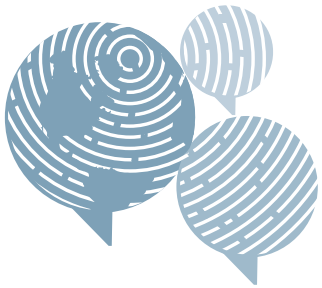
Judging Grammaticality with Tree Substitution Grammar Derivations

Matt Post



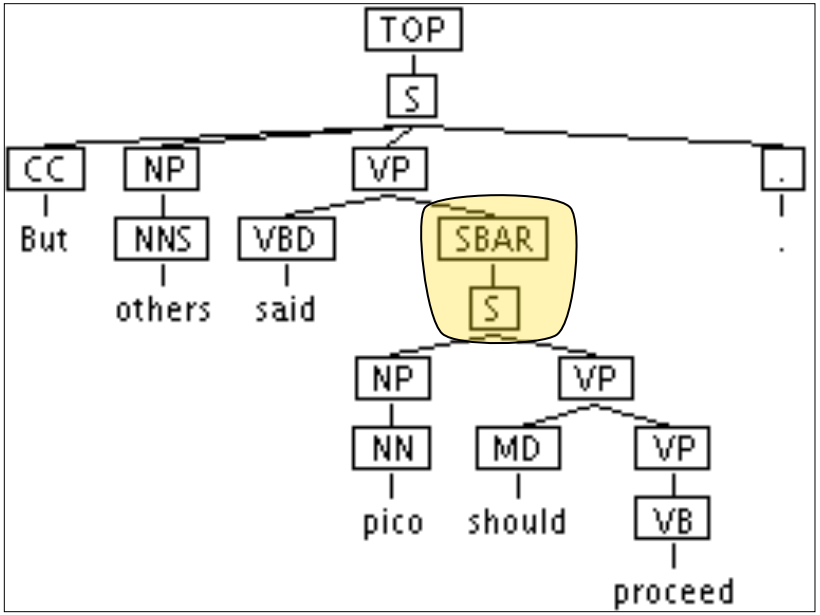
human language technology
center of excellence
JOHNS HOPKINS
UNIVERSITY

Overview

- Features from TSG derivations are helpful in distinguishing real text from "pseudo-negative" text

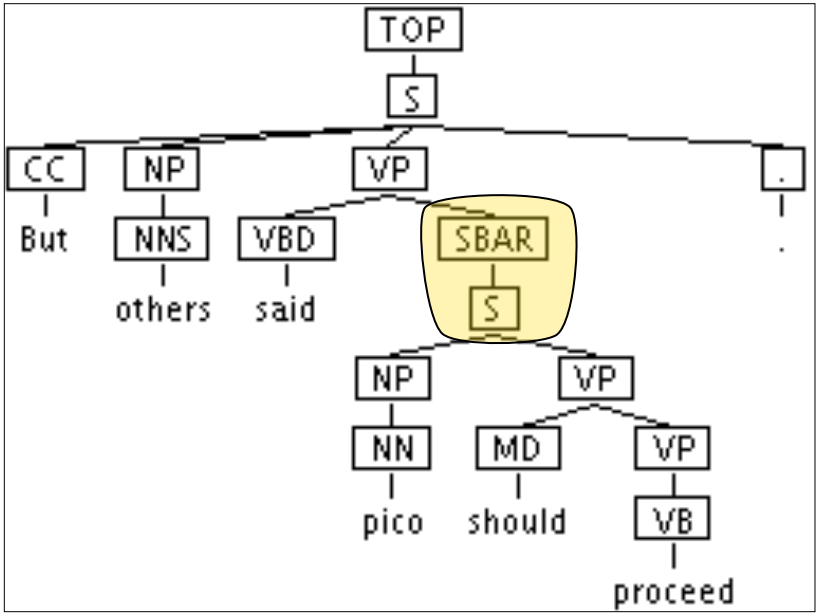
Standard CFG

Nonterminals rewrite as sequence of child nonterminals



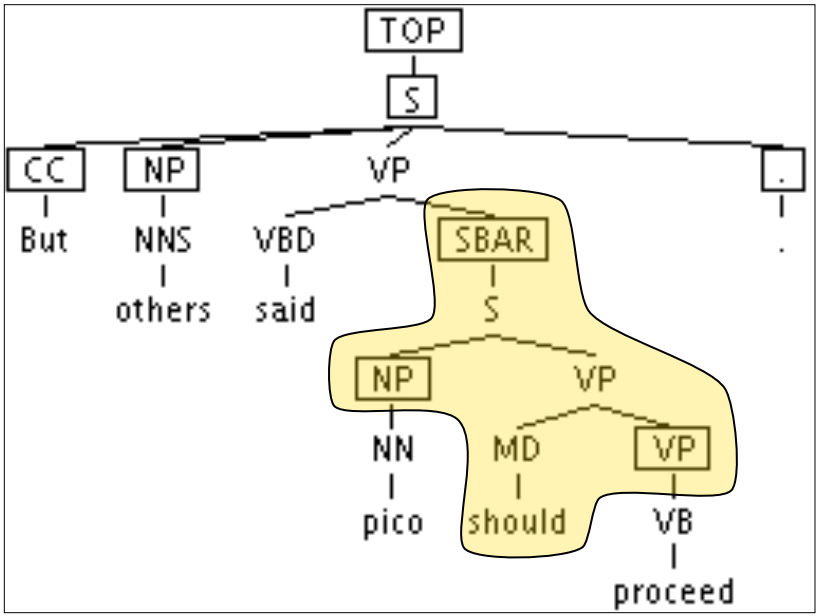
Standard CFG

Nonterminals rewrite as sequence of child nonterminals



TSG

Nonterminals rewrite as sequence of child fragments of arbitrary size



Motivation

- Language modeling is hard, and syntax isn't as easy a solution as one might hope:
  - parser scores aren't very helpful (Och et al., 2004)
  - but can be improved with task-specific training (Cherry & Quirk, 2008)

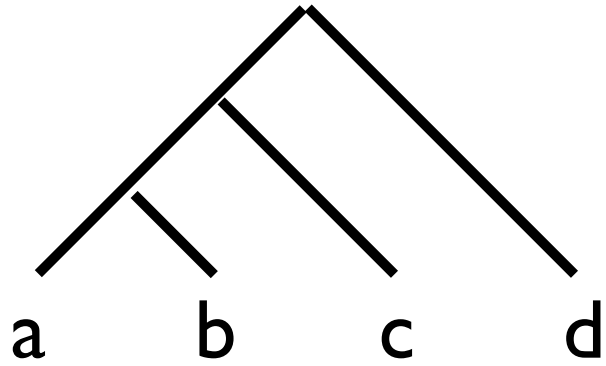
Motivation

- TSGs are an interesting formalism, but haven't been that helpful in parsing (though see Bansal & Klein, 2010)
- Tree features are helpful in related tasks, like parse reranking (Charniak & Johnson, 2005)
- Perhaps they can improve language modeling

Grammatical?

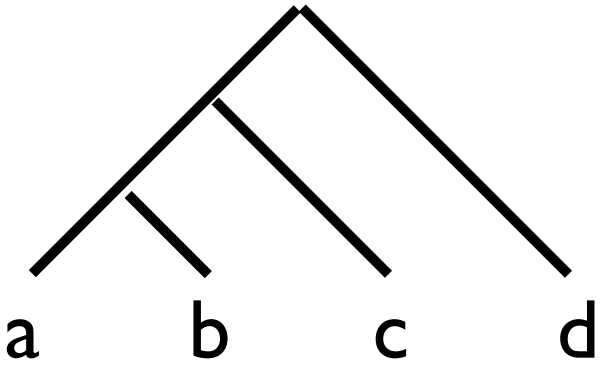
a b c d

Grammatical?

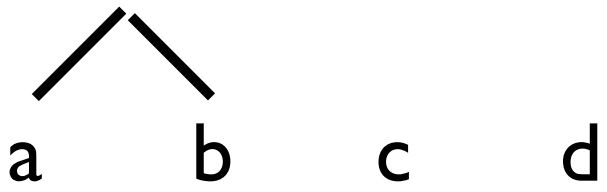


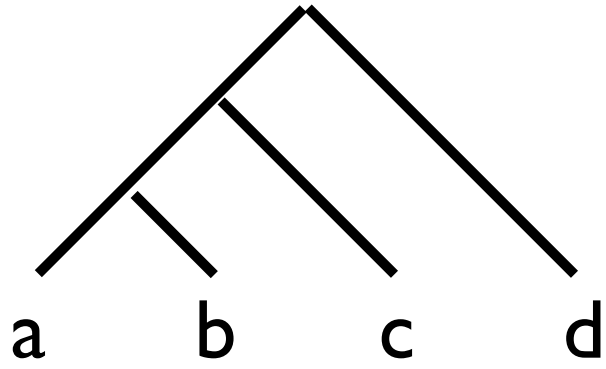
Grammatical?

a b c d

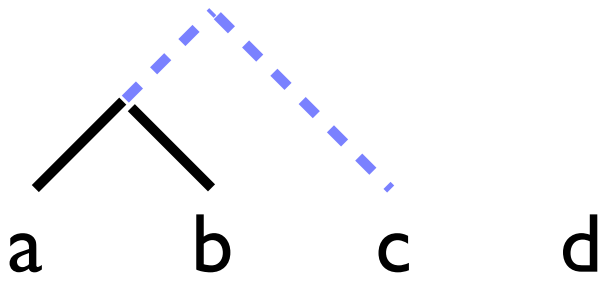


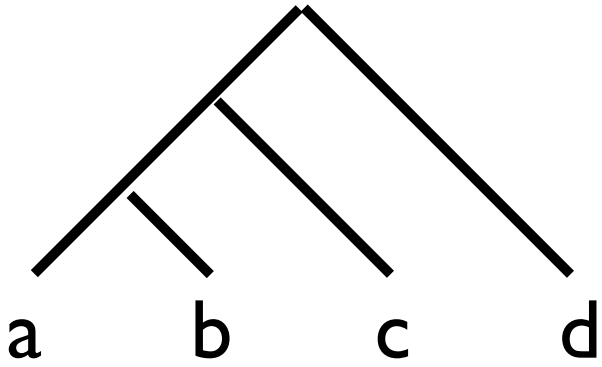
Grammatical?



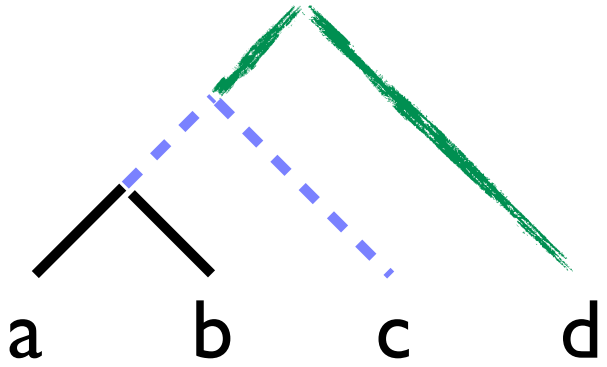


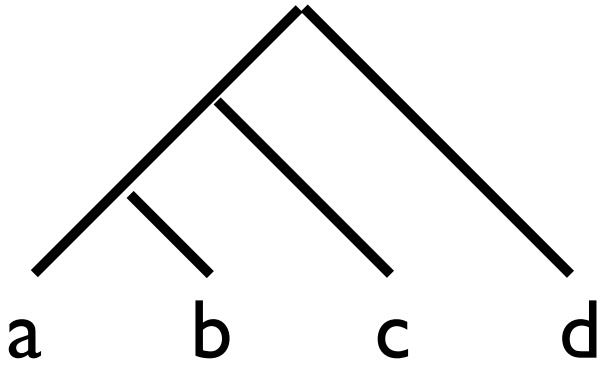
Grammatical?





Grammatical?

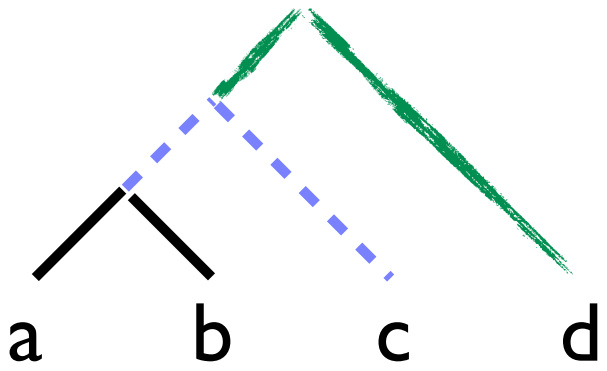




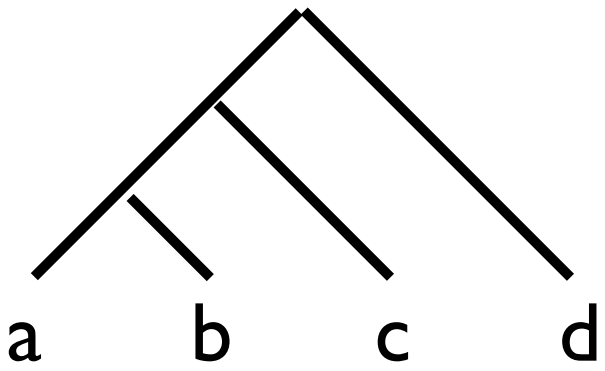
Grammatical?

increased likelihood of grammaticality →

many little fragments

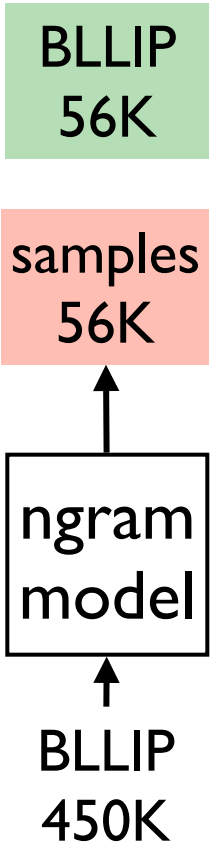


single large fragment



Experiments

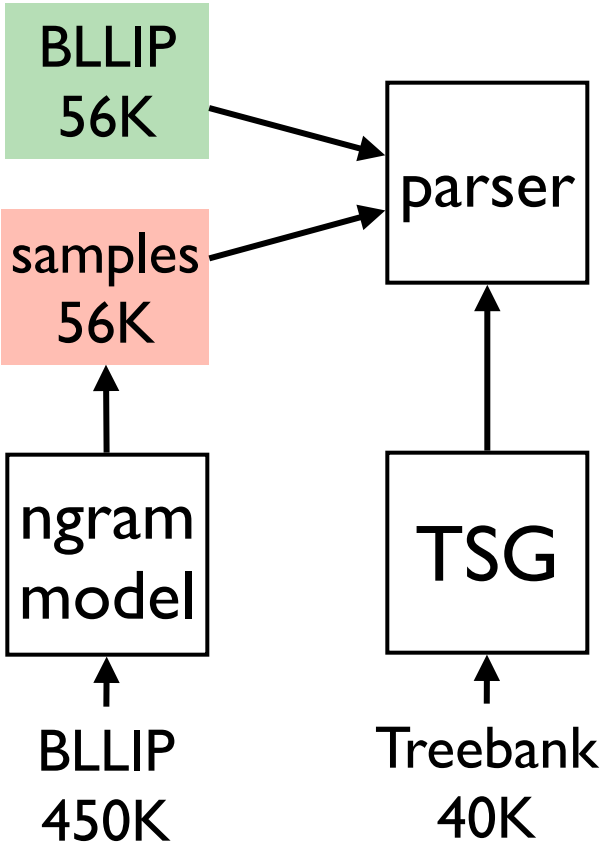
Cherry & Quirk (2008)



human language technology
center of excellence

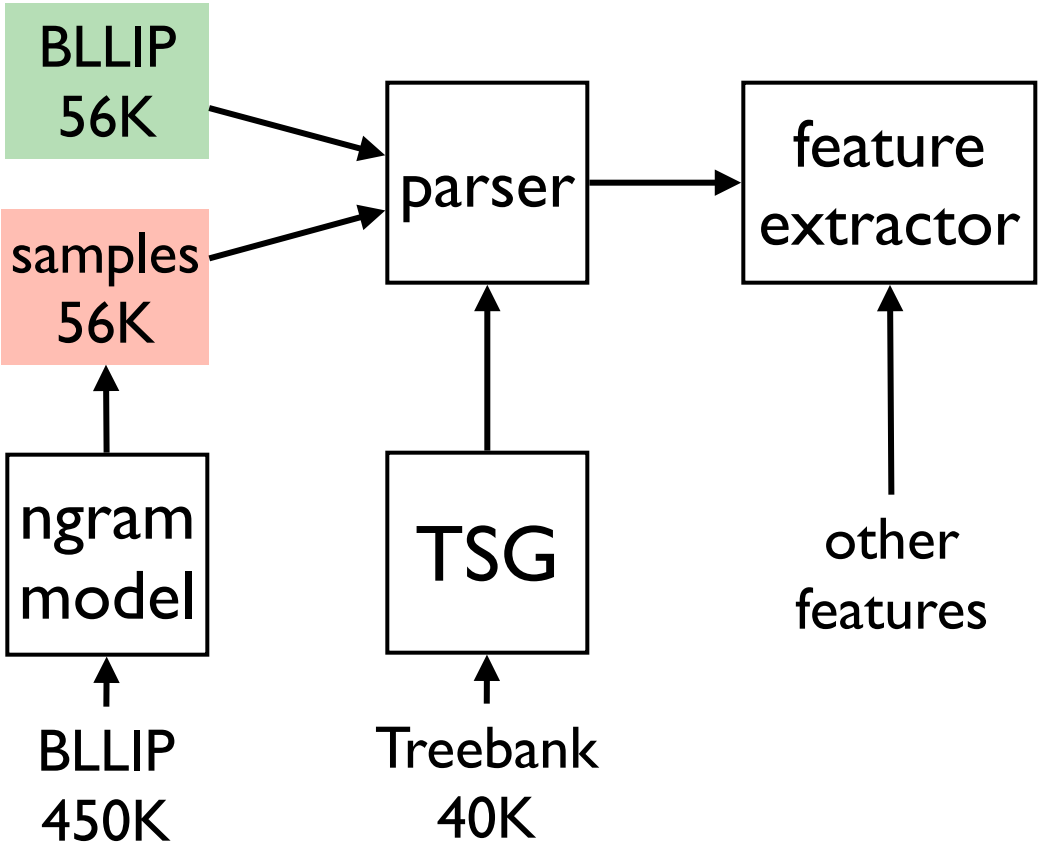
Experiments

Cherry & Quirk (2008)



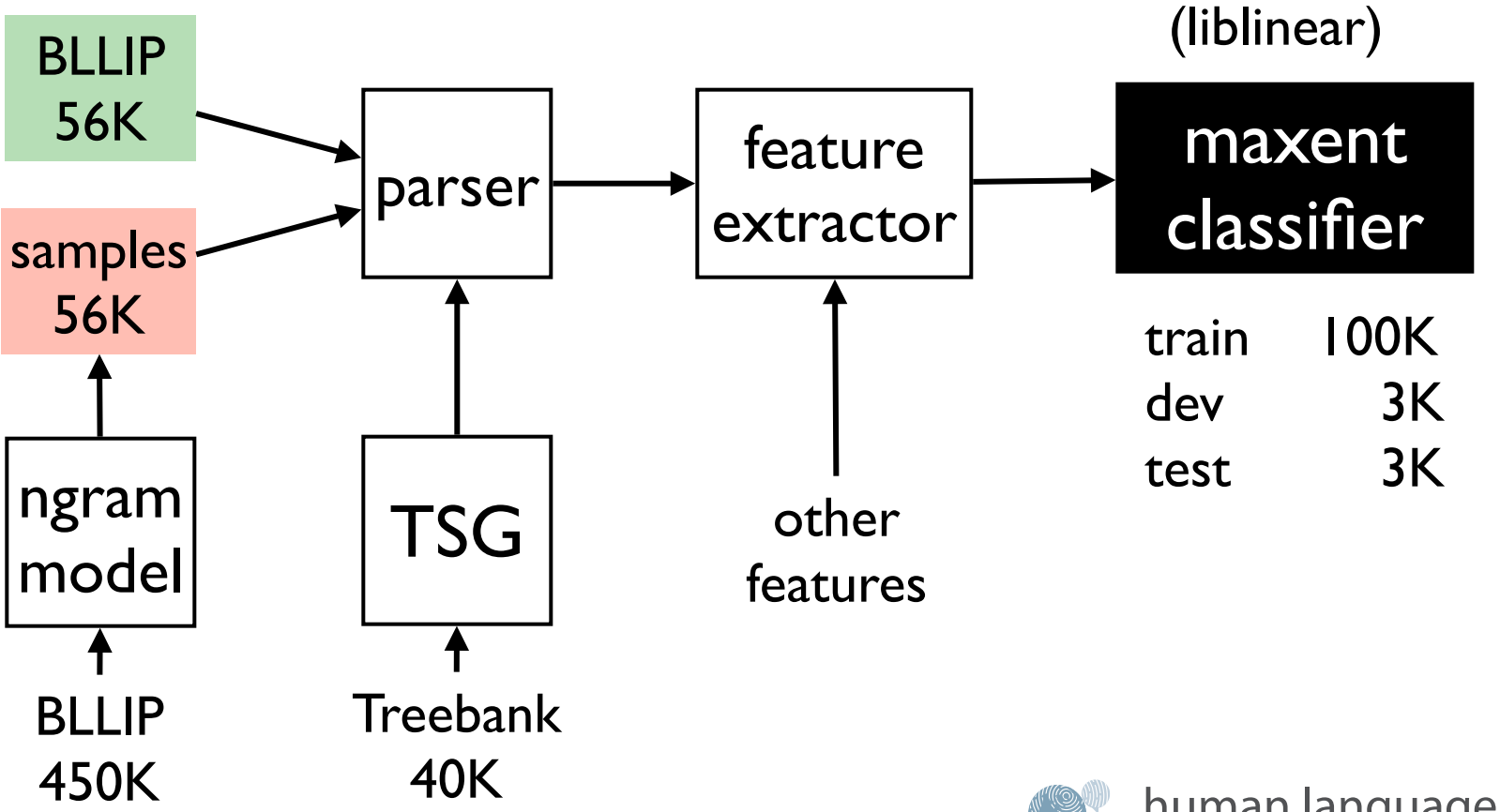
Experiments

Cherry & Quirk (2008)



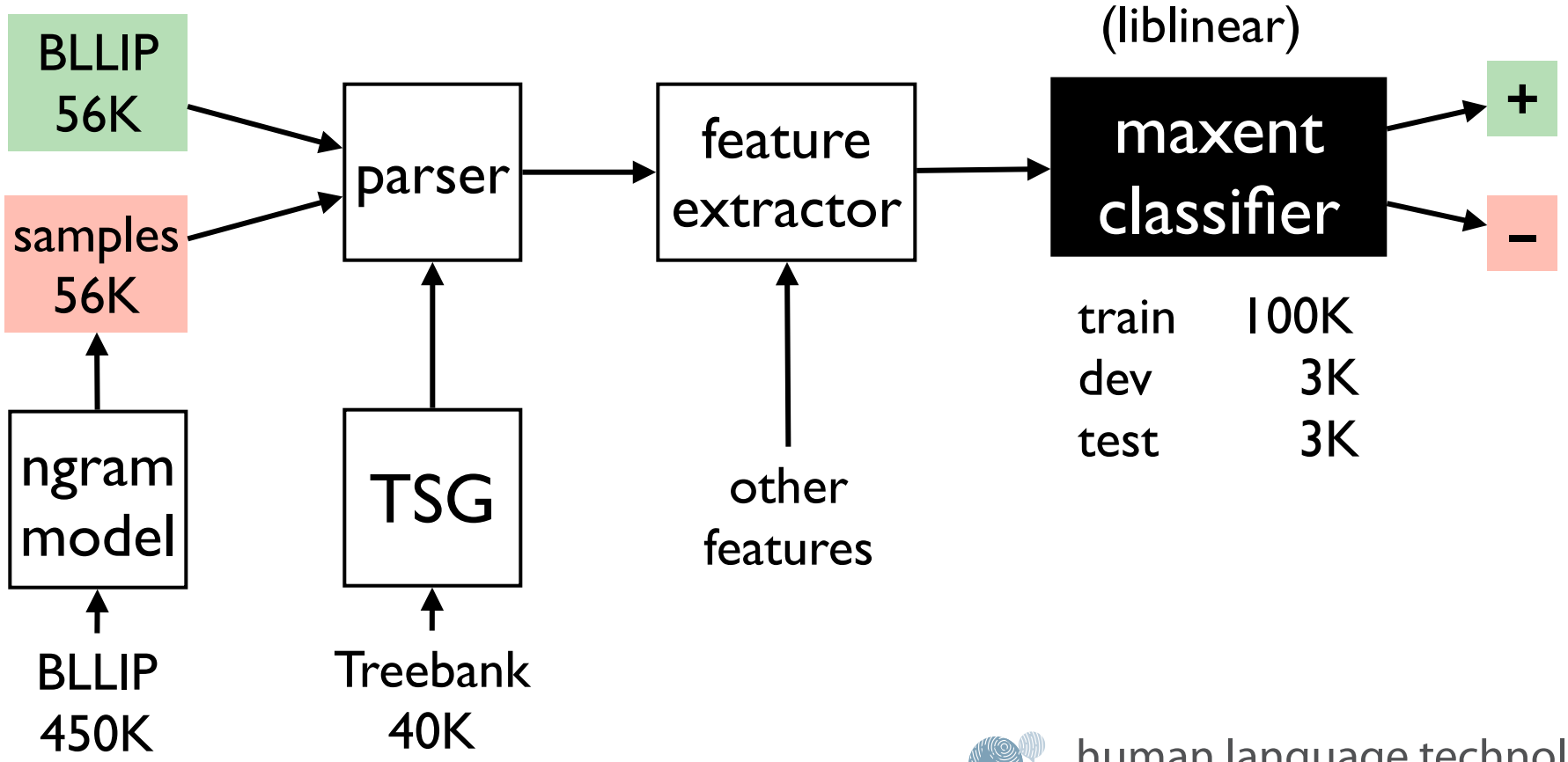
Experiments

Cherry & Quirk (2008)



Experiments

Cherry & Quirk (2008)



Feature sets

feature set	example
length	*l* 7

Feature sets

feature set	example
length	*17*
3-grams	*"he further praised"*

Feature sets

feature set	example
length	*17*
3-grams	*"he further praised"*
Charniak & Johnson (2005) reranking features	*number of nodes in the parse tree*

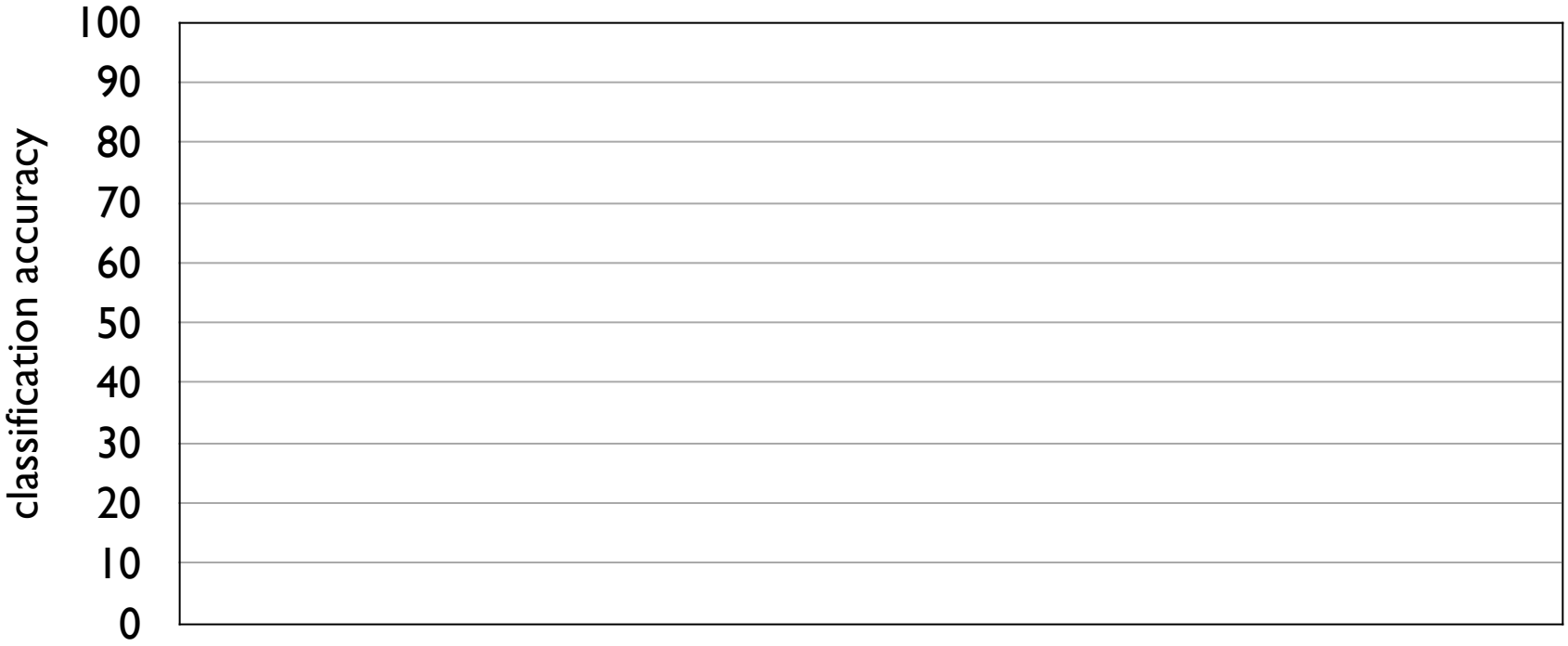
Feature sets

feature set	example
length	*17*
3-grams	*"he further praised"*
Charniak & Johnson (2005) reranking features	*number of nodes in the parse tree*
TSG aggregate features	*number of lexicalized fragments of rank 6*

Feature sets

feature set	example
length	*17*
3-grams	*"he further praised"*
Charniak & Johnson (2005) reranking features	*number of nodes in the parse tree*
TSG aggregate features	*number of lexicalized fragments of rank 6*
TSG fragments	(TOP (S NP (VP VBD said) NP SBAR) .)

Classification results

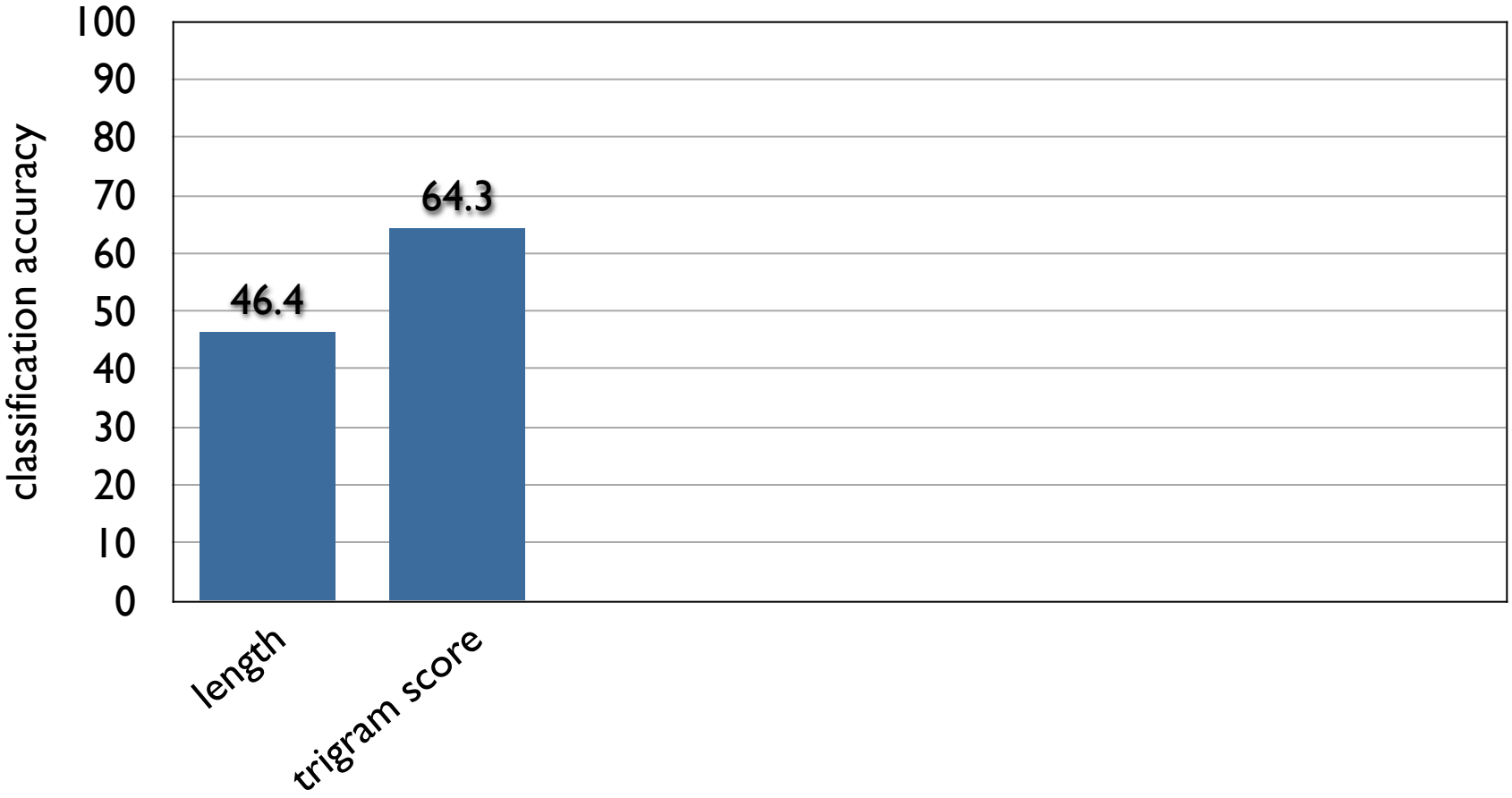


human language technology
center of excellence

Classification results

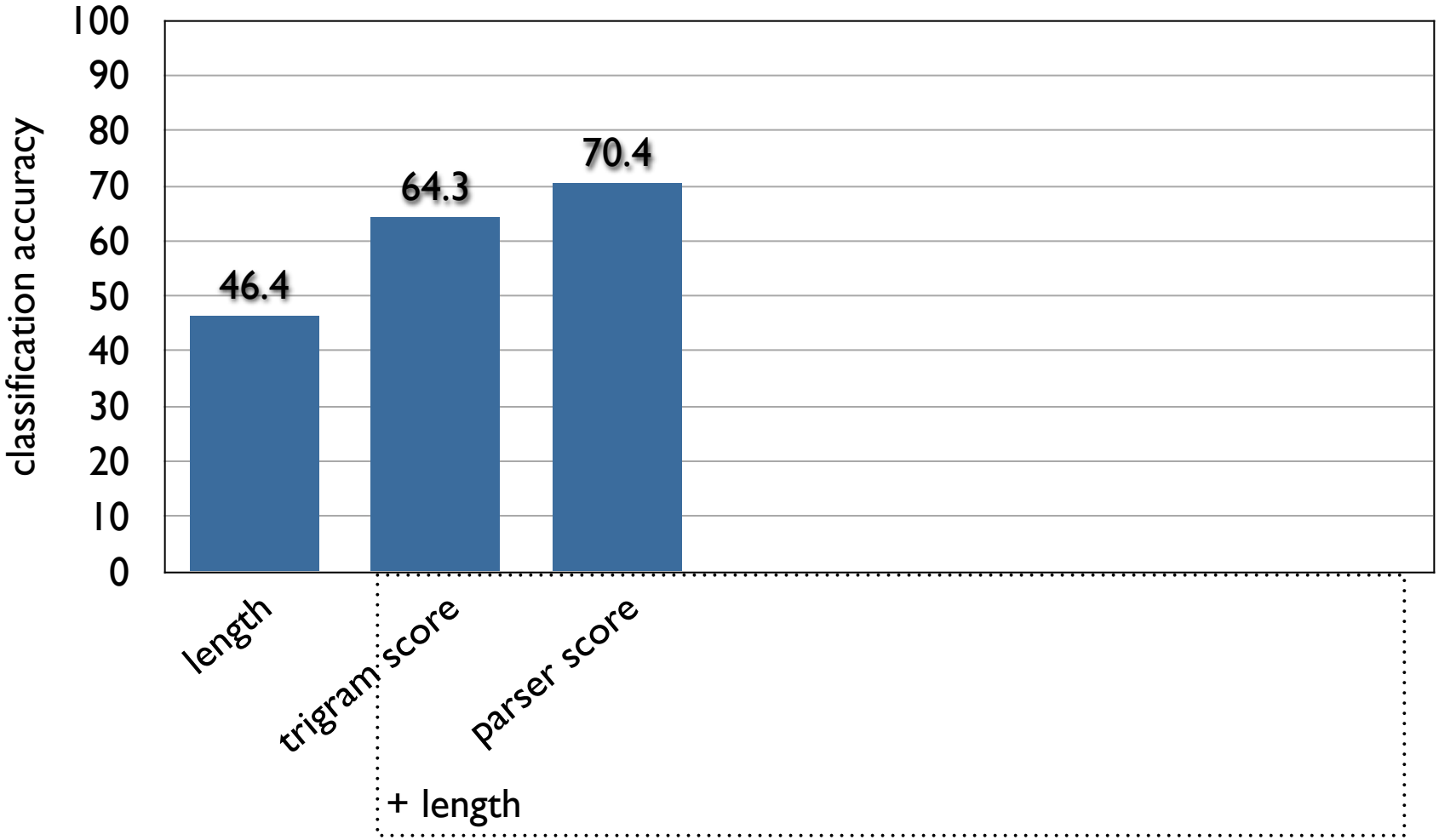
human language technology
center of excellence

Classification results



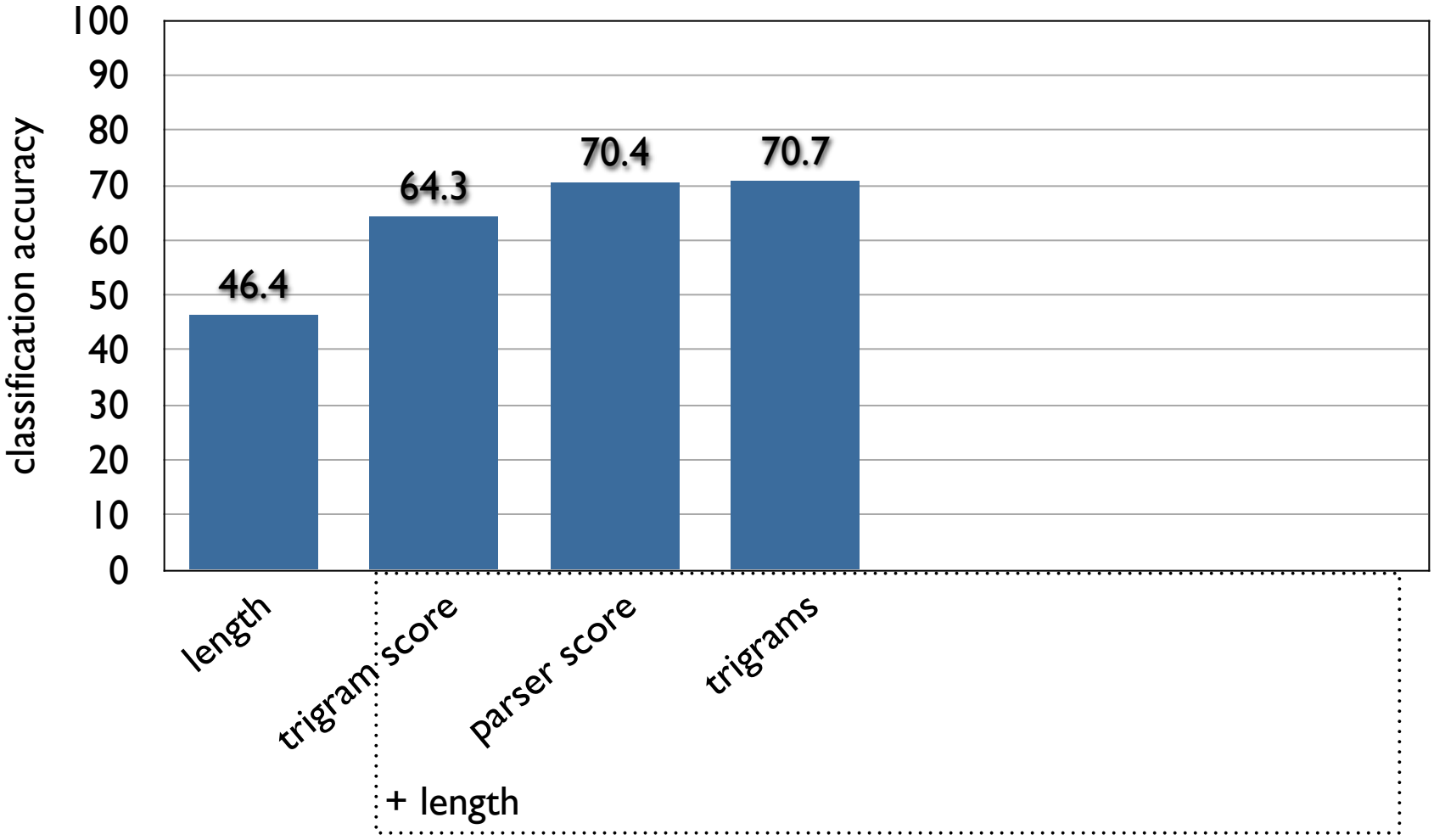
human language technology
center of excellence

Classification results

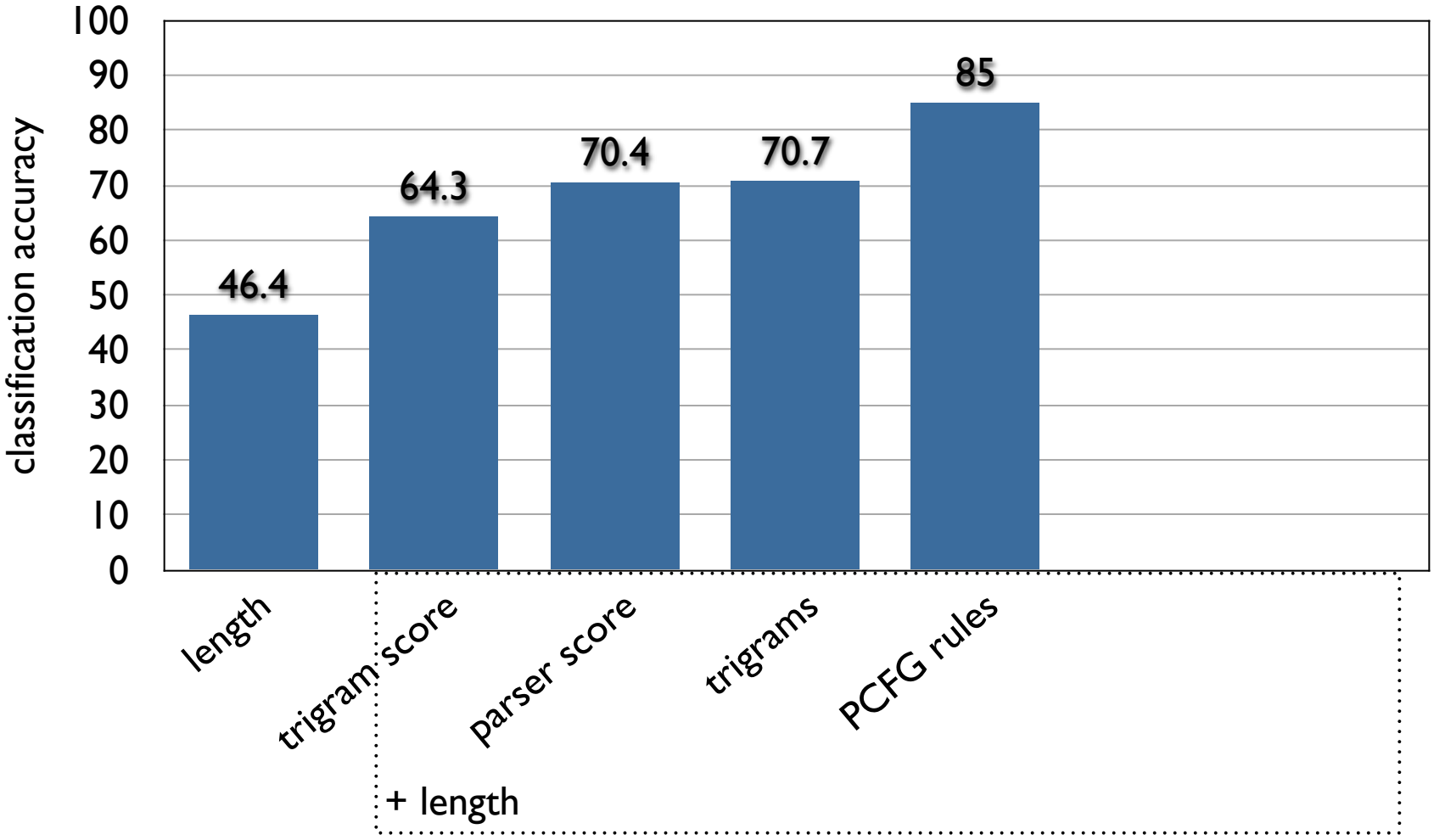


human language technology
center of excellence

Classification results

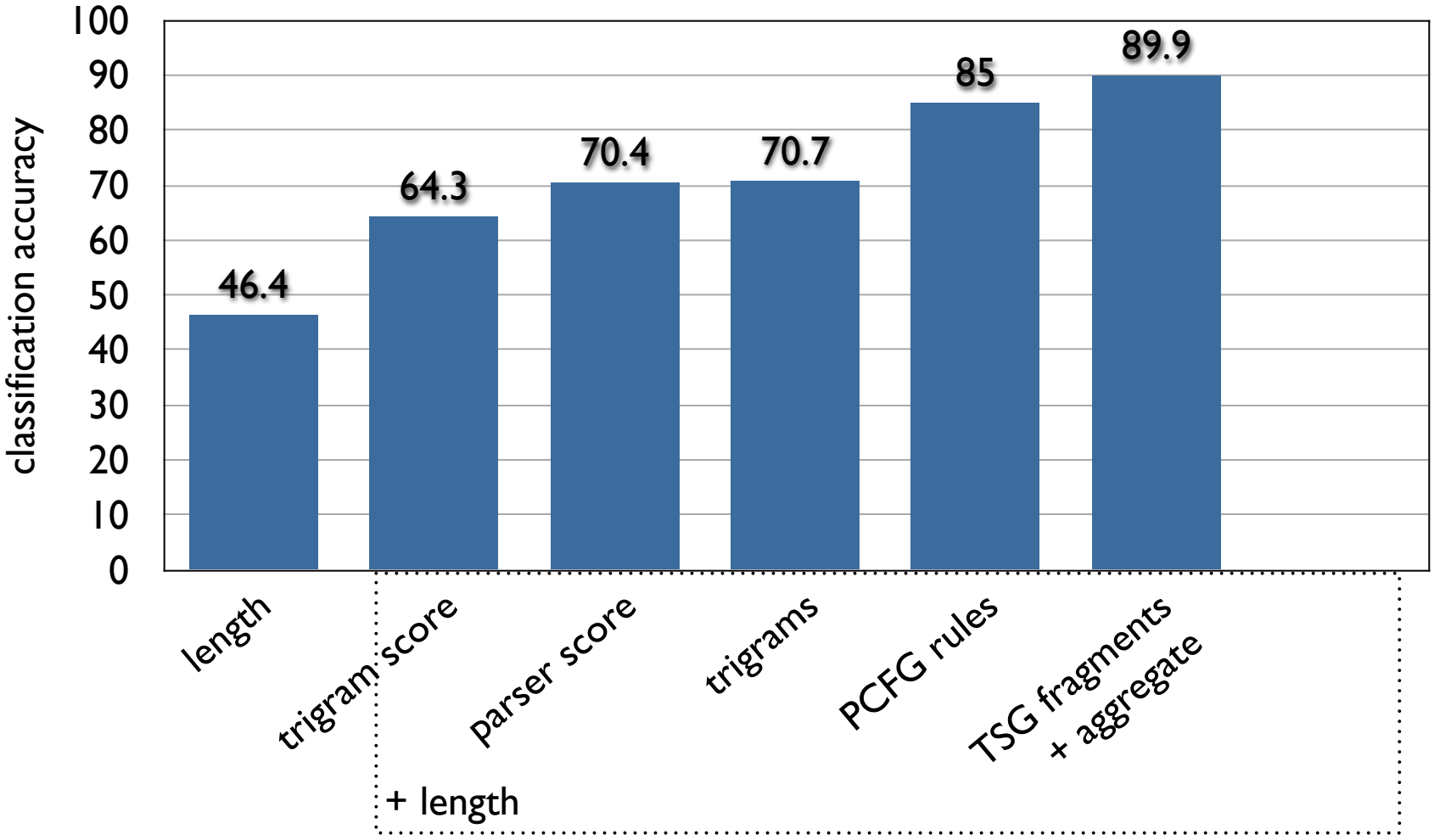


Classification results



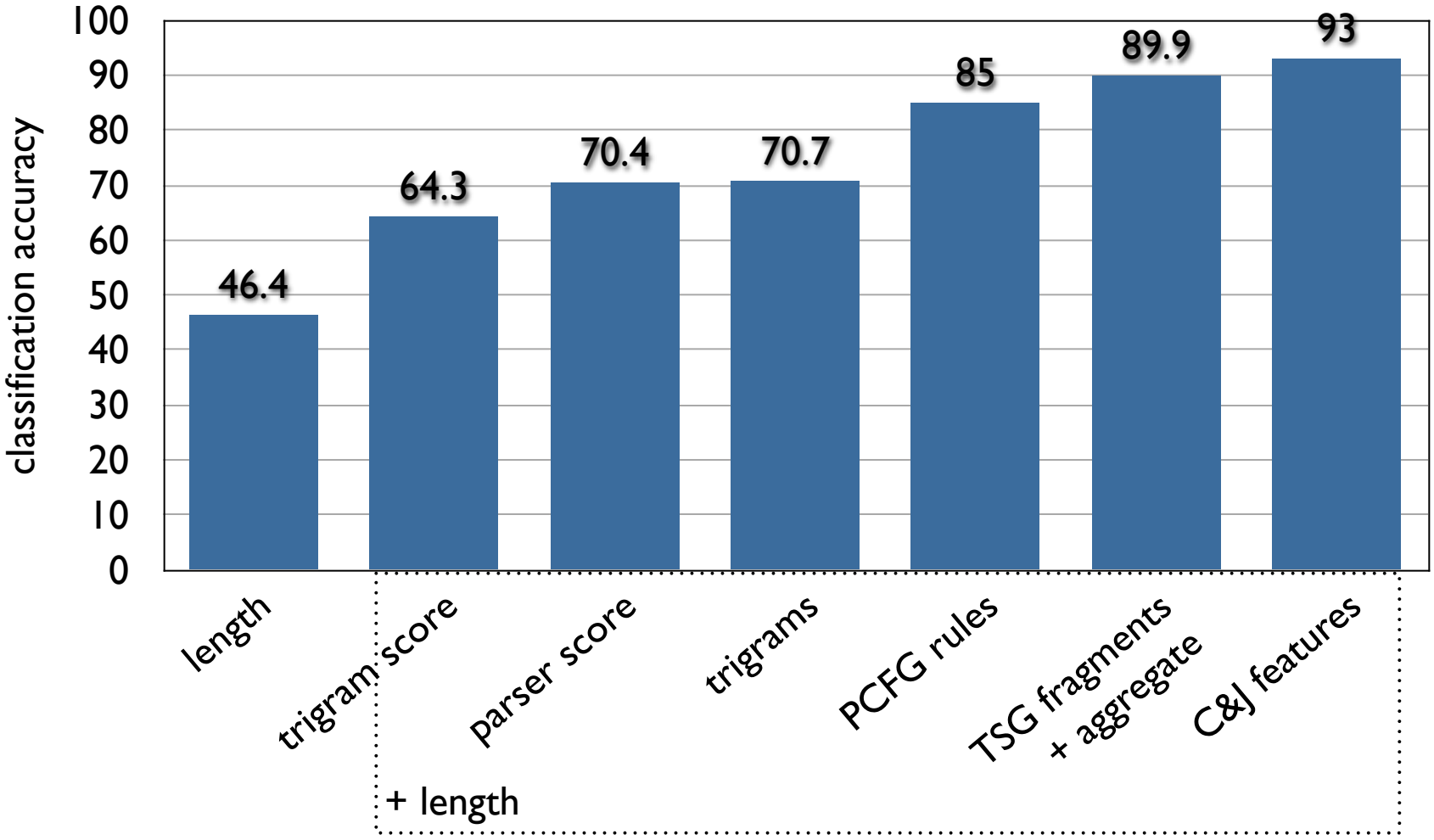
human language technology
center of excellence

Classification results



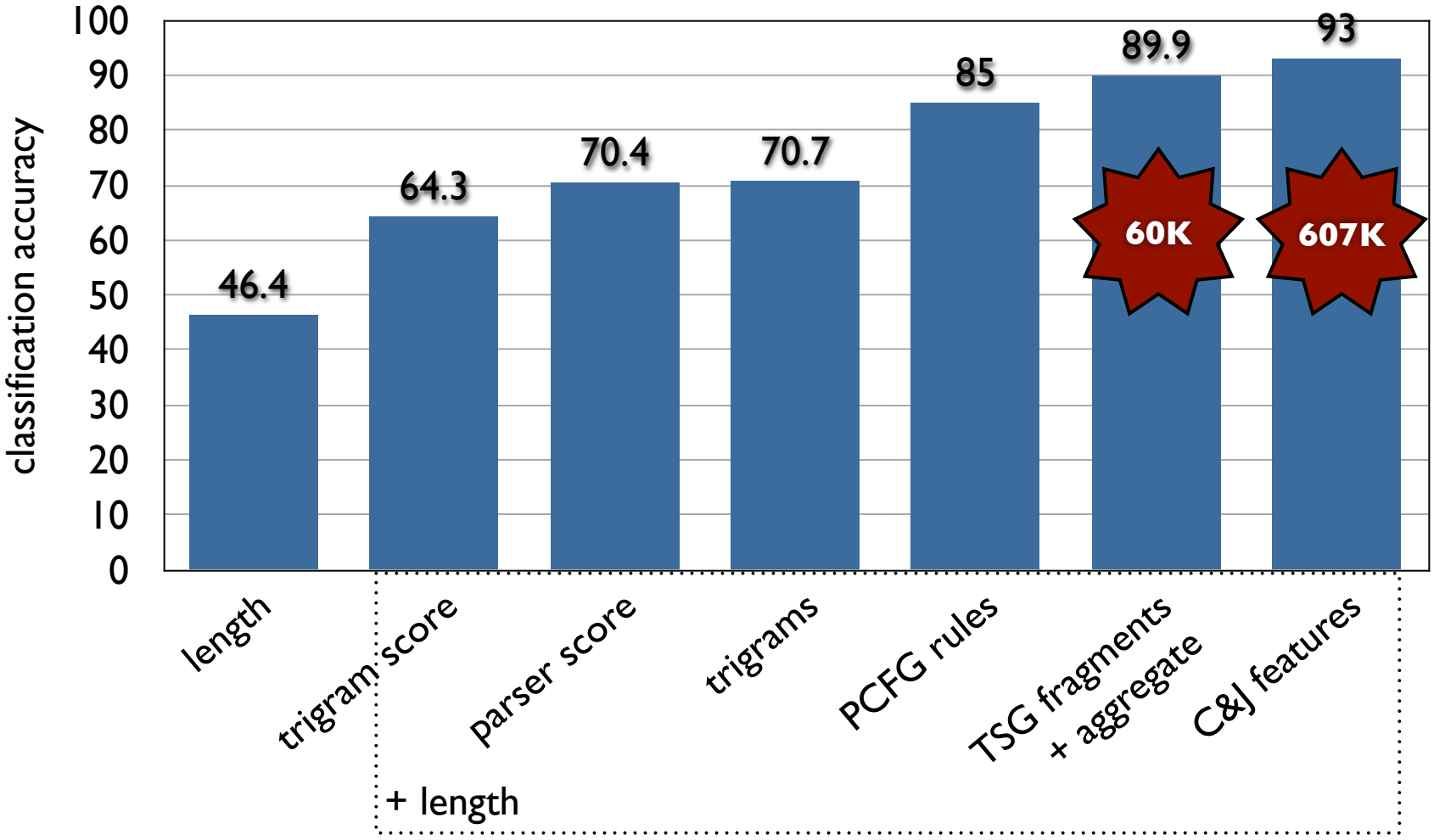
human language technology
center of excellence

Classification results

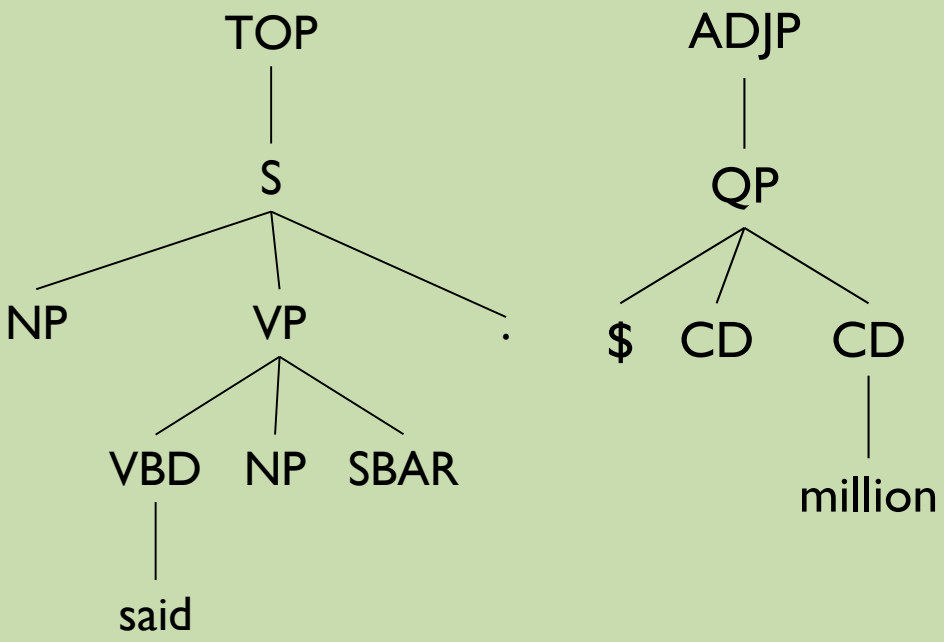


human language technology
center of excellence

Classification results

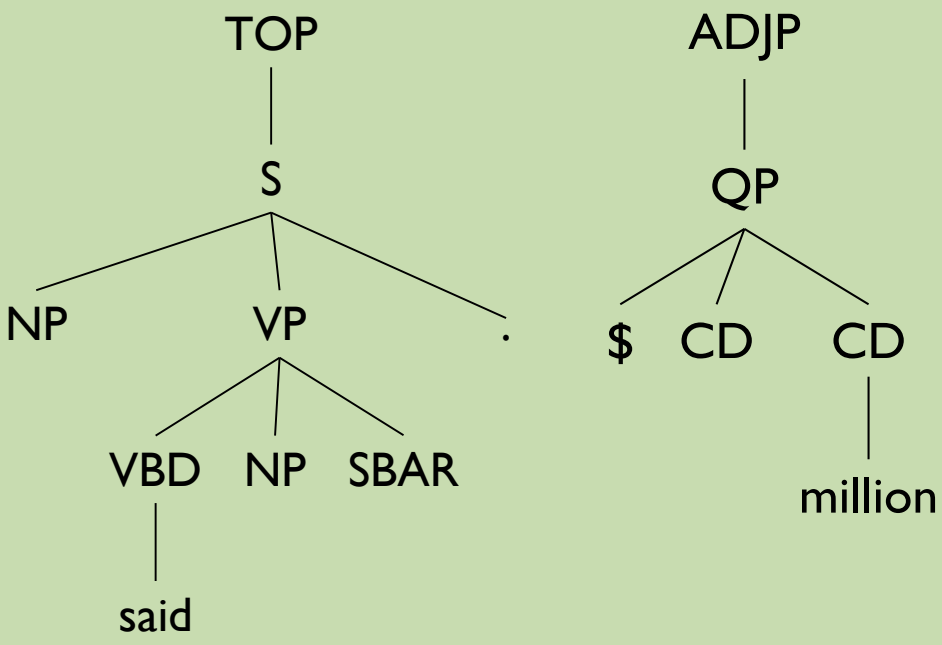


Features

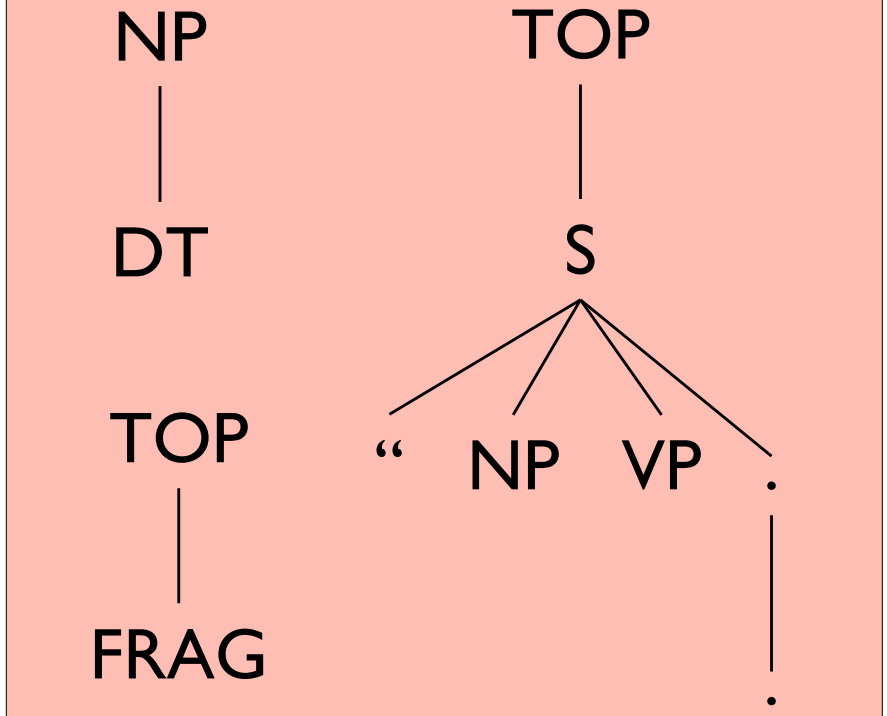


- fragments with two words
- fragments with a rank of 8

Features



- fragments with two words
- fragments with a rank of 8



- unlexicalized fragments

Conclusions

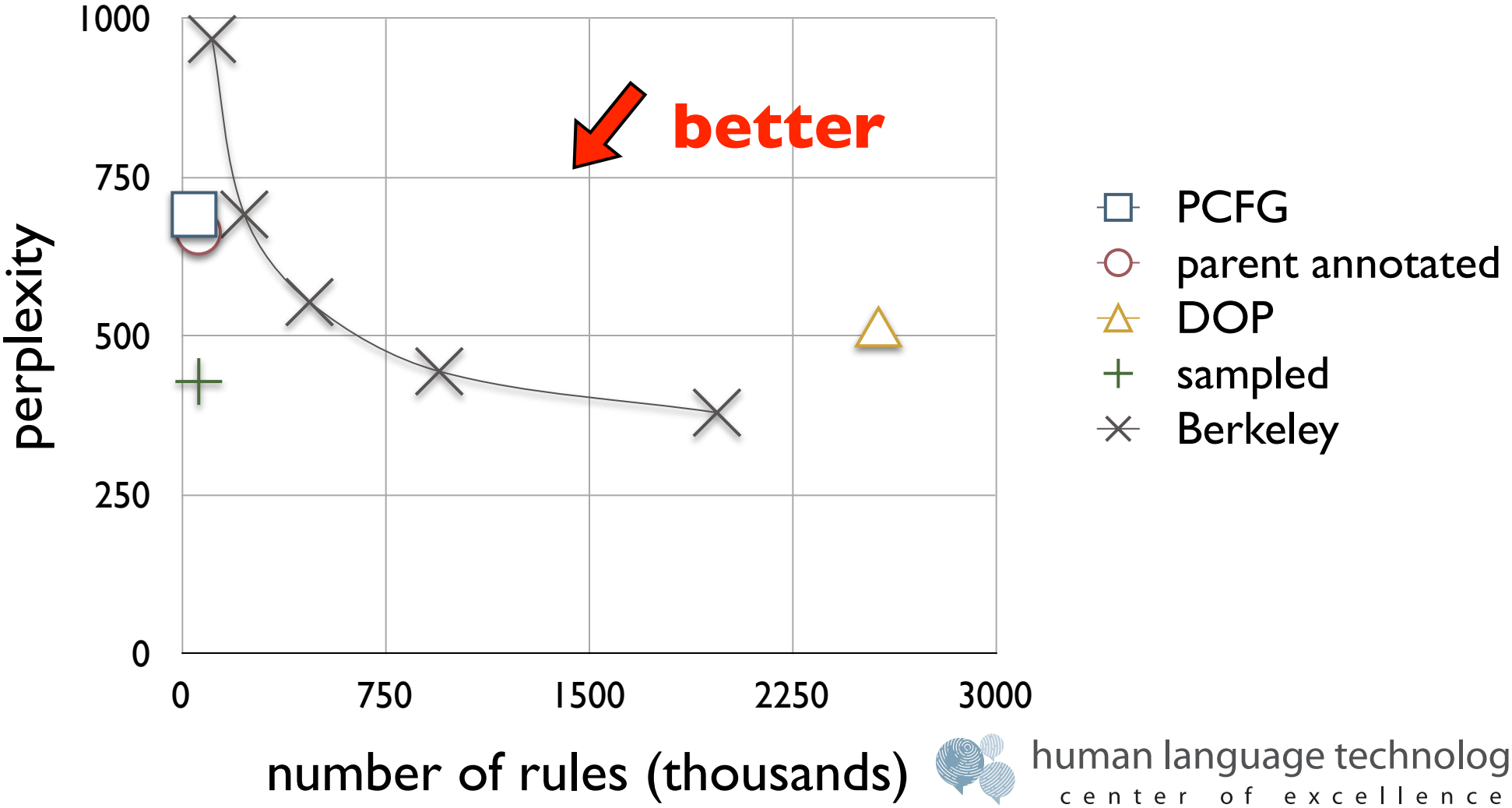
- TSGs performed well despite relatively poor parsing accuracy
- Shallow rules + coarse nonterminals => license for ungrammaticality
- Well below C&J features, but 1/10 the size, automatically discovered, and are a biproduct of the derivation

Future work

- Further baselines (implicit DOP representations, heuristic TSGs)
- Other discriminative tasks
  - more realistic negatives (Wagner et al., 2009; Tetreault et al., 2010)
  - MT classification and reranking
- Base measure changes

Thanks

Latent-variable TSGs



human language technology
center of excellence