TriS: A Statistical Sentence Simplifier with Log-linear Models and Margin-based Discriminative Training

Nguyen Bach, Qin Gao, Stephan Vogel, and Alex Waibel
Language Technologies Institute
Carnegie Mellon University

Why does Sentence Simplification matter?

John comes from England , works for IMF , and is an active hiker .

4th grade

Hanoi was eclipsed by Hue during the Nguyen Dynasty as the capital of Vietnam , and Hanoi served as the capital of French Indochina from 1902 to 1954 .

12th grade

Proposing a plan that cuts deeply in social programs to reduce the deficit could disappoint his Democratic supporters .

College level

Why does Sentence Simplification matter?

- Improve readability: help low-literacy readers

Source	Flesch	Grade Level
Comics	92	5th
Consumer ads in magazines	82	6th
Reader's Digest	65	8th-9th
New York Times	39	College
Standard auto insurance policy	10	College grad

John comes from England , works for IMF , and is an active hiker .

John comes from England.
John works for IMF .
John is an active hiker .

4th -> 2nd grade

Why does Sentence Simplification matter?

- Improve readability: help low-literacy readers

Source	Flesch	Grade Level
Comics	92	5th
Consumer ads in magazines	82	6th
Reader's Digest	65	8th-9th
New York Times	39	College
Standard auto insurance policy	10	College grad

John comes from England , works for IMF , and is an active hiker .

John comes from England.
John works for IMF .
John is an active hiker .

4th -> 2nd grade

Why does Sentence Simplification matter?

- Improve readability: help low-literacy readers

Source	Flesch	Grade Level
Comics	92	5th
Consumer ads in magazines	82	6th
Reader's Digest	65	8th-9th
New York Times	39	College
Standard auto insurance policy	10	College grad

John comes from England , works for IMF , and is an active hiker .

John comes from England.
John works for IMF .
John is an active hiker .

4th -> 2nd grade

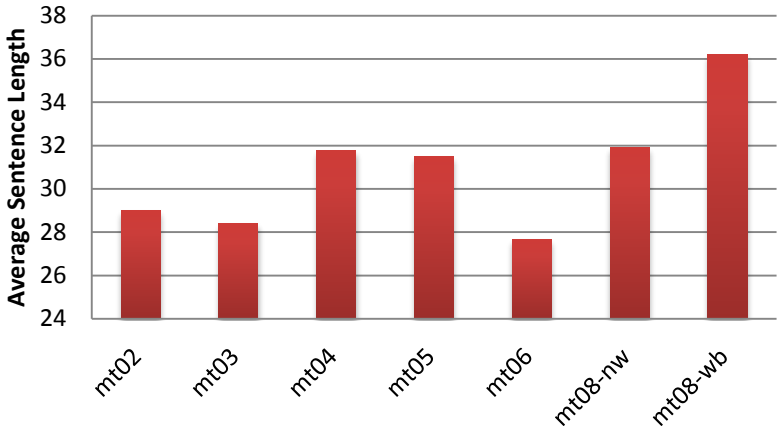
Why does Sentence Simplification matter?

- Improve readability
  - Help low-literacy readers
- Simplification for QA, IE tasks

  S: John comes from England , works for IMF , and is an active hiker .

  Q: Where does John come from?

  A: England
  - What is the relationship between John and IMF?
  - work_for(John, IMF)
- MT can benefit too



What's a simple sentence?

- A simple sentence in English (Klebanov, Knight, and Marcu, 2004)
  - Contains a single and independent clause
  - Structures: Subject + Verb + Object

- Examples
  - I have a dog.
  - The girl ran into her room.
  - The young boy climbed a tall tree.

Problem Definition

- Given an English sentence, return a simplified hypothesis which **contains several simple sentences and preserves the original meaning.**

- Underlying assumptions
  - Each simple sentence has 1S 1V 1O

e	*s*
John comes from England , works for IMF , and is an active hiker .	**John comes from England .** **John works for IMF .** **John is an active hiker .**
Hanoi was eclipsed by Hue during the Nguyen Dynasty as the capital of Vietnam , and Hanoi served as the capital of French Indochina from 1902 to 1954 .	**Hanoi was eclipsed by Hue during the Nguyen Dynasty as the capital of Vietnam .** **Hanoi served as the capital of French Indochina from 1902 to 1954 .**

Model

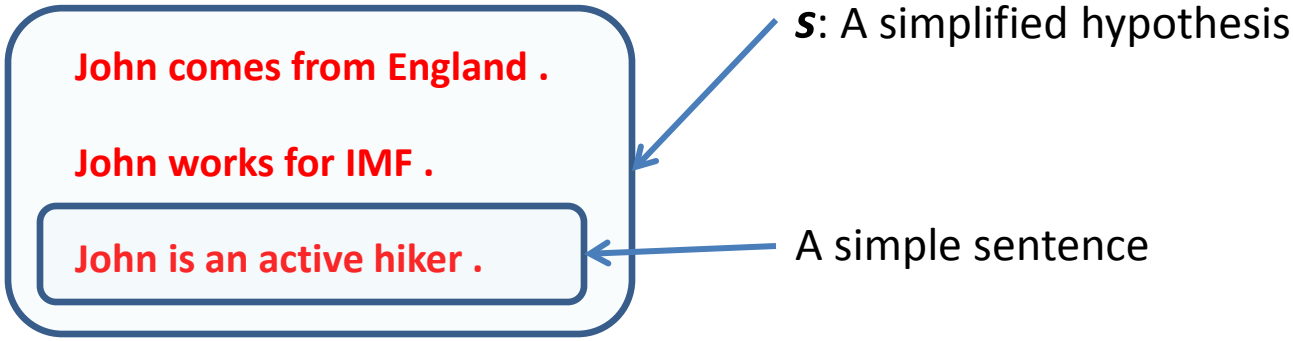
Original sentence

feature weight vector

$$p(s|e) = \frac{\exp(w^T f(e,s))}{\sum_{s'} \exp(w^T f(e,s'))}$$

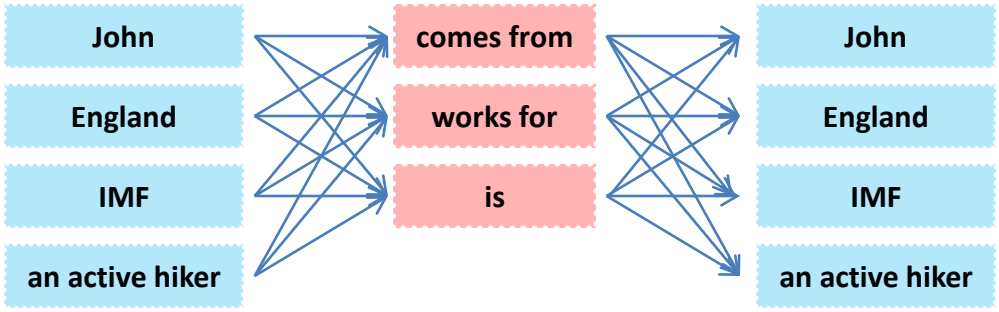
simplified candidate; ***s*** contains multiple simple sentences

feature weight vector



Constructing simple sentences

John	comes from	England	,	works for	IMF	, and	is	an active hiker	.
NP	VP	NP		VP	NP		VP	NP	



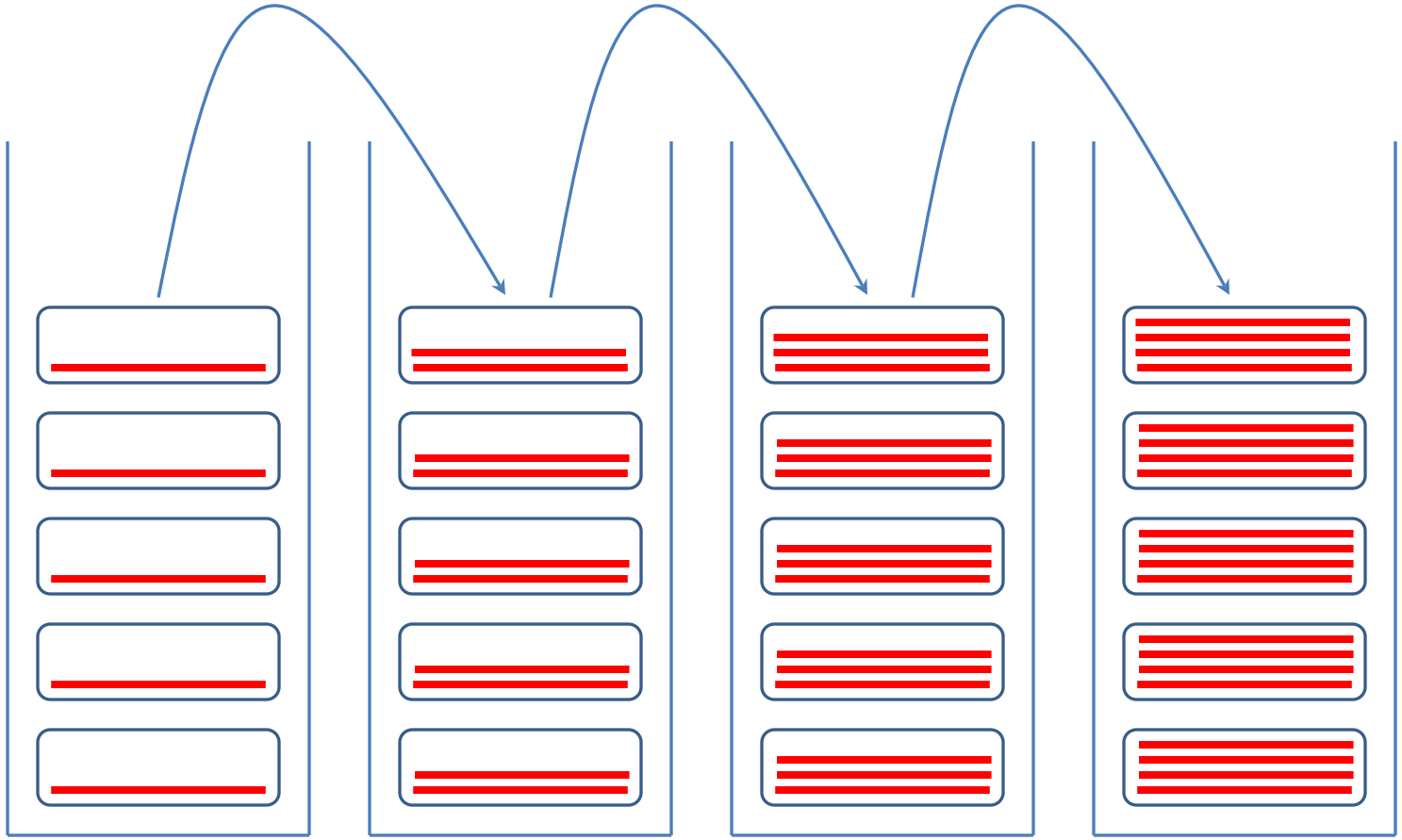
Subject **Verb** **Object**

Decoding: Beam Search

John comes from England , works for IMF , and is an active hiker .

Left-Right decoding according to objects

Decoding: beam search



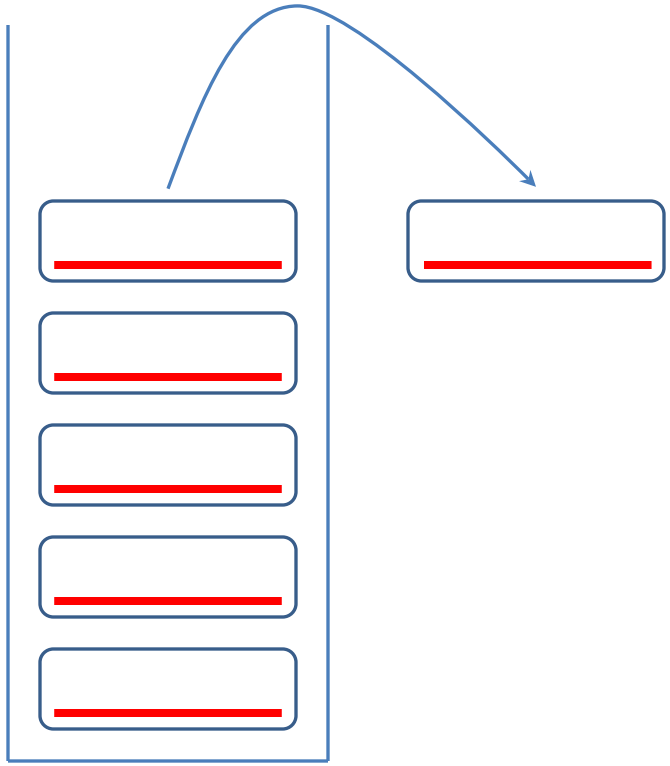
...

1
simple
sentence

2
simple
sentences

3
simple
sentences

4
simple
sentences



1
simple
sentence

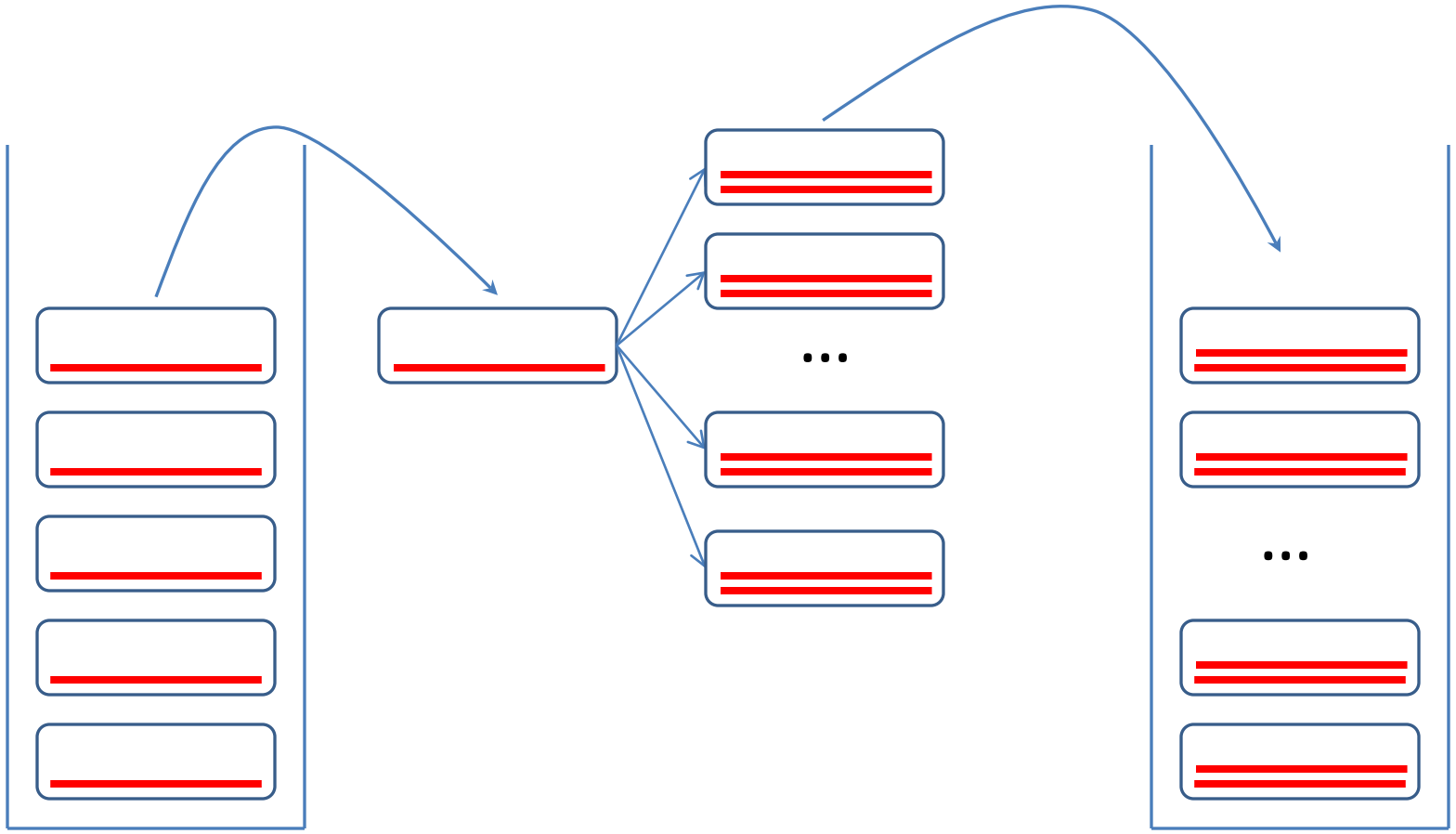
How many partial hypotheses in stack 1?

Exhaustive search: m^2n

m: number of NPs
n: number of VPs

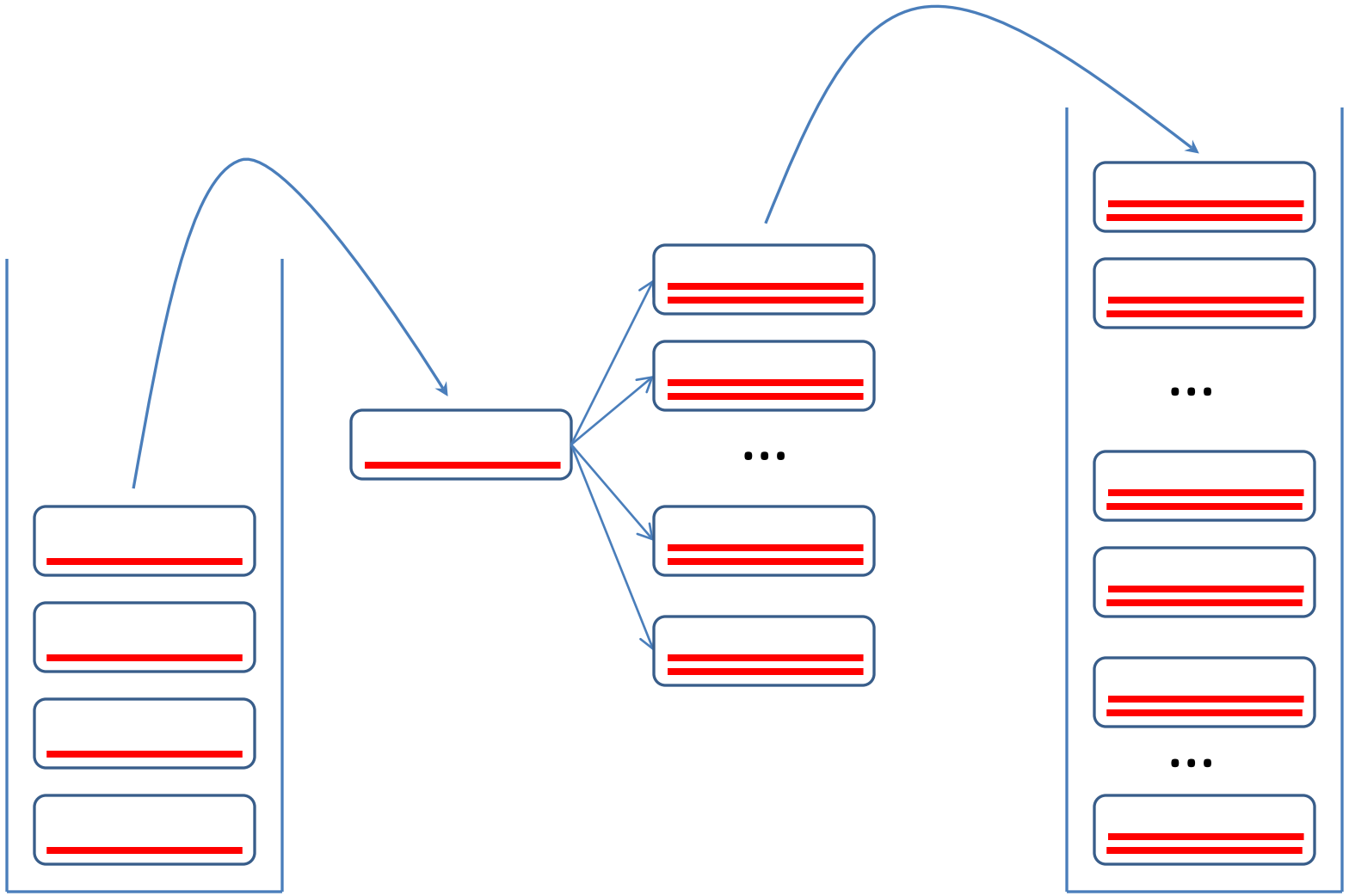
What is the maximum number of simple sentence per hypothesis?

n+1



1
simple
sentence

2
simple
sentences



1
simple
sentence

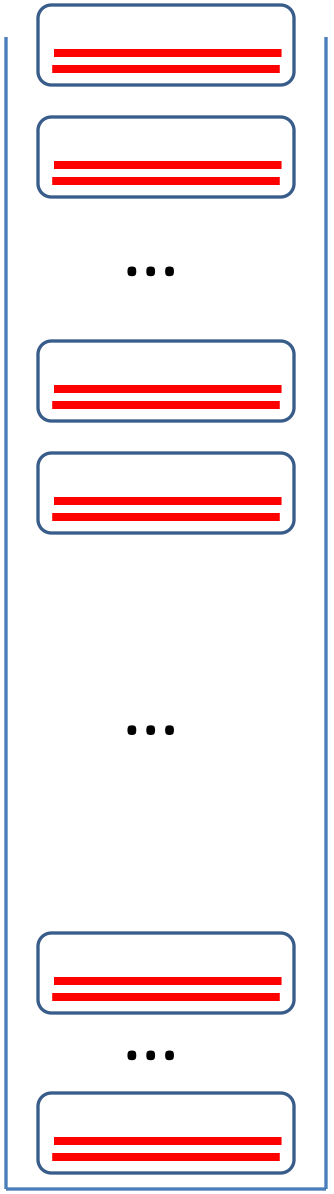
2
simple
sentences

How many partial hypotheses in stack 2?

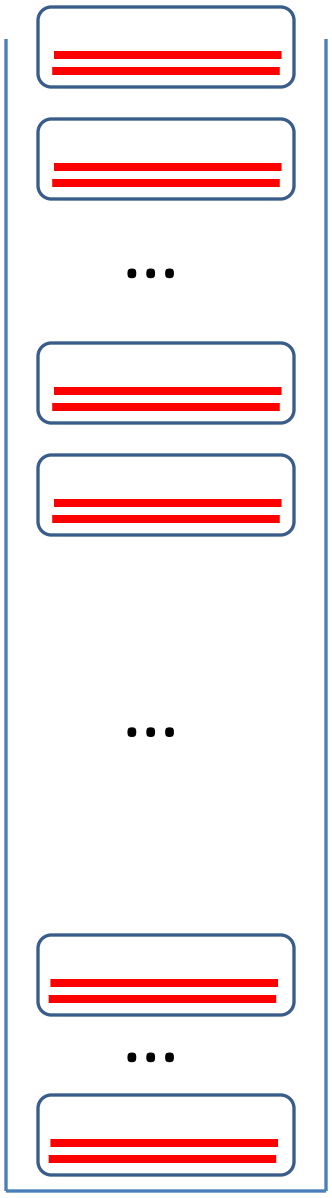
Exhaustive search: $(m^2n)^2$

m: number of NPs
n: number of VPs

1
simple
sentence



2
simple
sentences



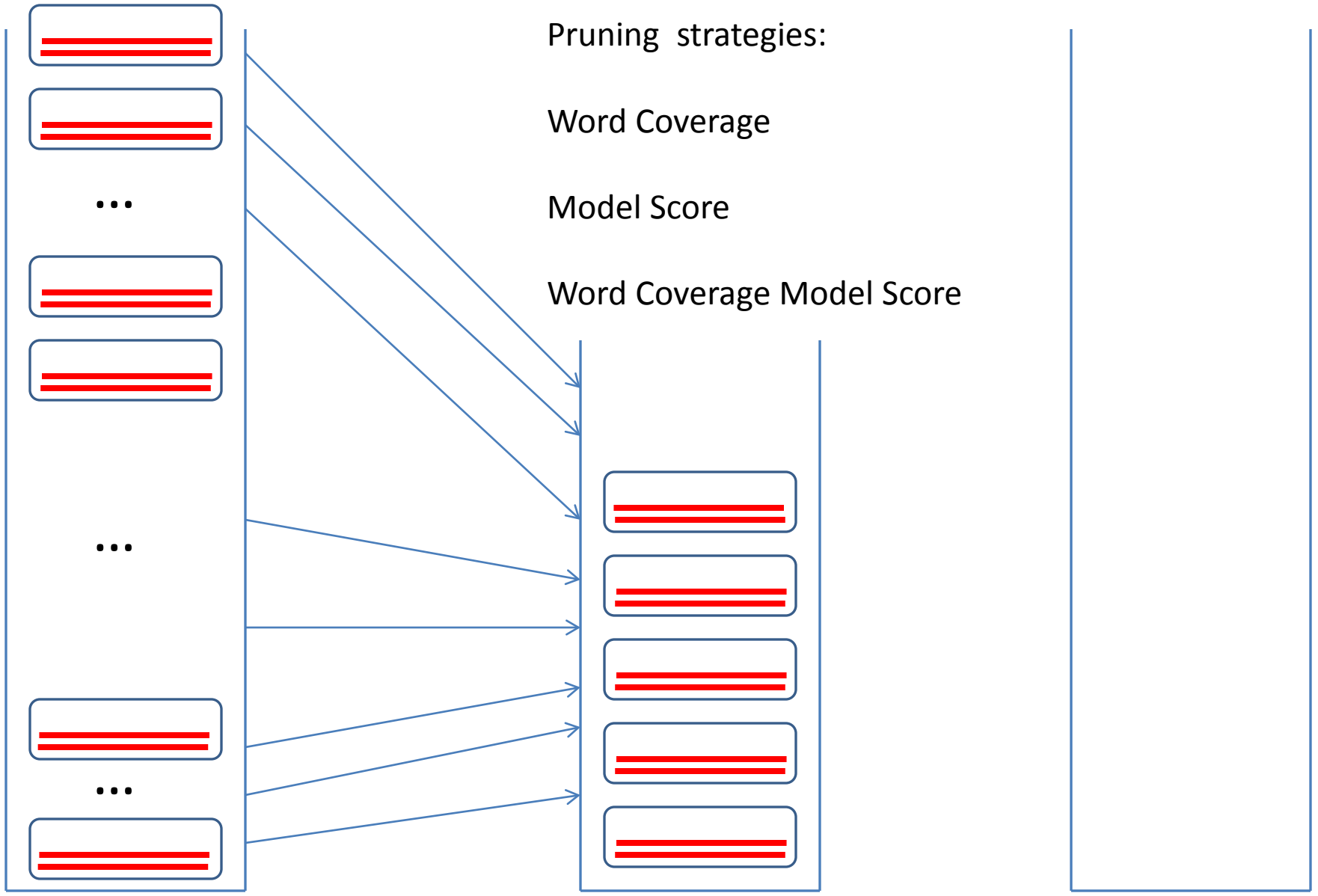
2
simple
sentences

How many partial hypotheses in stack 3?

Exhaustive search: $(m^2n)^3$

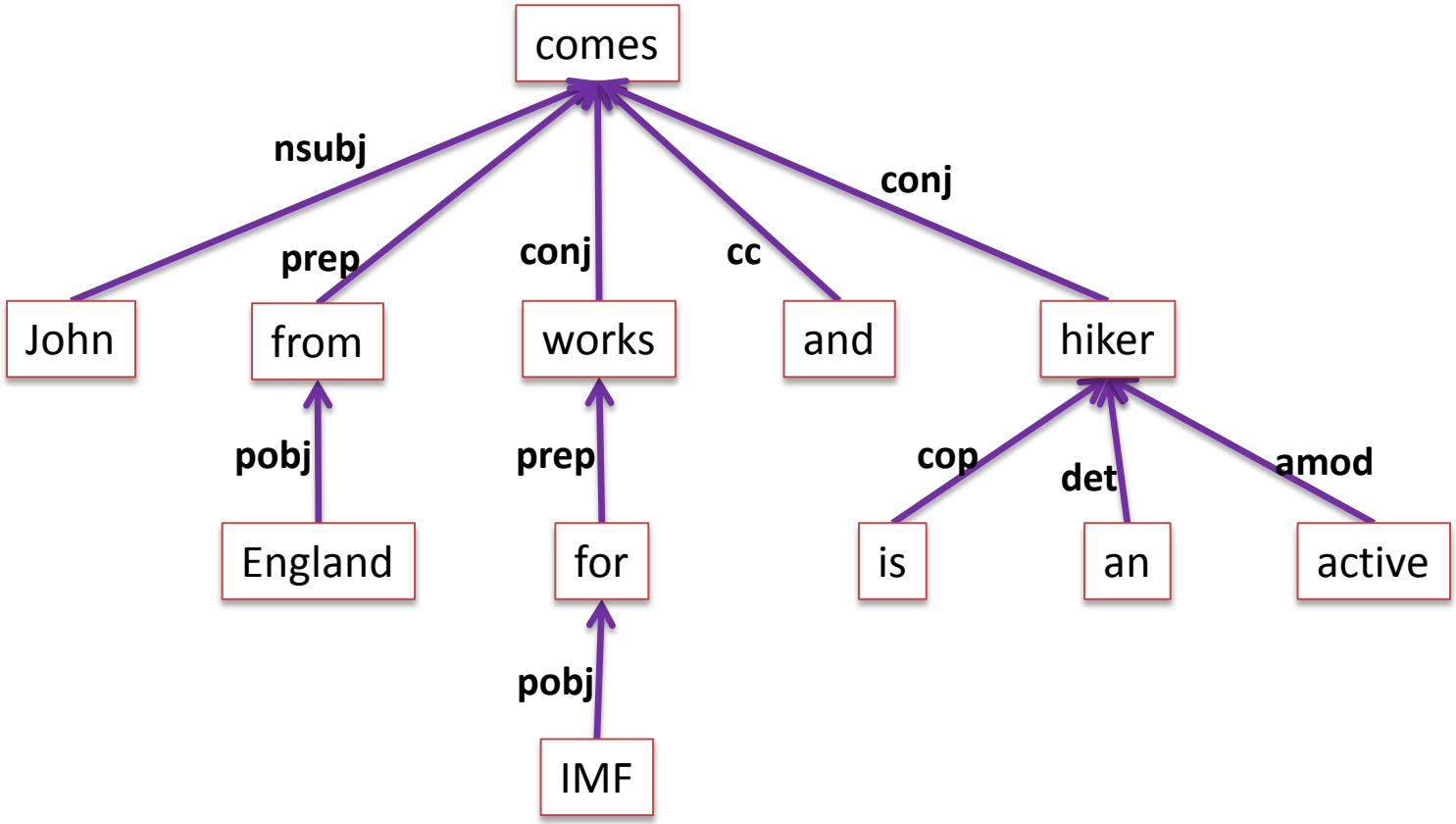
m: number of NPs
n: number of VPs

3
simple
sentences


Pruning strategies:
Word Coverage
Model Score
Word Coverage Model Score
2
simple
sentences
3
simple
sentences


Sort by model score and
extract k-best list
2
simple
sentences
3
simple
sentences
4
simple
sentences

Dependency Structure Features

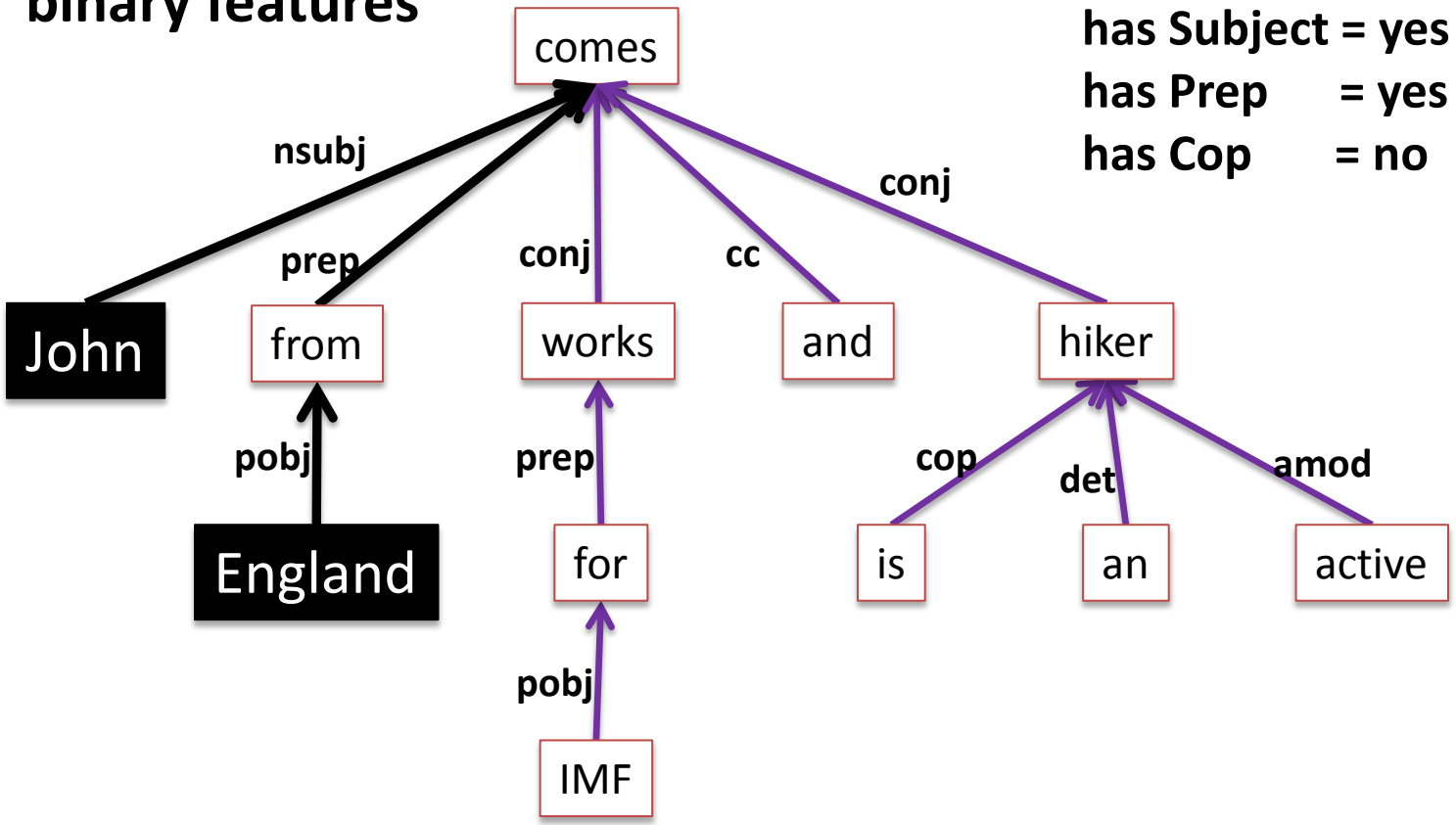


Dependency Structure Features

Typed dependency binary features

John : England

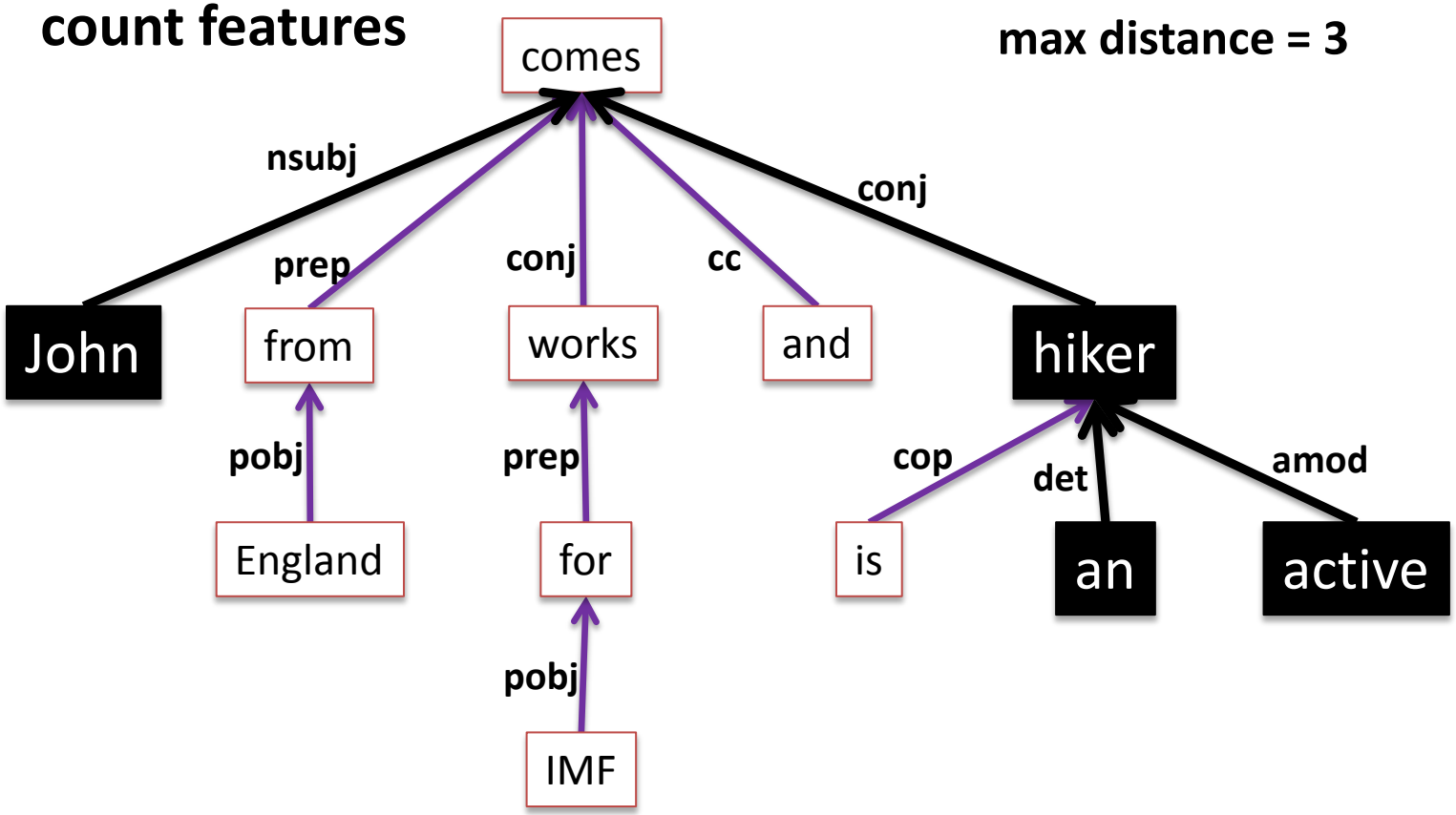
has Object = yes
has Subject = yes
has Prep = yes
has Cop = no



Dependency Structure Features

Dependency chain count features

John : an active hiker
min distance = 2
max distance = 3



Other Features

- Count
  - How many simple sentences per hypothesis?
  - How many words in subject, verb and object?
- Syntactic
  - Crossing SBAR; Single pronoun; PP attachment; …
- Readability
  - Flesch, Kincaid, Fog
- Word coverage
  - NP, VP, edit distance

- Total 177 features

Training

$$p(s|e) = \frac{\exp(w^T . f(e, s))}{\sum_{s'} \exp(w^T . f(e, s'))}$$

- $argmax_{s'}$
- Optimize ***w*** subject to a loss function
- Loss functions
  - Average n-gram co-occurrence F score
  - ROUGE-n

Training: EM-like Algorithm

Parameters *w*

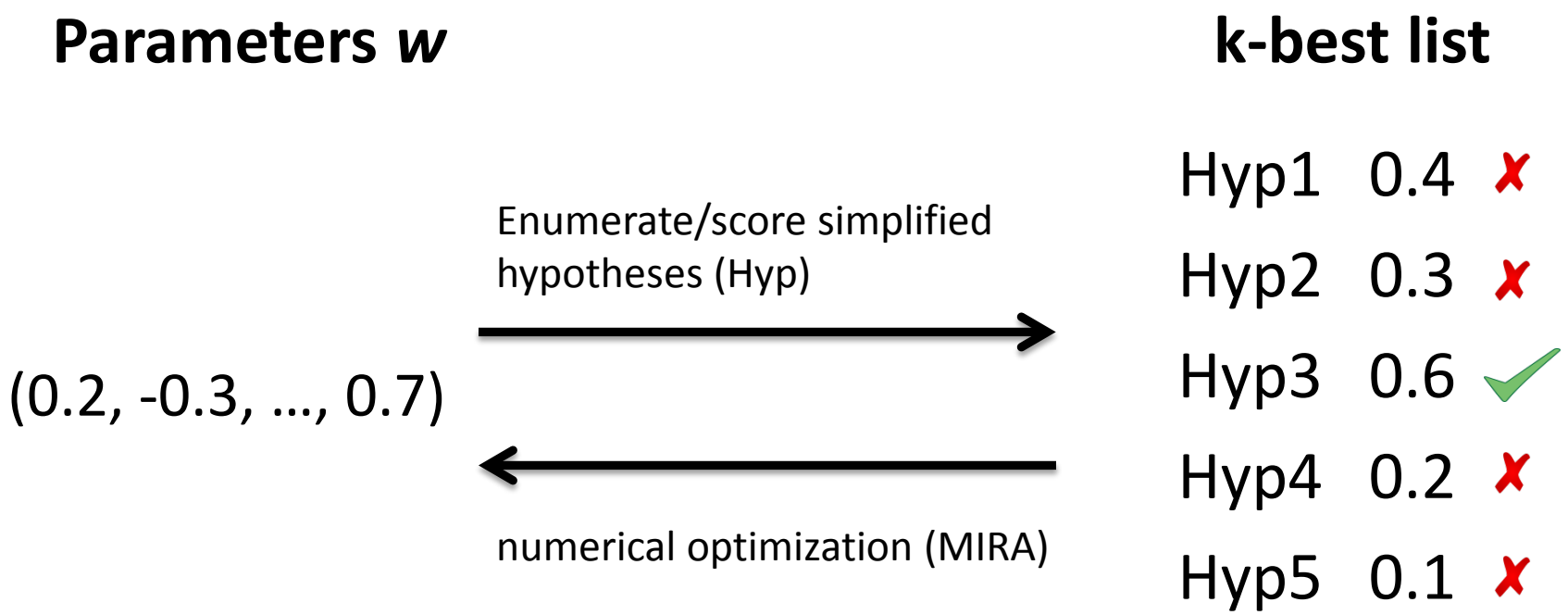
(1, 1, …, 1)

Enumerate/score simplified hypotheses (Hyp)

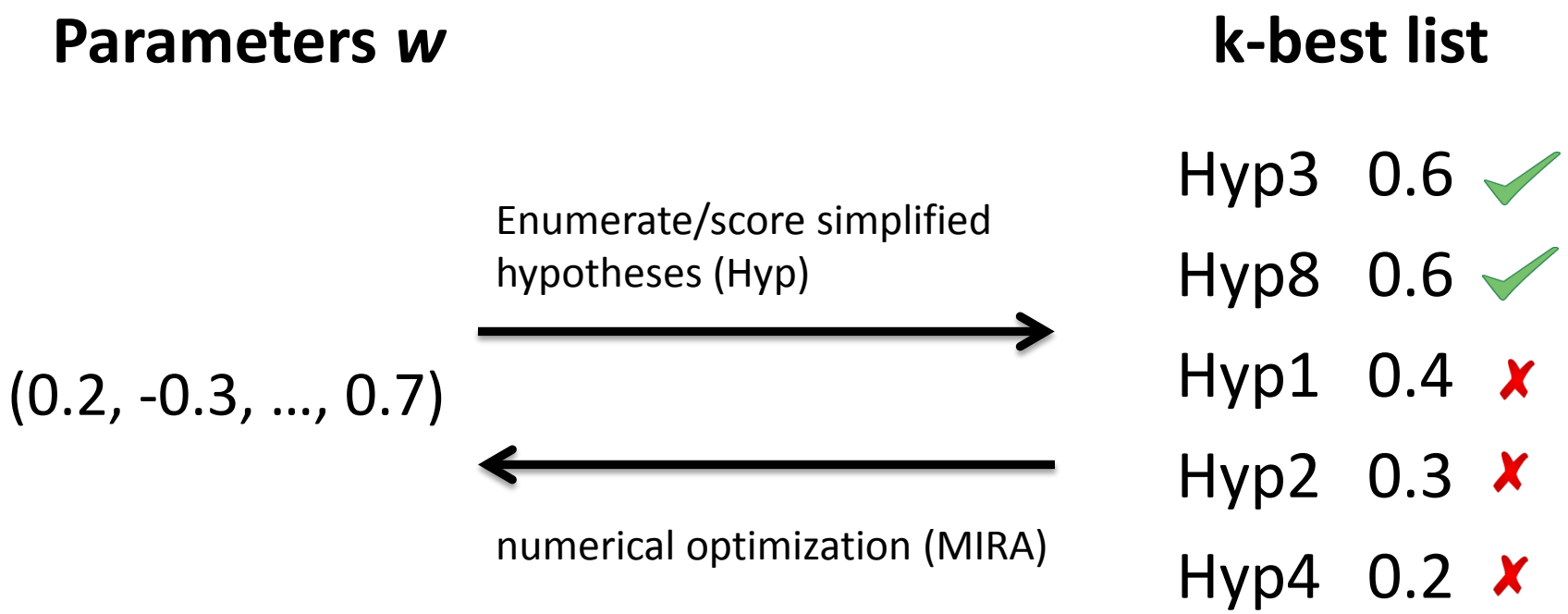
k-best list

Hyp1	0.4	✗
Hyp2	0.3	✗
Hyp3	0.6	✓
Hyp4	0.2	✗
Hyp5	0.1	✗

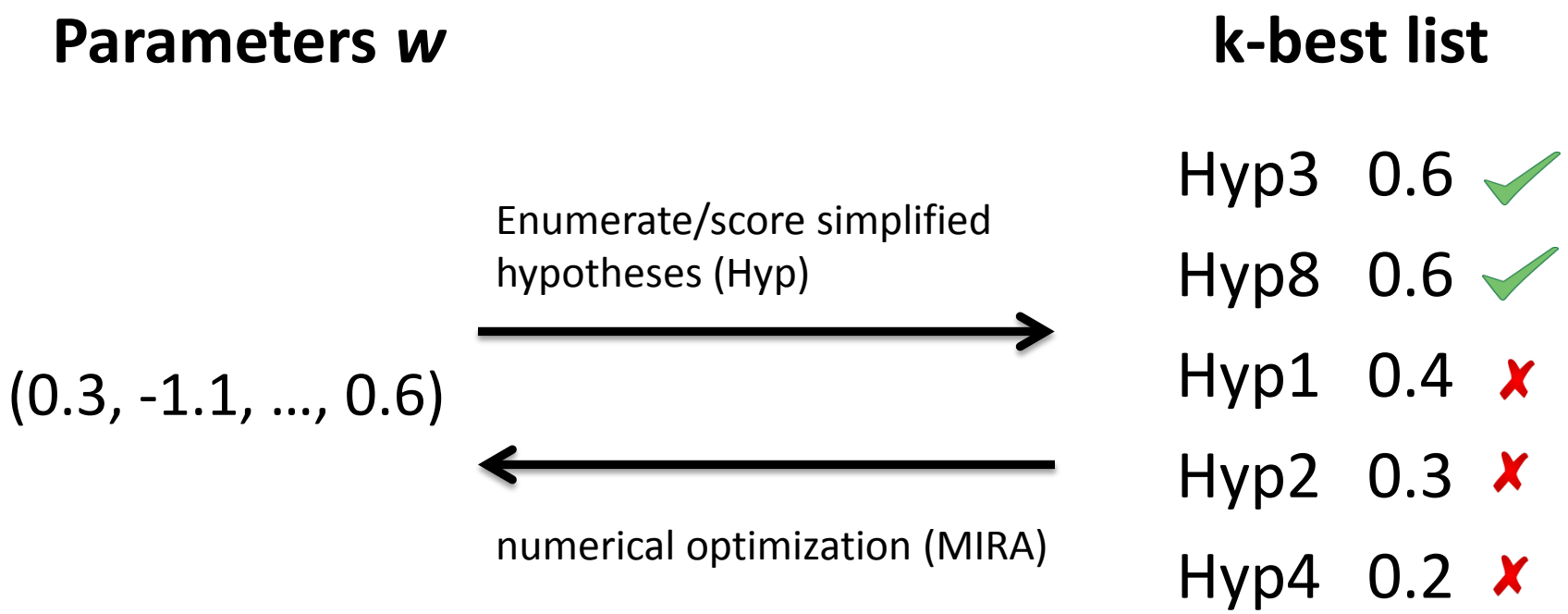
Training: EM-like Algorithm



Training: EM-like Algorithm



Training: EM-like Algorithm



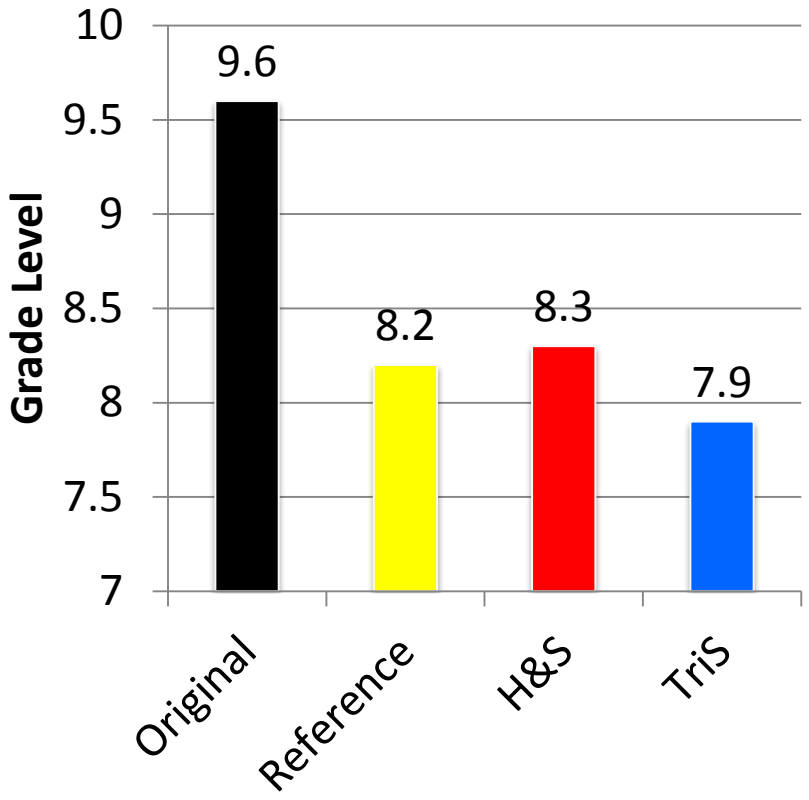
Experiment Setup

- Metrics:
  - Flesch-Kincaid grade level: the number of years of education generally required to understand a text.
  - ROUGE-n: n-gram co-occurrence between hypothesis and reference
- Data
  - Training set: 754 sentences
  - Unseen test set: 100 sentences
  - 70% Wikipedia, 25% NY Times, 5% synthetic

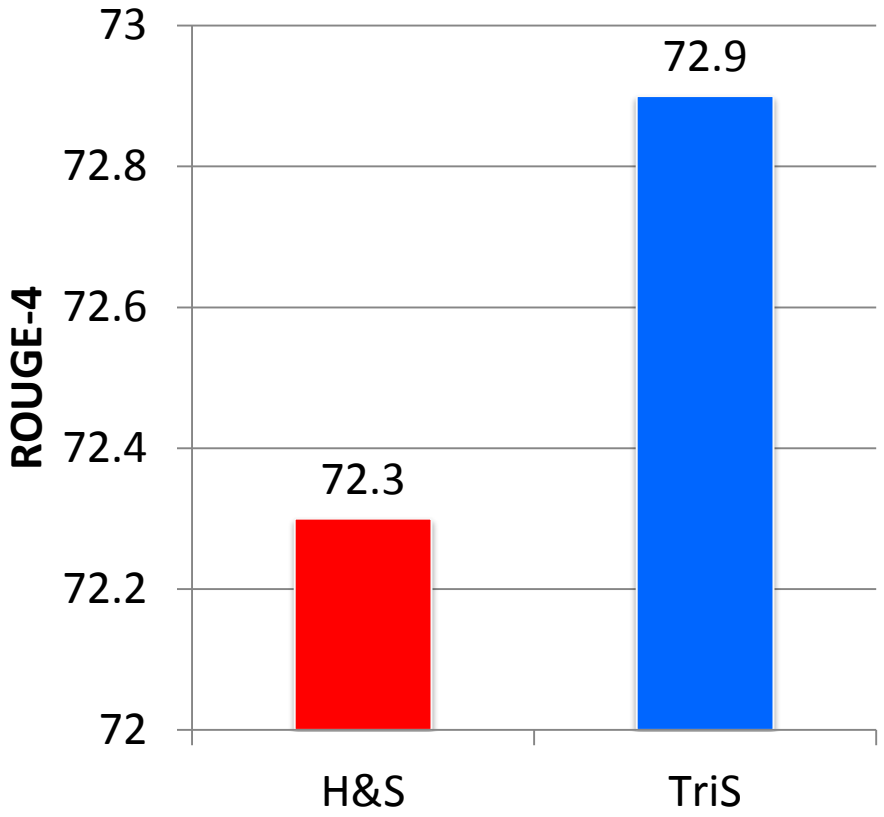
Experiment 1: Bring down grade level

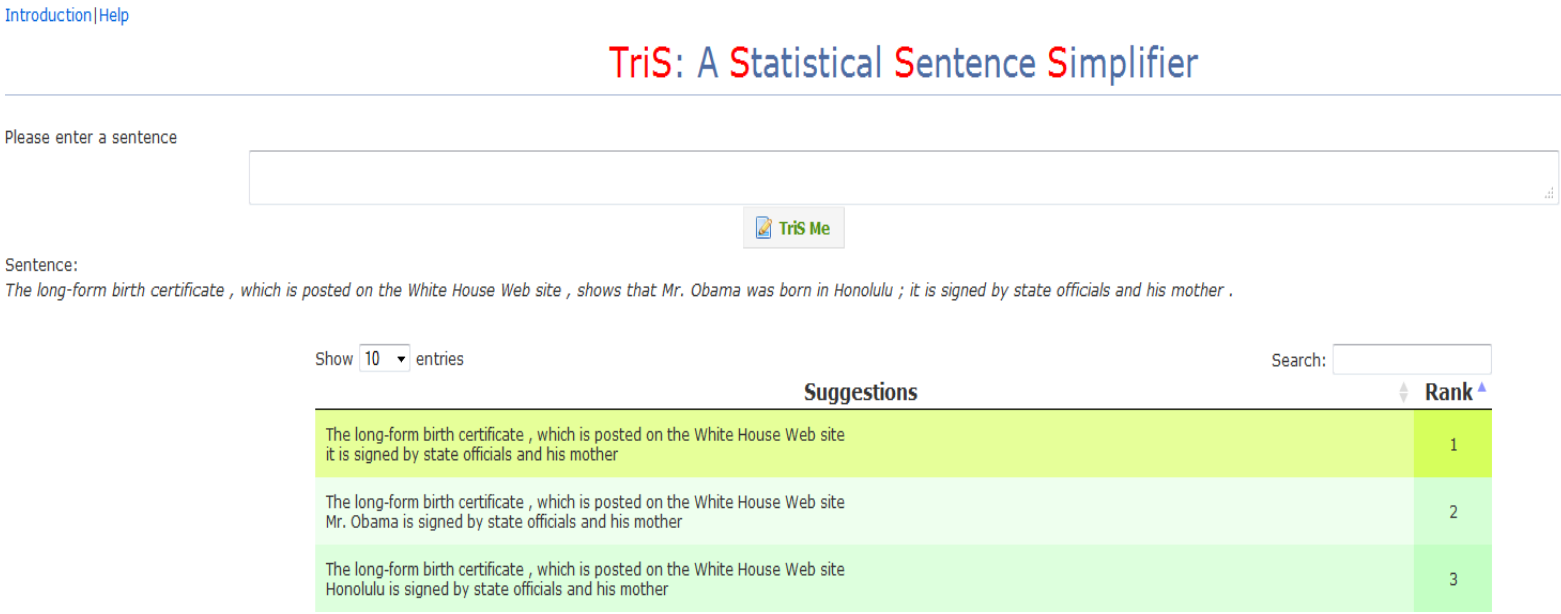
H&S (Heilman & Smith, 2010): rule-based

TriS (our system): statistical-based



Experiment 2: Improve Automatic Score ROUGE





http://bit.ly/TrisMe

Email nbach@cs.cmu.edu for source code

Original

Warren Weaver was an American scientist and is widely recognized as one of the pioneers of machine translation .

Simplification

Warren Weaver was an American scientist .

Warren Weaver is widely recognized as one of the pioneers of machine translation .

Reference

Warren Weaver was an American scientist .

Warren Weaver is widely recognized as one of the pioneers of machine translation .

Original

An elderly Georgian woman was scavenging for copper to sell as scrap when she accidentally sliced through an underground cable and cut off Internet services to all of neighbouring Armenia , it emerged on Wednesday .

Simplification

An elderly Georgian woman was scavenging for copper to sell .

scrap cut off Internet services to all of neighbouring Armenia .

Reference

An elderly Georgian woman was scavenging for copper to sell as scrap .

she accidentally sliced through an underground cable .

she cut off Internet services to all of neighbouring Armenia .

it emerged on Wednesday .

Contributions

- Algorithmic
  - Decoding with beam search algorithm
  - Discriminative training algorithm
  - Dependency structure features
- Experimental
  - Reduce the grade level
  - Outperforms a state-of-the-art rule-based system

Thank You!