

# Cross-lingual CCG Induction

Kilian Evang

TreeGraSP meeting #1

2018-02-21

## Joint Work with



Johan Bos



Lasha Abzianidze

Photos: LingPhot/Eleri

Overview

CCG

Derivation Projection

Status of Project

# Combinatory Categorical Grammar (CCG)

- grammar formalism known for its linguistic elegance and computational effectiveness
- widely used in semantic parsing

# Problem

- CCGs used in semantic parsing either
  - hand-written, or
  - extracted from a large annotated corpus (CCGbank)
- → generalizing to languages other than English requires much human effort

# Approach

- exploit parallel corpora
- align words
- project derivations (syntax trees) from English to target language
- extract lexicon, train parser

# CCG

Demo: `https://texttheater.net/ccgweb`

# CCG

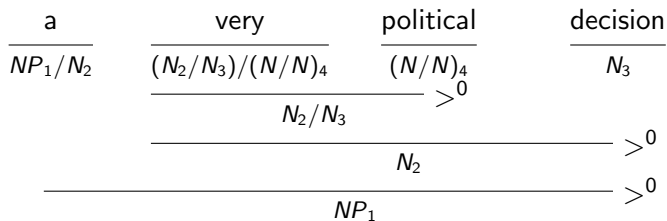
Different languages need different categories, e.g.

- postnominal adjectives in French, Italian
- main vs. subordinate clause order in German, Dutch
- pro-drop in Italian
- questions without *do*-support

How to learn these categories?



# Derivation Projection: Example 1



# Derivation Projection: Example 1

$$\frac{NP_1/N_2}{a}$$

une

$$\frac{(N_2/N_3)/(N/N)_4}{very}$$

décision

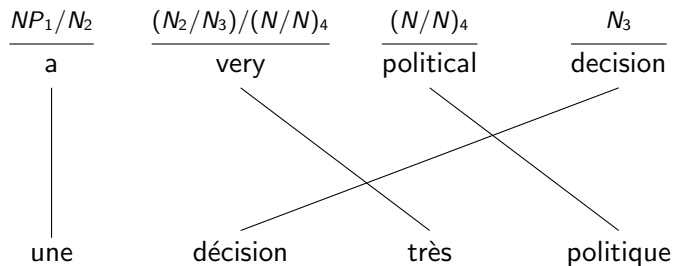
$$\frac{(N/N)_4}{political}$$

très

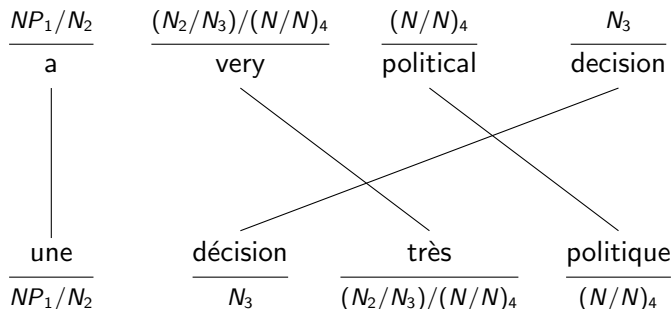
$$\frac{N_3}{decision}$$

politique

# Derivation Projection: Example 1



# Derivation Projection: Example 1



# Derivation Projection: Example 1

une  

---

 $NP_1/N_2$

décision  

---

 $N_3$

très  

---

 $(N_2/N_3)/(N/N)_4$

politique  

---

 $(N/N)_4$

# Derivation Projection: Example 1

une  

---

 $NP_1/N_2$

décision  

---

 $N_3$

très  

---

 $(N_2 \setminus N_3)/(N \setminus N)_4$

politique  

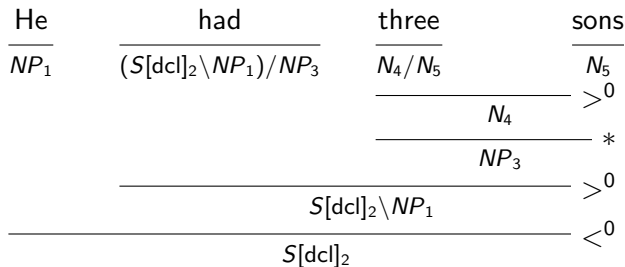
---

 $(N \setminus N)_4$

# Derivation Projection: Example 1

une	décision	très	politique
$NP_1/N_2$	$N_3$	$(N_2 \setminus N_3)/(N \setminus N)_4$	$(N \setminus N)_4$
			$>^0$
		$N_2 \setminus N_3$	
			$<^0$
		$N_2$	
			$>^0$
		$NP_1$	

## Derivation Projection: Example 2





# Derivation Projection: Example 2

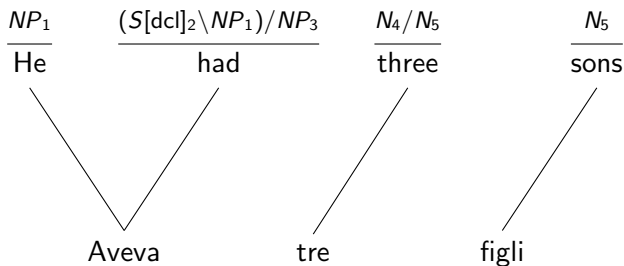
$$\begin{array}{cccc}
 \frac{NP_1}{\text{He}} & \frac{(S[\text{dcl}]_2 \backslash NP_1) / NP_3}{\text{had}} & \frac{N_4 / N_5}{\text{three}} & \frac{N_5}{\text{sons}}
 \end{array}$$

Aveva

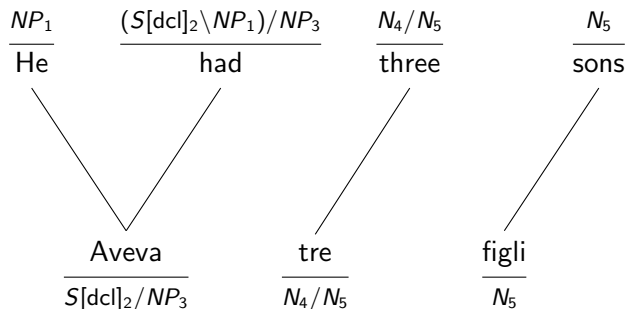
tre

figli

# Derivation Projection: Example 2



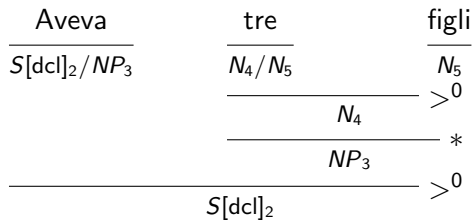
# Derivation Projection: Example 2



## Derivation Projection: Example 2

$$\frac{\text{Aveva}}{S[\text{dcl}]_2 / NP_3}$$
$$\frac{\text{tre}}{N_4 / N_5}$$
$$\frac{\text{figli}}{N_5}$$

# Derivation Projection: Example 2



# Limits

- non-literal translations
- divergent (non-isomorphic) translations
- bad word alignments

→ projection fails or produces lexical items that generalize poorly

# Research Questions

1. To what extent does derivation projection work even with divergent translations?
2. Where it does work, to what extent do the projected derivations make linguistic sense?
3. To what extent are projected derivations suitable as training data for parsing?

# Contributions

1. Web-based annotation platform (100%)
2. Quadrilingual evaluation corpus (50%)
3. Evaluation (almost ready to start)



# 1. To what extent does derivation projection work even with divergent translations?

Apply to evaluation corpus, identify causes of projection failures.

## 2. Where it does work, to what extent do the projected derivations make linguistic sense?

Apply to evaluation corpus, compare with human annotations.

### 3. To what extent are projected derivations suitable as training data for parsing?

Apply to unannotated parallel data, train, parse evaluation corpus, compare. Of particular interest: does the parser learn to use categories that are rare or nonexistent in the source language?

Thank you!