Weakly-Supervised Bayesian Learning of a CCG Supertagger

Dan Garrette, Chris Dyer, Jason Baldridge, Noah A. Smith

UT-Austin, CMU

Type-Supervised Learning

Can we learn a CCG supertagger from only?:

- Unannotated text
- Incomplete dictionary: word ↦ {tags}
- **Universal CCG principles**

Principle #1: Simplicity

Small > Big

buy := $(s_b\backslash np)/np$ appears **342** times

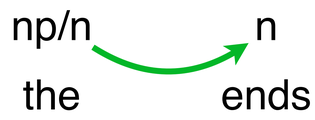
buy := $(((s_b\backslash np)/pp)/pp)/np$ appears **once**

Modifier > Non-modifier

$(s\backslash np)/(s\backslash np)$ more likely than $(s_b\backslash np)/(s_{adj}\backslash np)$

Principle #2: Connectivity

Connecting > Not Connecting



connects

doesn't connect

Transition Priors

Use CCG principles to inform **transition priors** for HMM parameter inference

$$P(t \rightarrow u) = \underbrace{\lambda \cdot P(u)}_{\text{simple is good}} + \underbrace{(1-\lambda) \cdot P(t \rightarrow u)}_{\text{connecting is good}}$$

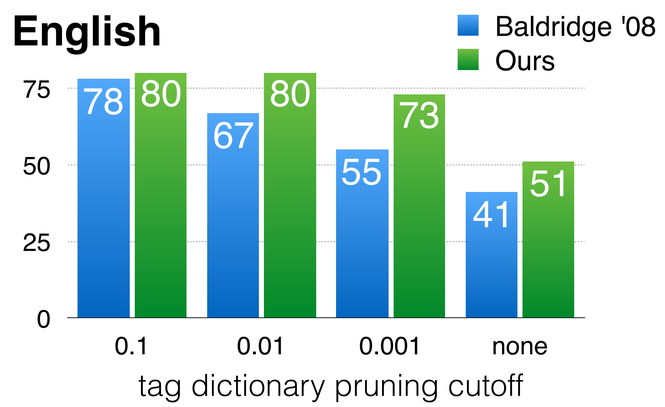
Weighted Tag Grammar

Tag prior **P(u)** defined by:

$a \rightarrow \{s, np, n, \ldots\}$	$p_{atom}(a) \times p_{term}$
$A \rightarrow B \,/\, B$	$\overline{p_{term}} \times p_{fwd} \times p_{mod}$
$A \rightarrow B \,/\, C$	$\overline{p_{term}} \times p_{fwd} \times \overline{p_{mod}}$
$A \rightarrow B \setminus B$	$\overline{p_{term}} \times \overline{p_{fwd}} \times p_{mod}$
$A \rightarrow B \setminus C$	$\overline{p_{term}} \times \overline{p_{fwd}} \times \overline{p_{mod}}$

where $\overline{p} = (1-p)$

Per-Token Supertag Accuracy



Similar results for **Chinese** and **Italian**