Constructive Type-Logical Supertagging with Self-Attention Networks

Konstantinos Kogkalidis Michael Moortgat Tejaswini Deoskar

Utrecht University

Dutch Research Council

Supertagging $\sim$ "almost parsing"

Assigning categorial types to words in context

Problems with established practice (RNN-based sequence classification)

Fixed set of labels $\implies$ can't predict **unseen types**

Class imbalance $\implies$ trouble predicting **rare types**

Type-Logical Grammars

Words are typed variables of a functional program

- Constants A
  $\{\mathsf{NP}, \mathsf{S}, \dots\}$
- Functions carrying **dependency** information
  $\{\mathsf{NP}_{su} \to \mathsf{S}, \mathsf{NP}_{obj} \to (\mathsf{NP}_{su} \to \mathsf{S}), \mathsf{S}_{mod} \to \mathsf{S}, \dots\}$

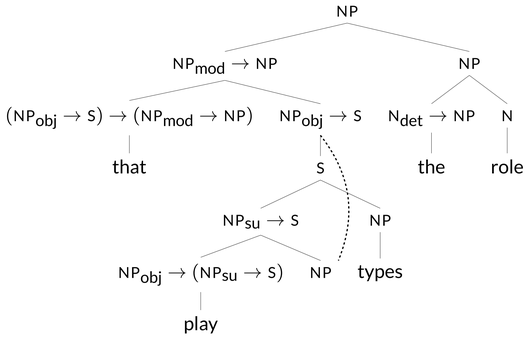
Type Syntax Inductive Scheme $\equiv$ **CFG**

$$\mathcal{T} := \quad \mathsf{A} \quad | \quad \mathsf{T}^1_d \to \mathsf{T}^2$$

Sentence Syntax Function application & abstraction

Parse $\equiv$ $\text{Proof}_{\text{MILL}}$ $\equiv$ Program $\equiv$ λ-term

"the role that types play"




```
(that λx.((play x) types))(the role)
```

λ-terms may guide vectorial **semantic composition**

Attentive supertaggers correctly assign types **unseen** during training

Sparser categorial grammars are learnable

Data

Automatically extracted type sequences from **written Dutch treebank** (Lassy-small)

- 65 000 annotated sentences
- 1 million words
- 30 POS & syntactic tags
- 22 dependency labels

Lexicon

Refined type system leads to highly descriptive but very sparse types

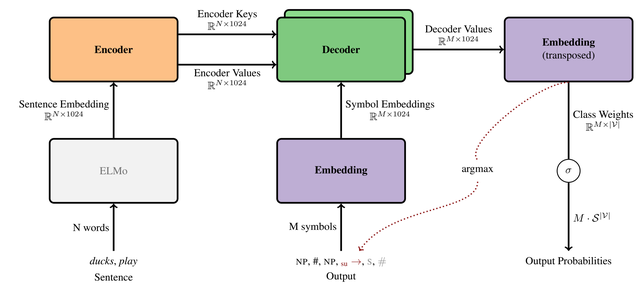
Feature, not bug!

- 6000 unique types
- 80% **rare** ($<$ 10 occurrences)
- 50% appear **once**

Approach

- Unfold complex types to sequences of atomic types and binary connectives.
- Words transduced to their unfolded representations.
- No hard-coded type lexicon, but **inductive construction** of any type in context.

Long-range dependences resolved by Transformer-like encoder/decoder stack with **intra-attention**:



Results

Model	% Type Accuracy				
	Unseen	Rare	Medium	Common	Overall
Predictive Baseline	–	23.9	59.0	89.9	87.2
Constructive	**19.2**	**45.7**	**65.6**	**89.9**	**88.0**

- **Generalization** to rare and unseen types.
- Constructed types are **well-formed** — perfect acquisition of type syntax.
- Phrasal **self-consistency** — good grasp of sentence/proof structure.
- **Non-trivial** new types but limited over-generation.