^n (t_i \mid t_1, \ldots t_{i-1}, w_1, \ldots w_n)$
  seq2seq (late 10s)



$* := \Diamond^{relcl}(\Diamond^{obj1} pron \multimap ssub) \multimap \Box^{mod}(np \multimap np)$

(pre-)history

$$p(t_1, \ldots t_n \mid w_1, \ldots w_n) \approx$$

- $\prod_i^n (t_i \mid w_i)$
  co-occurrence-based statistical models (90s)
- $\prod_i^n (t_i \mid w_{i-\kappa} \ldots w_{i+\kappa})$
  window-based n-gram models (00s), ffns (early 10s)
- $\prod_i^n (t_i \mid w_1, \ldots w_n)$
  sequence encoders (mid 10s)
- $\prod_i^n (t_i \mid t_1, \ldots t_{i-1}, w_1, \ldots w_n)$
  seq2seq (late 10s)

$\Box^{det}(n \multimap np)$ $\quad n \quad * \quad np \quad \Diamond^{obj} pron \multimap \Diamond^{su} np \multimap ssub$

$\uparrow$

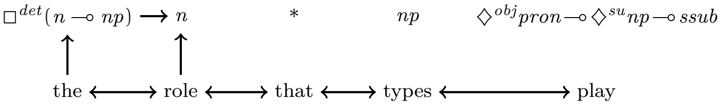
the $\longleftrightarrow$ role $\longleftrightarrow$ that $\longleftrightarrow$ types $\longleftrightarrow$ play

$$* := \Diamond^{relcl}(\Diamond^{obj1} pron \multimap ssub) \multimap \Box^{mod}(np \multimap np)$$

(pre-)history

$p(t_1, \ldots t_n \mid w_1, \ldots w_n) \approx$

- $\prod_i^n (t_i \mid w_i)$
  co-occurrence-based statistical models (90s)
- $\prod_i^n (t_i \mid w_{i-\kappa} \ldots w_{i+\kappa})$
  window-based n-gram models (00s), ffns (early 10s)
- $\prod_i^n (t_i \mid w_1, \ldots w_n)$
  sequence encoders (mid 10s)
- $\prod_i^n (t_i \mid t_1, \ldots t_{i-1}, w_1, \ldots w_n)$
  seq2seq (late 10s)

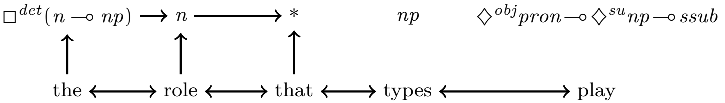


$* := \diamond^{relcl}(\diamond^{obj1} pron \multimap ssub) \multimap \square^{mod}(np \multimap np)$

(pre-)history

$p(t_1, \ldots t_n \mid w_1, \ldots w_n) \approx$

- $\prod_i^n (t_i \mid w_i)$
  co-occurrence-based statistical models (90s)
- $\prod_i^n (t_i \mid w_{i-\kappa} \ldots w_{i+\kappa})$
  window-based n-gram models (00s), ffns (early 10s)
- $\prod_i^n (t_i \mid w_1, \ldots w_n)$
  sequence encoders (mid 10s)
- $\prod_i^n (t_i \mid t_1, \ldots t_{i-1}, w_1, \ldots w_n)$
  seq2seq (late 10s)

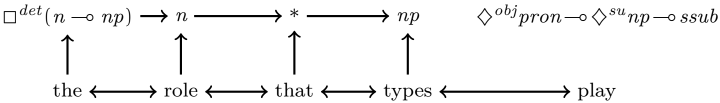


$* := \Diamond^{relcl}(\Diamond^{obj1} pron \multimap ssub) \multimap \Box^{mod}(np \multimap np)$

(pre-)history

$p(t_1, \ldots t_n \mid w_1, \ldots w_n) \approx$

- $\prod_i^n (t_i \mid w_i)$
  co-occurrence-based statistical models (90s)
- $\prod_i^n (t_i \mid w_{i-\kappa} \ldots w_{i+\kappa})$
  window-based n-gram models (00s), ffns (early 10s)
- $\prod_i^n (t_i \mid w_1, \ldots w_n)$
  sequence encoders (mid 10s)
- $\prod_i^n (t_i \mid t_1, \ldots t_{i-1}, w_1, \ldots w_n)$
  seq2seq (late 10s)

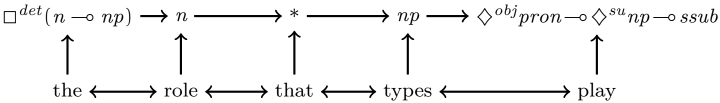


$* := \Diamond^{relcl}(\Diamond^{obj1} pron \multimap ssub) \multimap \Box^{mod}(np \multimap np)$

(pre-)history

$p(t_1, \ldots t_n \mid w_1, \ldots w_n) \approx$

- $\prod_i^n (t_i \mid w_i)$
  co-occurrence-based statistical models (90s)
- $\prod_i^n (t_i \mid w_{i-\kappa} \ldots w_{i+\kappa})$
  window-based n-gram models (00s), ffns (early 10s)
- $\prod_i^n (t_i \mid w_1, \ldots w_n)$
  sequence encoders (mid 10s)
- $\prod_i^n (t_i \mid t_1, \ldots t_{i-1}, w_1, \ldots w_n)$
  seq2seq (late 10s)



$* := \Diamond^{relcl}(\Diamond^{obj1} pron \multimap ssub) \multimap \Box^{mod}(np \multimap np)$

(pre-)history

$p(t_1, \ldots t_n \mid w_1, \ldots w_n) \approx$

- $\prod_i^n (t_i \mid w_i)$
  co-occurrence-based statistical models (90s)
- $\prod_i^n (t_i \mid w_{i-\kappa} \ldots w_{i+\kappa})$
  window-based n-gram models (00s), ffns (early 10s)
- $\prod_i^n (t_i \mid w_1, \ldots w_n)$
  sequence encoders (mid 10s)
- $\prod_i^n (t_i \mid t_1, \ldots t_{i-1}, w_1, \ldots w_n)$
  seq2seq (late 10s)

what have we done?

- more arrows (=more context)
- auto-regression (price: temporal delay)
- what about the co-domain?

Intermezzo: the curse(?) of sparsity

The majority of unique categories in standard datasets are rare

the "*fix*": ignore rare categories

- small penalty in accuracy
- less so for coverage..
- meta: sparse grammars = bad

the **fix**: decompose categories & build them up during decoding

- ↯ unlimited ~~power~~ generalization
- meta: sparse grammars = ok

Intermezzo: the curse(?) of sparsity

The majority of unique categories in standard datasets are rare

the "*fix*": ignore rare categories

- small penalty in accuracy
- less so for coverage..
- meta: sparse grammars = bad

the **fix**: decompose categories & build them up during decoding

- ↯ unlimited ~~power~~ generalization
- meta: sparse grammars = ok

Intermezzo: the curse(?) of sparsity

The majority of unique categories in standard datasets are rare

the "*fix*": ignore rare categories

- small penalty in accuracy
- less so for coverage..
- meta: sparse grammars = bad

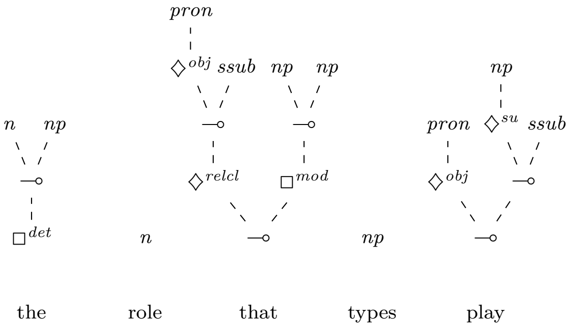
the **fix**: decompose categories & build them up during decoding

- ϟ unlimited ~~power~~ generalization
- meta: sparse grammars = ok

Modern Times

$p(\sigma_1, \ldots \sigma_m \mid w_1, \ldots w_n) \approx$

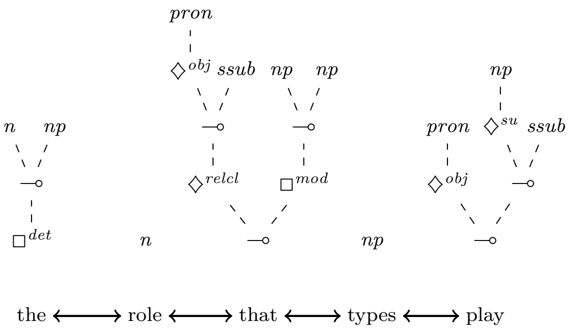
- $\prod_i^m (\sigma_i \mid \sigma_1, \ldots \sigma_{i-1}, w_1, \ldots w_n)$
  sequential constructive (w/ Moortgat & Deoskar, 2019)
- $\prod_i^m (\sigma_i \mid \mathrm{anc}(\sigma_i), w_1, \ldots w_n)$
  tree-recursive (Prange et. al 2020)



Modern Times

$p(\sigma_1, \ldots \sigma_m \mid w_1, \ldots w_n) \approx$

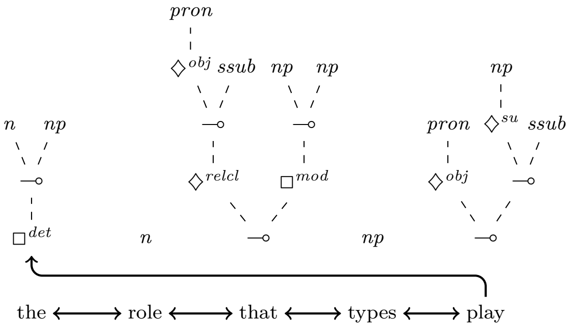
- $\prod_i^m (\sigma_i \mid \sigma_1, \ldots \sigma_{i-1}, w_1, \ldots w_n)$
  sequential constructive (w/ Moortgat & Deoskar, 2019)
- $\prod_i^m (\sigma_i \mid \text{anc}(\sigma_i), w_1, \ldots w_n)$
  tree-recursive (Prange et. al 2020)



Modern Times

$p(\sigma_1, \ldots \sigma_m \mid w_1, \ldots w_n) \approx$

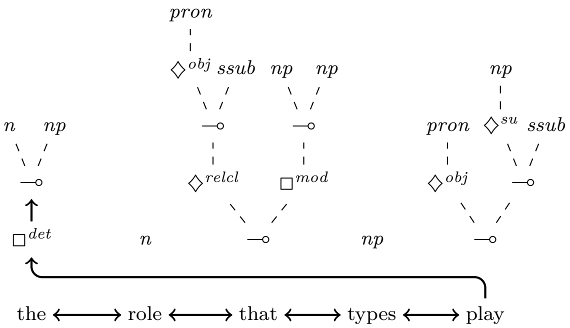
- $\prod_i^m (\sigma_i \mid \sigma_1, \ldots \sigma_{i-1}, w_1, \ldots w_n)$
  sequential constructive (w/ Moortgat & Deoskar, 2019)
- $\prod_i^m (\sigma_i \mid \mathrm{anc}(\sigma_i), w_1, \ldots w_n)$
  tree-recursive (Prange et. al 2020)



Modern Times

$p(\sigma_1, \ldots \sigma_m \mid w_1, \ldots w_n) \approx$

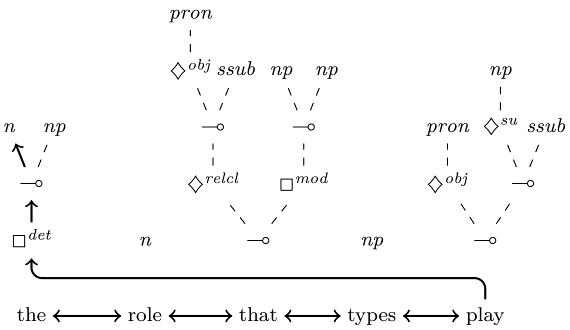
- $\prod_i^m (\sigma_i \mid \sigma_1, \ldots \sigma_{i-1}, w_1, \ldots w_n)$
  sequential constructive (w/ Moortgat & Deoskar, 2019)
- $\prod_i^m (\sigma_i \mid \text{anc}(\sigma_i), w_1, \ldots w_n)$
  tree-recursive (Prange et. al 2020)



Modern Times

$p(\sigma_1, \ldots \sigma_m \mid w_1, \ldots w_n) \approx$

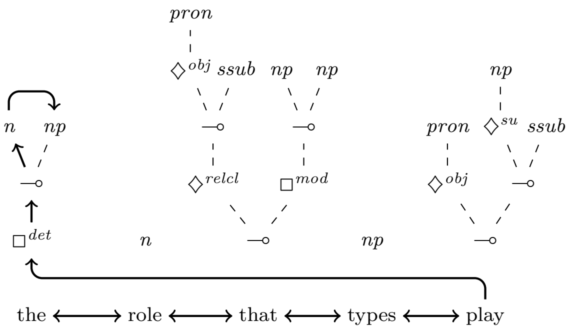
- $\prod_i^m (\sigma_i \mid \sigma_1, \ldots \sigma_{i-1}, w_1, \ldots w_n)$
  sequential constructive (w/ Moortgat & Deoskar, 2019)
- $\prod_i^m (\sigma_i \mid \mathrm{anc}(\sigma_i), w_1, \ldots w_n)$
  tree-recursive (Prange et. al 2020)



Modern Times

$p(\sigma_1, \ldots \sigma_m \mid w_1, \ldots w_n) \approx$

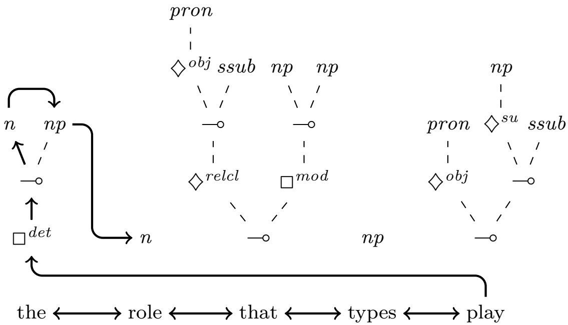
- $\prod_i^m(\sigma_i \mid \sigma_1, \ldots \sigma_{i-1}, w_1, \ldots w_n)$
  sequential constructive (w/ Moortgat & Deoskar, 2019)
- $\prod_i^m(\sigma_i \mid \mathrm{anc}(\sigma_i), w_1, \ldots w_n)$
  tree-recursive (Prange et. al 2020)



Modern Times

$p(\sigma_1, \ldots \sigma_m \mid w_1, \ldots w_n) \approx$

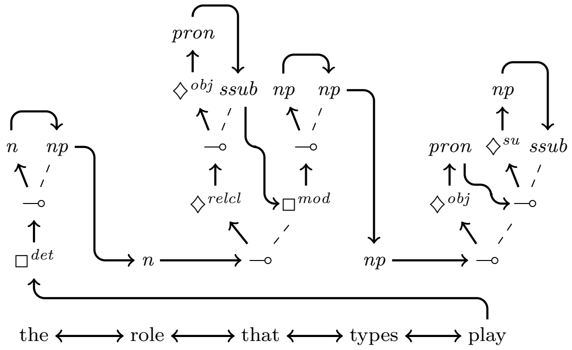
- $\prod_i^m (\sigma_i \mid \sigma_1, \ldots \sigma_{i-1}, w_1, \ldots w_n)$
  sequential constructive (w/ Moortgat & Deoskar, 2019)
- $\prod_i^m (\sigma_i \mid \mathrm{anc}(\sigma_i), w_1, \ldots w_n)$
  tree-recursive (Prange et. al 2020)



Modern Times

$p(\sigma_1, \ldots \sigma_m \mid w_1, \ldots w_n) \approx$

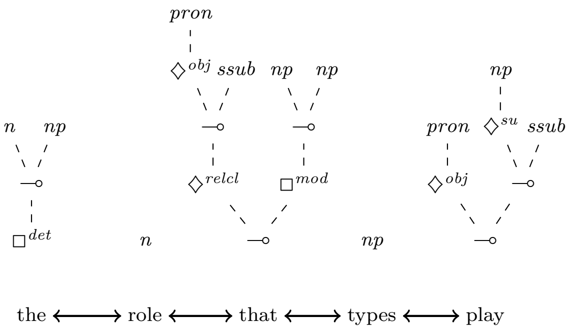
- $\prod_i^m(\sigma_i \mid \sigma_1, \ldots \sigma_{i-1}, w_1, \ldots w_n)$
  sequential constructive (w/ Moortgat & Deoskar, 2019)
- $\prod_i^m(\sigma_i \mid \text{anc}(\sigma_i), w_1, \ldots w_n)$
  tree-recursive (Prange et. al 2020)



Modern Times

$p(\sigma_1, \ldots \sigma_m \mid w_1, \ldots w_n) \approx$

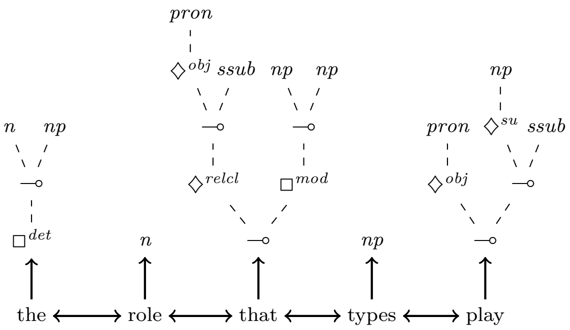
- $\prod_i^m(\sigma_i \mid \sigma_1, \ldots \sigma_{i-1}, w_1, \ldots w_n)$
  sequential constructive (w/ Moortgat & Deoskar, 2019)
- $\prod_i^m(\sigma_i \mid \mathrm{anc}(\sigma_i), w_1, \ldots w_n)$
  tree-recursive (Prange et. al 2020)



Modern Times

$p(\sigma_1, \ldots \sigma_m \mid w_1, \ldots w_n) \approx$

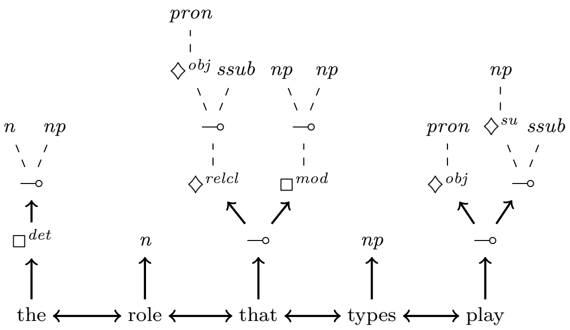
- $\prod_i^m (\sigma_i \mid \sigma_1, \ldots \sigma_{i-1}, w_1, \ldots w_n)$
  sequential constructive (w/ Moortgat & Deoskar, 2019)
- $\prod_i^m (\sigma_i \mid \text{anc}(\sigma_i), w_1, \ldots w_n)$
  tree-recursive (Prange et. al 2020)



Modern Times

$p(\sigma_1, \ldots \sigma_m \mid w_1, \ldots w_n) \approx$

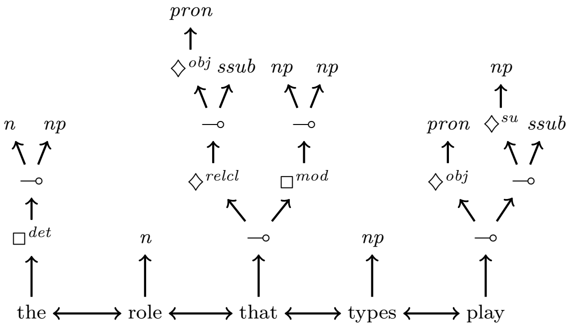
- $\prod_i^m(\sigma_i \mid \sigma_1, \ldots \sigma_{i-1}, w_1, \ldots w_n)$
  sequential constructive (w/ Moortgat & Deoskar, 2019)
- $\prod_i^m(\sigma_i \mid \text{anc}(\sigma_i), w_1, \ldots w_n)$
  tree-recursive (Prange et. al 2020)



Modern Times

$p(\sigma_1, \ldots \sigma_m \mid w_1, \ldots w_n) \approx$

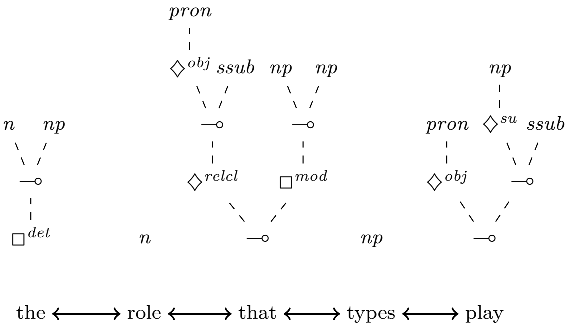
- $\prod_i^m (\sigma_i \mid \sigma_1, \ldots \sigma_{i-1}, w_1, \ldots w_n)$
  sequential constructive (w/ Moortgat & Deoskar, 2019)
- $\prod_i^m (\sigma_i \mid \text{anc}(\sigma_i), w_1, \ldots w_n)$
  tree-recursive (Prange et. al 2020)



Post-modernity

neither sequence nor tree but sequence of trees

$p(\sigma_1, \dots \sigma_m \mid w_1, \dots w_n) \approx \prod_i^m (\sigma_i \mid \sigma_j : \text{depth}(\sigma_j) < \text{depth}(\sigma_i), w_1, \dots w_n)$

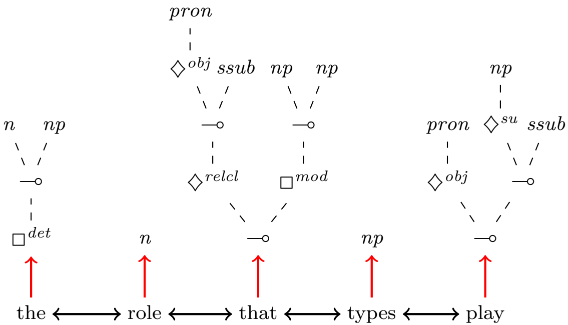


(encode)

Post-modernity

neither sequence nor tree but sequence of trees

$p(\sigma_1, \ldots \sigma_m \mid w_1, \ldots w_n) \approx \prod_i^m (\sigma_i \mid \sigma_j : \text{depth}(\sigma_j) < \text{depth}(\sigma_i), w_1, \ldots w_n)$

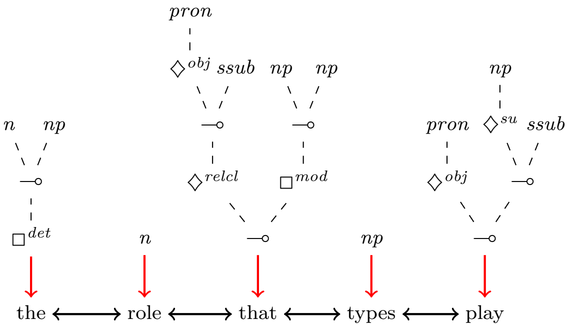


(predict)

Post-modernity

neither sequence nor tree but sequence of trees

$p(\sigma_1, \ldots \sigma_m \mid w_1, \ldots w_n) \approx \prod_i^m (\sigma_i \mid \sigma_j : \mathrm{depth}(\sigma_j) < \mathrm{depth}(\sigma_i), w_1, \ldots w_n)$

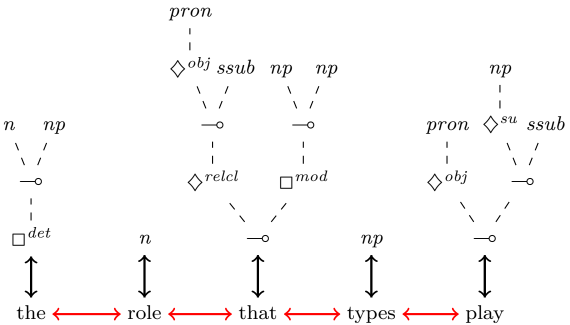


(feedback)

Post-modernity

neither sequence nor tree but sequence of trees

$p(\sigma_1, \ldots \sigma_m \mid w_1, \ldots w_n) \approx \prod_i^m (\sigma_i \mid \sigma_j : \text{depth}(\sigma_j) < \text{depth}(\sigma_i), w_1, \ldots w_n)$

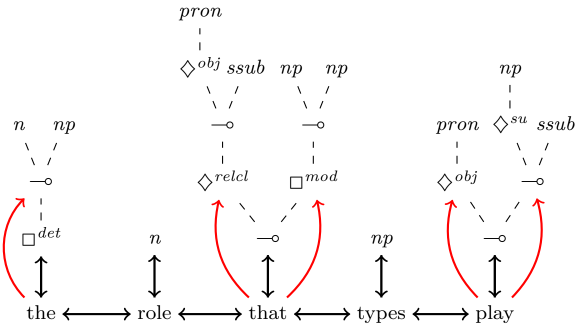


(contextualize)

Post-modernity

neither sequence nor tree but sequence of trees

$p(\sigma_1, \ldots \sigma_m \mid w_1, \ldots w_n) \approx \prod_i^m (\sigma_i \mid \sigma_j : \text{depth}(\sigma_j) < \text{depth}(\sigma_i), w_1, \ldots w_n)$

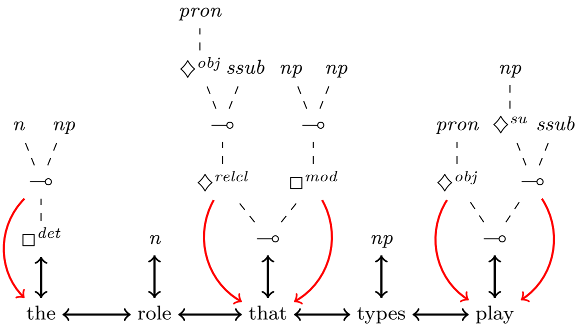


(predict)

Post-modernity

neither sequence nor tree but sequence of trees

$p(\sigma_1, \ldots \sigma_m \mid w_1, \ldots w_n) \approx \prod_i^m (\sigma_i \mid \sigma_j : \text{depth}(\sigma_j) < \text{depth}(\sigma_i), w_1, \ldots w_n)$

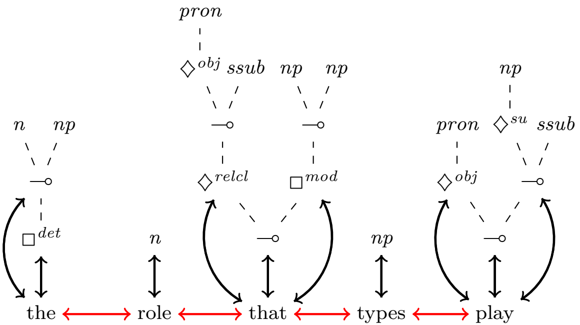


(feedback)

Post-modernity

neither sequence nor tree but sequence of trees

$p(\sigma_1, \dots \sigma_m \mid w_1, \dots w_n) \approx \prod_i^m (\sigma_i \mid \sigma_j : \text{depth}(\sigma_j) < \text{depth}(\sigma_i), w_1, \dots w_n)$

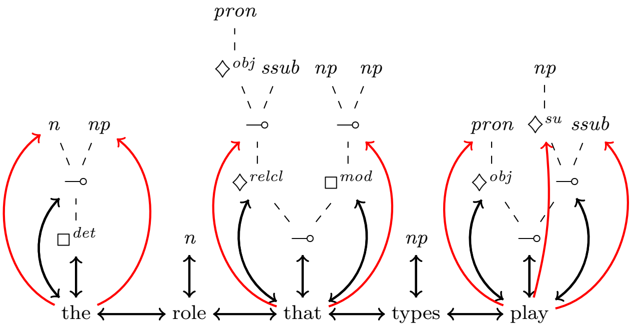


(contextualize)

Post-modernity

neither sequence nor tree but sequence of trees

$p(\sigma_1, \ldots \sigma_m \mid w_1, \ldots w_n) \approx \prod_i^m (\sigma_i \mid \sigma_j : \text{depth}(\sigma_j) < \text{depth}(\sigma_i), w_1, \ldots w_n)$



(predict)

DL Jargon

- *contextualize: states → states*
  universal transformer encoder w/ relative distance weights
  (many-to-many, update states with neighborhood context)
- *predict: state → nodes*
  token classification w/ unary tree node embeddings
  (one-to-many, predict fringe nodes from current state)
- *feedback: nodes → state*
  heterogeneous dynamic graph attention
  (many-to-one, update state with last predicted nodes)

Color coded summary

decoder	seq2seq[t]	seq2seq[σ]	tree	dynamic graph
codomain	fixed	open	constrained	constrained
context	left	preorder (global)	ancestors (local)	levels (global)
complexity	# words	# symbols	tree depth	tree depth
treeness	ignored	implicit	explicit	explicit
sequencess	explicit	misaligned	ignored	explicit
search?	✓	✓	✗	?

legend

- green = good
- yellow = meh
- red = bad

Table with numbers

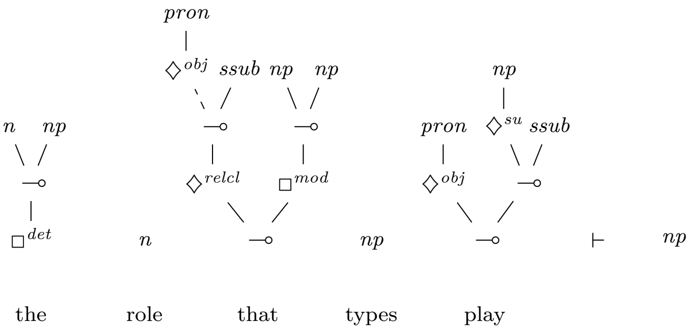
model	accuracy (%) overall	frequent	uncommon	rare	unseen
Æthel *(van Benthem calculus & dependency modalities, nl)*					
Sequential Transformer	83.67	84.55	64.70	50.58	24.55
this work	93.67	94.72	73.45	53.83	15.78
TLGBank *(Lambek calculus & control modalities, fr)*					
ELMo LSTM	93.20	95.10	75.19	25.85	–
this work	95.93	96.40	81.48	55.37	7.26
CCGbank *(Combinatory Categorial Grammar, en)*					
Sequential RNN	95.10	95.48	65.76	26.02	0.00
Tree Recursive	96.09	96.44	68.10	37.40	3.03
Attentive Convolutions	96.25	96.64	71.04	–	–
this work	96.29	96.61	72.06	34.45	4.55
CCGrebank *(ditto, improved version)*					
Sequential RNN	94.44	94.93	66.90	27.41	1.23
Tree Recursive	94.70	95.11	68.86	36.76	4.94
this work	95.07	95.45	71.40	37.19	3.70

Proof Nets 101

Proof Frame A bi-colored sequence of decomposed formula assignments.

- \+ (sub-)formulas we *have*
- \- (sub-)formulas we *need*

$\multimap$ *preserves* result and *inverts* argument polarity



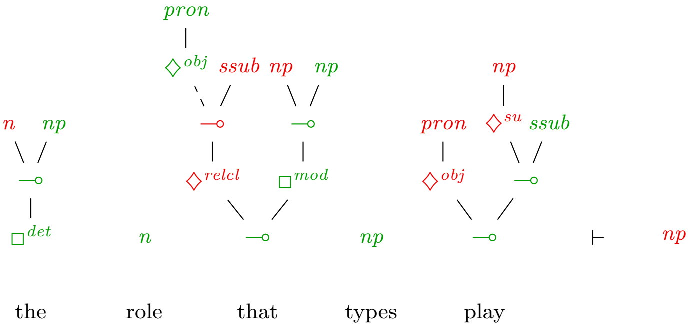
Proof Nets 101

Proof Frame A bi-colored sequence of decomposed formula assignments.

+ (sub-)formulas we *have*

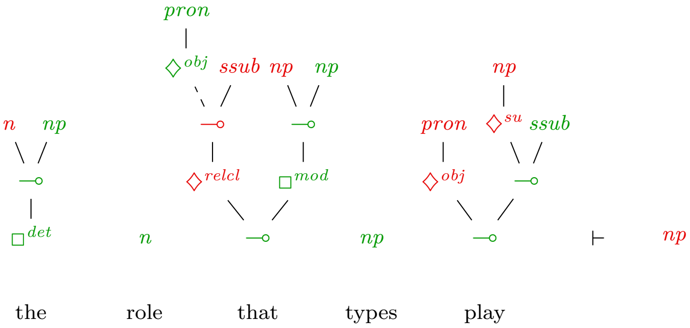
- (sub-)formulas we *need*

$\multimap$ *preserves* result and *inverts* argument polarity



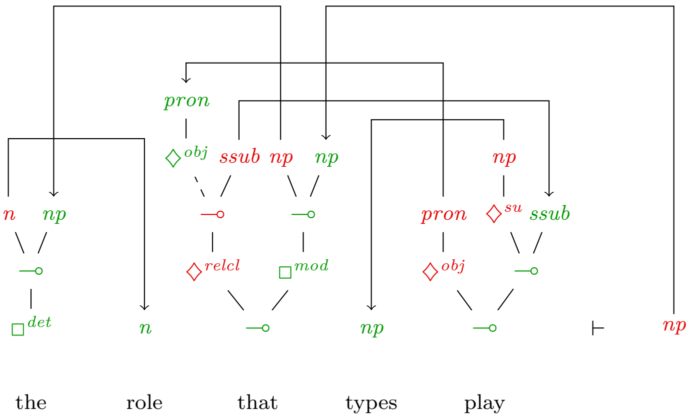
Proof Nets 101

Proof Structure A proof frame & a bijection between + and − atoms



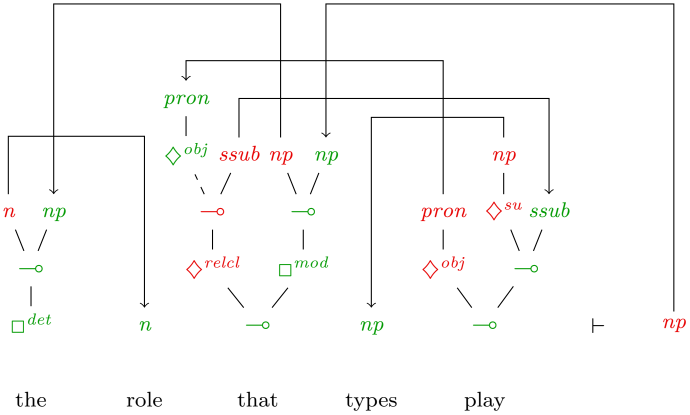
Proof Nets 101

Proof Structure A proof frame & a bijection between + and − atoms



Proof Nets 101

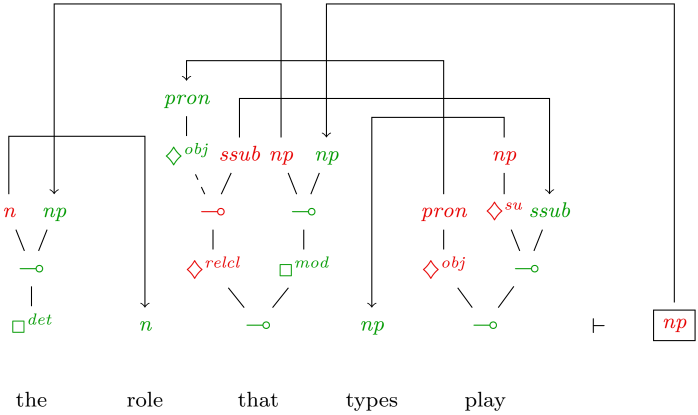
Proof Net
≡ proof, a proof structure you can navigate



Proof Nets 101

Proof Net

≡ proof, a proof structure you can navigate

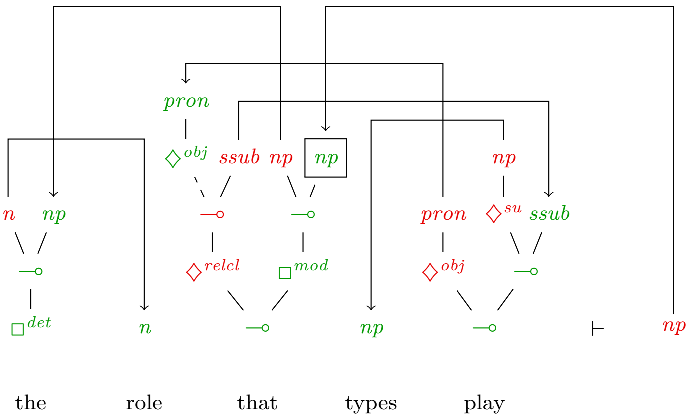


???

Proof Nets 101

Proof Net

≡ proof, a proof structure you can navigate

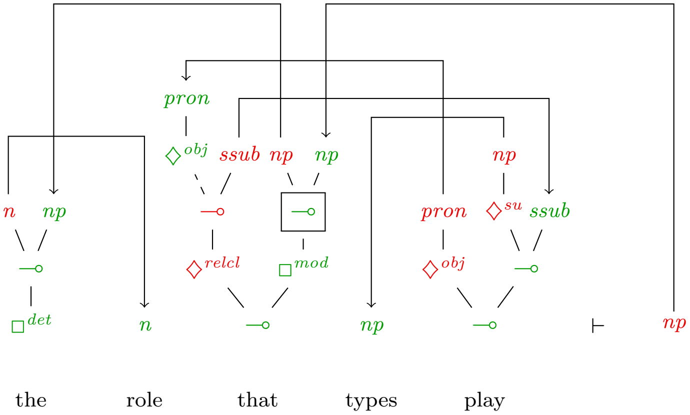


???

Proof Nets 101

Proof Net

≡ proof, a proof structure you can navigate

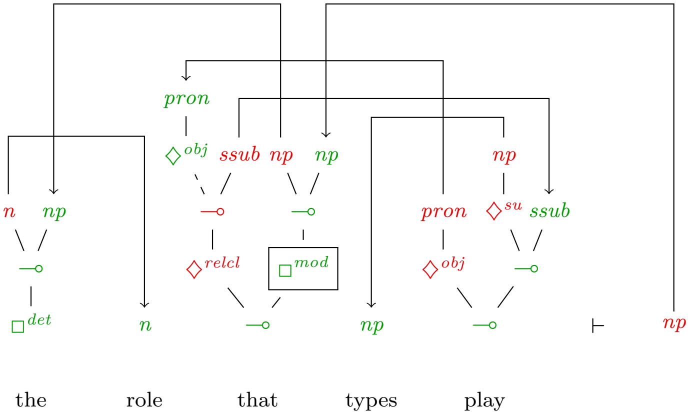


??? ???

Proof Nets 101

Proof Net

≡ proof, a proof structure you can navigate

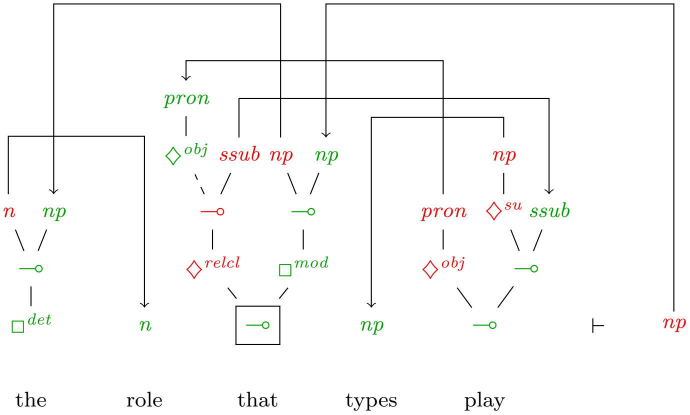


$\blacktriangledown^{mod}$(???) ???

Proof Nets 101

Proof Net

≡ proof, a proof structure you can navigate

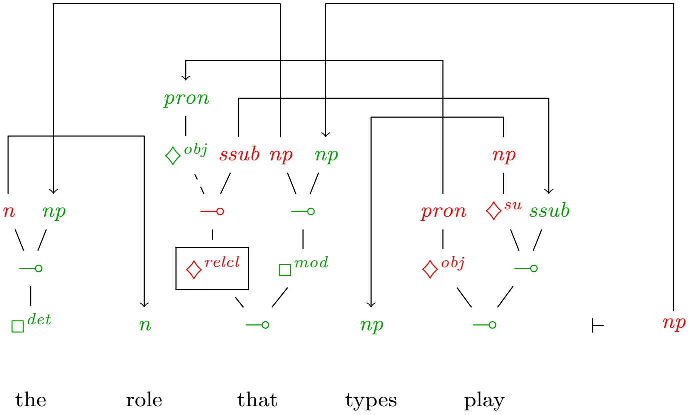


$\blacktriangledown^{mod}$(that ???) ???

Proof Nets 101

Proof Net

≡ proof, a proof structure you can navigate

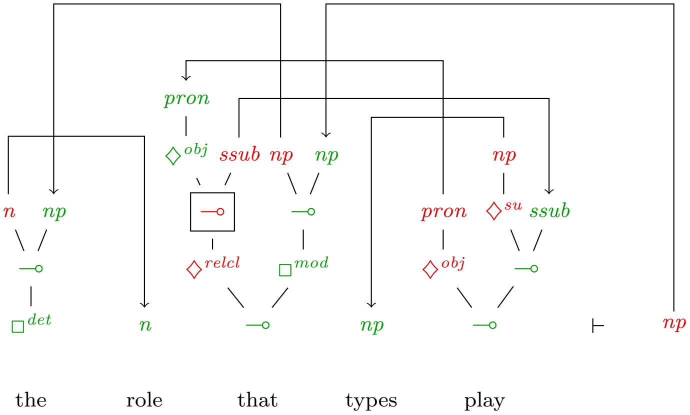


$\blacktriangledown^{mod}$(that $\triangle^{relcl}$(???)) ???

Proof Nets 101

Proof Net

≡ proof, a proof structure you can navigate

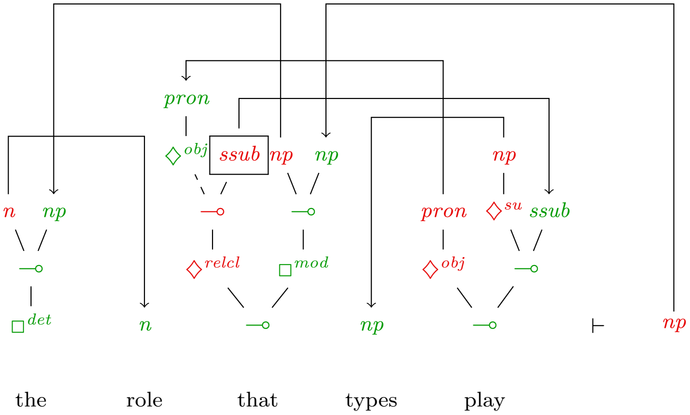


$\blacktriangledown^{mod}$(that $\triangle^{relcl}$ (λx.???)) ???

Proof Nets 101

Proof Net

≡ proof, a proof structure you can navigate

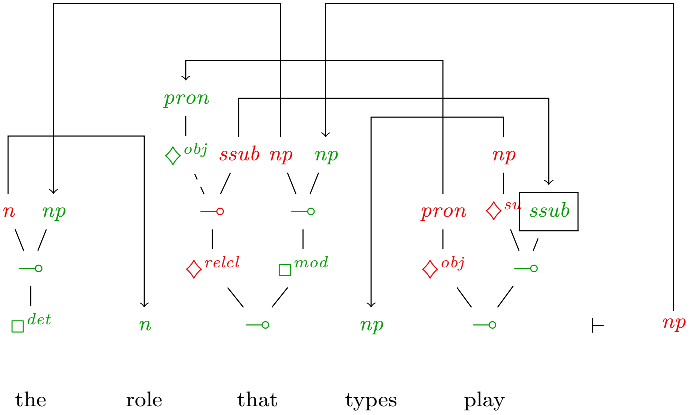


$\blacktriangledown^{mod}$(that $\vartriangle^{relcl}$ (λx.???)) ???

Proof Nets 101

Proof Net

≡ proof, a proof structure you can navigate

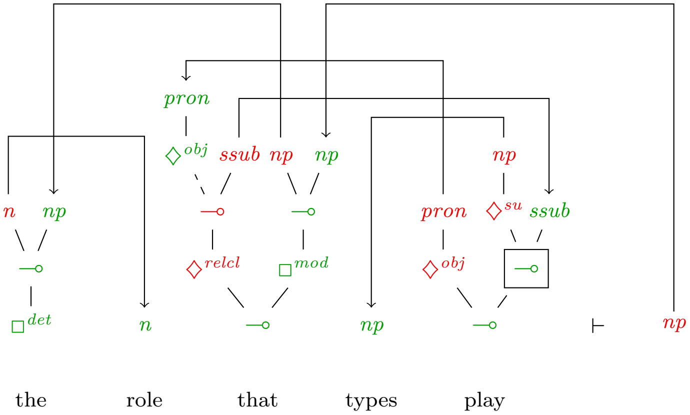


▼^{mod}(that △^{relcl} (λx.???)) ???

Proof Nets 101

Proof Net

≡ proof, a proof structure you can navigate

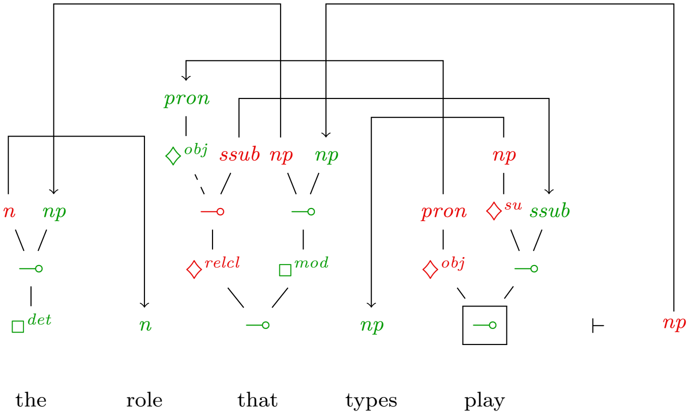


$\blacktriangledown^{mod}$(that $\vartriangle^{relcl}$ (λx.??? ???)) ???

Proof Nets 101

Proof Net

≡ proof, a proof structure you can navigate

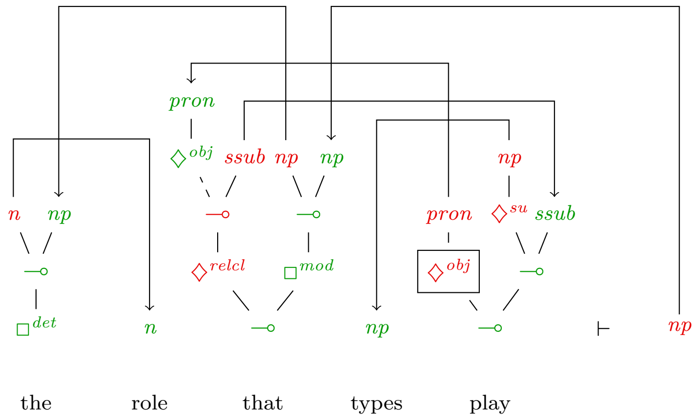


$\blacktriangledown^{mod}$(that $\triangle^{relcl}$ (λx.play ??? ???)) ???

Proof Nets 101

Proof Net

≡ proof, a proof structure you can navigate

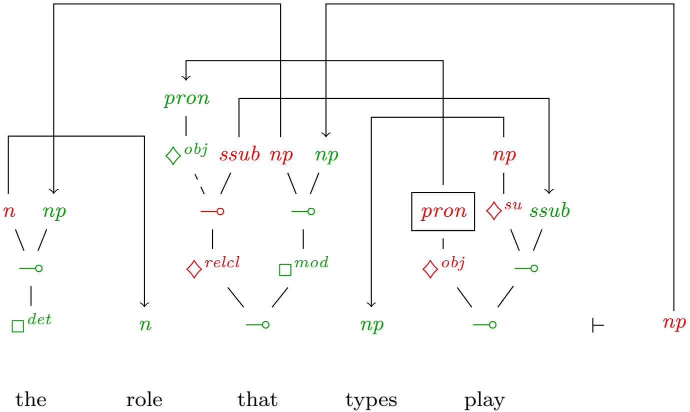


$\blacktriangledown^{mod}$(that $\triangle^{relcl}$ (λx.play $\triangledown^{obj}$??? ???)) ???

Proof Nets 101

Proof Net

≡ proof, a proof structure you can navigate

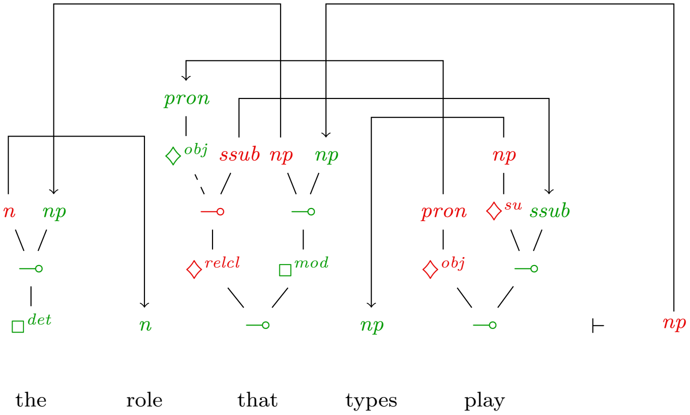


$\blacktriangledown^{mod}(\text{that}\ \triangle^{relcl}(\lambda x.\text{play}\ \triangledown^{obj}???\ ???))\ ???$

Proof Nets 101

Proof Net

≡ proof, a proof structure you can navigate

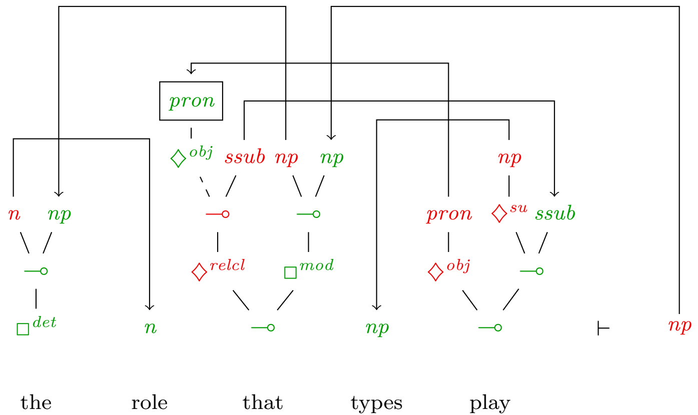


$\blacktriangledown^{mod}$(that $\triangle^{relcl}$ (λx.play $\triangledown^{obj}$??? ???)) ???

Proof Nets 101

Proof Net

≡ proof, a proof structure you can navigate

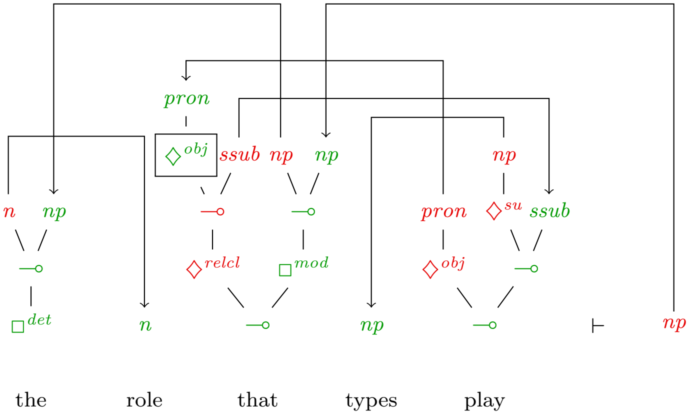


$\blacktriangledown^{mod}$(that $\triangle^{relcl}$ (λx.play $\triangledown^{obj}$??? ???)) ???

Proof Nets 101

Proof Net

≡ proof, a proof structure you can navigate

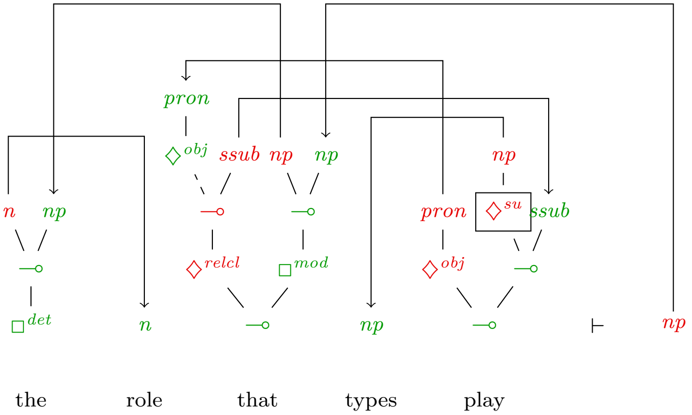


$\blacktriangledown^{mod}$(that $\triangle^{relcl}$ (λx.play $\triangle^{obj}$ $\triangledown^{obj}$x ???)) ???

Proof Nets 101

Proof Net

≡ proof, a proof structure you can navigate

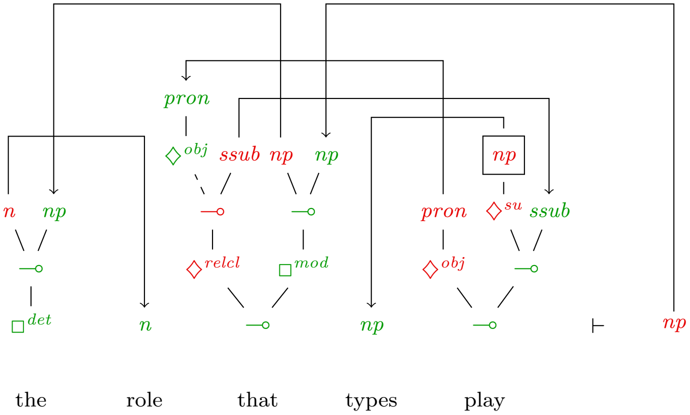


$\blacktriangledown^{mod}(\text{that}\ \ \triangle^{relcl}(\lambda x.\text{play}\ \ \triangle^{obj}\ \triangledown^{obj}x\ \ \triangle^{su}???))\ ???$

Proof Nets 101

Proof Net

≡ proof, a proof structure you can navigate

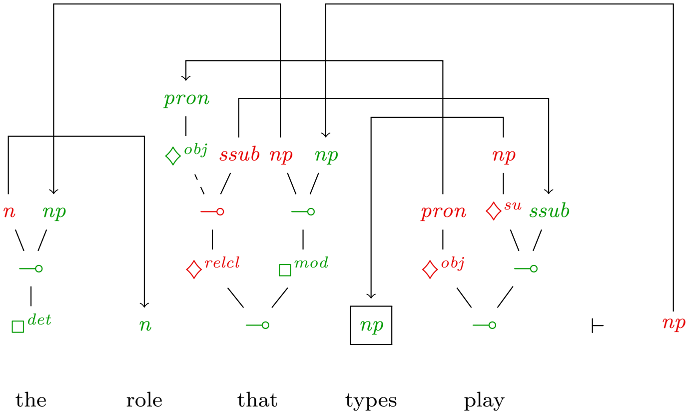


$\blacktriangledown^{mod}$(that $\triangle^{relcl}$ (λx.play $\triangle^{obj}$ $\triangledown^{obj}$x $\triangle^{su}$???)) ???

Proof Nets 101

Proof Net

≡ proof, a proof structure you can navigate

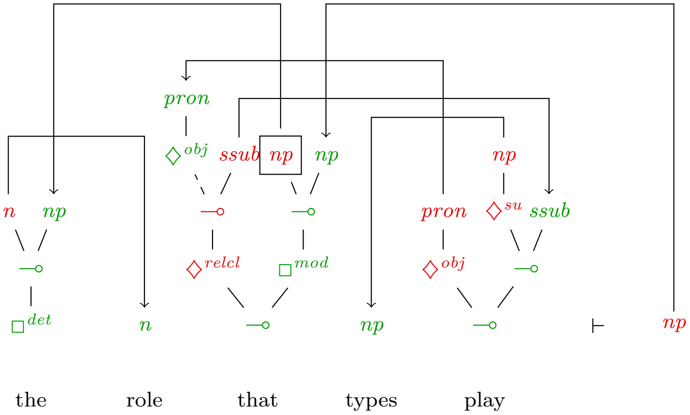


$\blacktriangledown^{mod}$(that $\triangle^{relcl}$ (λx.play $\triangle^{obj}$ $\triangledown^{obj}$x $\triangle^{su}$ types)) ???

Proof Nets 101

Proof Net

≡ proof, a proof structure you can navigate

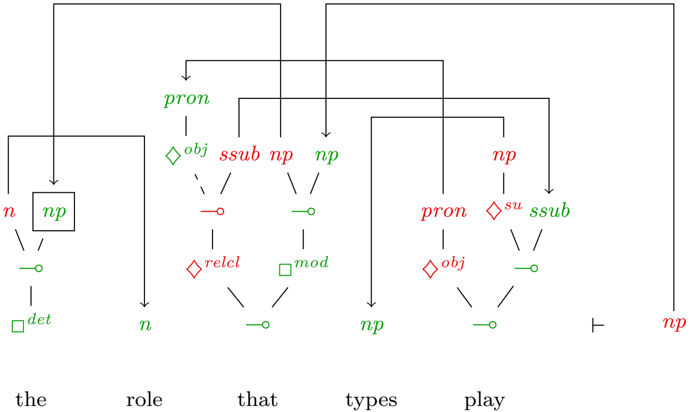


$\blacktriangledown^{mod}$(that $\triangle^{relcl}$ (λx.play $\triangle^{obj}$ $\triangledown^{obj}$x $\triangle^{su}$ types)) ???

Proof Nets 101

Proof Net

≡ proof, a proof structure you can navigate

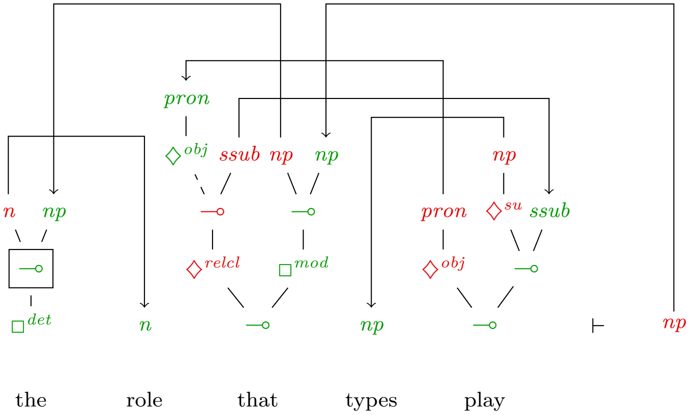


$\blacktriangledown^{mod}$(that $\triangle^{relcl}$ (λx.play $\triangle^{obj}$ $\triangledown^{obj}$x $\triangle^{su}$ types)) ???

Proof Nets 101

Proof Net

≡ proof, a proof structure you can navigate

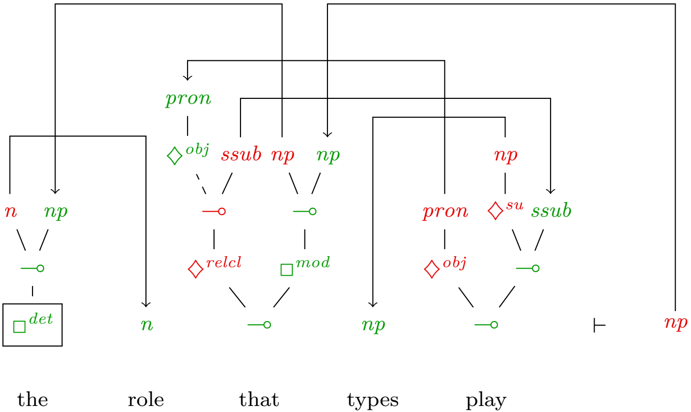


$\blacktriangledown^{mod}$(that $\triangle^{relcl}$ (λx.play $\triangle^{obj}$ $\triangledown^{obj}$x $\triangle^{su}$ types)) (??? ???)

Proof Nets 101

Proof Net

≡ proof, a proof structure you can navigate

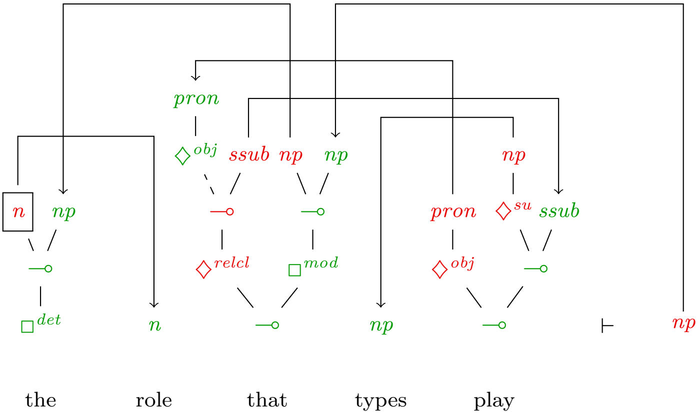


$\blacktriangledown^{mod}(\text{that } \triangle^{relcl}(\lambda x.\text{play } \triangle^{obj} \triangledown^{obj} x \ \triangle^{su} \text{types})) \ (\blacktriangledown^{det}\text{the ???})$

Proof Nets 101

Proof Net

≡ proof, a proof structure you can navigate

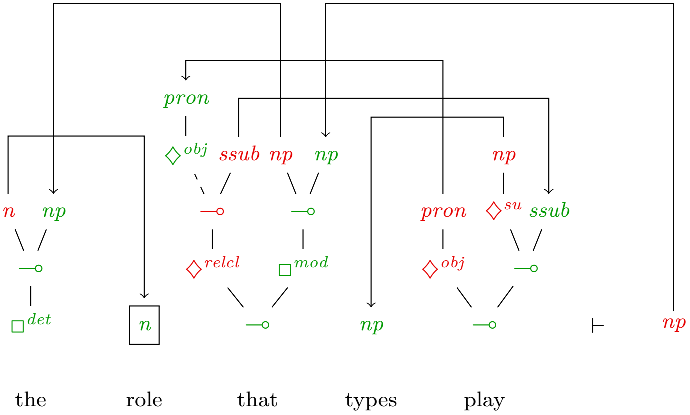


$\blacktriangledown^{mod}(\text{that}\ \vartriangle^{relcl}(\lambda x.\text{play}\ \vartriangle^{obj}\triangledown^{obj}x\ \vartriangle^{su}\text{types}))\ (\blacktriangledown^{det}\text{the ???})$

Proof Nets 101

Proof Net

≡ proof, a proof structure you can navigate

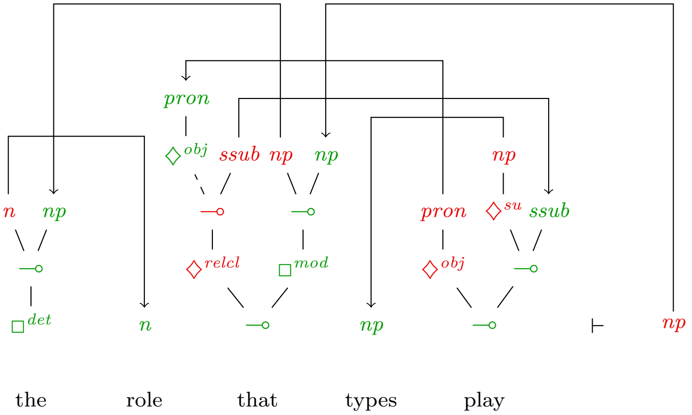


$\blacktriangledown^{mod}$(that $\triangle^{relcl}$ (λx.play $\triangle^{obj}$ $\triangledown^{obj}$x $\triangle^{su}$ types)) ($\blacktriangledown^{det}$the role)

Proof Nets 101

Proof Net

≡ proof, a proof structure you can navigate

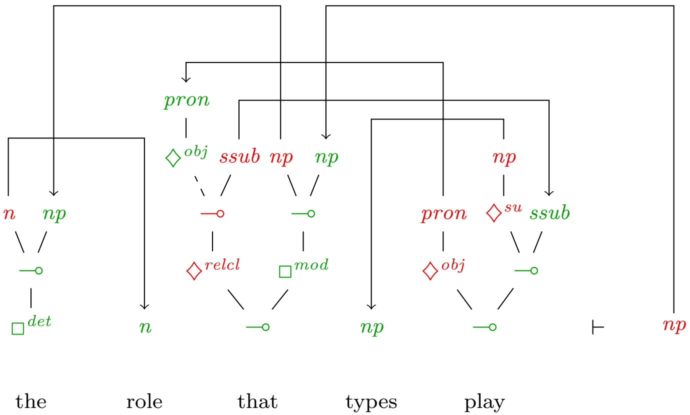


$\overset{\eta}{\rightsquigarrow}\ \blacktriangledown^{mod}(\text{that}\ \ \triangle^{relcl}\ \lambda \text{x}.(\text{play x}\ \ \triangle^{su}\ \text{types}))\ \ (\blacktriangledown^{det}\text{the role})$

Proof Nets 101

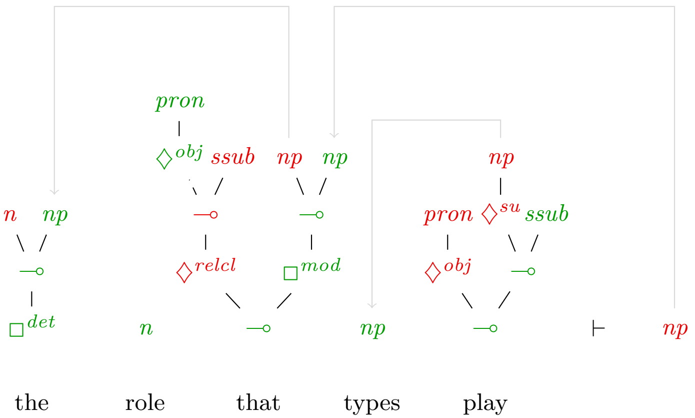
Proof Net

≡ proof, a proof structure you can navigate

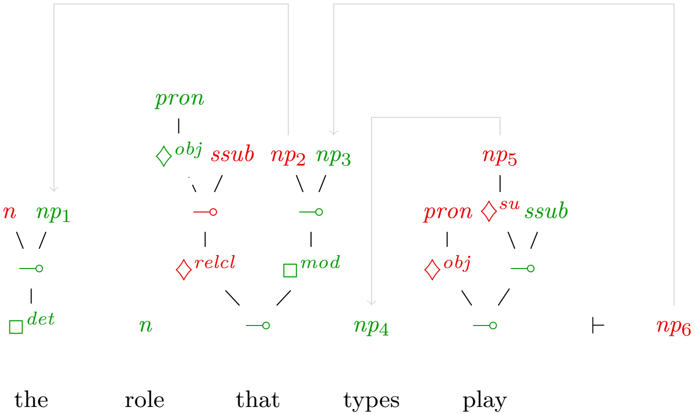


the bad news: (# *atoms*)! combinations to consider

Proof Nets 102: neural this time



Proof Nets 102: neural this time



Proof Nets 102: neural this time

Goal

	np_2	np_5	np_6
np_1	✗		
np_3			✗
np_4		✗	

Proof Nets 102: neural this time

Cost

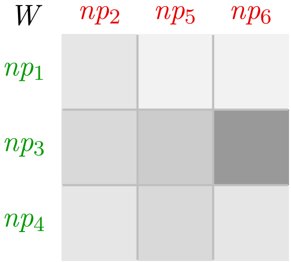
	np_2	np_5	np_6
np_1			
np_3			
np_4			

Goal

	np_2	np_5	np_6
np_1	✗		
np_3			✗
np_4		✗	

Proof Nets 102: neural this time

Cost

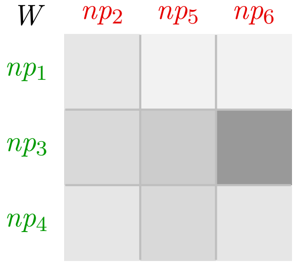


Goal

	np_2	np_5	np_6
np_1	✗		
np_3			✗
np_4		✗	

Proof Nets 102: neural this time

Cost



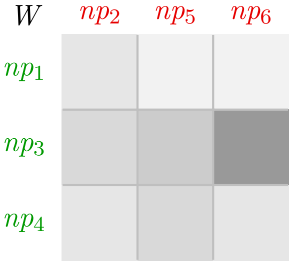
Goal

	np_2	np_5	np_6
np_1	✗		
np_3			✗
np_4		✗	

LAP: Find bijection $f : P \to N$ s.t. $\sum_{p \in P} Cost(p, f(p))$ max.

Proof Nets 102: neural this time

Cost



Goal

	np_2	np_5	np_6
np_1	✗		
np_3			✗
np_4		✗	

LAP: Find bijection $f\colon P \to N$ s.t. $\sum_{p \in P} Cost\,(p, f(p))$ max.

? boundedness

? backprop

Proof Nets 102: neural this time

Sinkhorn-Knopp

iterative row/column-wise normalization $\rightsquigarrow$ bistochasticity

in the log scale:

$$\mathsf{LSE} : \mathbb{R}^n \to \mathbb{R}$$

$$\mathsf{LSE}(\mathbf{x}) = x^* + log\left(\sum_i \mathsf{exp}(x_i - x^*)\right), \quad x^* := \mathsf{max}(\mathbf{x})$$

$$\mathsf{norm}_1 : \mathbb{R}^n \to \mathbb{R}^n$$

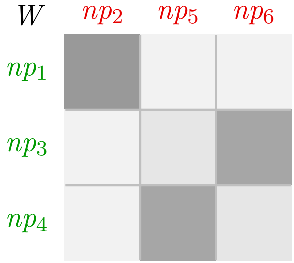
$$\mathsf{norm}_1(\mathbf{x}) = \mathbf{x} - \mathsf{LSE}(\mathbf{x})$$

$$\mathsf{norm}_2 : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$$

$$\mathsf{norm}_2(X) = \mathsf{norm}_1\left(\mathsf{norm}_1(X)^\top\right)^\top$$

Proof Nets 102: neural this time

Cost



Goal

	np_2	np_5	np_6
np_1	✗		
np_3			✗
np_4		✗	

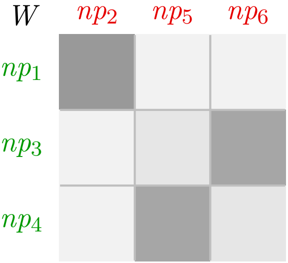
LAP: Find bijection $f\colon P \to N$ s.t. $\sum_{p \in P} Cost\,(p, f(p))$ max.

✓ boundedness : *negative in the log scale*

? backprop

Proof Nets 102: neural this time

Cost



Goal

	np_2	np_5	np_6
np_1	✗		
np_3			✗
np_4		✗	

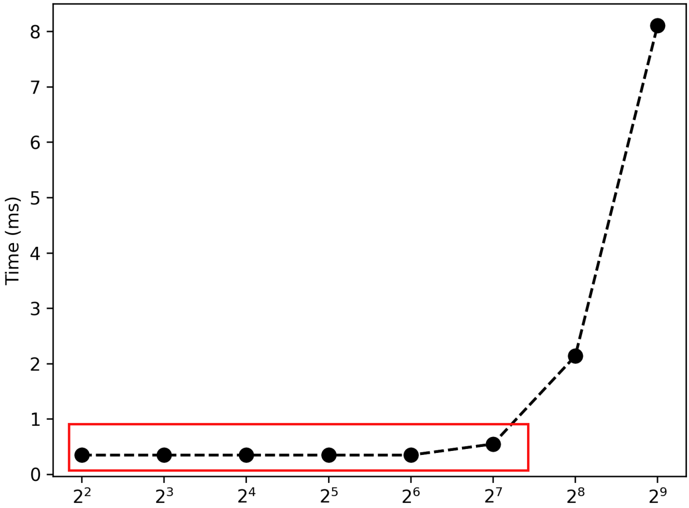
LAP: Find bijection $f\colon P \rightarrow N$ s.t. $\sum_{p \in P} Cost(p, f(p))$ max.

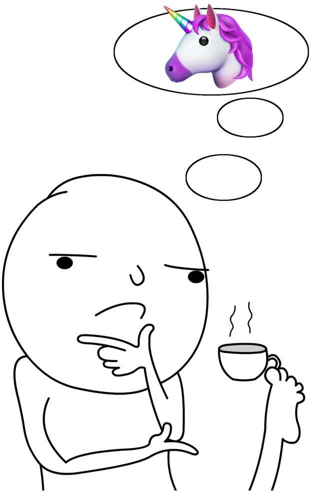
✓ boundedness : *negative in the log scale*

✓ backprop : *NLL /w straight-through estimator*

A note on complexity

Forward pass of 64 matrix-batches, 3 Sinkhorn iterations





constant decoding + constant linking = ???