Typed Supertags and Semantic Parses for Dutch

Konstantinos Kogkalidis[1], Michael Moortgat[1], Richard Moot[2]

January 2020

[1]Utrecht University, [2]Université de Montpellier

Overview

- λ Type-Logical Grammars *why types?*
- λ Type System *how types?*
- λ Type Lexicon
- λ Semantic Parses
- λ Usecases

Type-Logical Grammars

Idea

Words assigned *formulas*, parsing a process of *deduction*.

Syntax

Structural Well-Formedness $\equiv$ Formal Derivability

Curry-Howard Isomorphism

Propositions $\equiv$ Types

Proofs $\equiv$ Functional Programs

Syntax-Semantics Interface

Parse $\equiv$ Proof $\equiv$ Computational Terms

Type System (1/2)

IILL

$$\mathcal{T} := A \mid T_1 \multimap T_2$$

$A \in \mathcal{A}$ :: Atoms denoting complete phrases

N, NP, S, . . .

$T_1 \multimap T_2$ :: Linear functor from T_1 to T_2

NP ⊸ S, NP ⊸ NP, NP ⊸ NP ⊸ S, (NP ⊸ NP) ⊸ (NP ⊸ NP)

$$\frac{\Gamma \vdash M : A \multimap B \quad \Delta \vdash N : A}{\Gamma, \Delta \vdash (M\ N) : B}\ E$$

$$\frac{\Gamma, x : A \vdash M : B}{\Gamma \vdash \lambda x.M : A \multimap B}\ I$$

Type System (2/2)

Modal Decoration Refinement

$$\mathcal{T} := A \mid T_1 \multimap T_2 \mid \diamond^d T_1 \multimap T_2 \mid \Box^d T_1 \multimap T_2$$

$d \in \mathcal{D}$ :: Dependency relations

subj, dobj, mod, body . . .

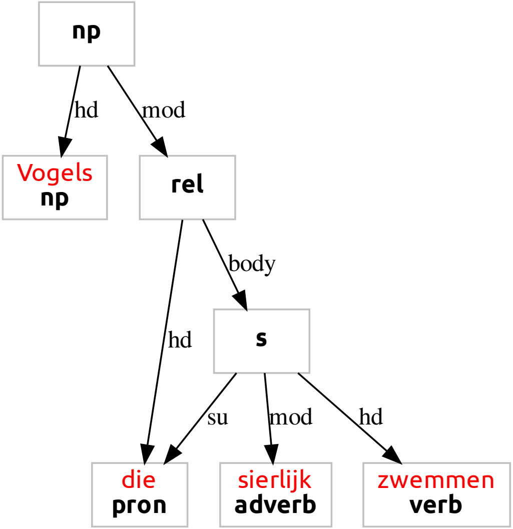
$\diamond^d T_1 \multimap T_2$ :: .. selecting complement T_1 under dependency domain d

$\diamond^{subj}$ NP $\multimap$ S, $\diamond^{dobj}$ NP $\multimap$ $\diamond^{subj}$ NP $\multimap$ S, . . .

$\Box^d T_1 \multimap T_2$ :: .. transforming T_1 by projecting dependency d

$\Box^{det}$ N $\multimap$ NP, $\Box^{mod}$ ($\Box^{mod}$ NP $\multimap$ NP) $\multimap$ $\Box^{mod}$ NP $\multimap$ NP

Ducks, in Alpino



Ducks, Proven

$$\dfrac{\dfrac{\mathit{vogels}}{\mathit{NP}}\,\mathit{Ax} \qquad \dfrac{\dfrac{\mathit{die}}{\diamond^{\mathit{body}}(\diamond^{\mathit{subj}}\mathit{NP} \multimap S) \multimap \square^{\mathit{mod}}\mathit{NP} \multimap \mathit{NP}}\,\mathit{Ax} \qquad \dfrac{\dfrac{\dfrac{\dfrac{}{\diamond^{\mathit{subj}}\mathit{NP} \vdash \diamond^{\mathit{subj}}\mathit{NP}}\,\mathit{Ax} \quad \dfrac{\mathit{zwemmen}}{\diamond^{\mathit{subj}}\mathit{NP} \multimap S}\,\mathit{Ax}}{\diamond^{\mathit{subj}}\mathit{NP}, \text{zwemmen} \vdash S}\,\mathit{E} \qquad \dfrac{\mathit{sierlijk}}{\square^{\mathit{mod}}S \multimap S}\,\mathit{Ax}}{\diamond^{\mathit{subj}}\mathit{NP}, \text{sierlijk zwemmen} \vdash S}\,\mathit{E}}{\text{sierlijk zwemmen} \vdash \diamond^{\mathit{subj}}\mathit{NP} \multimap S}\,\mathit{I}}{\text{die sierlijk zwemmen} \vdash \square^{\mathit{mod}}\mathit{NP} \multimap \mathit{NP}}\,\mathit{E}}{\mathit{vogels}\left\langle \mathit{die}\left\langle \langle \mathit{sierlijk}\rangle^{\mathit{mod}}\ \mathit{zwemmen}\right\rangle^{\mathit{body}}\right\rangle^{\mathit{mod}} \vdash S}\,\mathit{E}$$

$\texttt{die}\,(\lambda x.\,(\texttt{sierlijk}\ (\texttt{zwemmen}\ x)))\ \texttt{vogels}$

ÆTHEL

Extraction

From Lassy Parses to IILL Types & Theorems

arxiv: `abs/1912.12635`

Resources

1. Type Lexicon: Word → Type Distribution
2. Proofs: Lassy DAG → IILL Proof

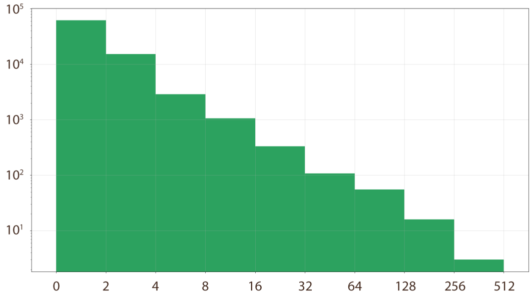
$\sim$ 97% *coverage*

Wikipedia subset publicly available at
`github.com/konstantinosKokos/aethel-public`

ÆTHEL: Lexicon

Stats

- ~900 000 word & type pairs
- 81 730 unique words
- 5 771 unique semantic types



Lexical Type Ambiguity Histogram (*log10-log2*)

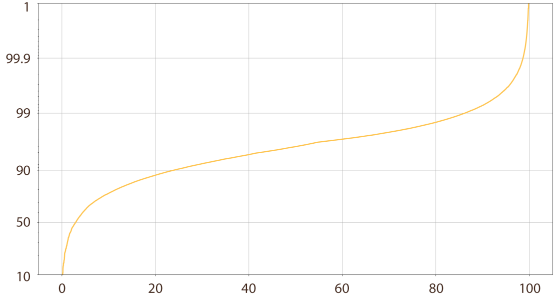
ÆTHEL: Proofs

Formats

- N.D. Proofs
- S.S. Proofs
- Linear Proofnets
- λ-terms

Stats

- 65 020 Lassy DAGs
- 72 263 IILL Proofs



Proof coverage w.r.t. most frequent types (*logit-linear*)

Usecases & Applications

- Supertagging with no type lexicon

arxiv: `abs/1905.13418`

- Parsing with type hints
- Type-aware language modeling
- Text to λ-term translation
- Semantic Compositionality
- ...?