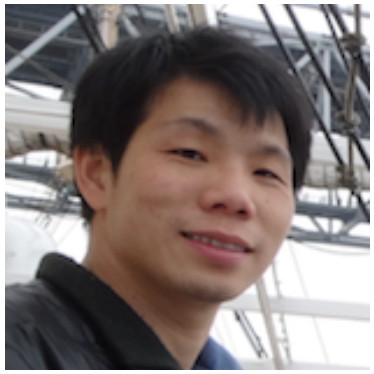
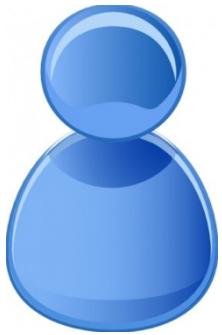
Cross-lingual Open Information Extraction

Sheng Zhang Kevin Duh Ben Van Durme

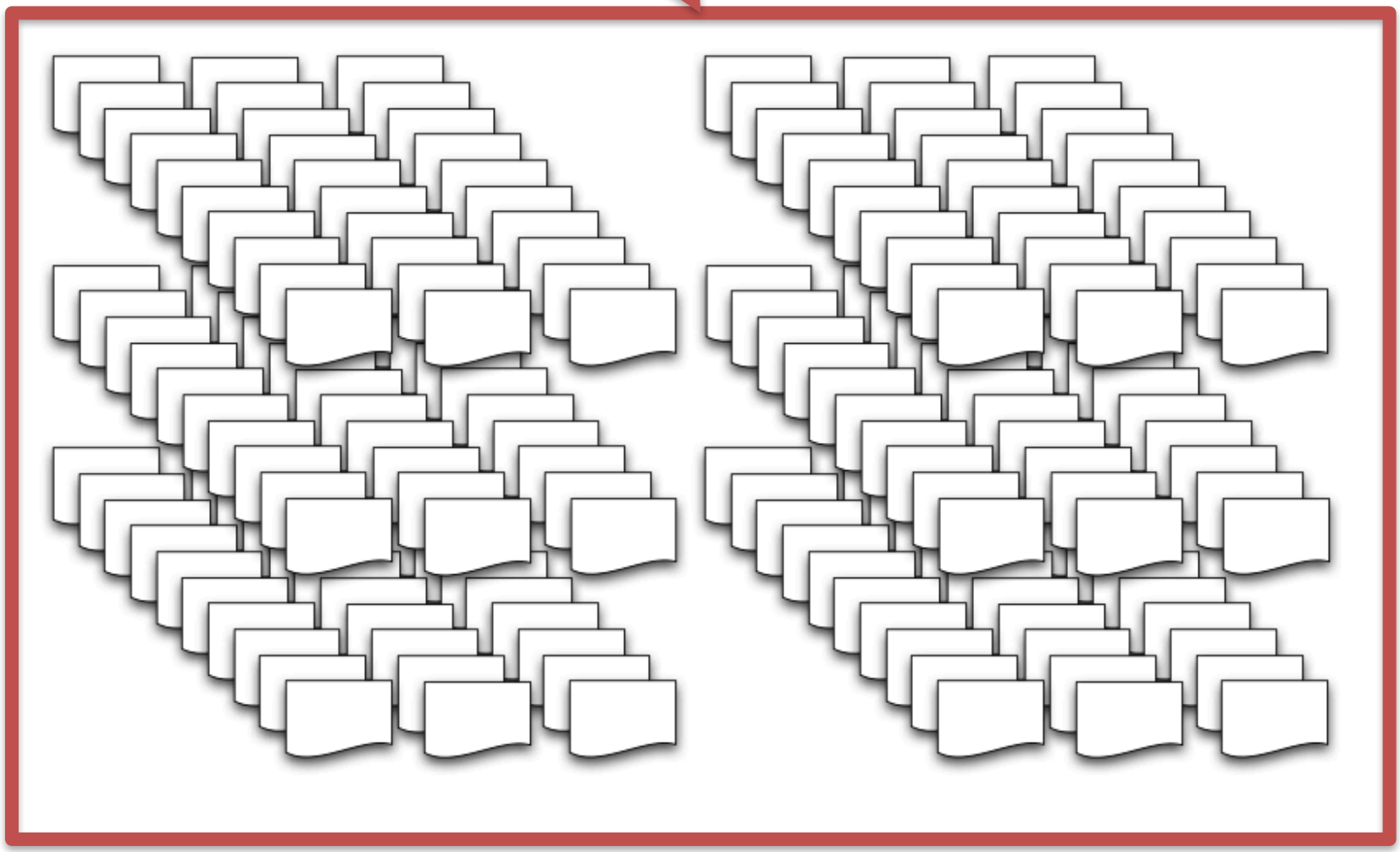




Johns Hopkins University

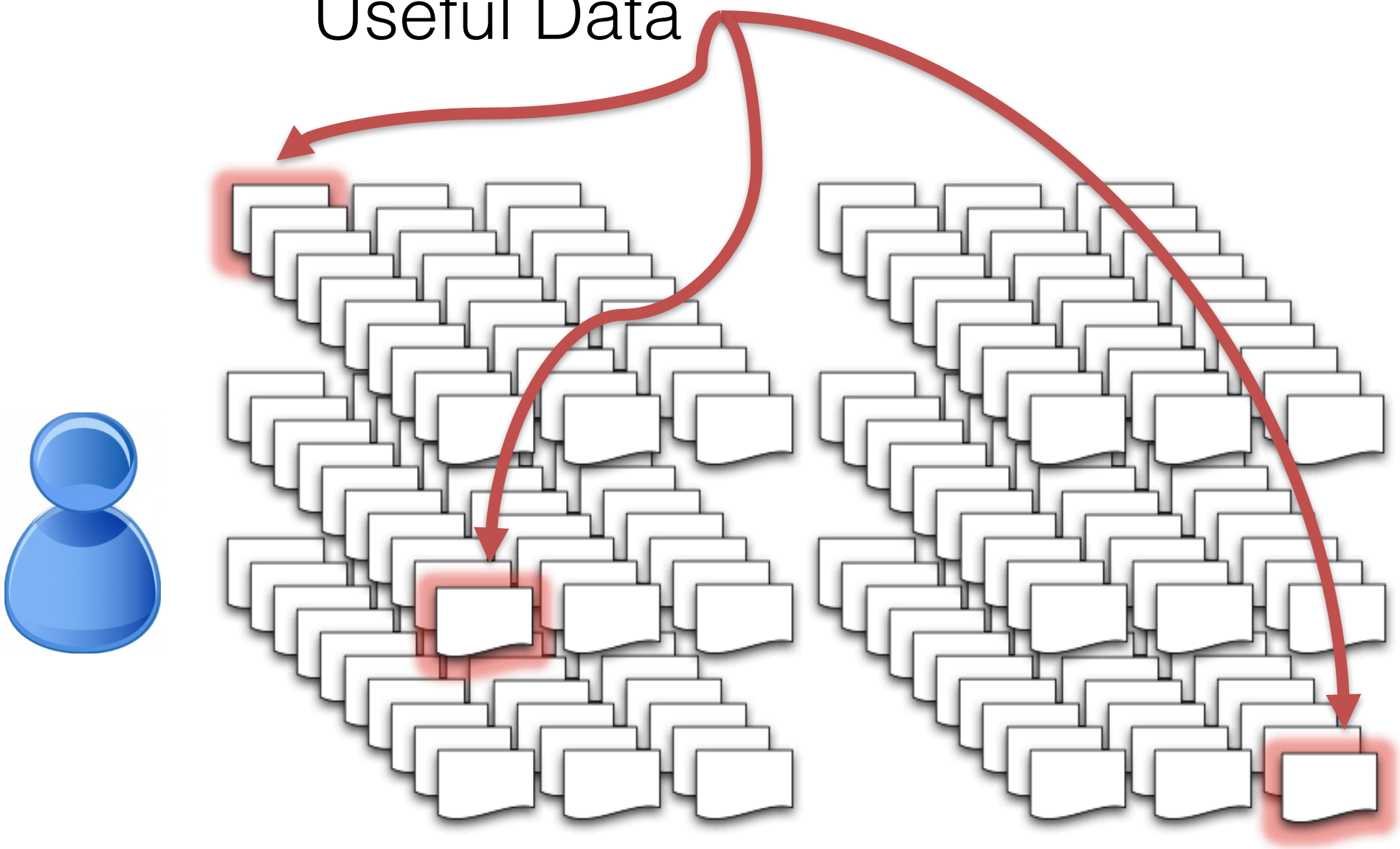


Analyst: User with complex information need

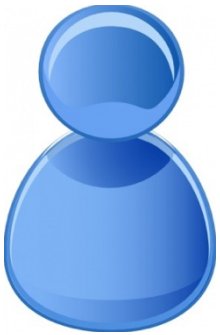
Data

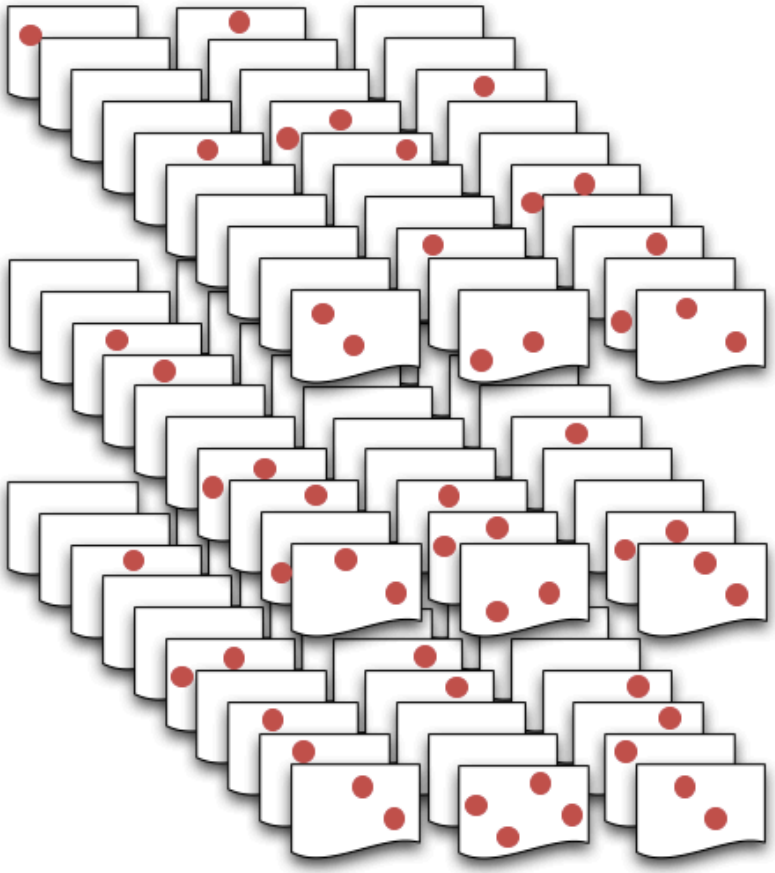


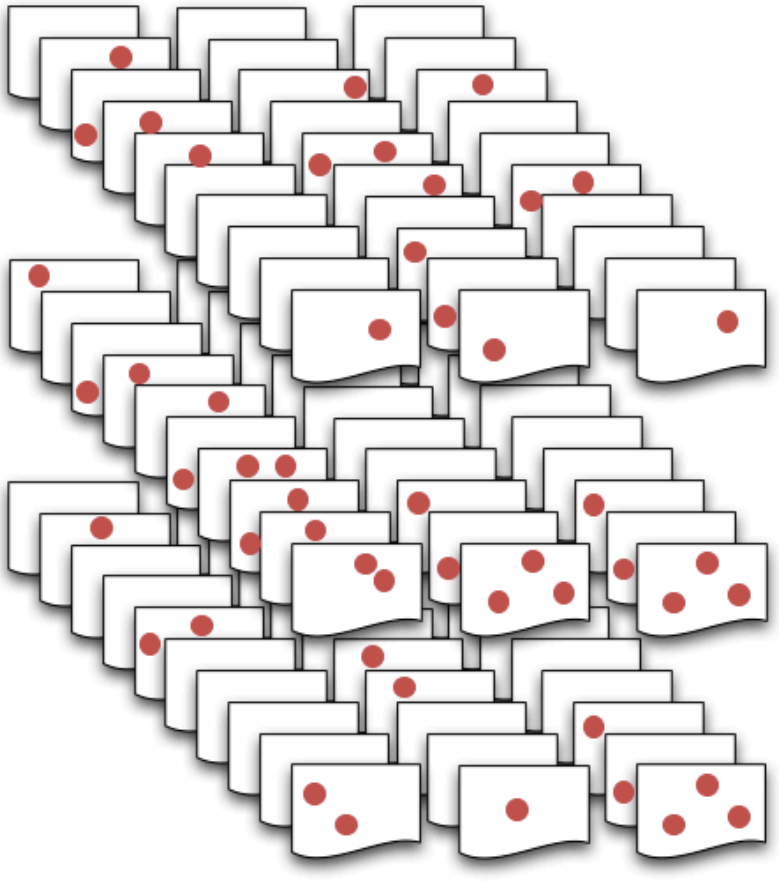
Useful Data



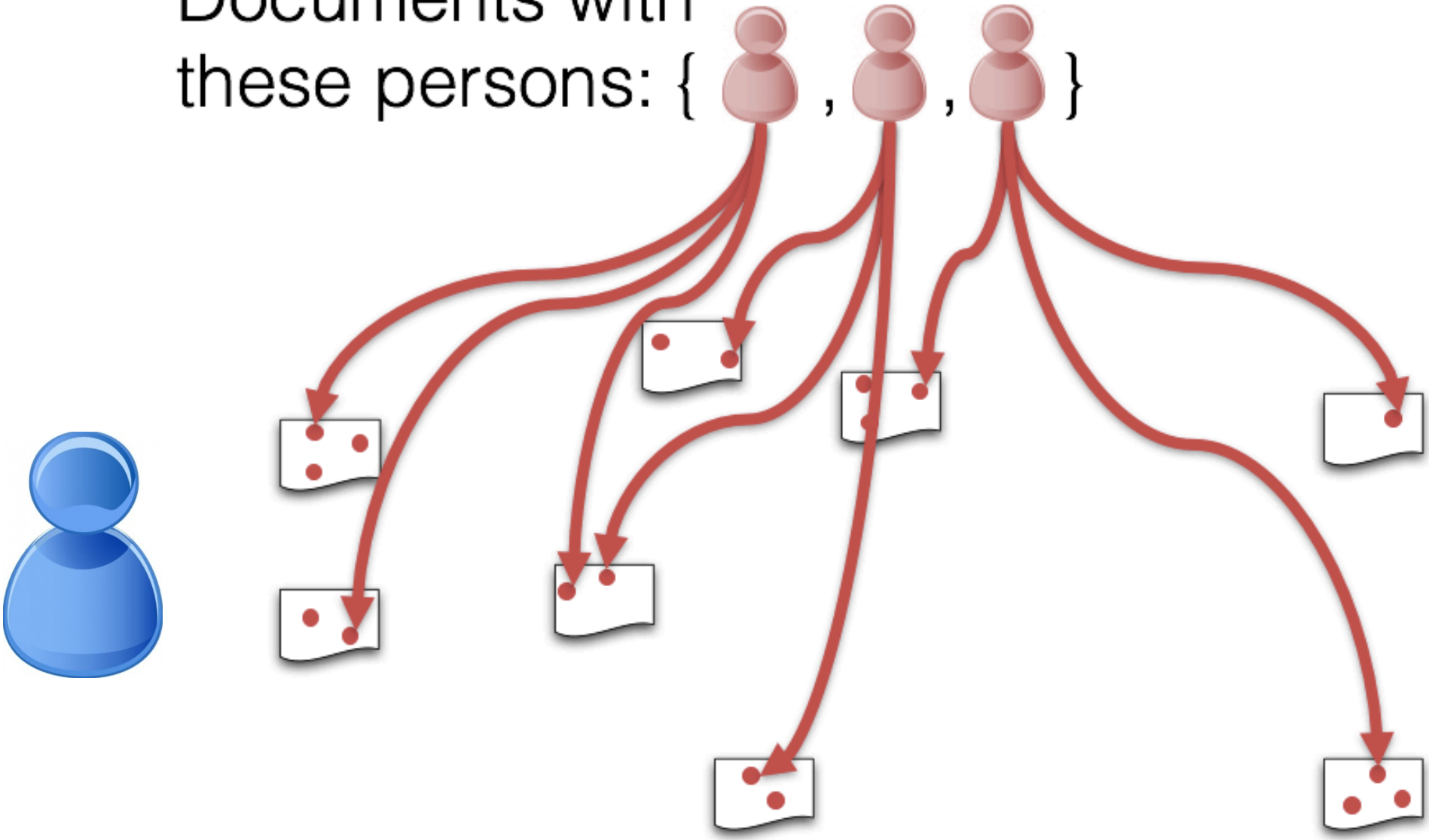
Entity Recognition



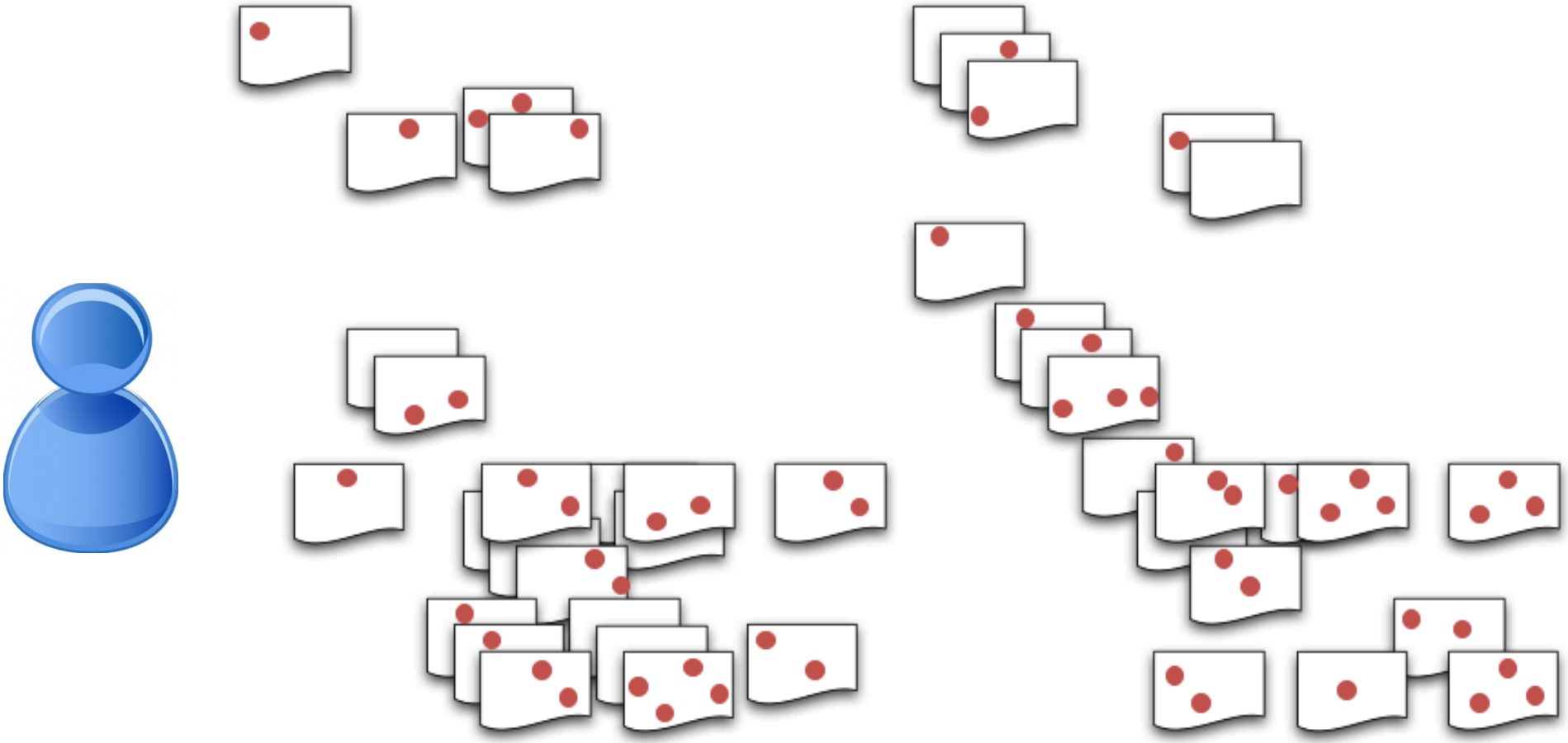




Documents with these persons: { , , }

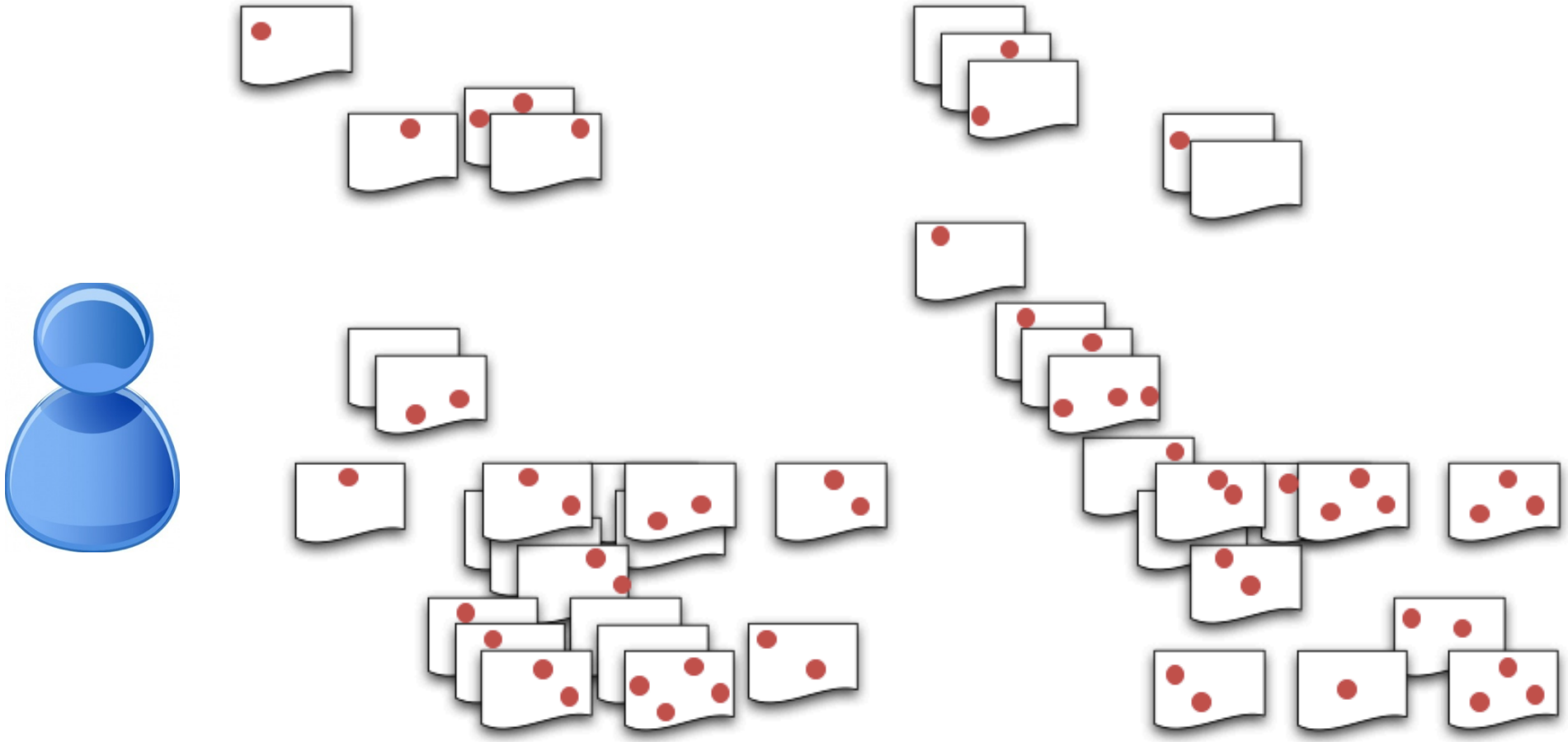


Documents with a:
CausalRelation(DISEASE, SYMPTOM)



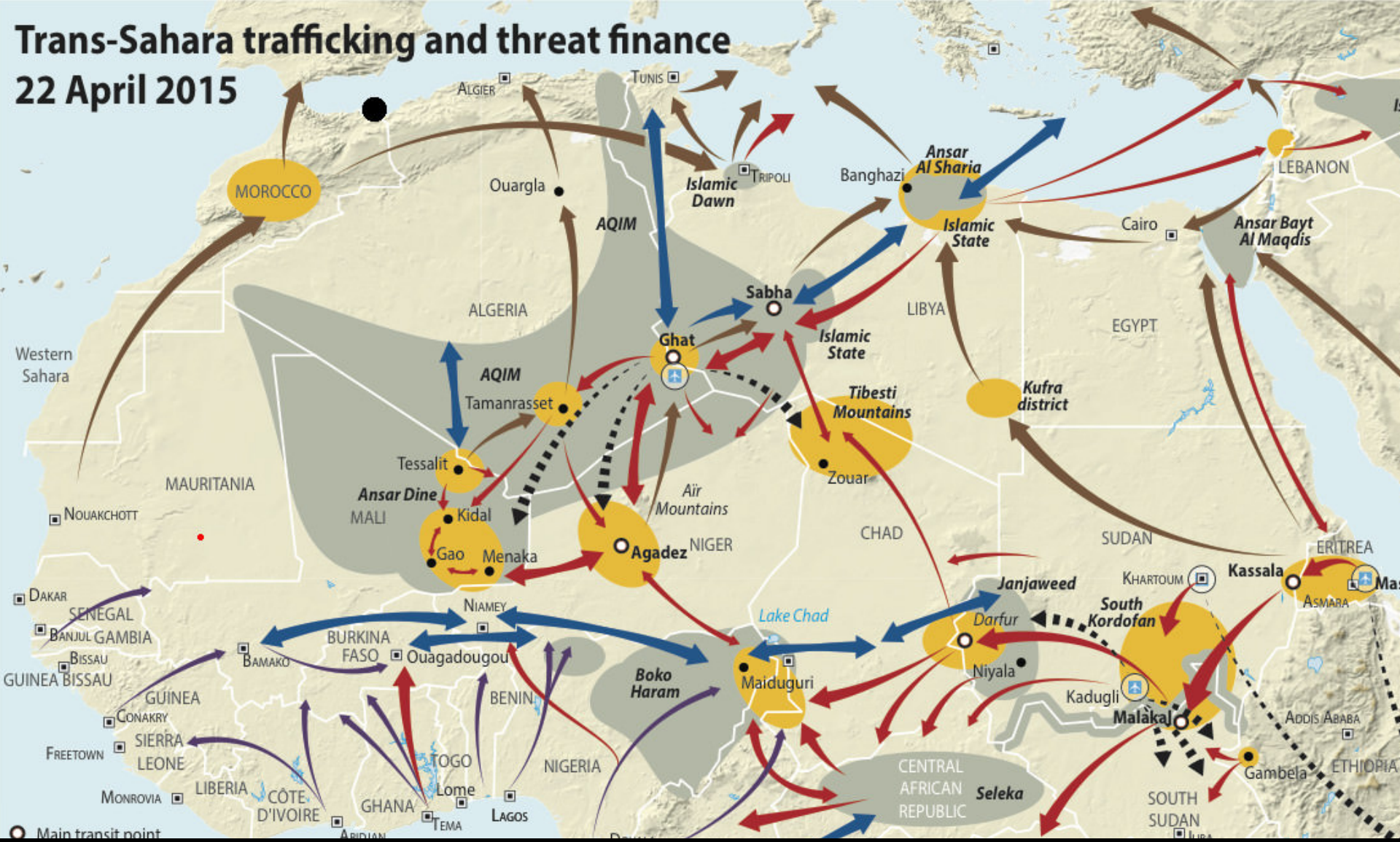
Documents with a:

CausalRelation(DISEASE, SYMPTOM)



Foreign Language Documents

Financial Analyst: Do I invest in commodity futures?



Intelligence Analyst: How are the terrorists connected?

Aid Worker: Which locations have immediate need?

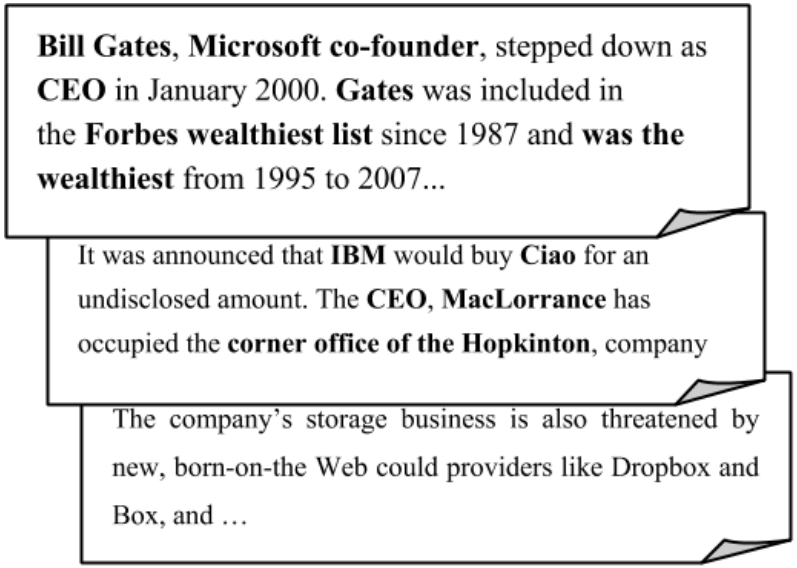
Outline

1. Motivation
2. Problem Definition
3. Pipeline vs. Joint Solution
4. Improved Joint Solution

Task Formulation

- Cross-lingual:
  - analyst speaks English, but document collection is in other languages
- Cross-lingual Information Retrieval?
  - Document unit is too large
- Cross-lingual Question Answering?
  - Difficulty in formulating questions
- Cross-lingual Information Extraction?
  - Close, but no fixed ontology

Information Extraction vs. Open Information Extraction



IE

Co-founder(Bill Gates, Microsoft)
Director-of (MacLorrance, Ciao)
Employee-of (MacLorrance, Ciao)
…

Open IE

(Bill Gate, be, Microsoft co-founder)
(Bill Gates, stepped down as, CEO)
(Bill Gates, was included in, the Forbes wealthiest list)
(Bill Gates, was, the wealthiest)
(IBM, would buy, Ciao)
(MacLorrance, has occupied, the corner office of the Hopkinton)
…

	IE	Open IE
Input	Sentences + Labeled relations	Sentences
Relation	Specified relations in advance	Free discovery
Extractor	Specified relations	Independent-relations

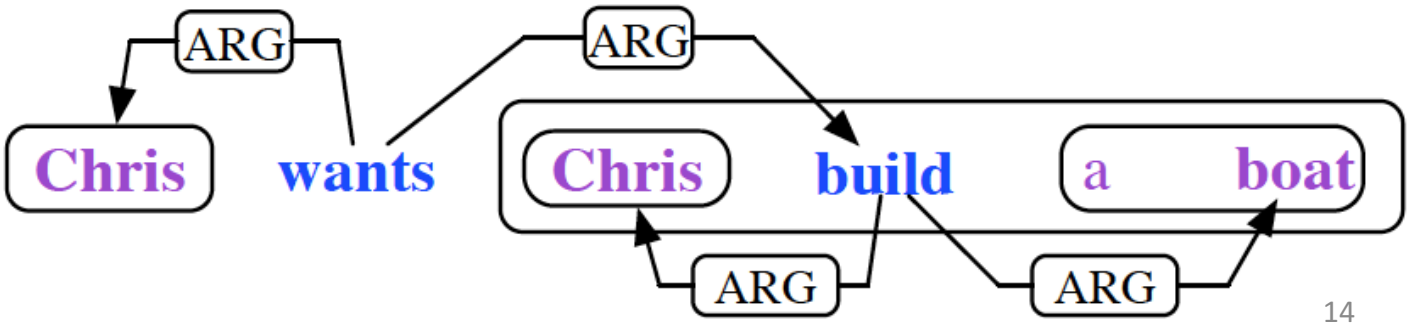
Figure from: Duc-Thuan Vo and Ebrahim Bagheri. (2016) "Open Information Extraction" https://arxiv.org/pdf/1607.02784.pdf

Input: Chinese sentence

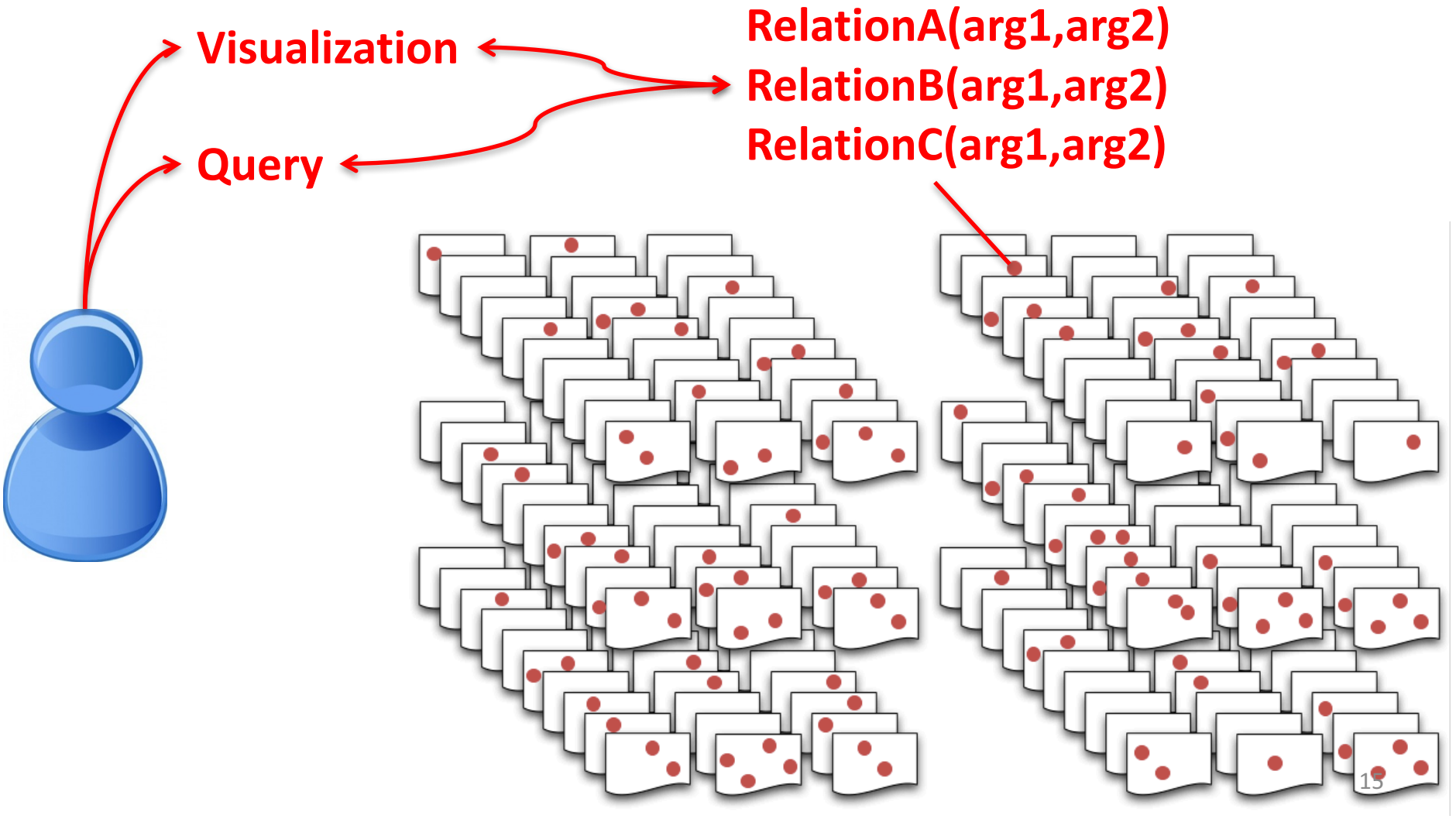
克里斯 想 造 一 艘 船 。

Cross-lingual
Open Information Extraction

Output: A set of English tuples, e.g. Relation(arg1,arg2)



Cross-lingual Open Information Extraction



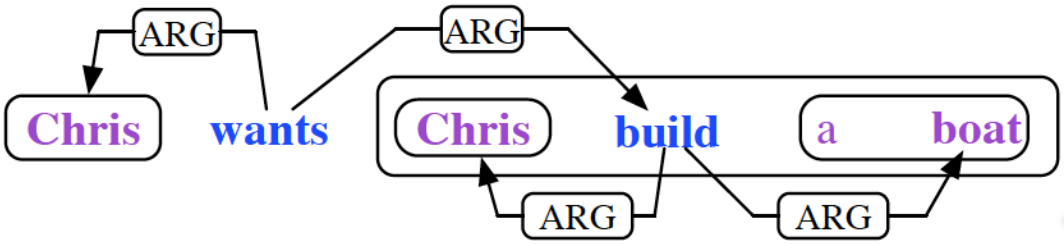
Assumptions

1. Training data: Chinese-English bitext
2. Monolingual Open IE system in English

BITEXT

克里斯 想 造 一 艘 船 。

Chris wants to build a boat



AUTOMATIC OPEN IE RESULT

Monolingual Open IE System

PredPatt: https://github.com/hltcoe/PredPatt

- Based on Universal Dependencies
- Rules for:
  1. identifying predicate root and argument root:
     e.g. nsubj(s, v), dobj(o, v)
  2. resolving arguments:
     Chris expects to visit Pat → nsubj(Chris,visit)
     Chris likes to sing and dance → nsubj(Chris,dance)
  3. phrase extraction:
     PredPatt finds structure in text → ?a finds ?b in ?c

Pierre Vinken , 61 years old , will join the board as a nonexecutive director Nov. 29 .

?a is/are 61 years old

?a: Pierre Vinken

?a will join ?b as ?c ?d

?a: Pierre Vinken , 61 years old

?b: the board

?c: a nonexecutive director

?d: Nov. 29

?a is/are nonexecutive

?a: a director

Mr. Vinken is chairman of Elsevier N.V. , the Dutch publishing group .

?a is chairman of ?b

?a: Mr. Vinken

?b: Elsevier N.V.

?a is/are the Dutch publishing group

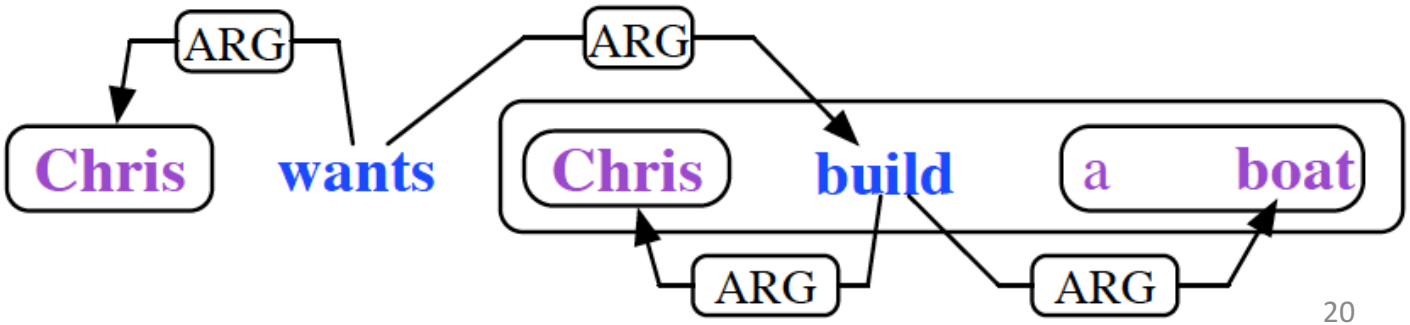
?a: Elsevier N.V.

Input: Chinese sentence

克里斯 想 造 一 艘 船 。

Cross-lingual
Open Information Extraction

Output: A set of English tuples, e.g. Relation(arg1,arg2)

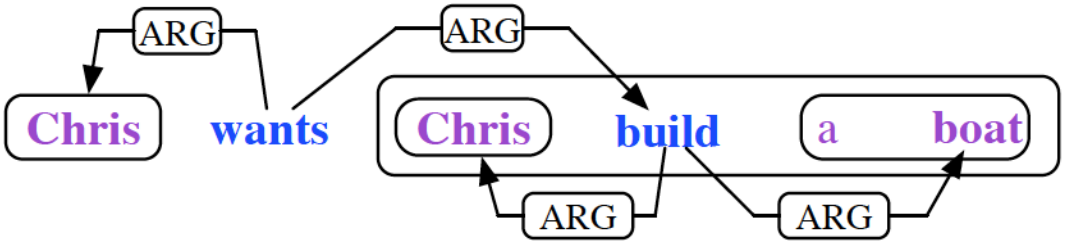


Outline

1. Motivation
2. Problem Definition
3. Pipeline vs. Joint Solution
4. Improved Joint Solution

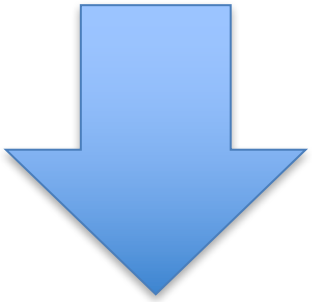
克里斯 想 造 一 艘 船 。

Chris wants to build a boat



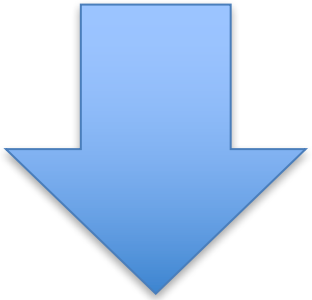
PIPELINE SOLUTION

克里斯想造一艘船。

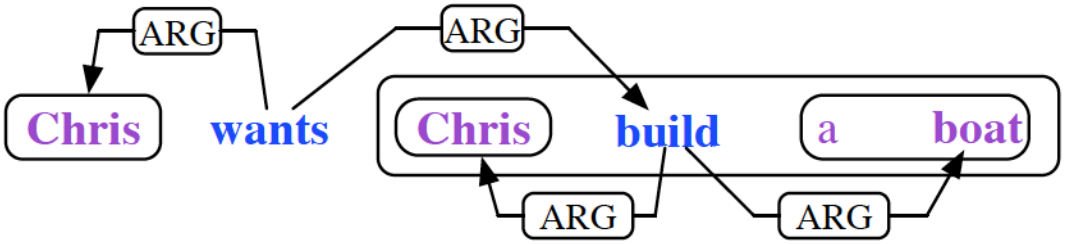


Machine Translation

Chris wants to build a boat

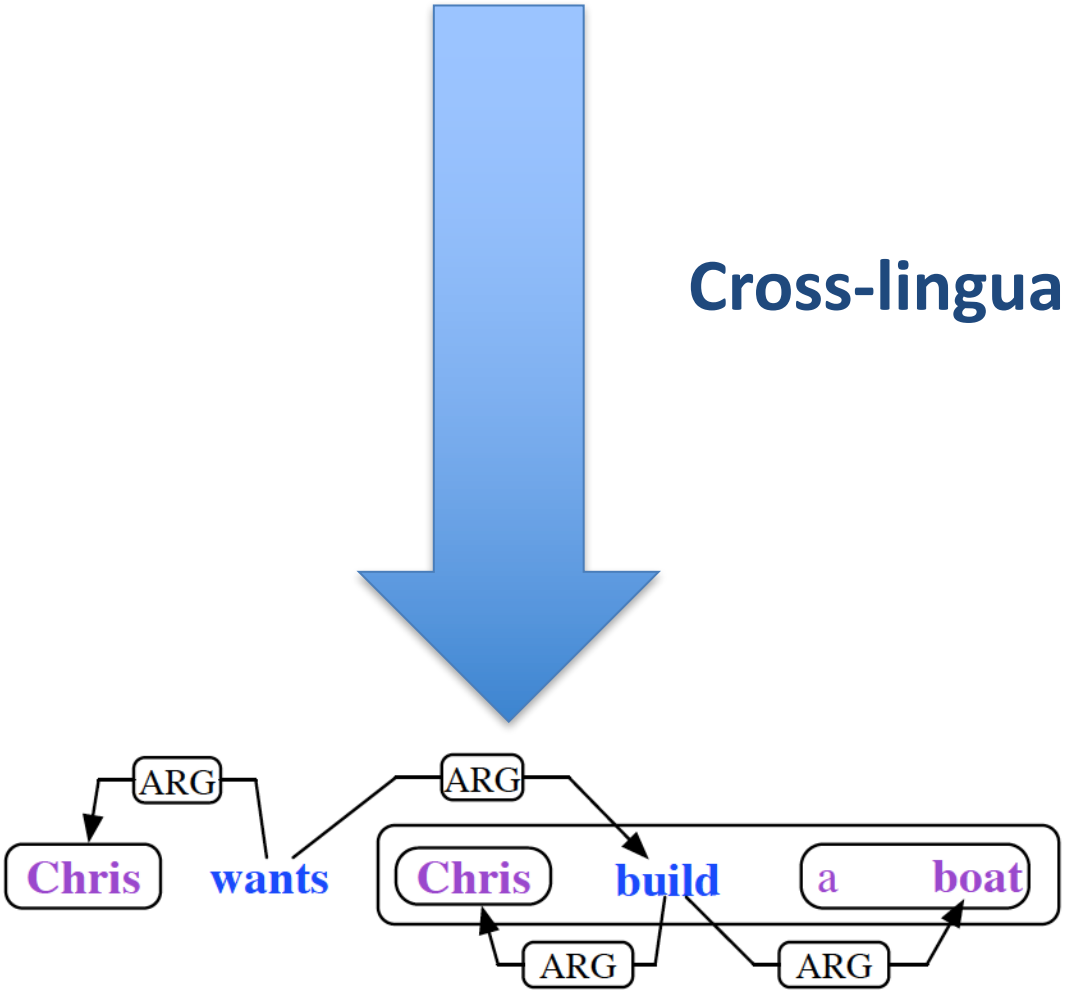


Dependency Parser + English Open IE



JOINT SOLUTION

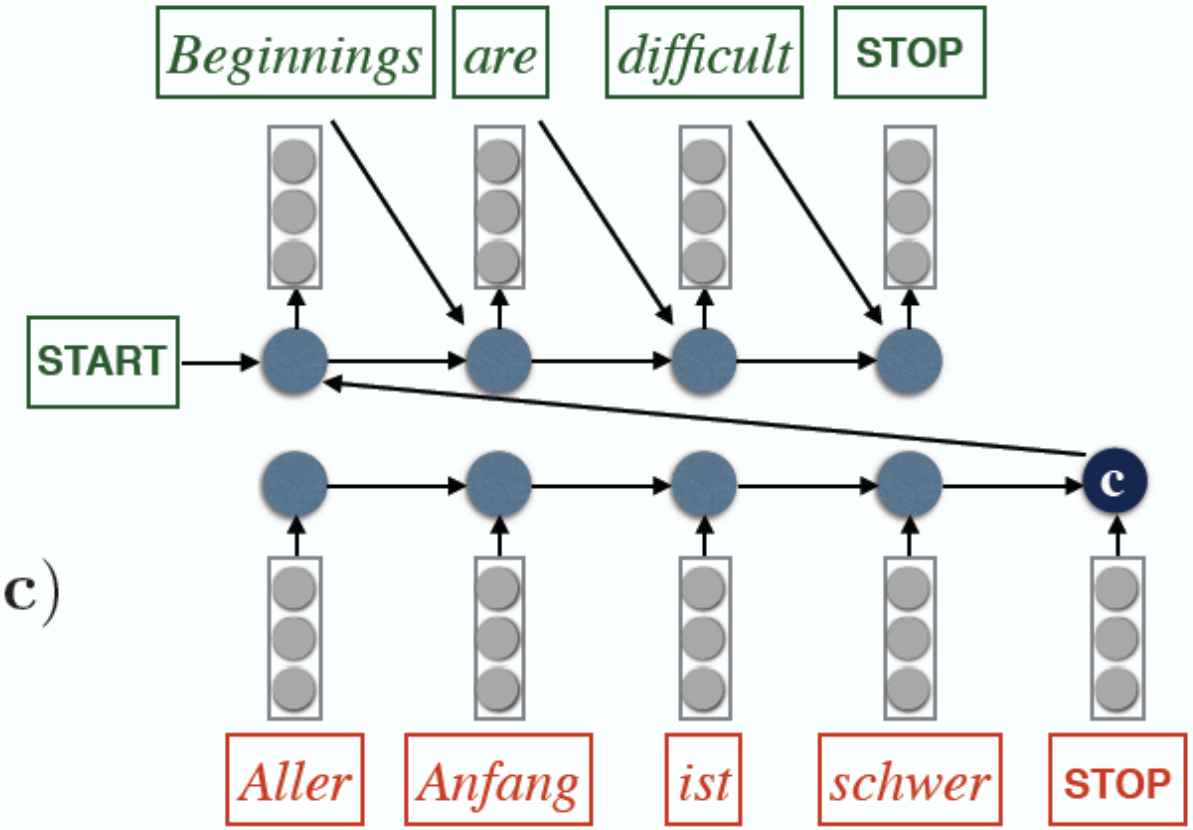
克里斯 想 造 一 艘 船 。



Cross-lingual Open IE

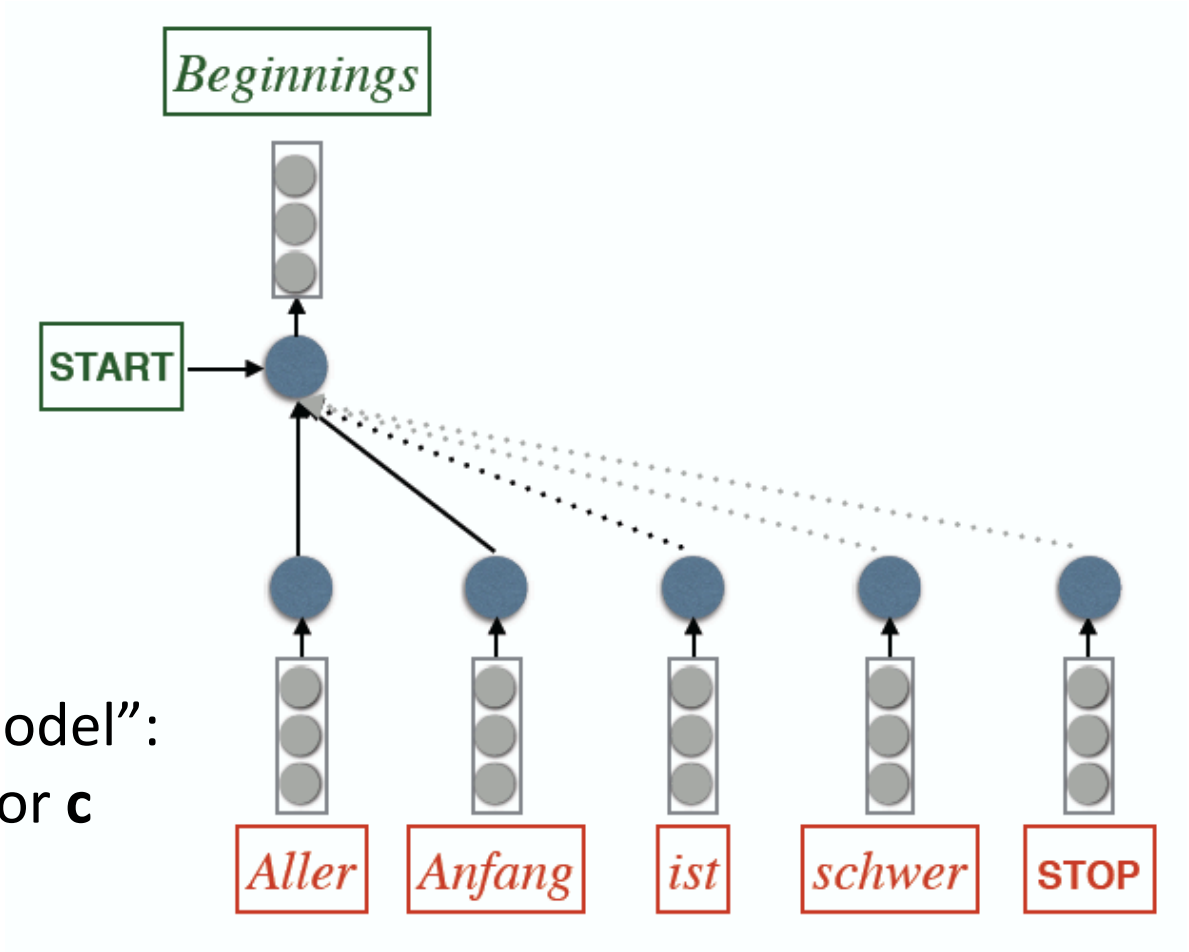
Neural Sequence-to-Sequence Model

$$\mathbf{c} = \mathrm{RNN}(\boldsymbol{x})$$
$$\boldsymbol{y} \mid \mathbf{c} \sim \mathrm{RNNLM}(\mathbf{c})$$



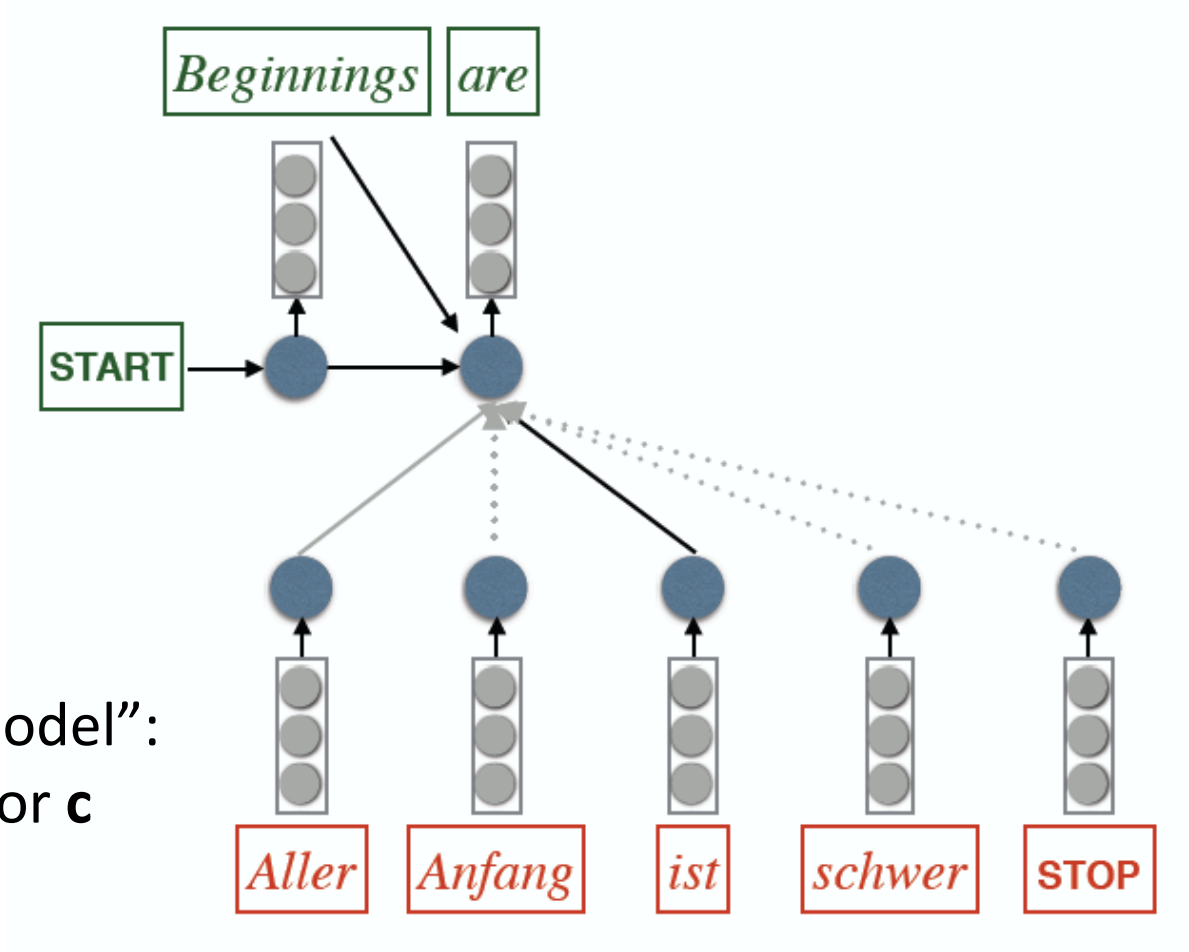
Cho (2014); Sutskever (2014); Bahdanau (2015)

Neural Sequence-to-Sequence Model



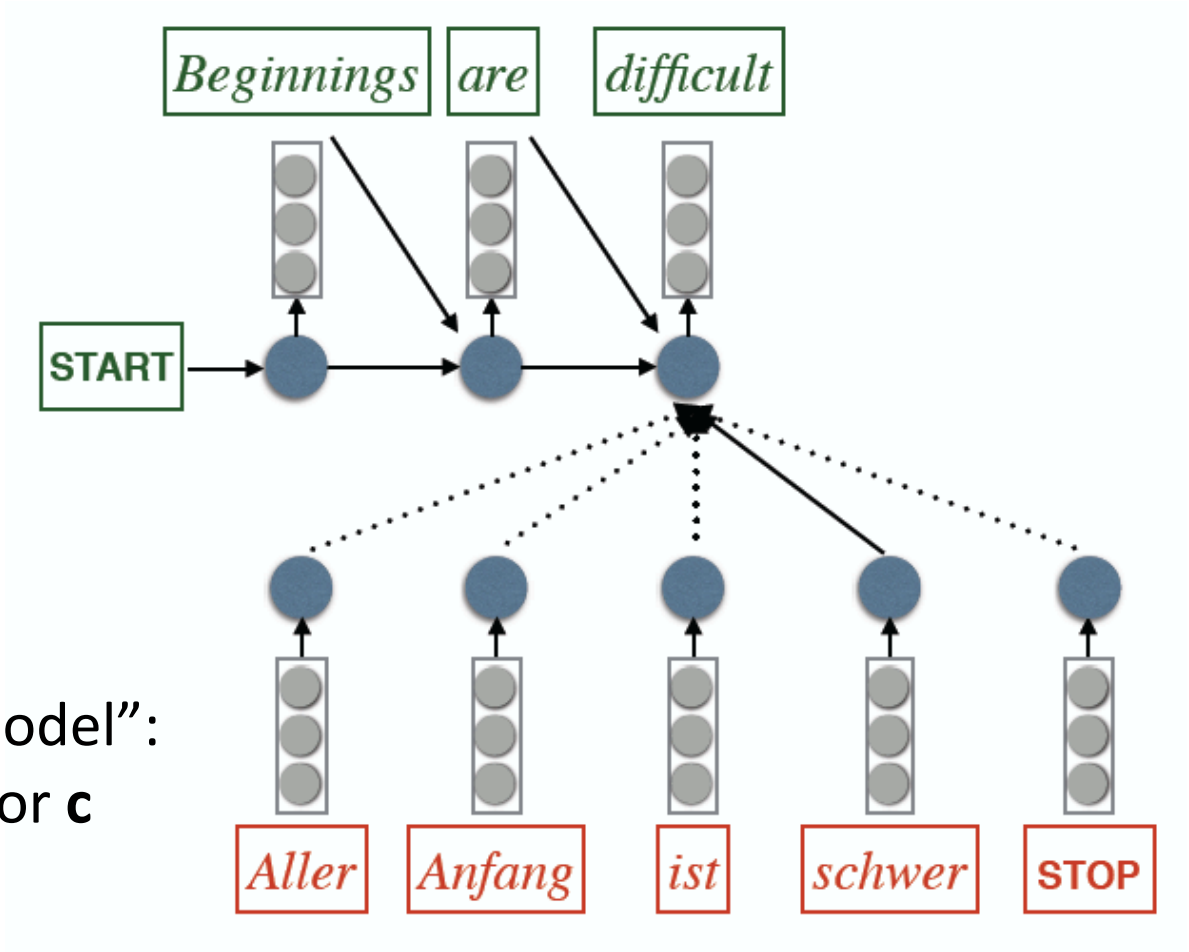
"Attention model":
Context vector **c**
is dynamic

Neural Sequence-to-Sequence Model



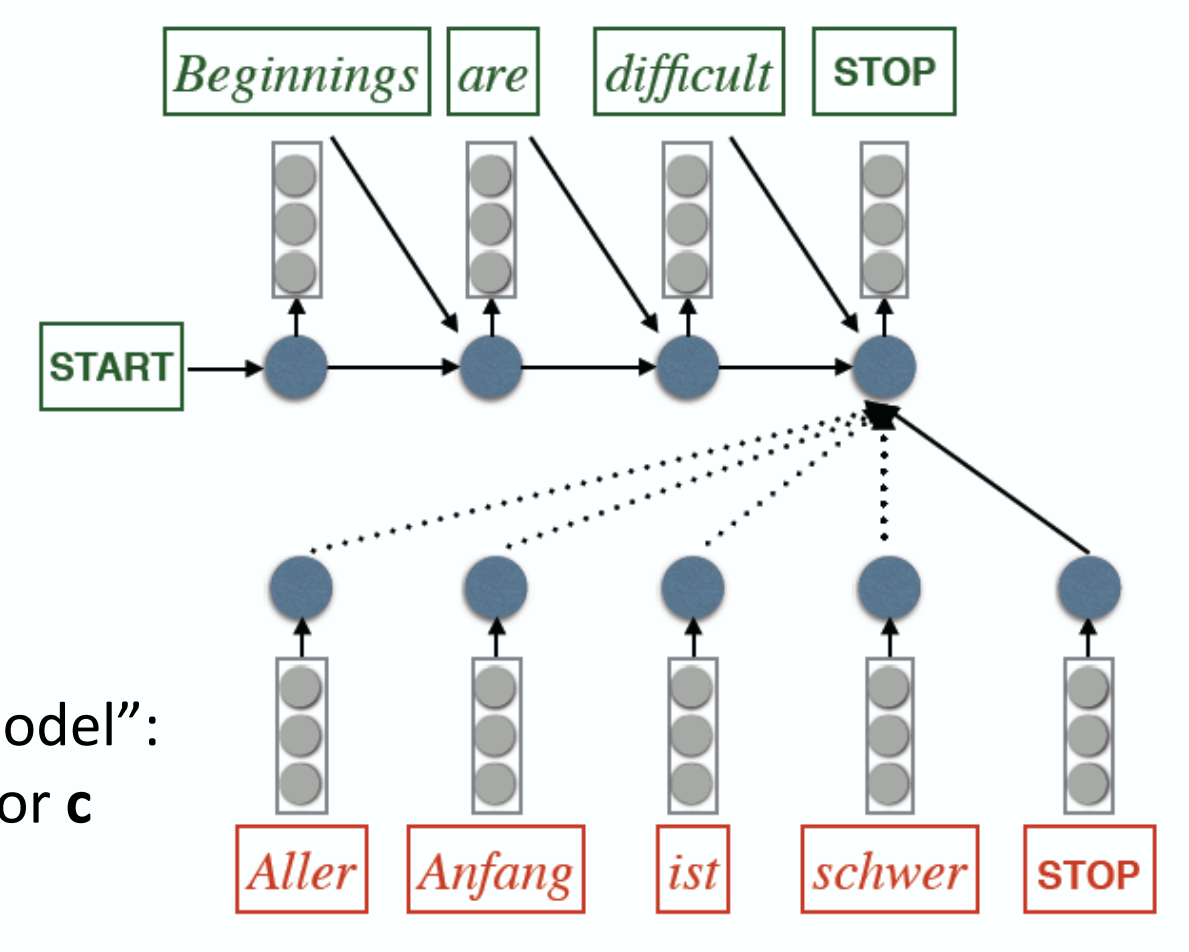
"Attention model":
Context vector **c**
is dynamic

Neural Sequence-to-Sequence Model



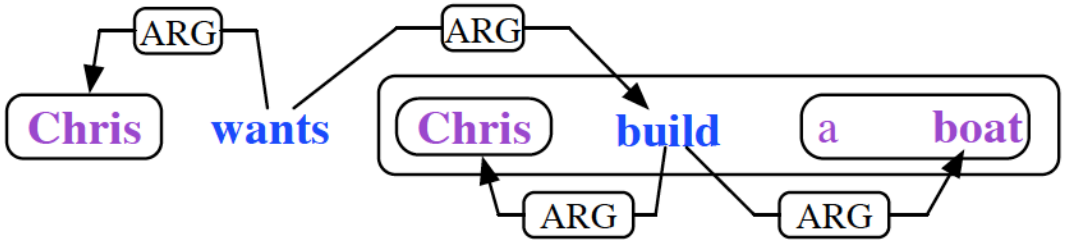
"Attention model":
Context vector **c**
is dynamic

Neural Sequence-to-Sequence Model

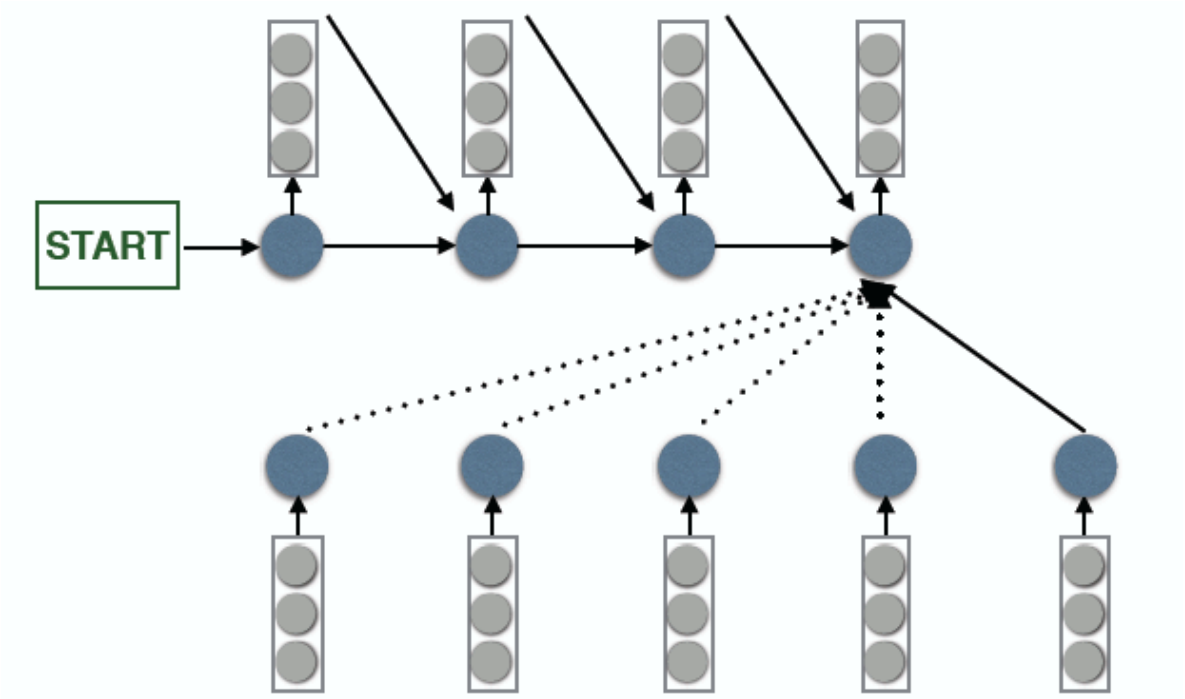


"Attention model":
Context vector **c**
is dynamic

Linearized OpenIE output as target



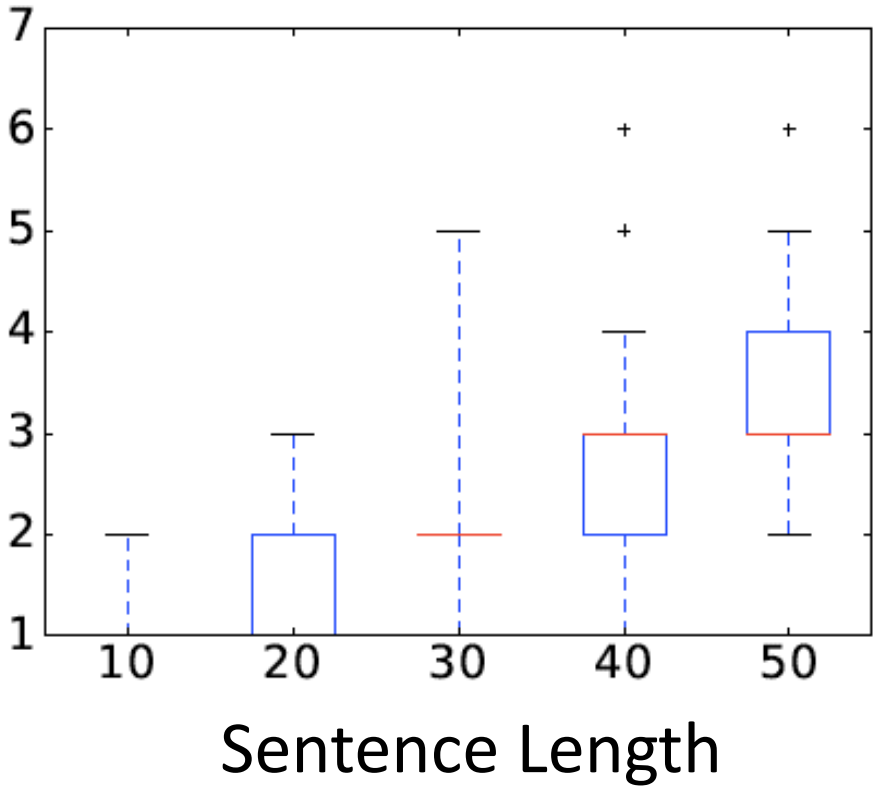
[(Chris:a_h) wants:p_h [(Chris:a_h) build:p_h (a:a boat:a_h)]]



Experiment Setting

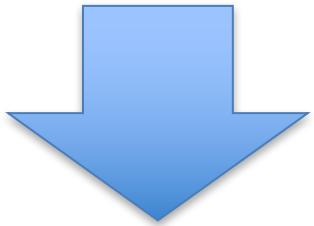
- 1 million sentence Chinese-English bitext (GALE project; mixed domain)

Predicates in OpenIE output



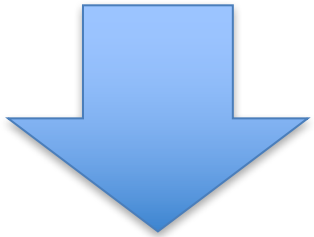
PIPELINE: BLEU=17.2 / PredicateF1=24.2

克里斯想造一艘船。

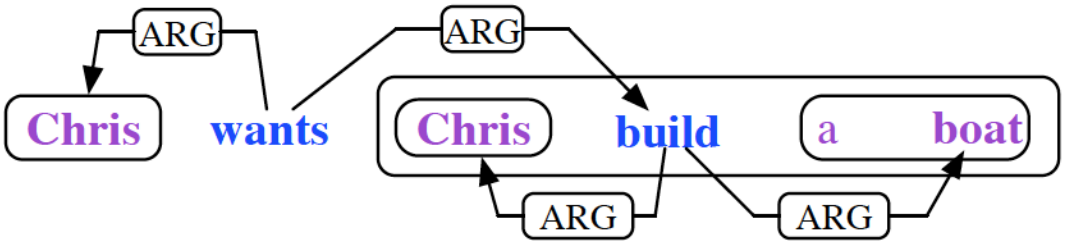


Machine Translation (Moses)

Chris wants to build a boat



Dependency Parser (Parsey) +
English Open IE (PredPatt)

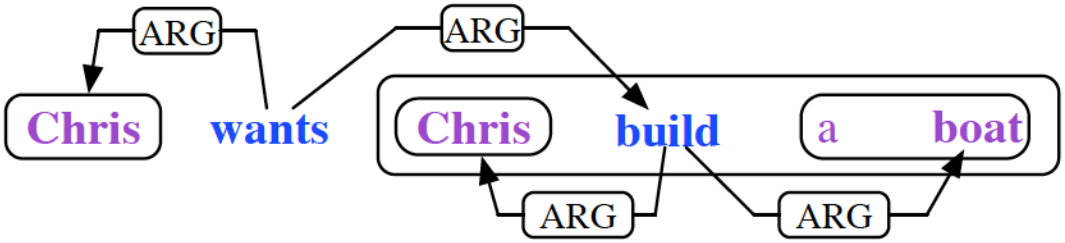


PIPELINE: BLEU=17.2 / PredicateF1=24.2

JOINT w/ Moses: BLEU=18.3 / PredicateF1=25.1

克里斯 想 造 一 艘 船 。

Phrase-based Machine Translation (Moses)



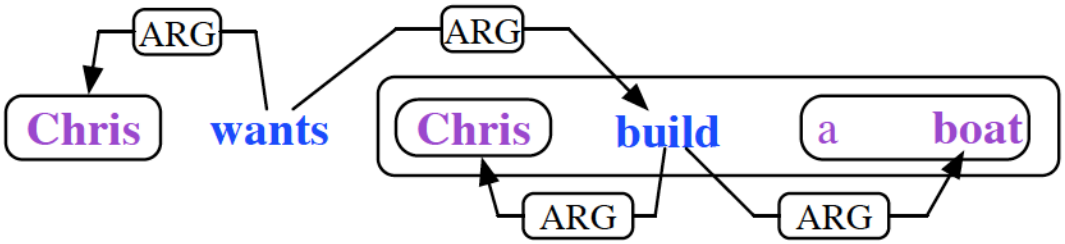
PIPELINE: BLEU=17.2 / PredicateF1=24.2

JOINT w/ Moses: BLEU=18.3 / PredicateF1=25.1

JOINT w/ Neural: BLEU=18.9 / PredicateF1=25.8

克里斯想造一艘船。

Neural Sequence-to-Sequence Model



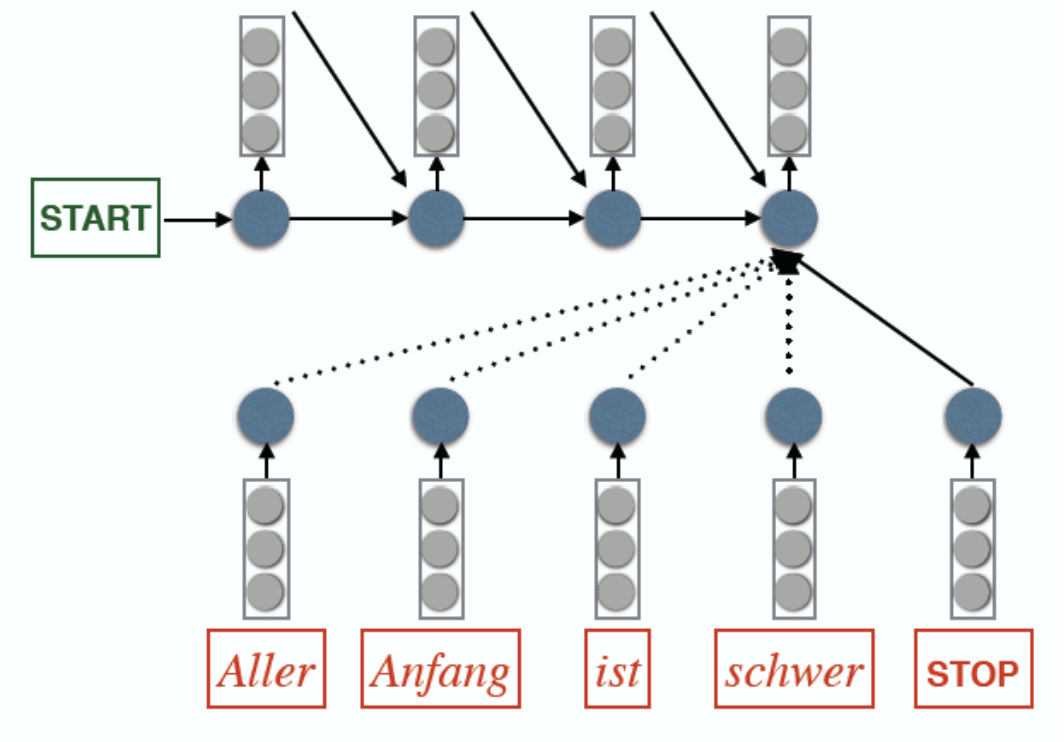
Outline

1. Motivation
2. Problem Definition
3. Pipeline vs. Joint Solution
4. Improved Joint Solution

Sequence generation vs. labeling

- Previously, treat word:label as single token

[(Chris:a_h) wants:p_h [(Chris:a_h) build:p_h (a:a boat:a_h)]]



Decompose generation and labeling

target words source words previous target words

$$P(Y, T \mid X) = \prod_{i=1}^{|Y|} P(y_i, t_i \mid y_{<i}, t_{<i}, X)$$

target labels previous labels

$$= \prod_{i=1}^{|Y|} P(y_i \mid y_{<i}, t_{\leq i}, X) P(t_i \mid y_{<i}, t_{<i}, X)$$

Decoder depends on t_i Predict label t_i

Decompose generation and labeling

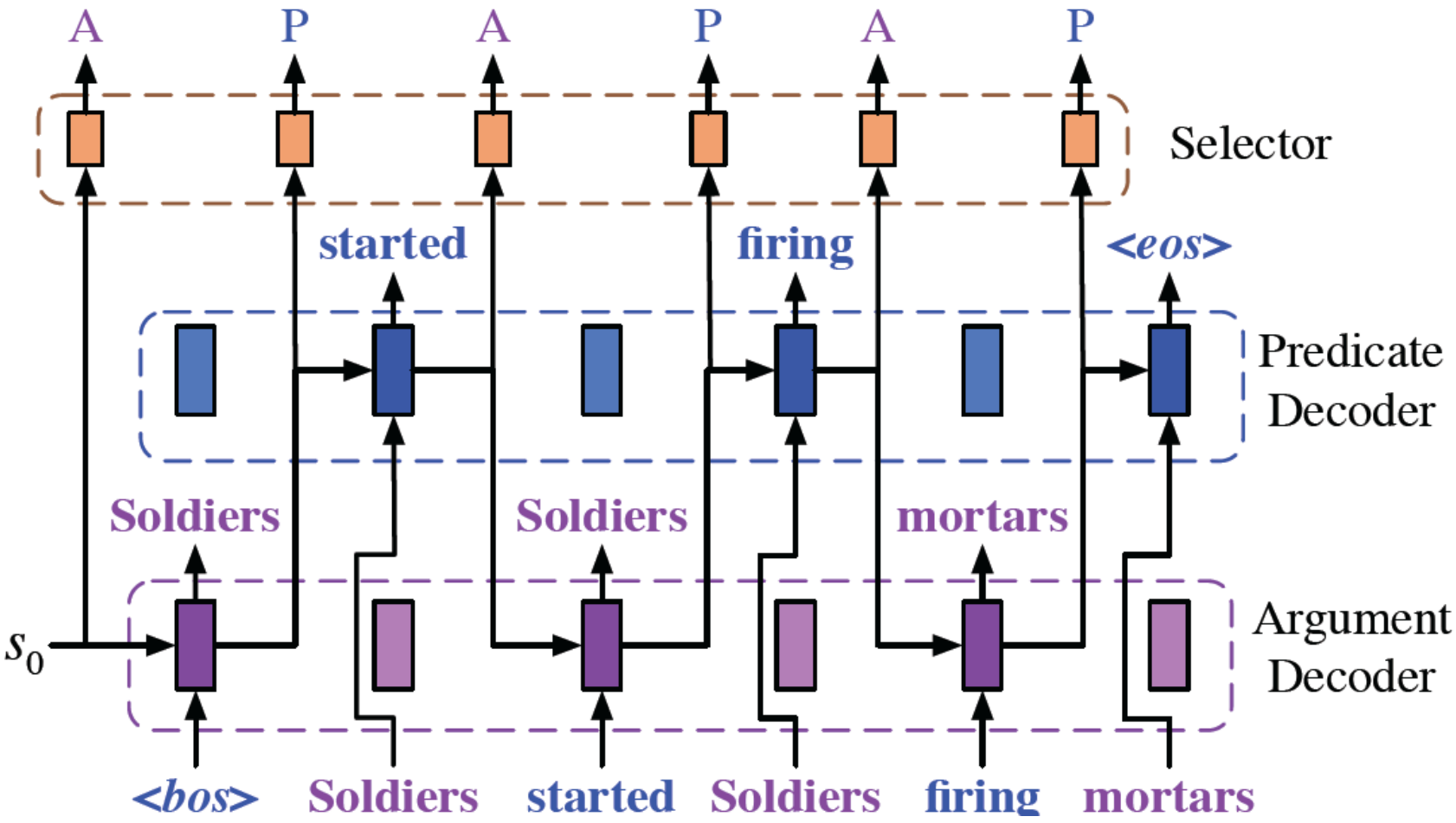
$$= \prod_{i=1}^{|Y|} P(y_i \mid y_{<i}, t_{\leq i}, X) P(t_i \mid y_{<i}, t_{<i}, X)$$

Decoder depends on t_i

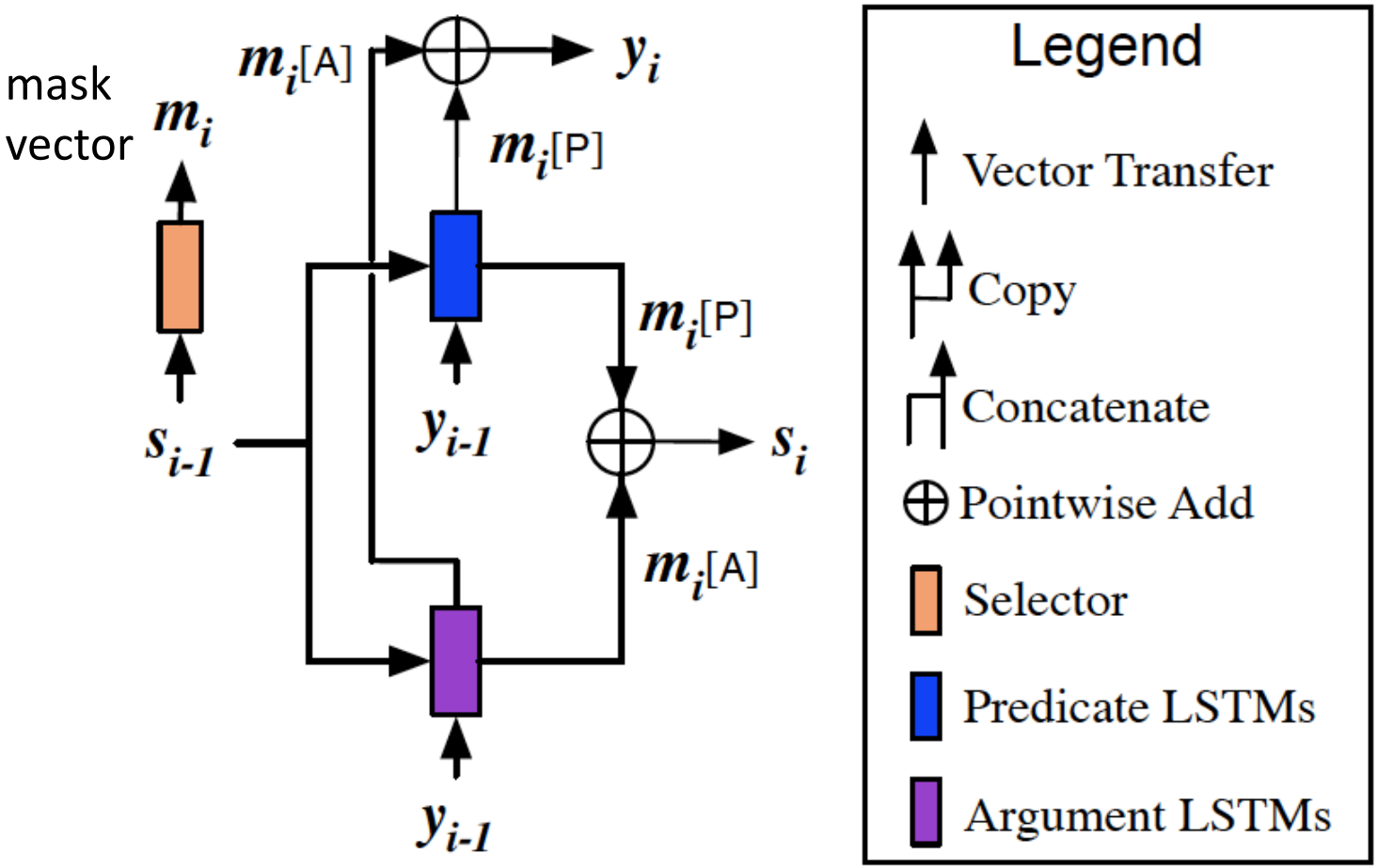
Predict label t_i

- Limits increase of target vocabulary
- Models generation process separately by type
  - Given previous word "wanted":
  - predicate decoder generates "to", "by"
  - argument decoder generates "a", "him"

Selective Decoding



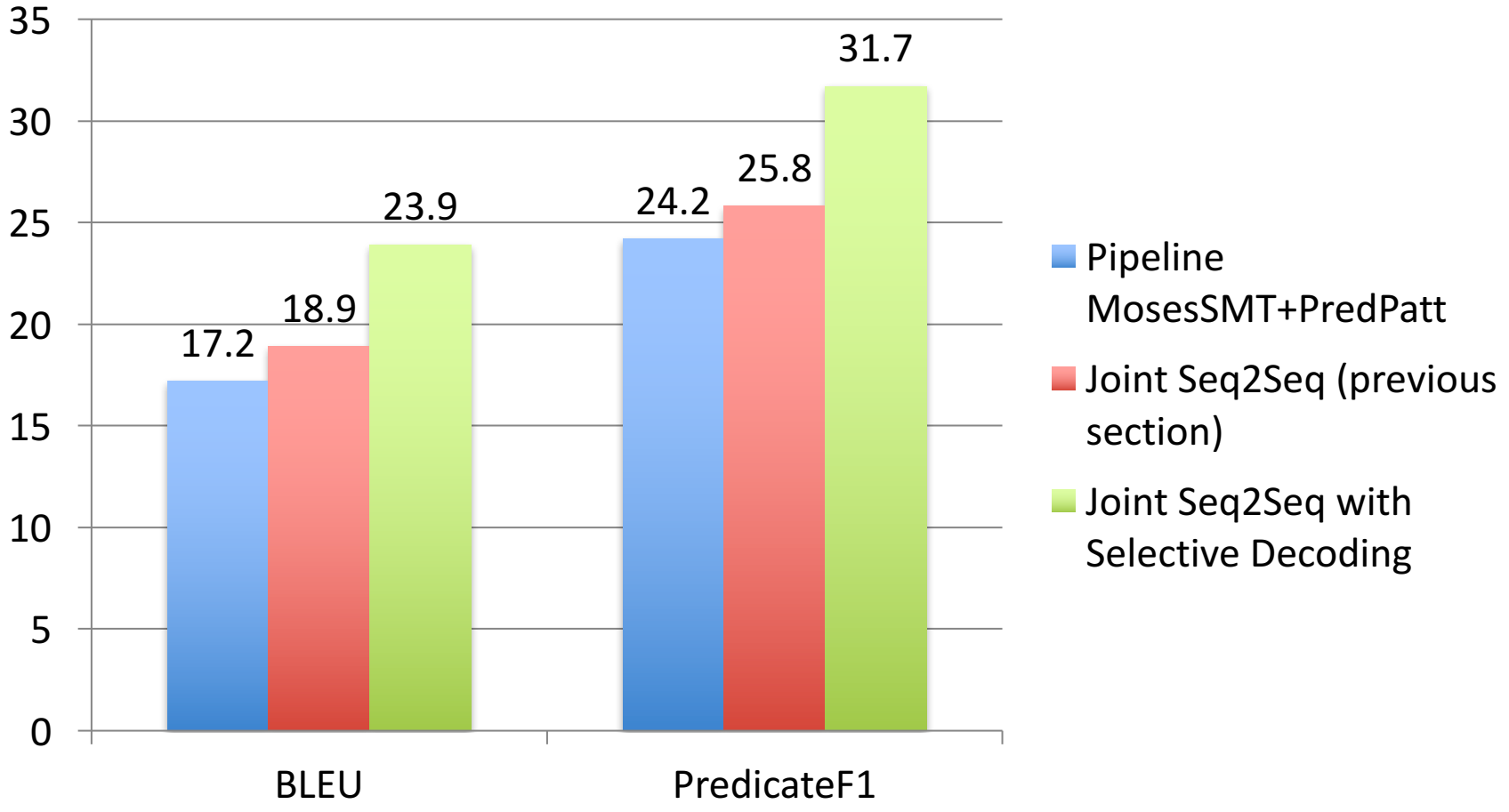
In detail: at each decoding step



Attention context (c_i) is omitted in the figure

$$s_i = \sum_{t_i \in \mathcal{T}} m_i[t_i] f_{t_i}(y_{i-1}, s_{i-1}, c_i)$$

Results on Chinese-English task



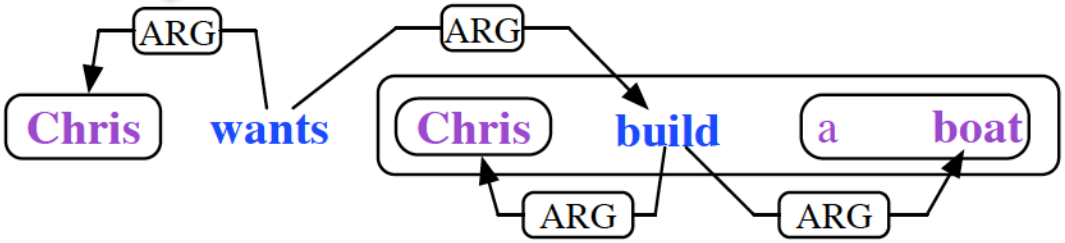
If we only care about MT (not MT+IE)

克里斯 想 造 一 艘 船 。

Standard seq2seq
BLEU = 24.92

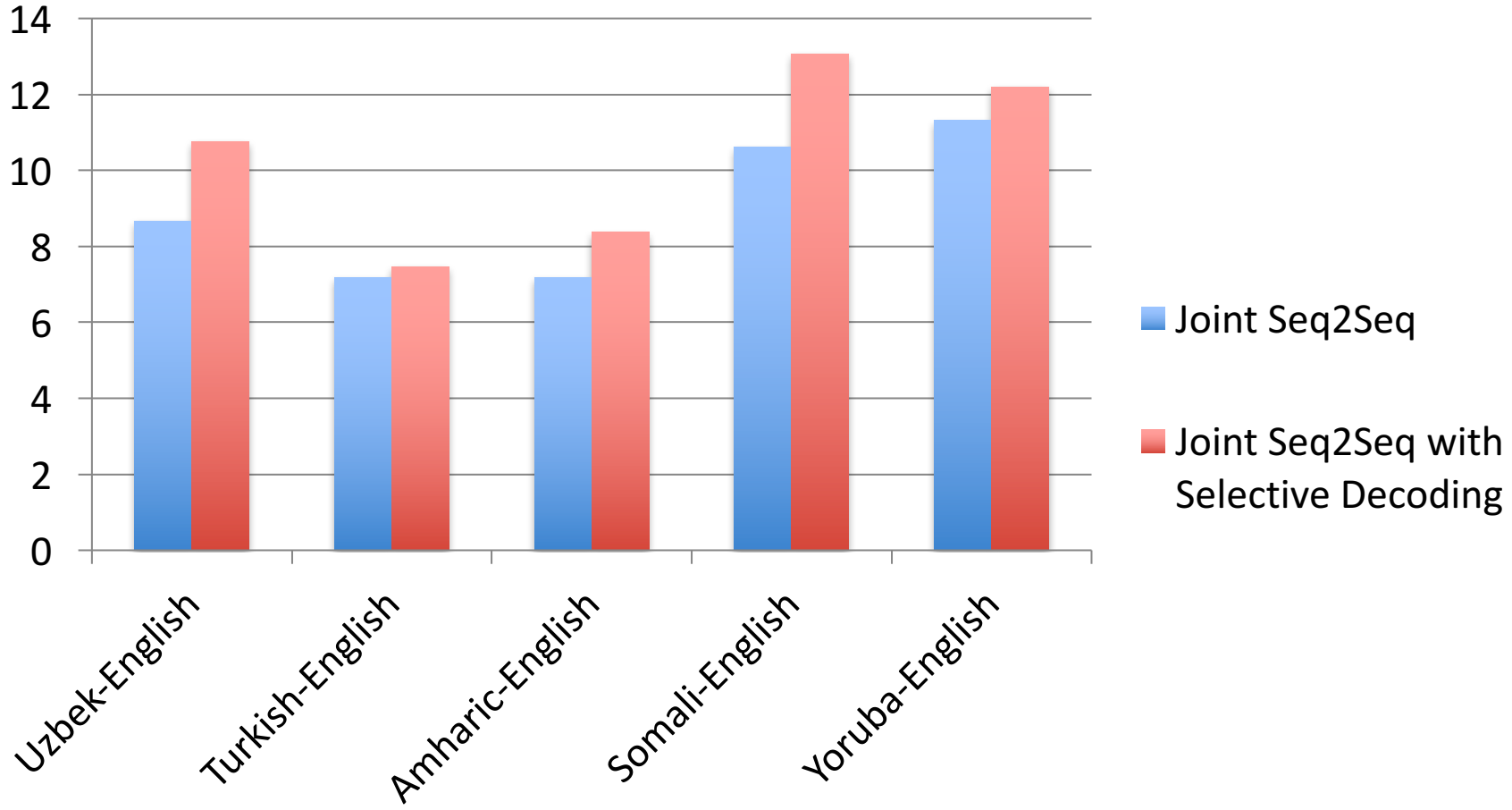
Chris wants to build a boat

Seq2Seq with selective decoding, then ignore labels
BLEU = 25.16



BLEU on Low-Resource Languages

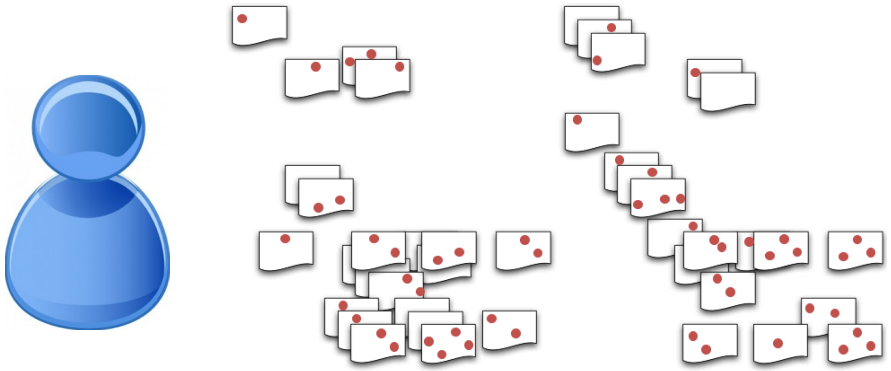
(Data from DARPA LORELEI Project)

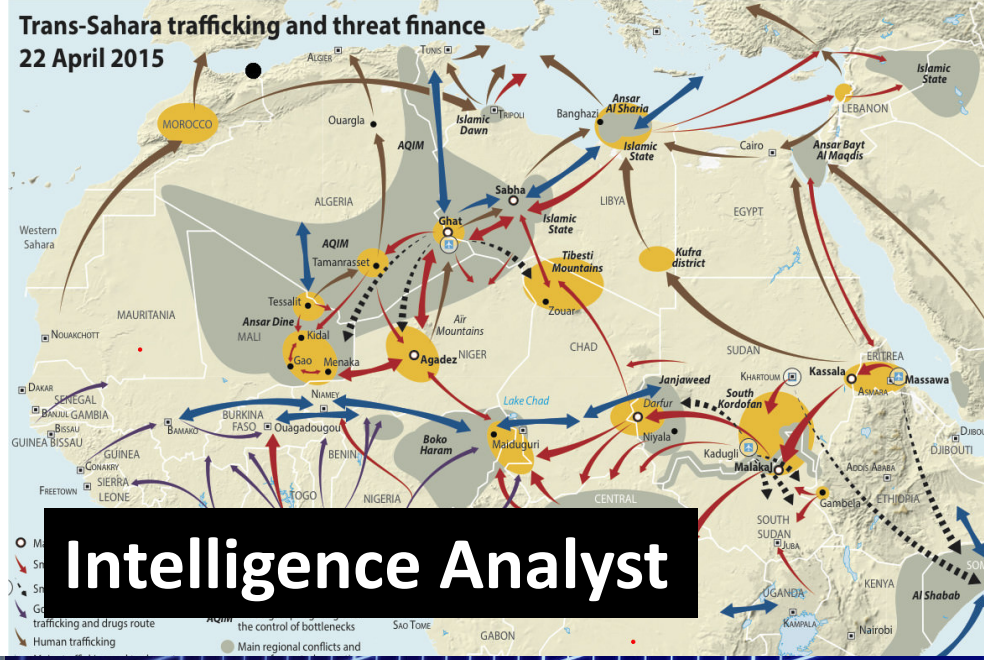


NOT SHOWN: The winner is less clear for Predicate F1 (8-14%)

Summary

Support users with complex information needs

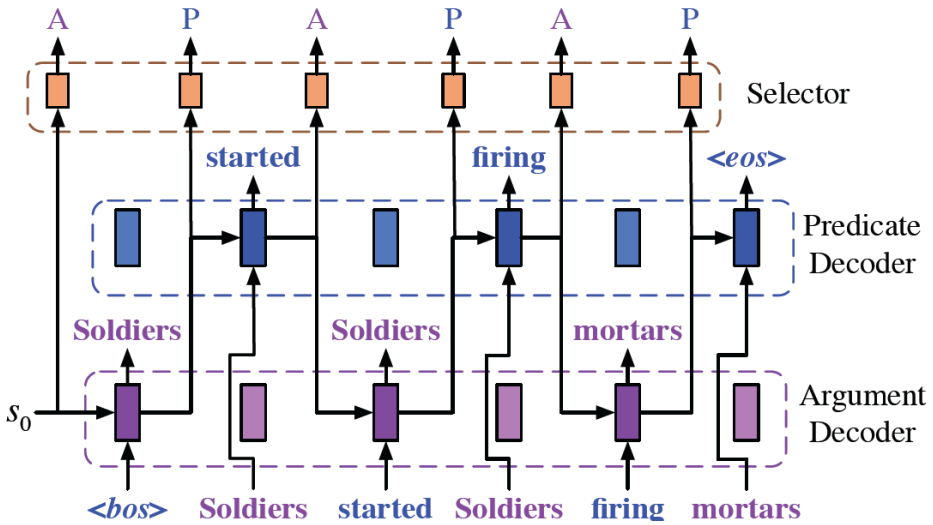




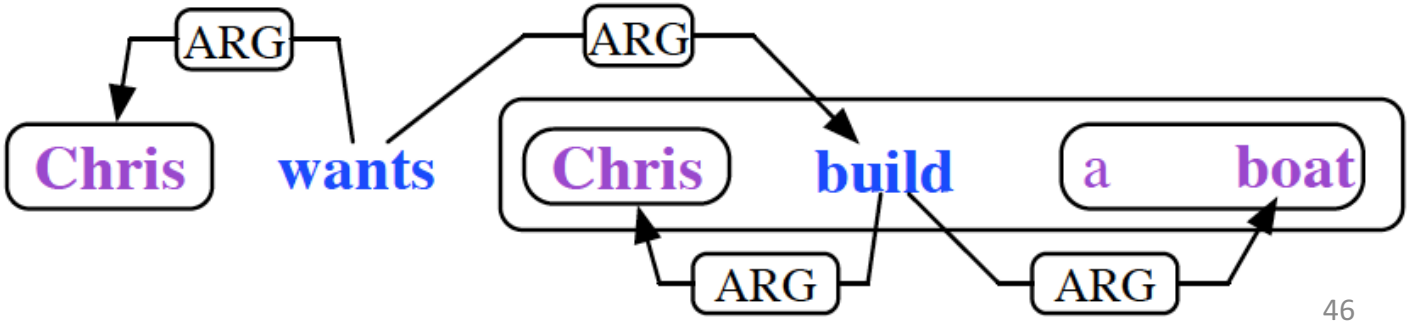
Input: Chinese sentence

克里斯 想 造 一 艘 船 。

Cross-lingual Open IE



Output: English tuples, e.g. Relation(arg1,arg2)



Next Steps

- Integration with analyst search engine
- Directly optimize IE objective, not likelihood
- Explore Selective Decoding for other problems

Thanks!

- To Learn More:
  - S. Zhang, K. Duh, B. Van Durme. “MT/IE: Cross-lingual Open Information Extraction” (EACL2017)
  - S. Zhang, K. Duh, B. Van Durme. “Selective Decoding for Cross-lingual Open Information Extraction” (IJCNLP2017)
  - Code on GitHub