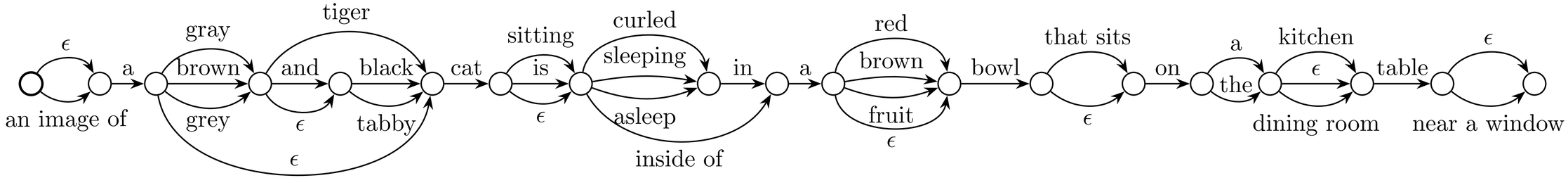
Multi-reference Training with Pseudo-References for Neural Translation and Text Generation

Renjie Zheng [1] Mingbo Ma [1,2] Liang Huang [1,2]

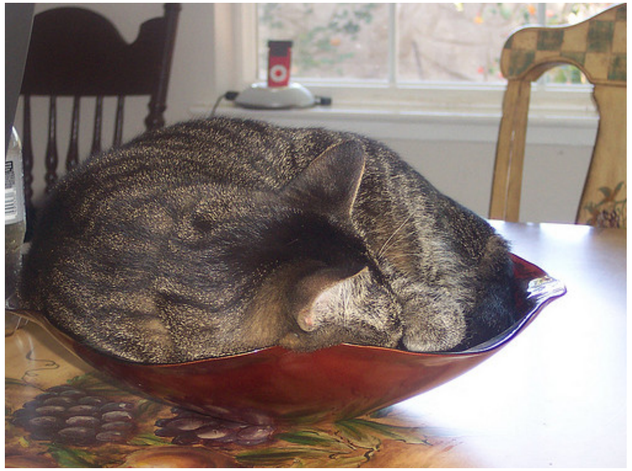
1 Oregon State University

2 Baidu Research



Multi-reference Data in Image Captioning

- MSCOCO Image Captioning Dataset
  - 114,287 images, each with 5 reference captions
- PASCAL-50S
  - 1,000 images, each with 50 references
- example from MSCOCO
  - *1:* A brown and black cat is sleeping in a bowl on a table
  - *2:* A grey tiger cat sleeping in a brown bowl on a table
  - *3:* An image of a cat sitting inside of a bowl on the kitchen table
  - *4:* A cat asleep in a fruit bowl on a dining room table
  - *5:* A gray tabby cat is curled in a red bowl that sits on a table near a window



Multi-reference Data in MT

- NIST Chinese-to-English and Arabic-to-English Evaluations (2002-2008)
  - ~10,000 Chinese (Arabic) sentences, each with 4 reference English translations
- example from NIST02

Yìnní	*chóngshēn*	*fǎndùi*	*wàigúo*	*jūndùi*	*jìnzhù*
印尼	重申	反对	外国	军队	进驻
Indonesia	reiterate	opposition	foreign	army	garrison

  - *1*. Indonesia reiterated its opposition to foreign military presence
  - *2*. Indonesia repeats its opposition against station of foreign troops in Indonesia
  - *3*. Indonesia reiterates opposition to garrisoning foreign armies
  - *4*. Indonesia reiterates : no foreign troops in country

Can we use multiple references in training?

- multiple references are always used in evaluations
- previous work under-utilizes available multiple references in training
  - *image captioning*: existing literatures simply sample one of them in each epoch
  - *machine translation:* no previous work using extra references during training

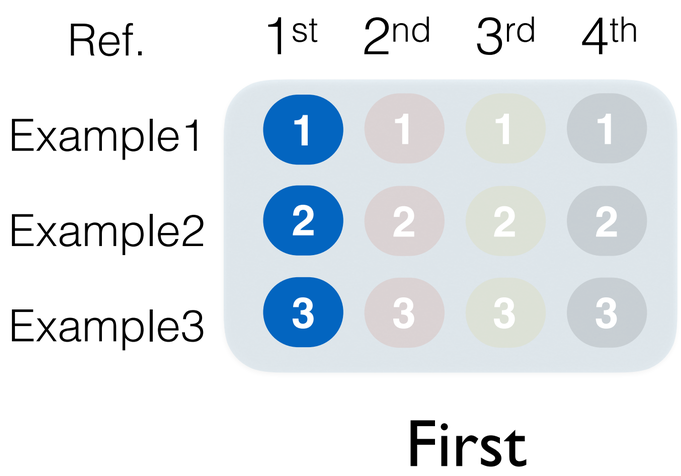
Can we use multiple references in training?

- multiple references are always used in evaluations
- previous work under-utilizes available multiple references in training
  - *image captioning*: existing literatures simply sample one of them in each epoch
  - *machine translation:* no previous work using extra references during training
  - *Q:* can we benefit more from training with multiple references?

Multi-reference Training

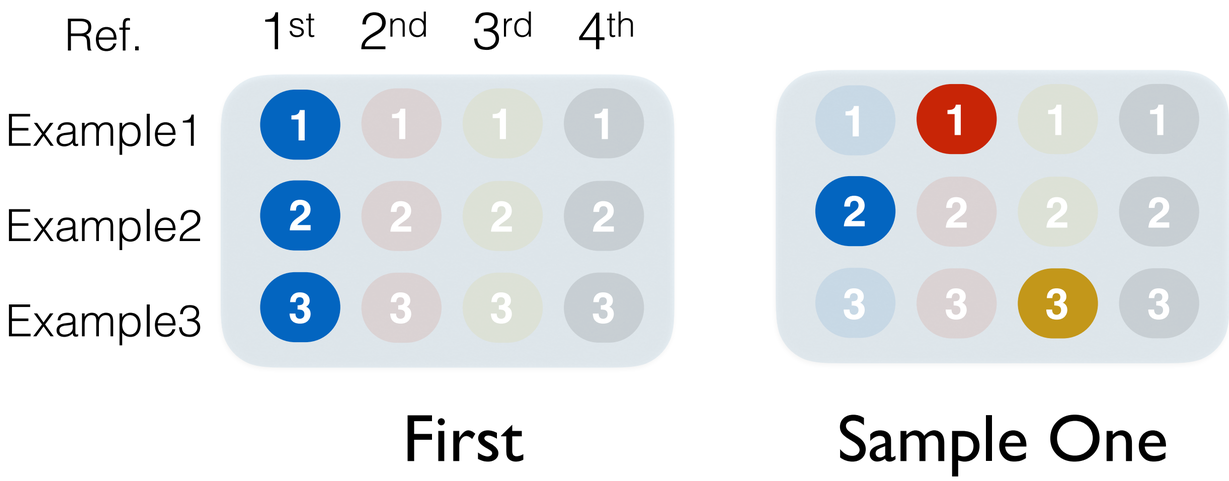
Multi-reference Training

- *First:* only uses first reference



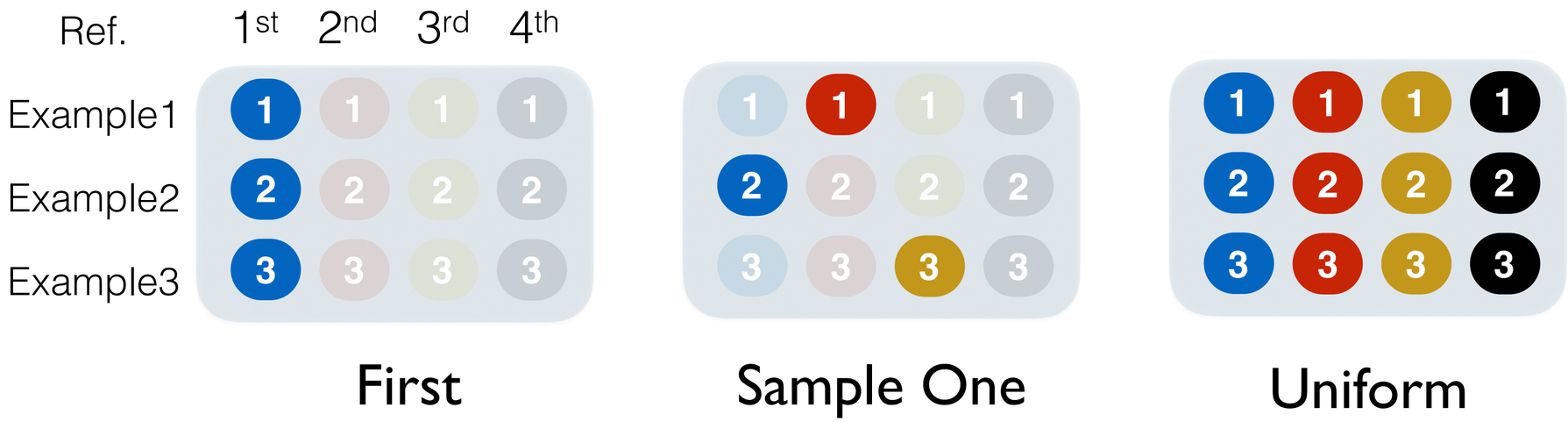
Multi-reference Training

- *First:* only uses first reference
- *Sample One:* samples one reference for each example in each epoch



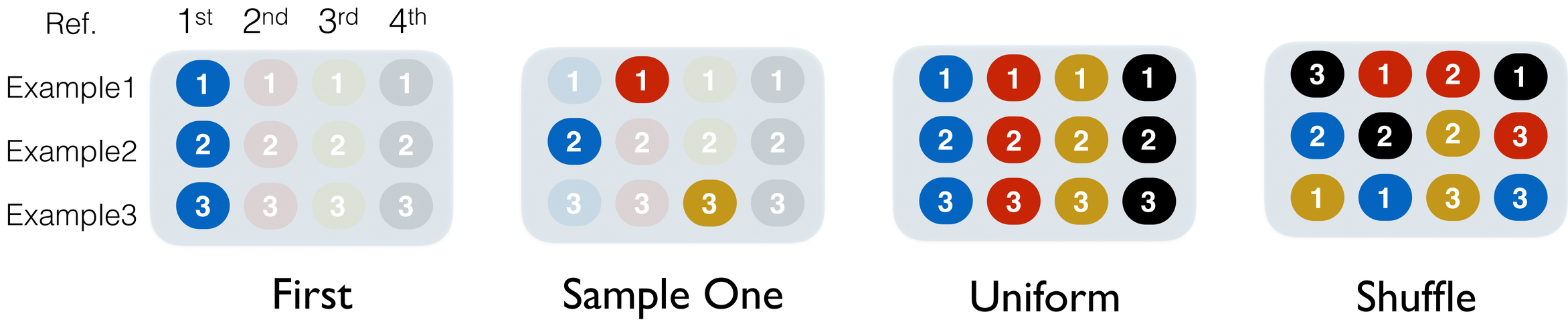
Multi-reference Training

- *First:* only uses first reference
- *Sample One:* samples one reference for each example in each epoch
- *Uniform:* utilizes all the references of each example



Multi-reference Training

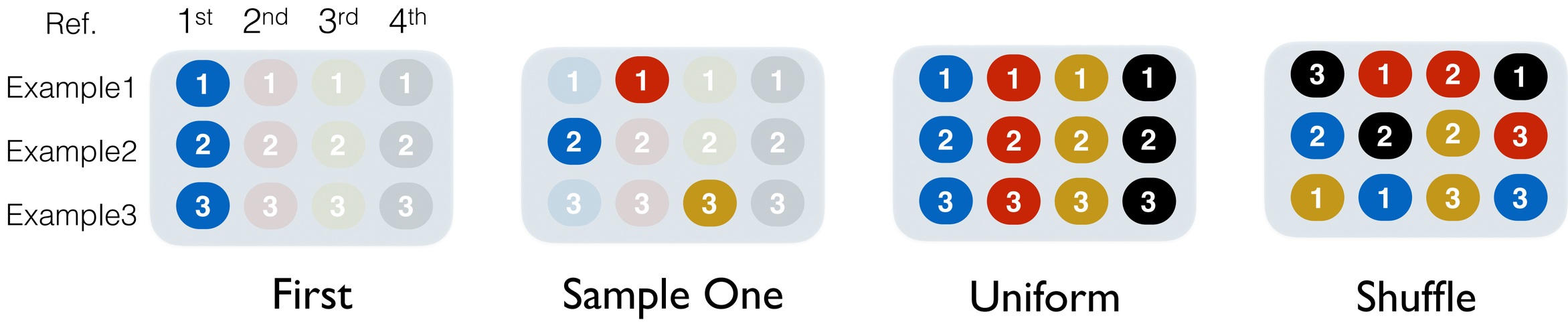
- *First:* only uses first reference
- *Sample One:* samples one reference for each example in each epoch
- *Uniform:* utilizes all the references of each example
- *Shuffle:* also utilizes all the references but shuffles all of them



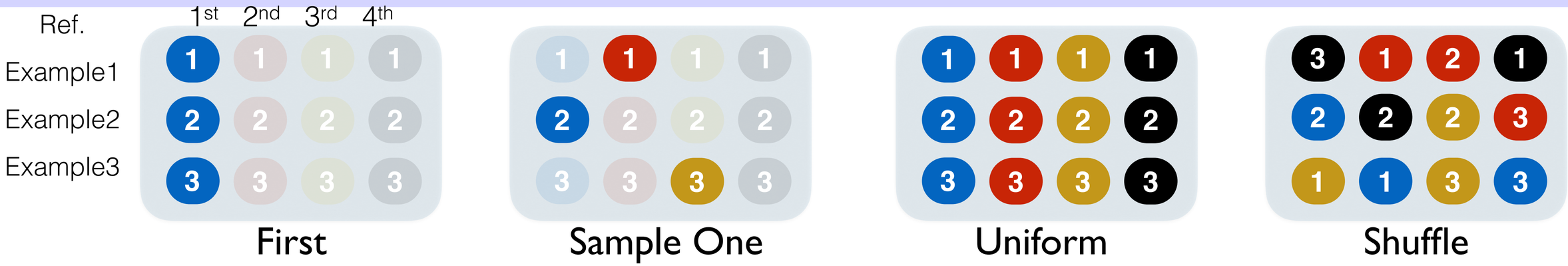
Multi-reference Training

- *First:* only uses first reference
- *Sample One:* samples one reference for each example in each epoch
- *Uniform:* utilizes all the references of each example
- *Shuffle:* also utilizes all the references but shuffles all of them

MT	Cap
New	Old
New	Old
New	New
New	New



Multi-reference Training Results



Machine Translation

# Refs pre-train	# Refs fine-tune	Method	BLEU
1		Pre-train	37.4
1	1	First	38.6
1	4	Sample One	38.8
1	4	Uniform	38.8
1	4	Shuffle	**38.9**

pre-train on 1M sentence pairs
fine-tune on 4,667 sents w/ 4 refs
dev results on 616 sents (NIST06)

Image Captioning

# Refs	Method	BLEU	CIDEr
1	First	26.3	79.1
5	Sample One	29.0	85.4
5	Uniform	30.1	89.8
5	Shuffle	**30.4**	**91.2**

train on MSCOCO 113,287 examples
dev results on 5,000 examples

But can we go beyond existing references?

- text generation potentially have exponentially many possible answers
  - 4 or 5 human references are only a tiny fraction of the whole space (Markus & Marcu 2012)

But can we go beyond existing references?

- text generation potentially have exponentially many possible answers
  - 4 or 5 human references are only a tiny fraction of the whole space (Markus & Marcu 2012)
  - *Q:* can we generate more pseudo-references based on the existing references?
  - *Q:* will these extra references improve generation quality?

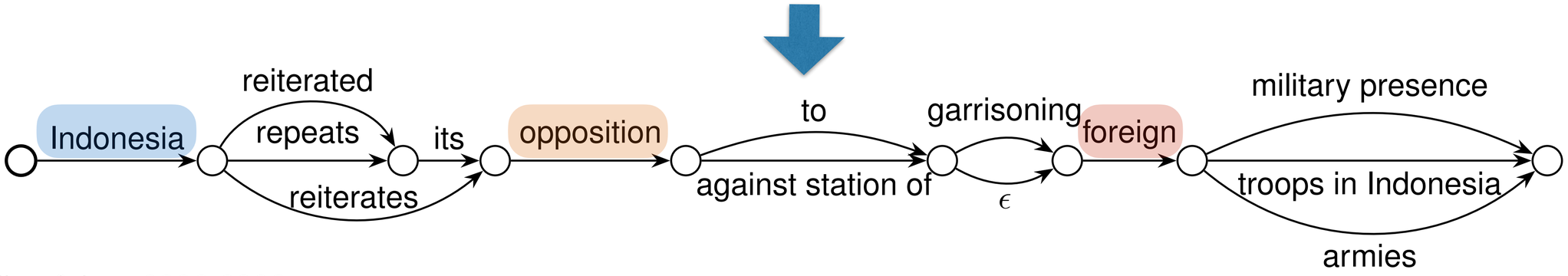
How to Generate more Pseudo-references?

- first idea: hard alignment
  - *step 1:* compress the existing references to a lattice by merging identical words
  - *step 2:* generate pseudo-references by traversing the lattice

Indonesia reiterated its opposition to foreign military presence

Indonesia repeats its opposition against station of foreign troops in Indonesia

Indonesia reiterates opposition to garrisoning foreign armies



Barzilay & Lee 2002;2003

Step 1: Compress References to Lattice

1. initialize

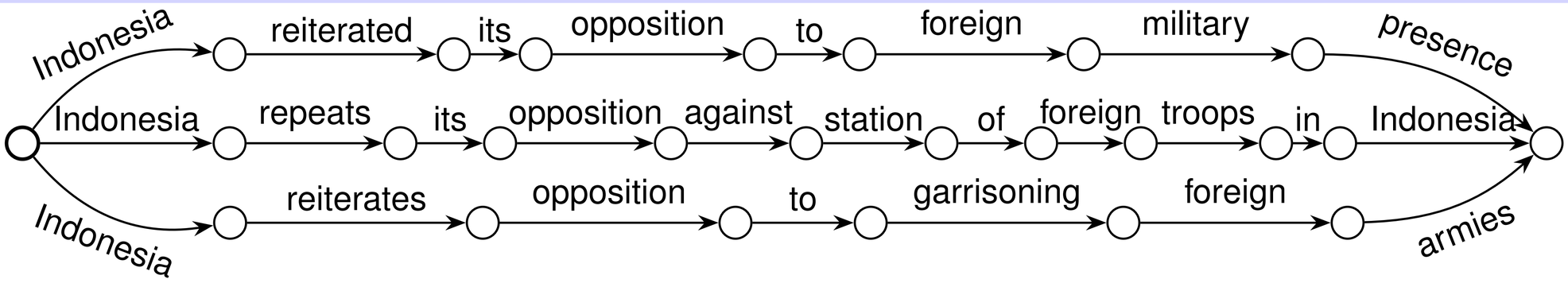
2.1 pick two sents

2.2 find identical words

2.3 merge

Step 1: Compress References to Lattice

(a)



1. initialize

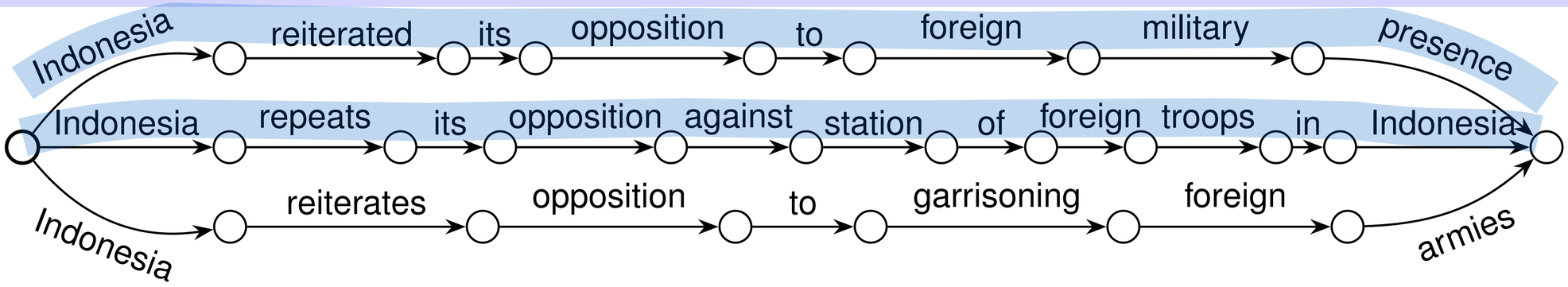
2.1 pick two sents

2.2 find identical words

2.3 merge

Step 1: Compress References to Lattice

(a)



1. initialize

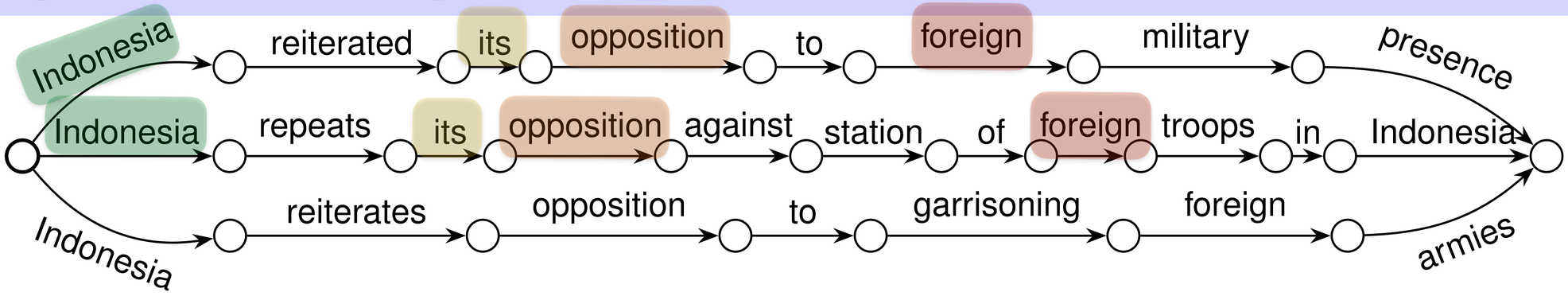
2.1 pick two sents

2.2 find identical words

2.3 merge

Step 1: Compress References to Lattice

(a)



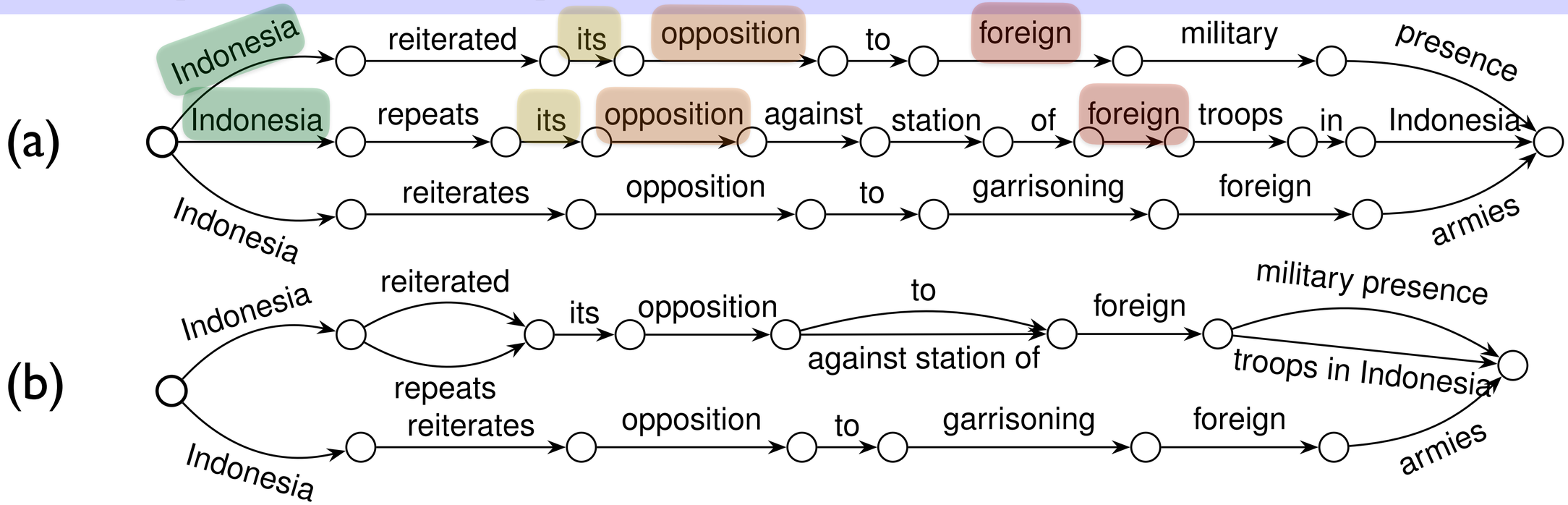
1. initialize

2.1 pick two sents

2.2 find identical words

2.3 merge

Step 1: Compress References to Lattice



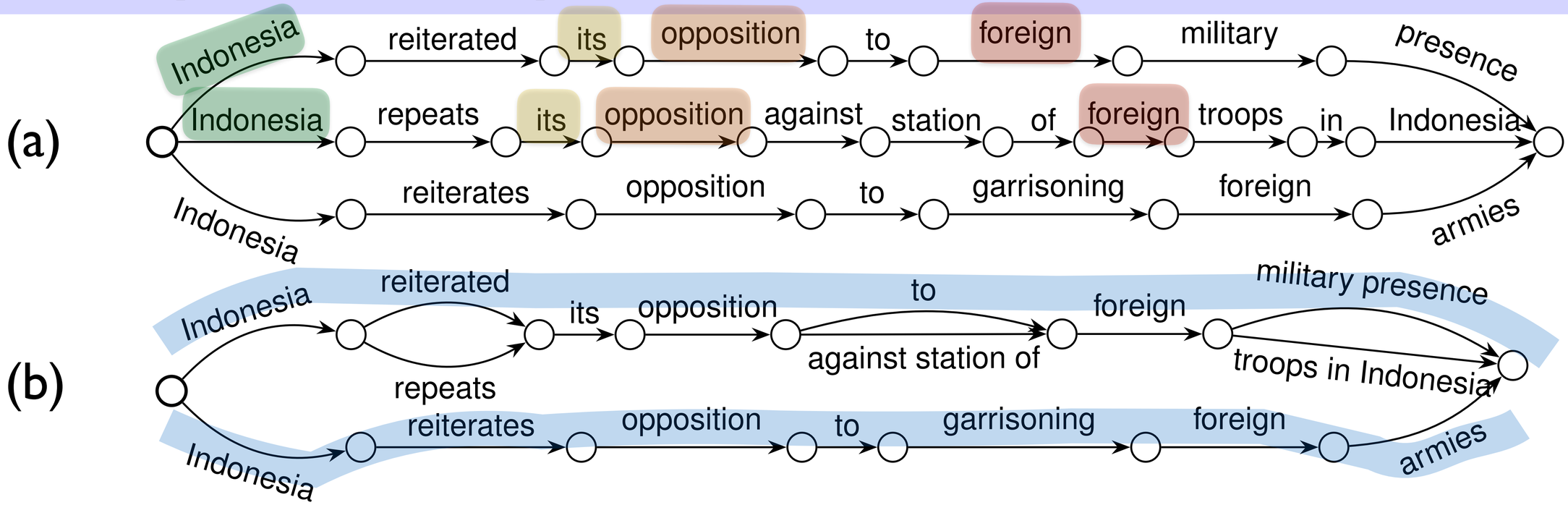
1. initialize

2.1 pick two sents

2.2 find identical words

2.3 merge

Step 1: Compress References to Lattice



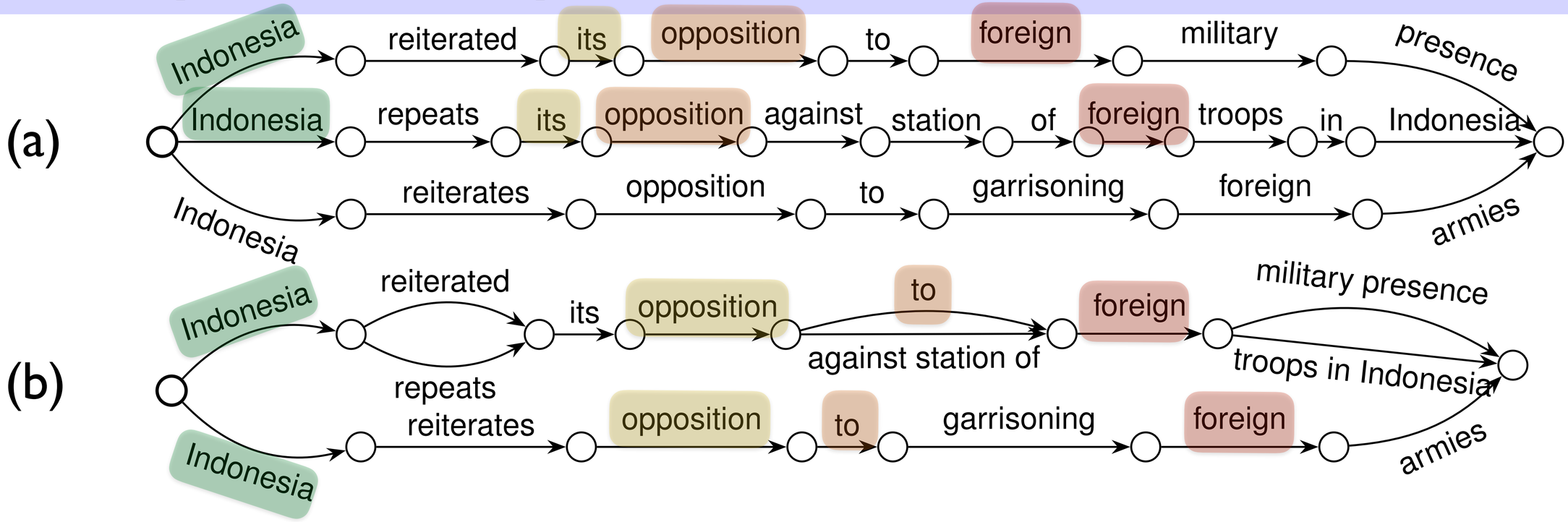
1. initialize

2.1 pick two sents

2.2 find identical words

2.3 merge

Step 1: Compress References to Lattice



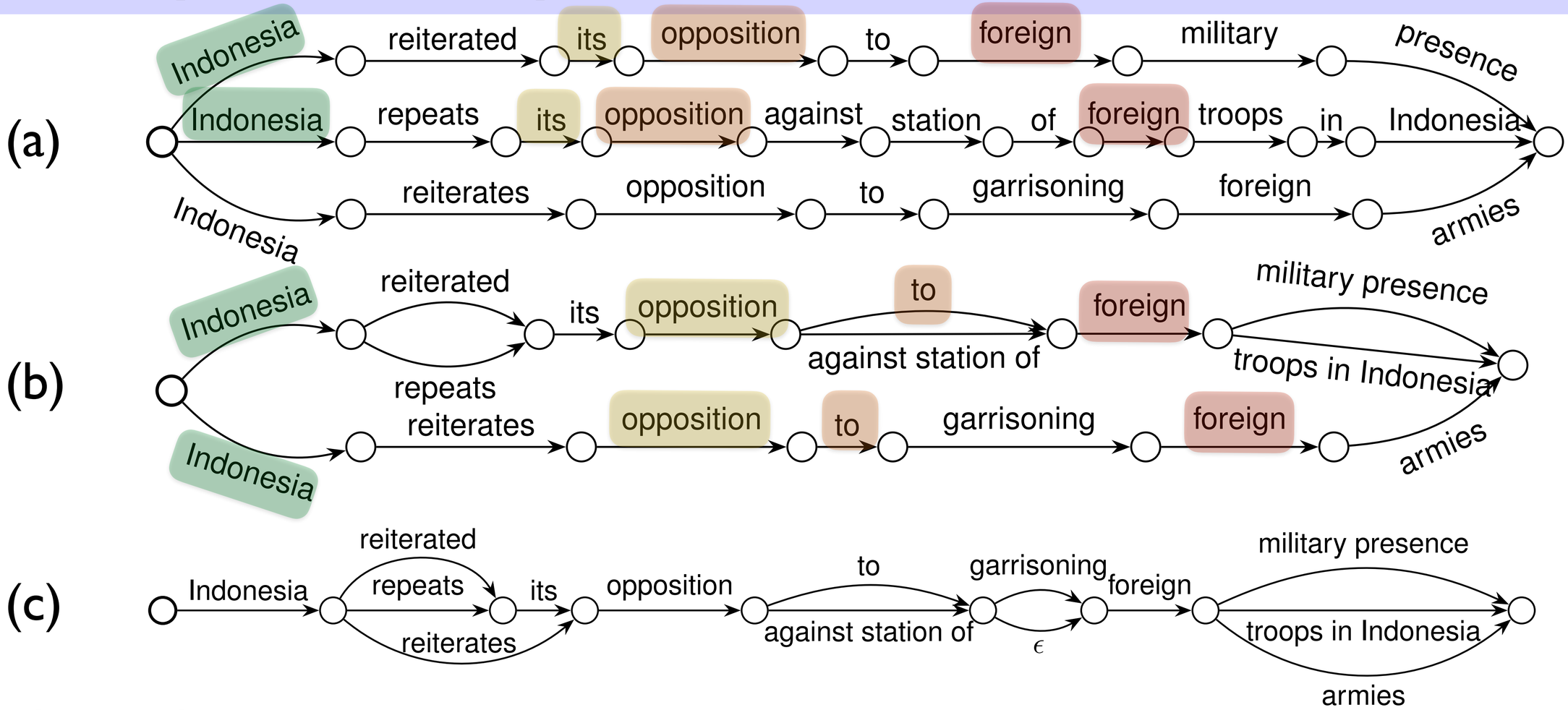
1. initialize

2.1 pick two sents

2.2 find identical words

2.3 merge

Step 1: Compress References to Lattice



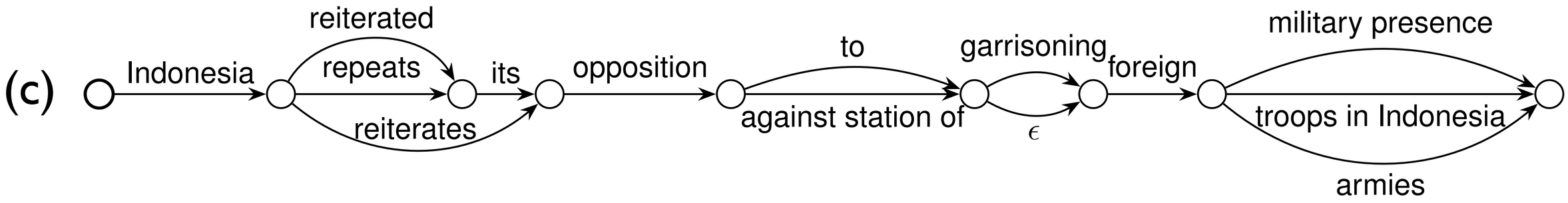
1. initialize

2.1 pick two sents

2.2 find identical words

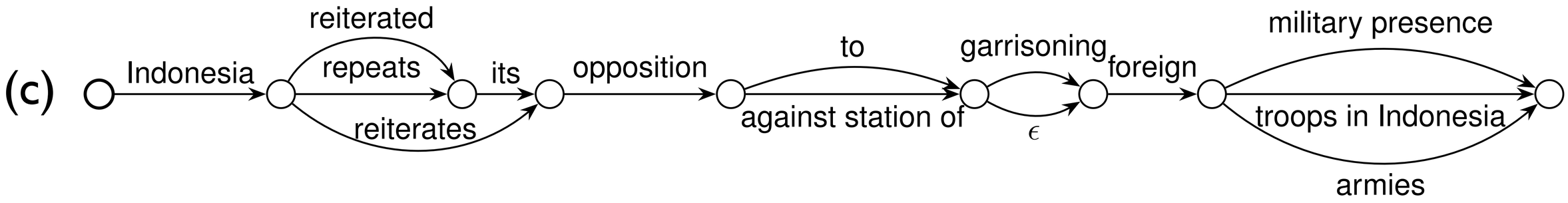
2.3 merge

Step 2: Generate Pseudo-references from Lattice



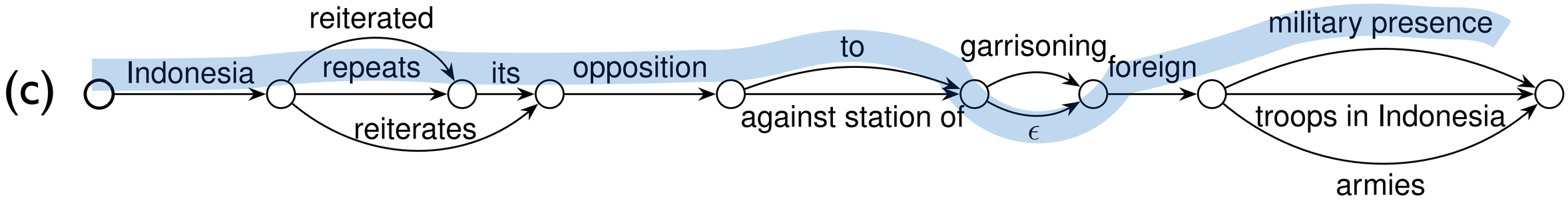
- first Idea: hard alignment
  - *step 1:* compress the references to lattice by merging the identical words
  - *step 2:* generate pseudo-references by traversing the lattice

Step 2: Generate Pseudo-references from Lattice



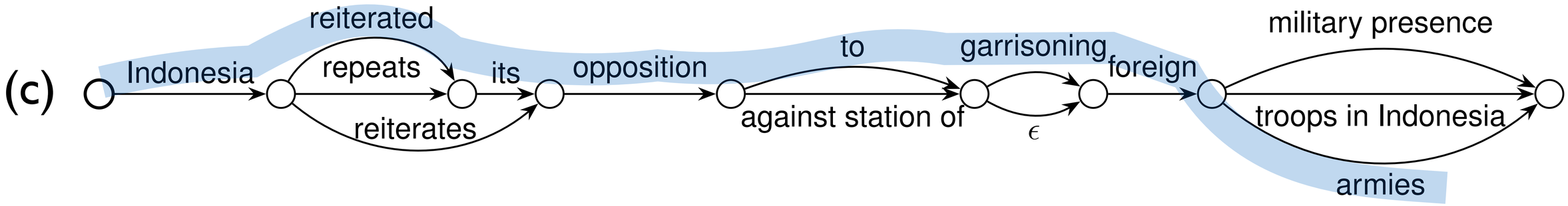
- first Idea: hard alignment
  - *step 1:* compress the references to lattice by merging the identical words
  - *step 2:* generate pseudo-references by traversing the lattice
    - 33 pseudo-references can be generated in total

Step 2: Generate Pseudo-references from Lattice



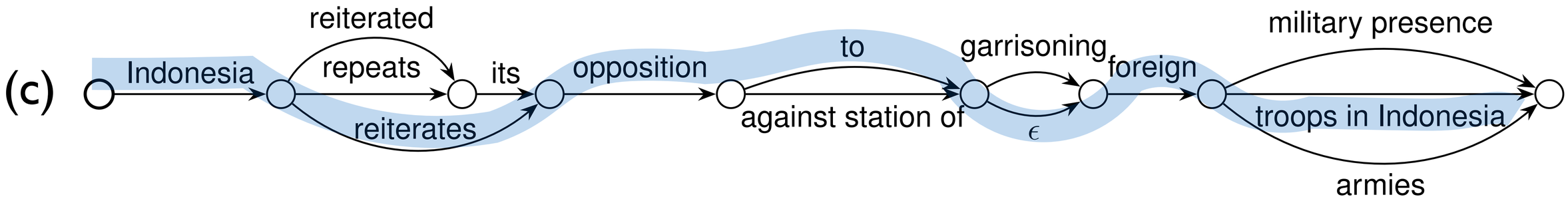
- first Idea: hard alignment
  - *step 1:* compress the references to lattice by merging the identical words
  - *step 2:* generate pseudo-references by traversing the lattice
    - 33 pseudo-references can be generated in total
      - Indonesia repeats its opposition to foreign military presence

Step 2: Generate Pseudo-references from Lattice



- first Idea: hard alignment
  - *step 1:* compress the references to lattice by merging the identical words
  - *step 2:* generate pseudo-references by traversing the lattice
    - 33 pseudo-references can be generated in total
      - Indonesia repeats its opposition to foreign military presence
      - Indonesia reiterated its opposition to garrisoning foreign armies

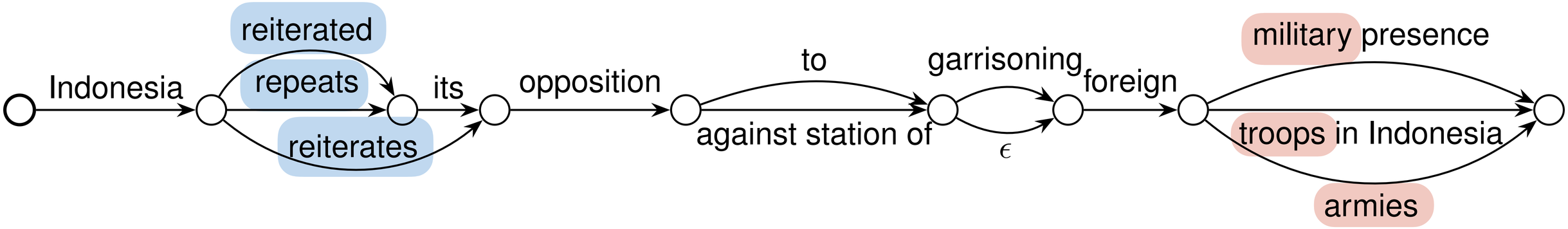
Step 2: Generate Pseudo-references from Lattice



- first Idea: hard alignment
  - *step 1:* compress the references to lattice by merging the identical words
  - *step 2:* generate pseudo-references by traversing the lattice
    - 33 pseudo-references can be generated in total
      - Indonesia repeats its opposition to foreign military presence
      - Indonesia reiterated its opposition to garrisoning foreign armies
      - Indonesia reiterates opposition to foreign troops in Indonesia
      - …

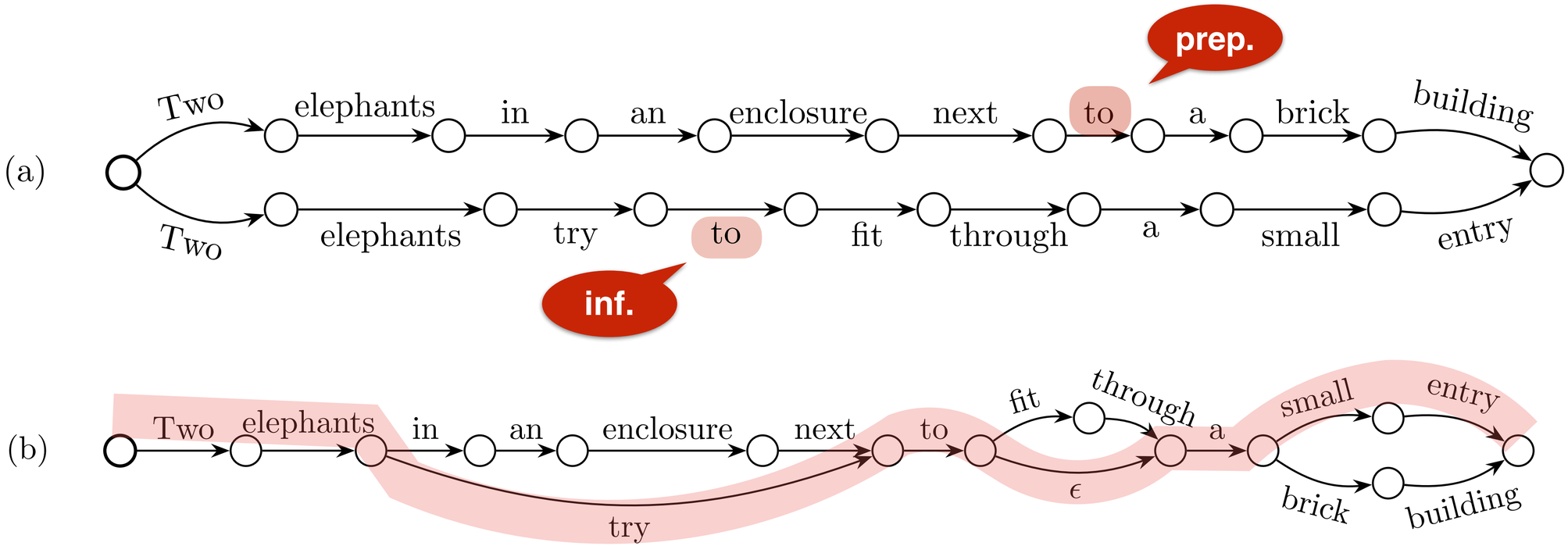
Limitation 1 of Hard Alignment: Synonyms?

- can not merge different words with similar meanings



Limitation 2 of Hard Alignment: Polysemy?

- may merge identical words with different meanings in different contexts

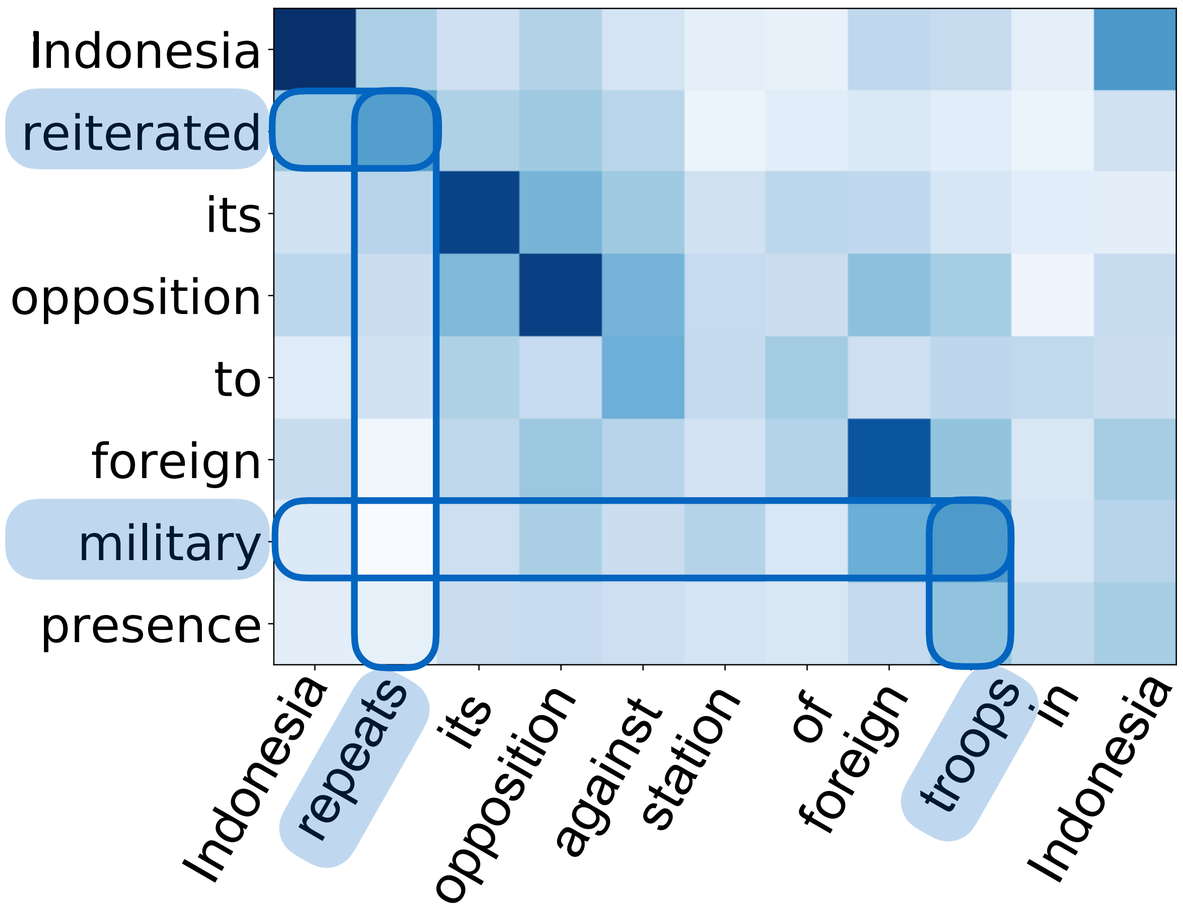


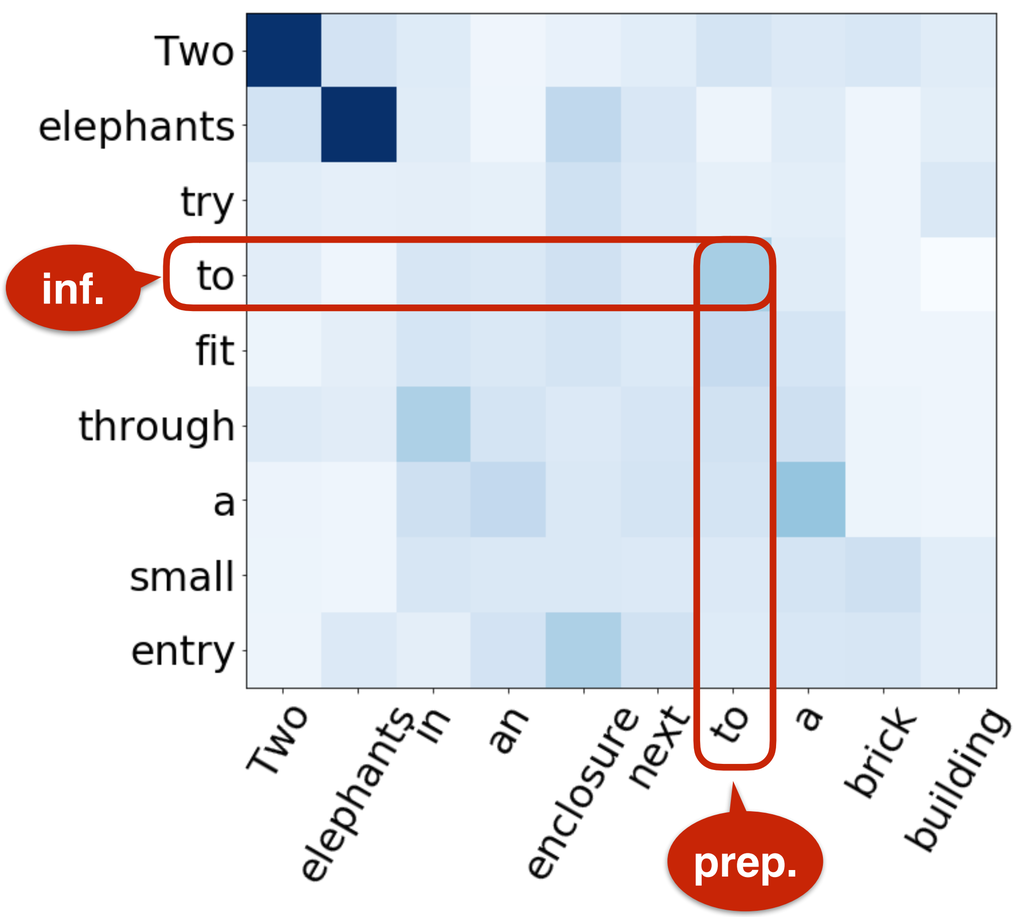
- wrong pseudo-reference:

Two elephants try to a small entry

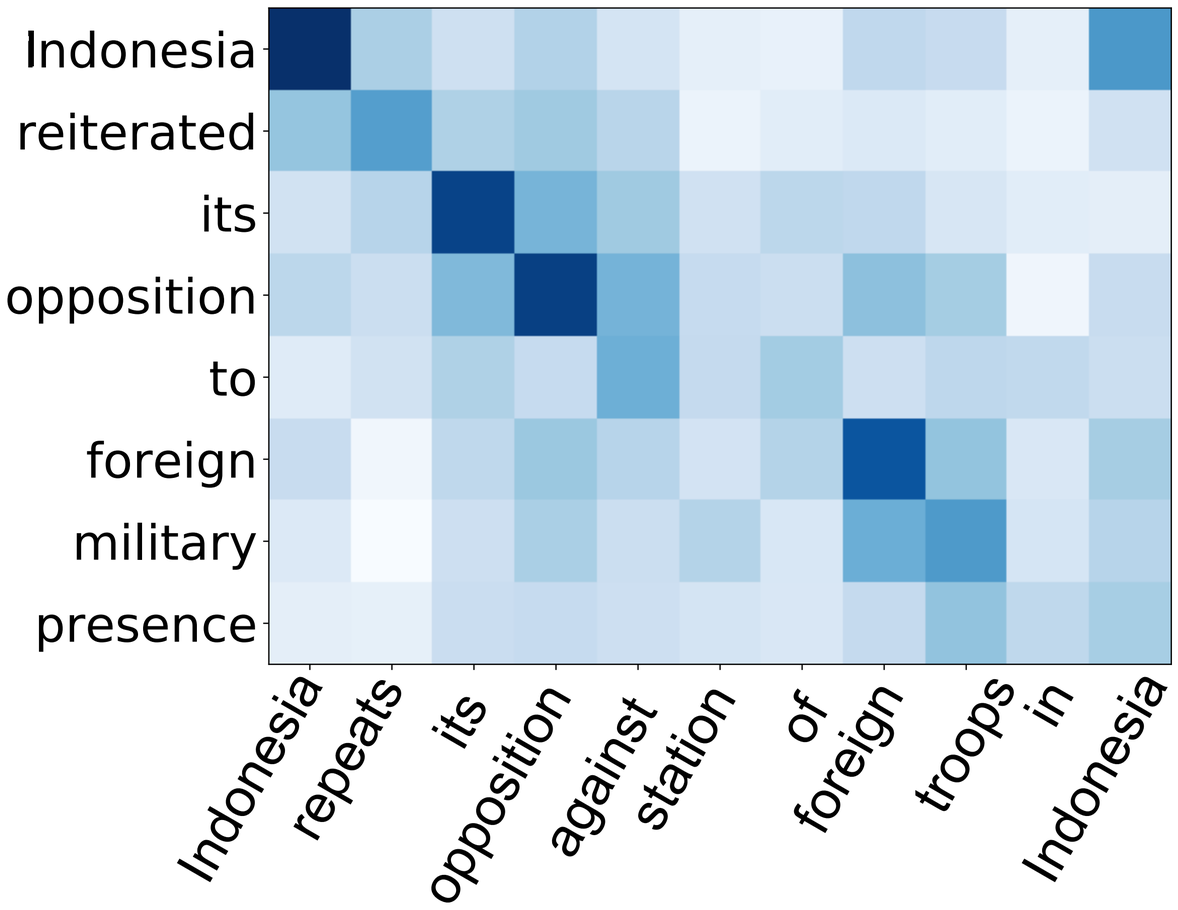
How to Solve these Limitations

- *solution:* identify synonyms and polysemes
  - need a semantic similarity matrix





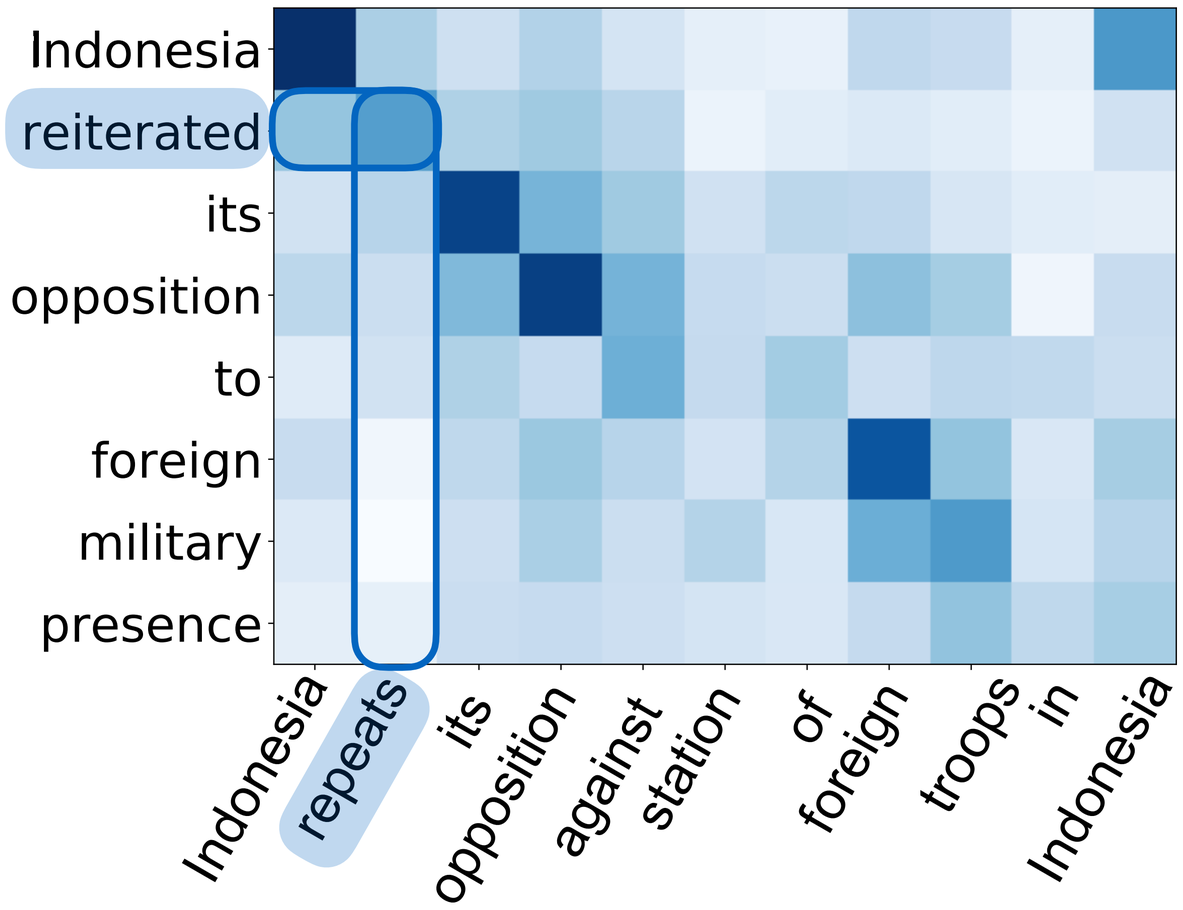
Solution: Identify Synonyms and Polysemes



Semantic Substitution Matrix

- train bi-directional Language Model
  - predict each word by given the surrounding words
- word pair similarity with cosine similarity

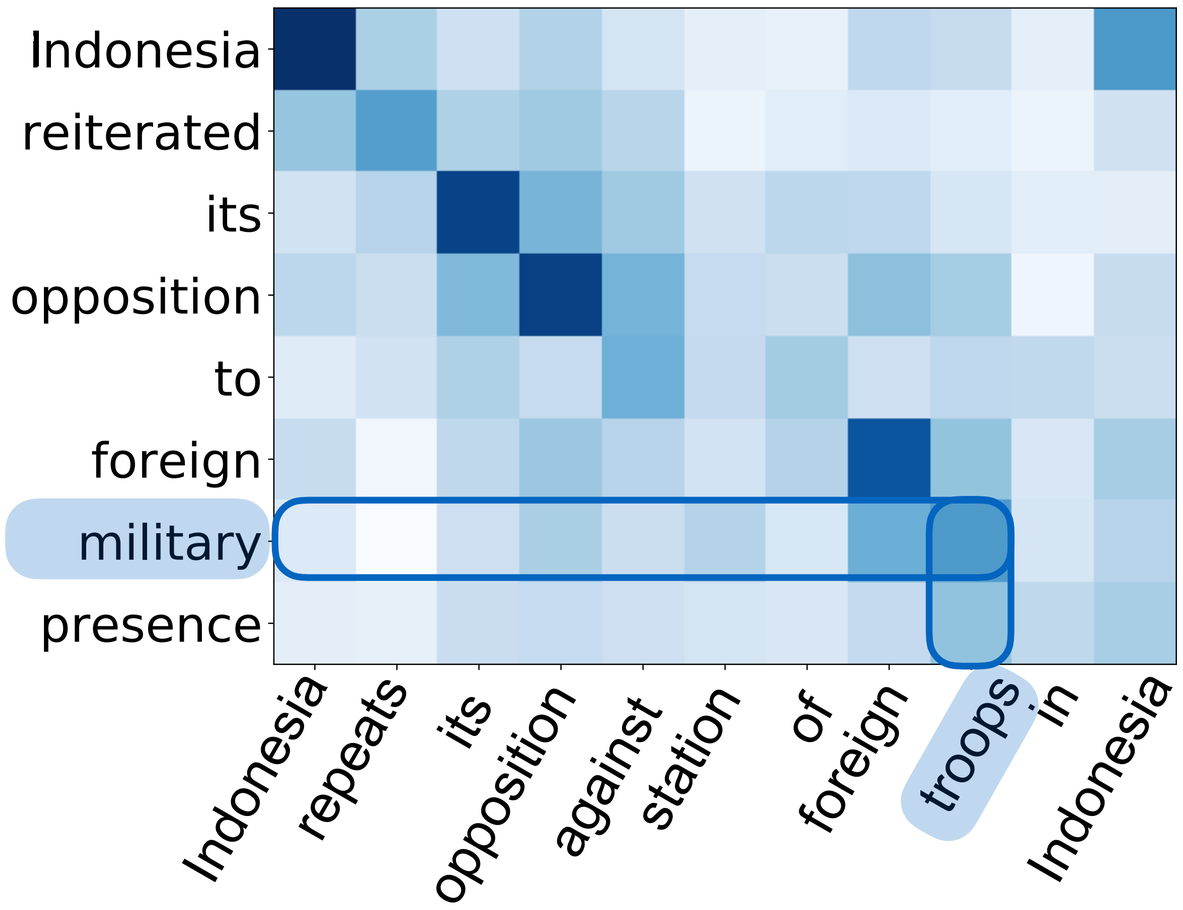
Solution: Identify Synonyms and Polysemes



Semantic Substitution Matrix

- train bi-directional Language Model
  - predict each word by given the surrounding words
- word pair similarity with cosine similarity
- **reiterated** and **repeats** have high similarity score

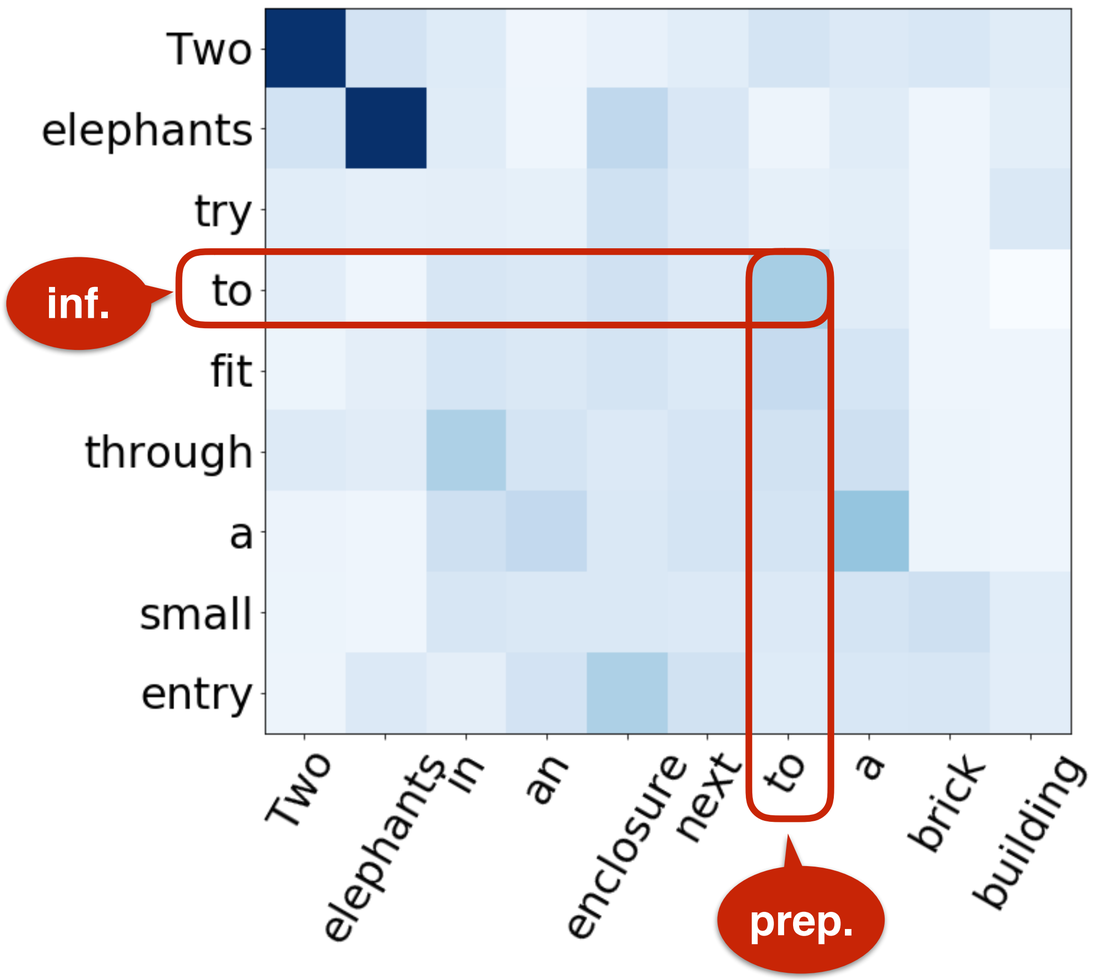
Solution: Identify Synonyms and Polysemes



Semantic Substitution Matrix

- train bi-directional Language Model
  - predict each word by given the surrounding words
- word pair similarity with cosine similarity
- **reiterated** and **repeats** have high similarity score
- **military** and **troops** have high similarity score

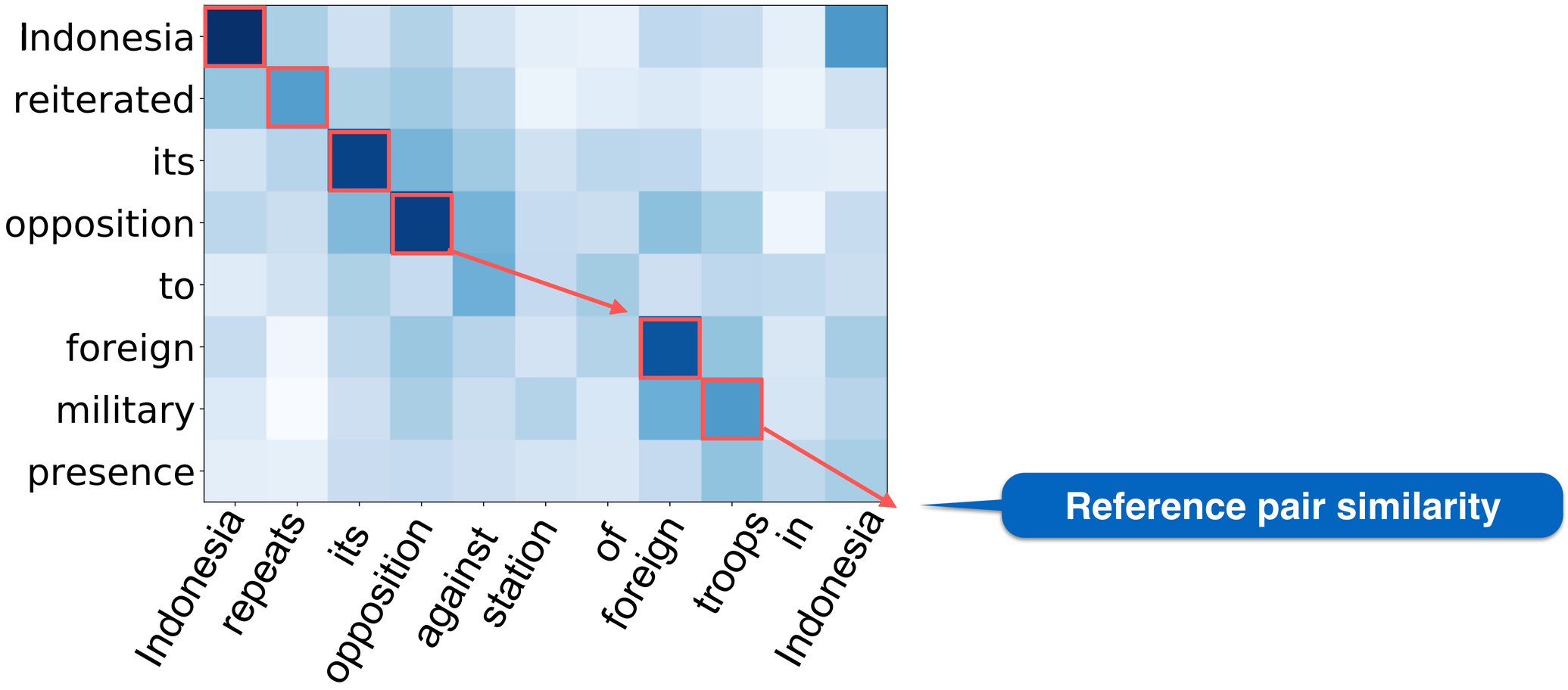
Solution: Identify Synonyms and Polysemes



Semantic Substitution Matrix

- train bi-directional Language Model
  - predict each word by given the surrounding words
- word pair similarity with cosine similarity
- **reiterated** and **repeats** have high similarity score
- **military** and **troops** have high similarity score
- two **to**'s have low similarity score

Soft Alignment with Dynamic Programming



Iterative Pairwise Soft Alignment

1. initialize

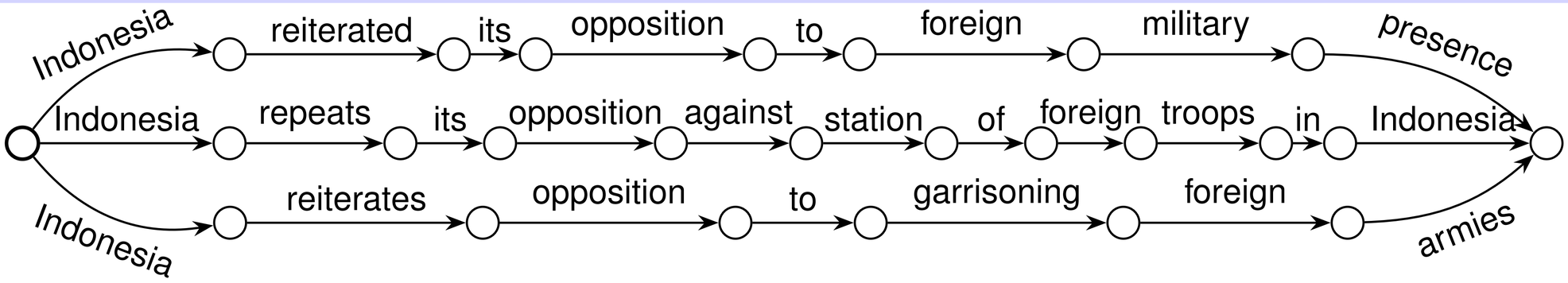
2.1 pick two sents

2.2 find aligned words

2.3 merge

Iterative Pairwise Soft Alignment

(a)



1. initialize

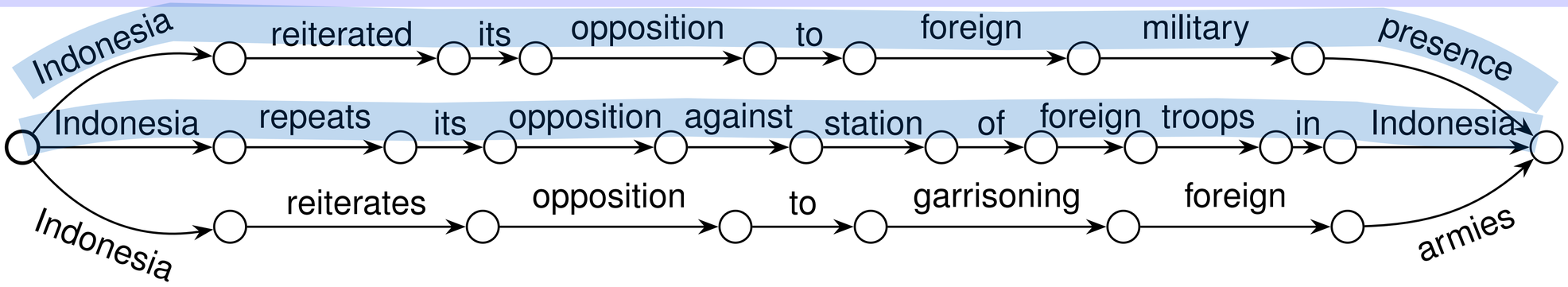
2.1 pick two sents

2.2 find aligned words

2.3 merge

Iterative Pairwise Soft Alignment

(a)



1. initialize

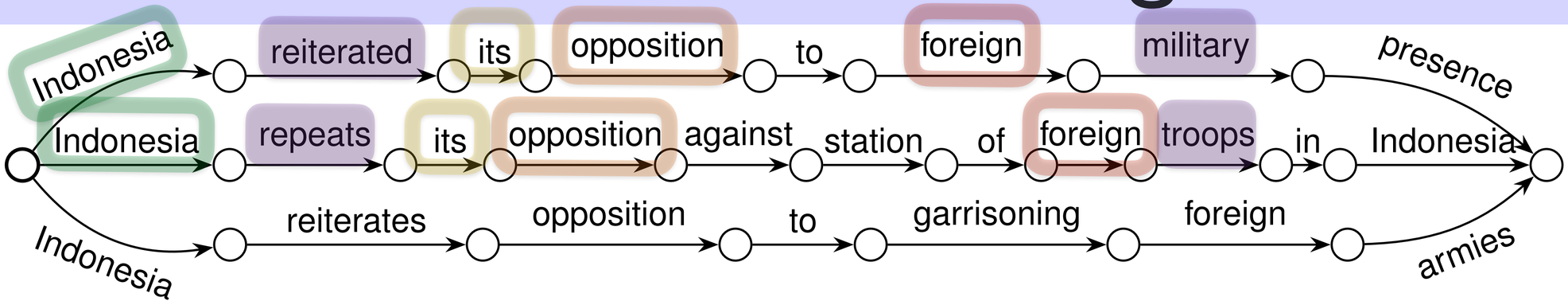
2.1 pick two sents

2.2 find aligned words

2.3 merge

Iterative Pairwise Soft Alignment

(a)



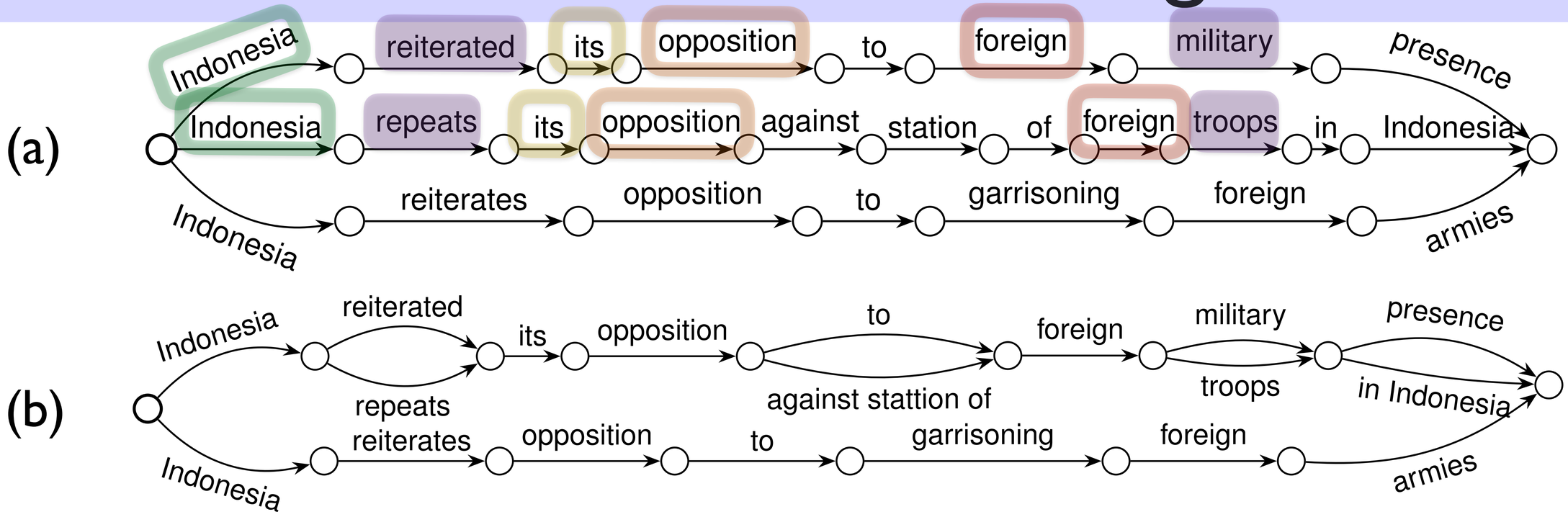
1. initialize

2.1 pick two sents

2.2 find aligned words

2.3 merge

Iterative Pairwise Soft Alignment



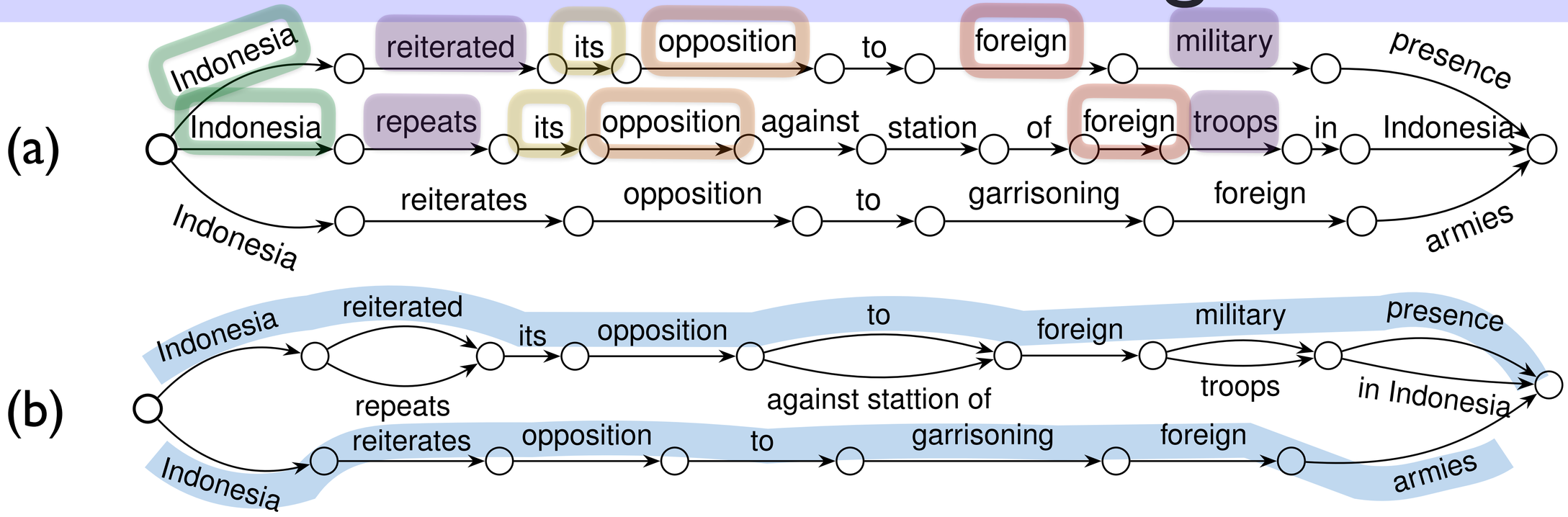
1. initialize

2.1 pick two sents

2.2 find aligned words

2.3 merge

Iterative Pairwise Soft Alignment



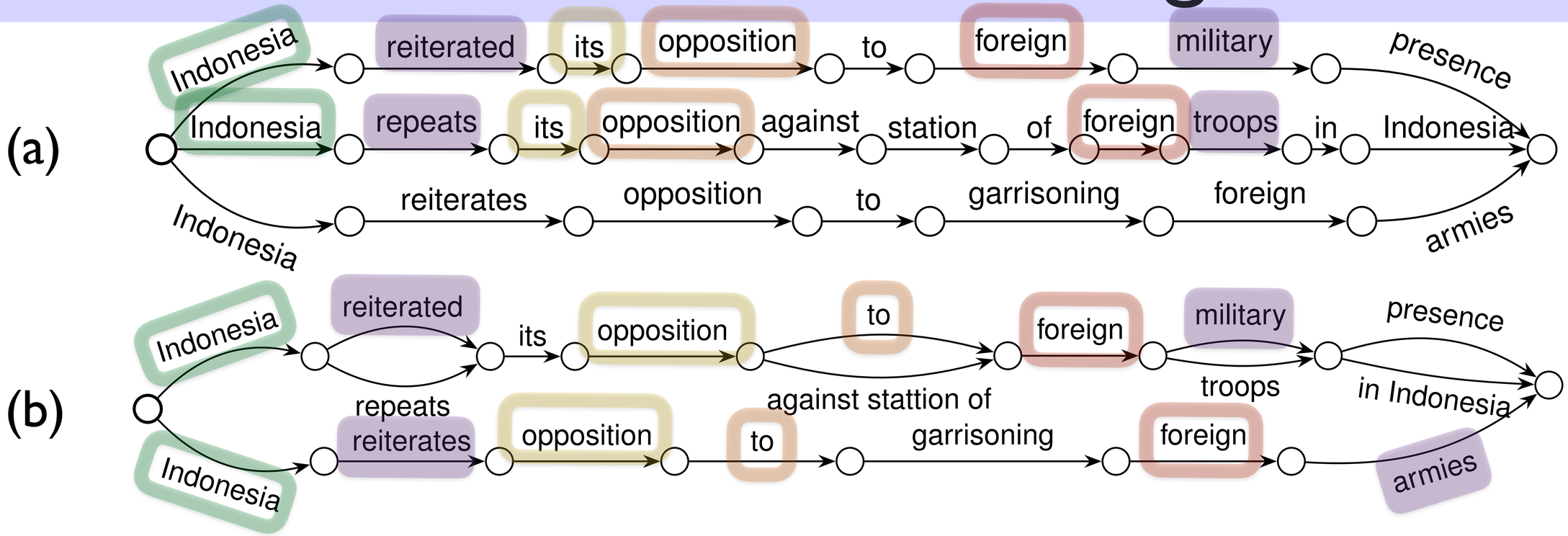
1. initialize

2.1 pick two sents

2.2 find aligned words

2.3 merge

Iterative Pairwise Soft Alignment



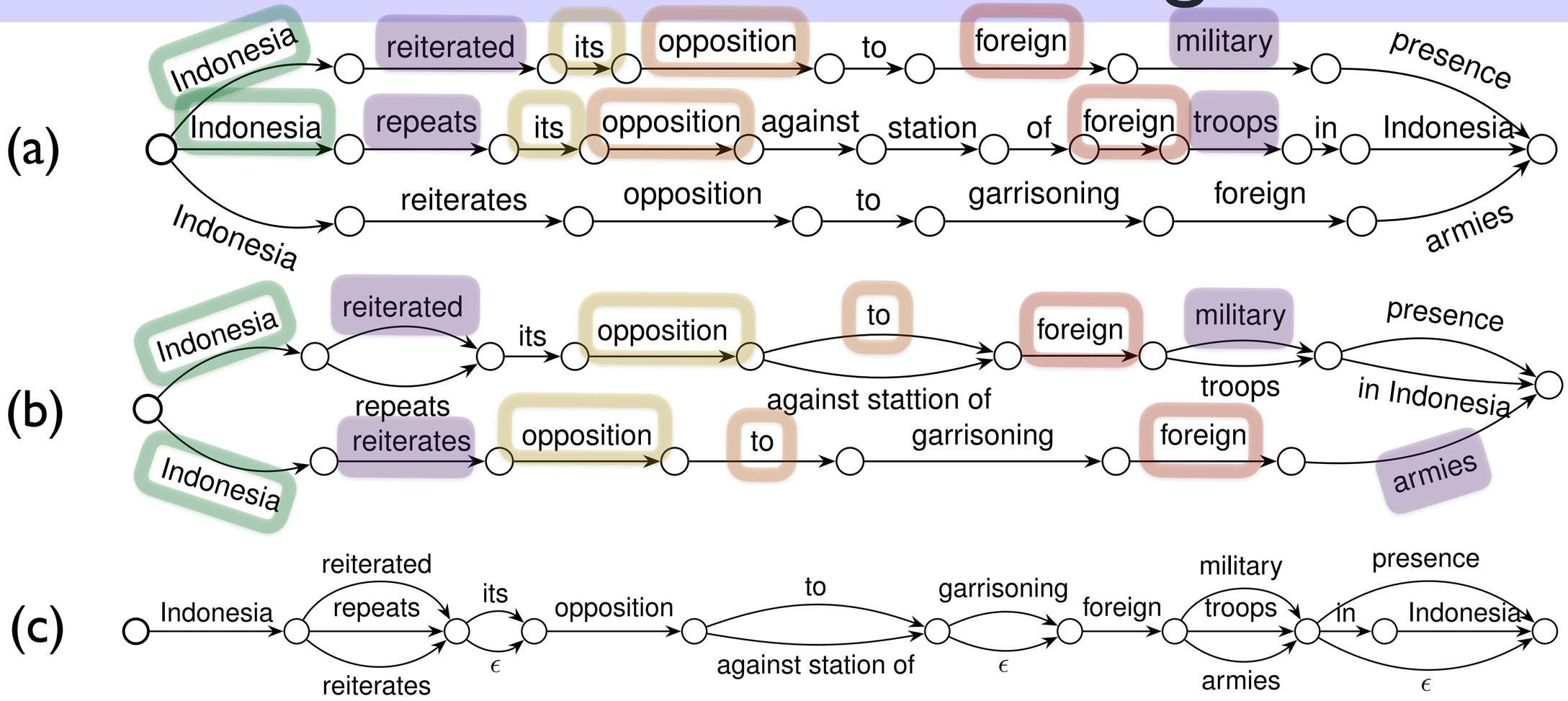
1. initialize

2.1 pick two sents

2.2 find aligned words

2.3 merge

Iterative Pairwise Soft Alignment



1. initialize

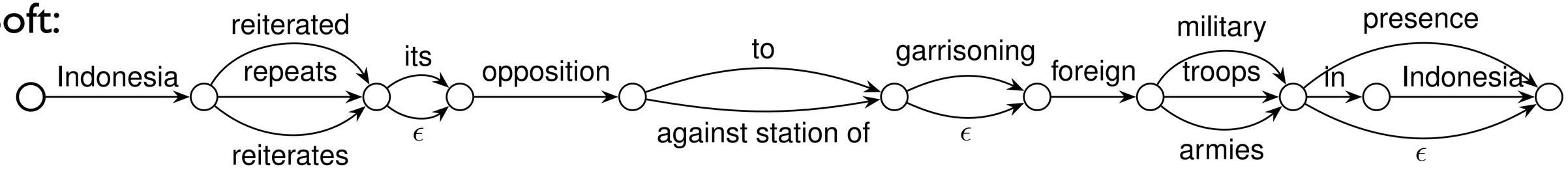
2.1 pick two sents

2.2 find aligned words

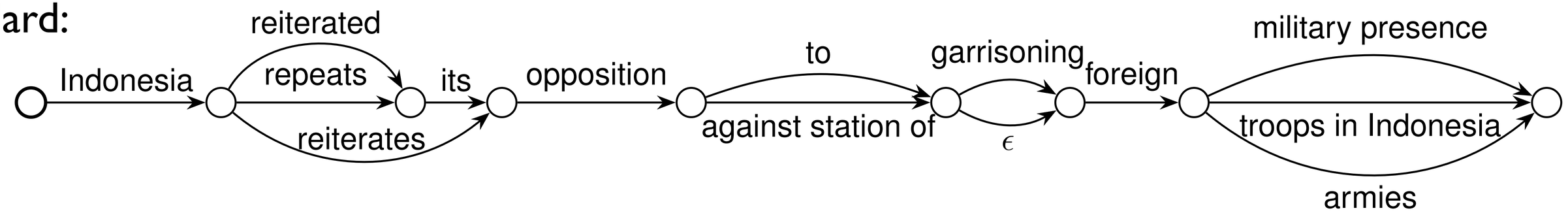
2.3 merge

Iterative Pairwise Soft Word Alignment

Soft:

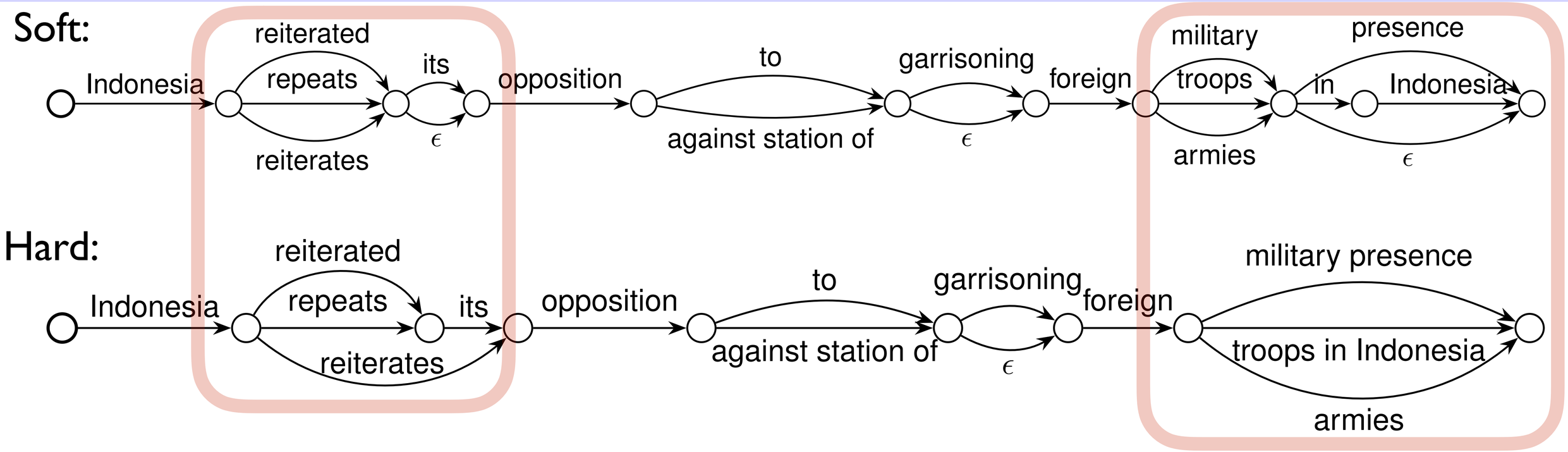


Hard:



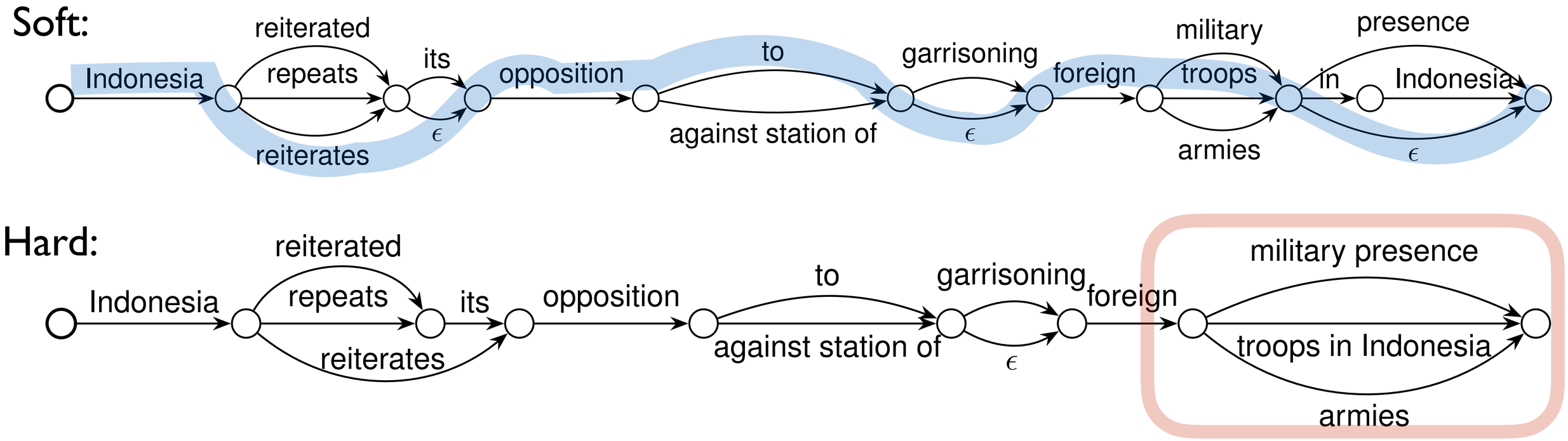
- 213 pseudo-refs generated (180 more than by hard alignment)

Iterative Pairwise Soft Word Alignment



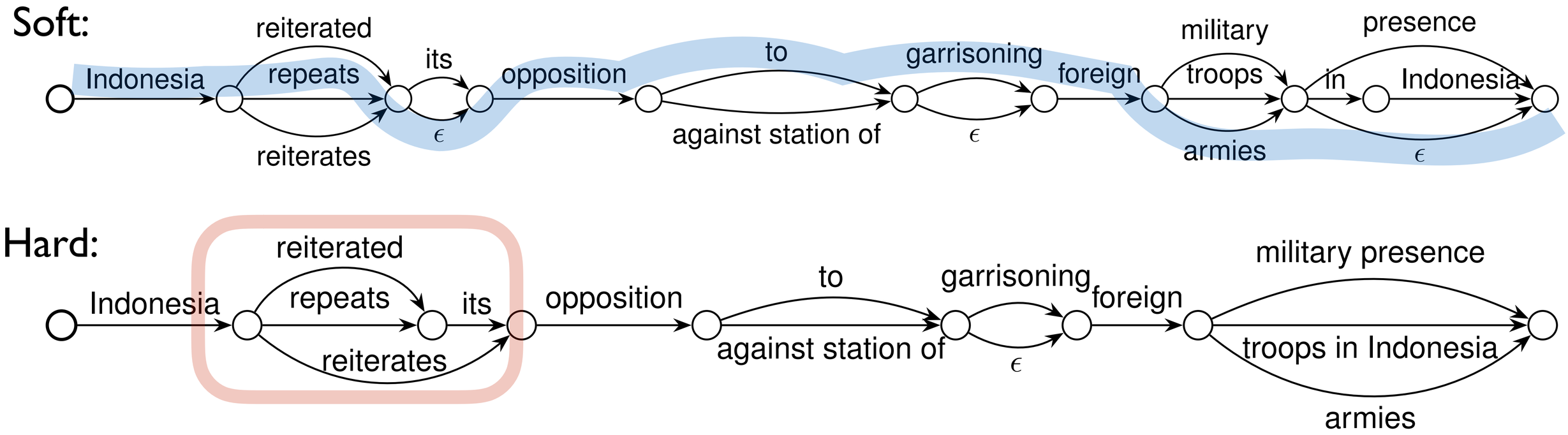
- 213 pseudo-refs generated (180 more than by hard alignment)

Iterative Pairwise Soft Word Alignment



- 213 pseudo-refs generated (180 more than by hard alignment)
  - Indonesia reiterates opposition to foreign troops

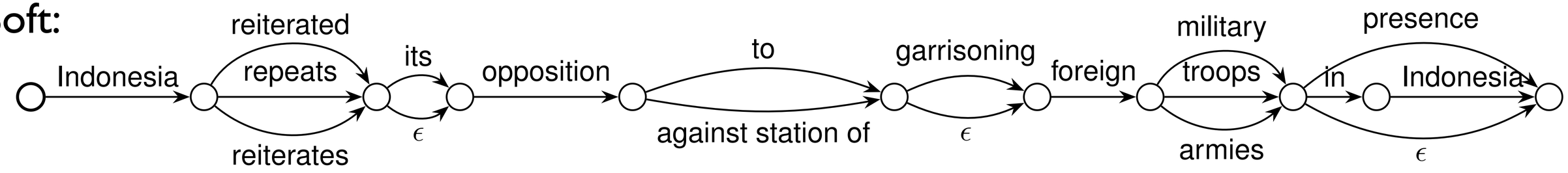
Iterative Pairwise Soft Word Alignment



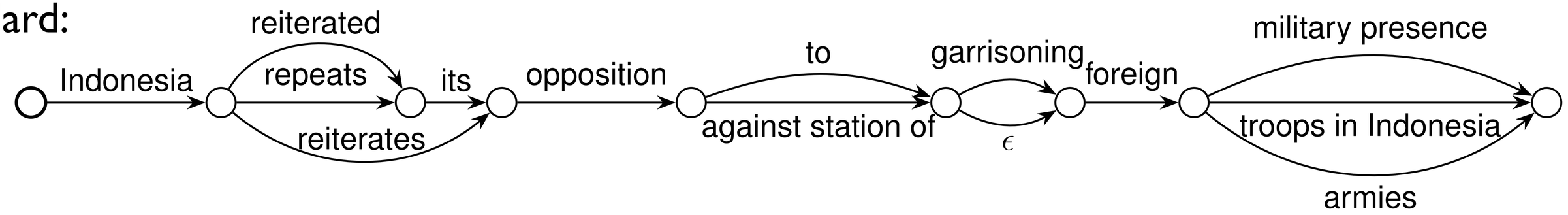
- 213 pseudo-refs generated (180 more than by hard alignment)
  - Indonesia reiterates opposition to foreign troops
  - Indonesia repeats opposition to garrisoning foreign armies

Iterative Pairwise Soft Word Alignment

Soft:

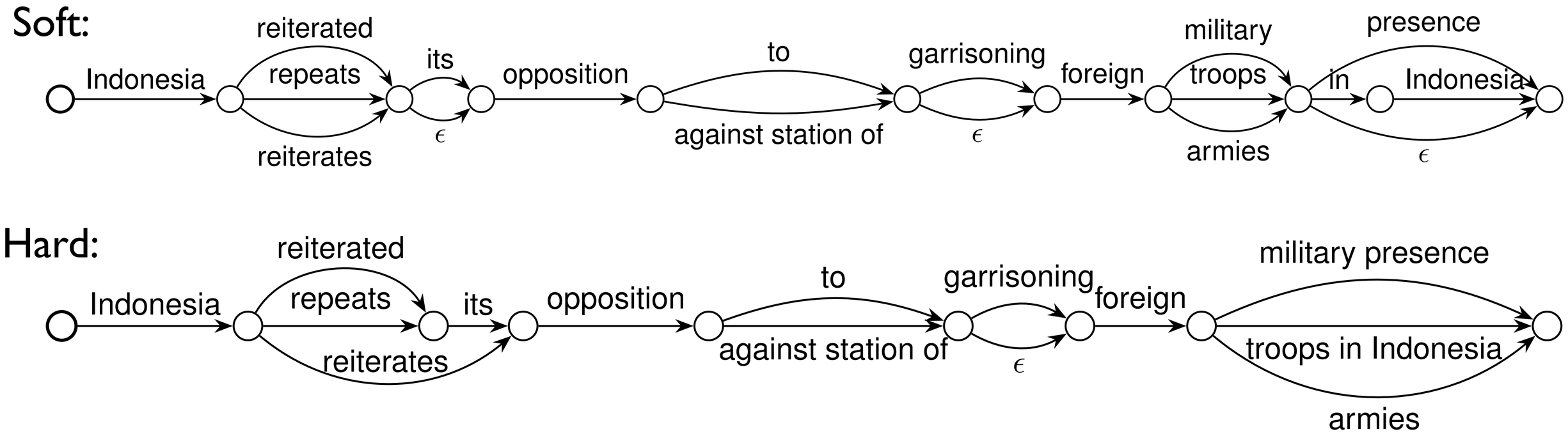


Hard:



- 213 pseudo-refs generated (180 more than by hard alignment)
  - Indonesia reiterates opposition to foreign troops
  - Indonesia repeats opposition to garrisoning foreign armies
  - …

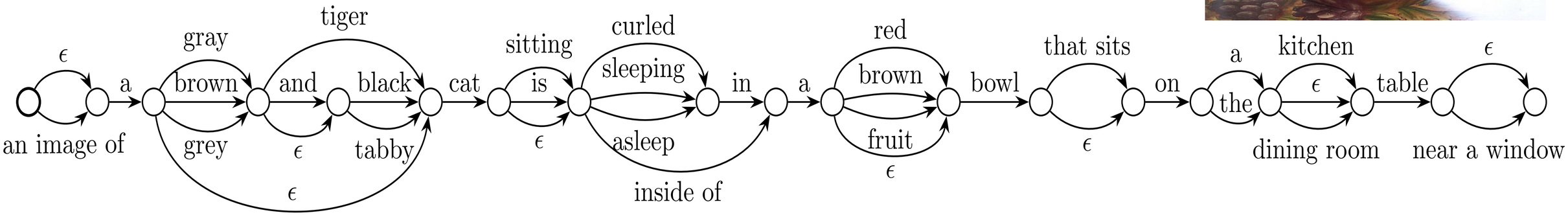
Iterative Pairwise Soft Word Alignment



- 213 pseudo-refs generated (180 more than by hard alignment)
  - Indonesia reiterates opposition to foreign troops
  - Indonesia repeats opposition to garrisoning foreign armies
  - …
- keep best *k* pseudo-refs according to BLEU score over original refs

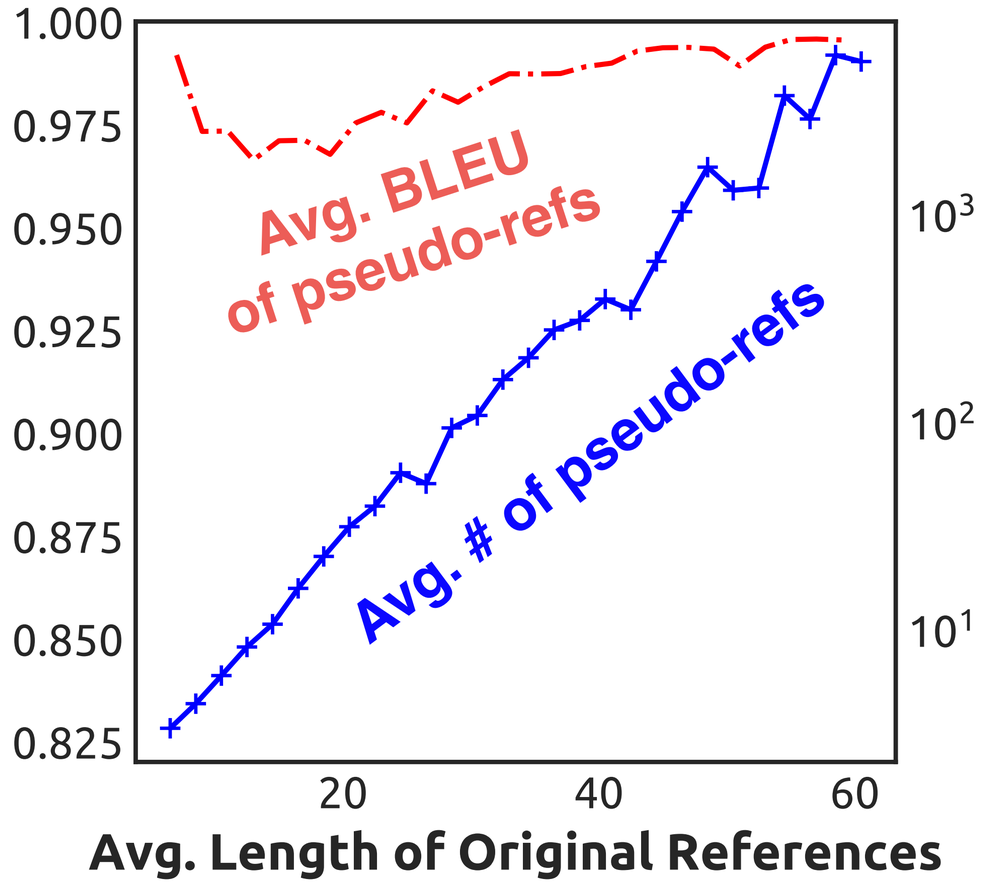
Image Captioning Example

- original references
  - A gray tabby cat is curled in a red bowl that sits on a table near a window
  - A brown and black cat is sleeping in a bowl on a table
  - A grey tiger cat sleeping in a brown bowl on a table
  - An image of a cat sitting inside of a bowl on the kitchen table
  - A cat asleep in a fruit bowl on a dining room table

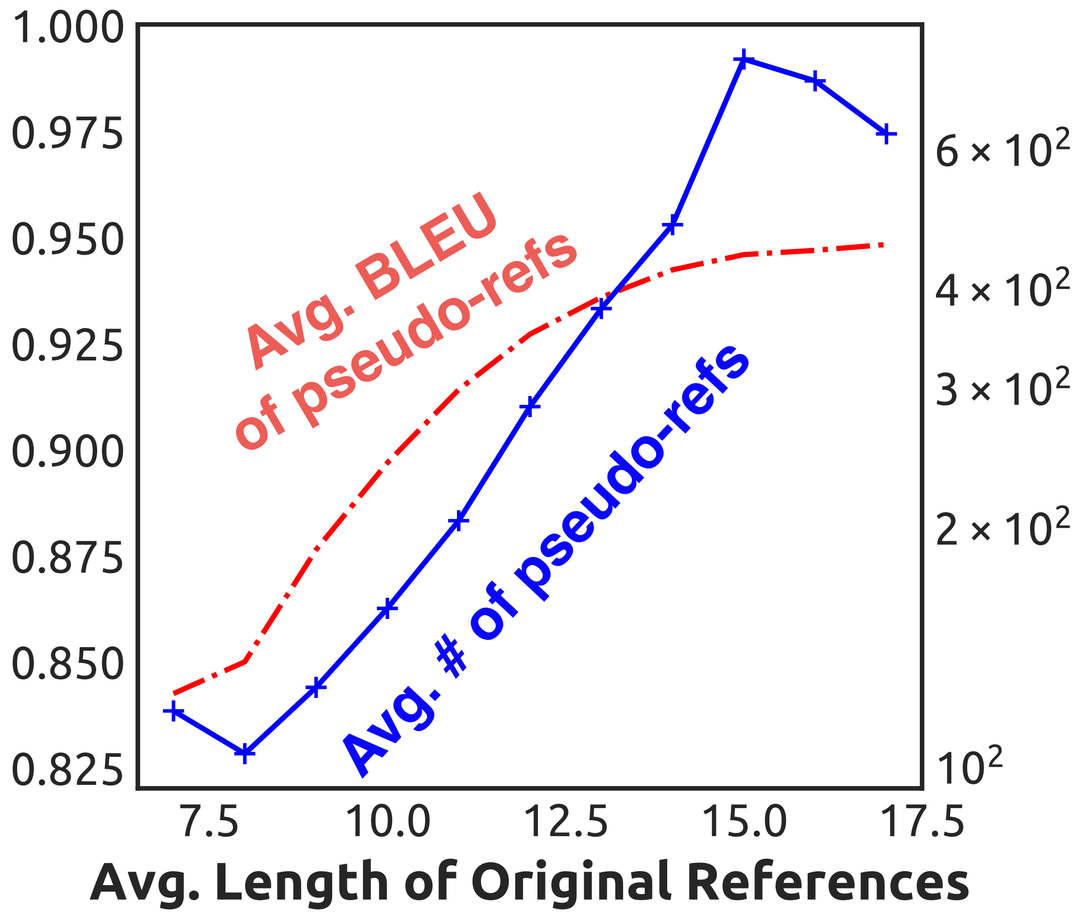


Rank	Pseudo-reference	BLEU
1	A grey tiger cat sleeping in a brown bowl on a table near a window	100%
. . .	. . .	. . .
48	A grey tiger cat sleeping in a fruit bowl on a table	97%
. . .	. . .	. . .
73,274	A grey and tabby cat inside of a red bowl on the dining room table	0%

Analysis of Generated References

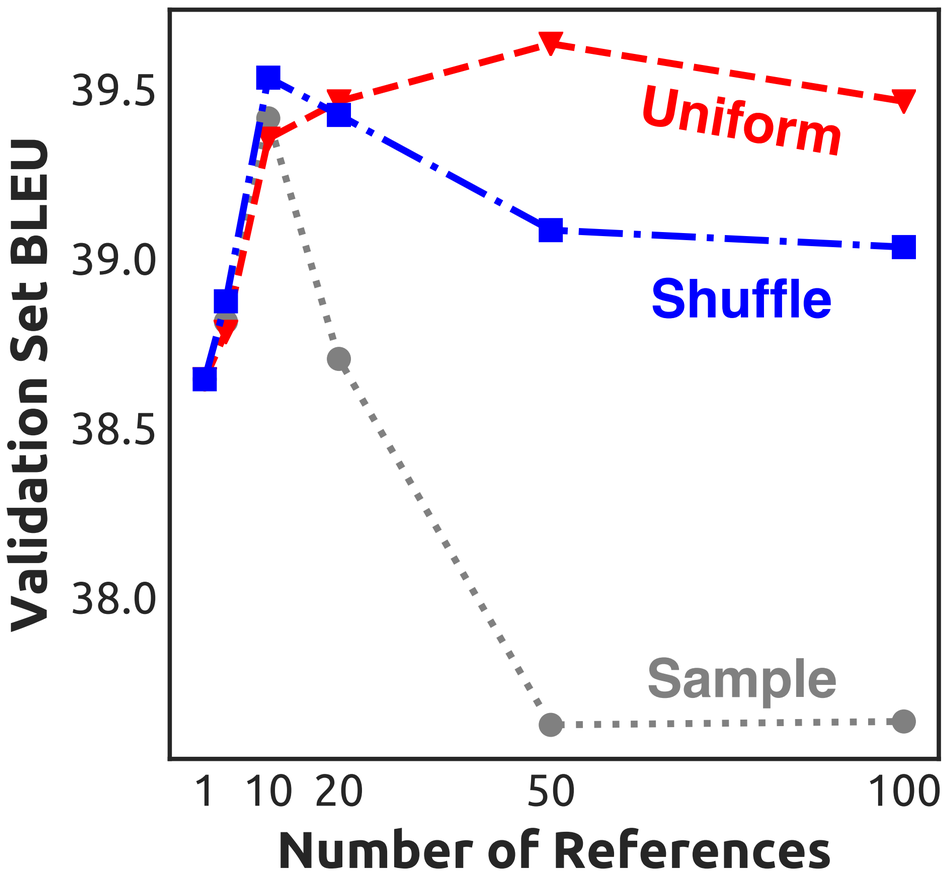


NIST Machine Translation

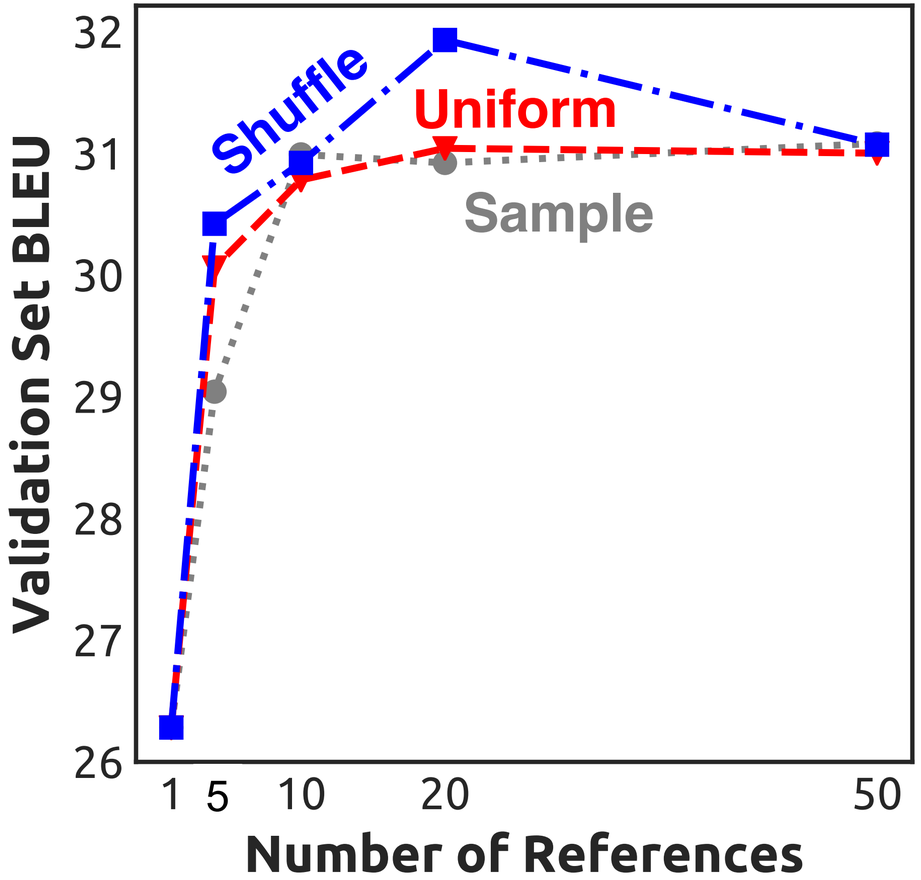


MSCOCO Image Captioning

Results on Validation Sets



NIST Machine Translation



MSCOCO Image Captioning

Results on Test set

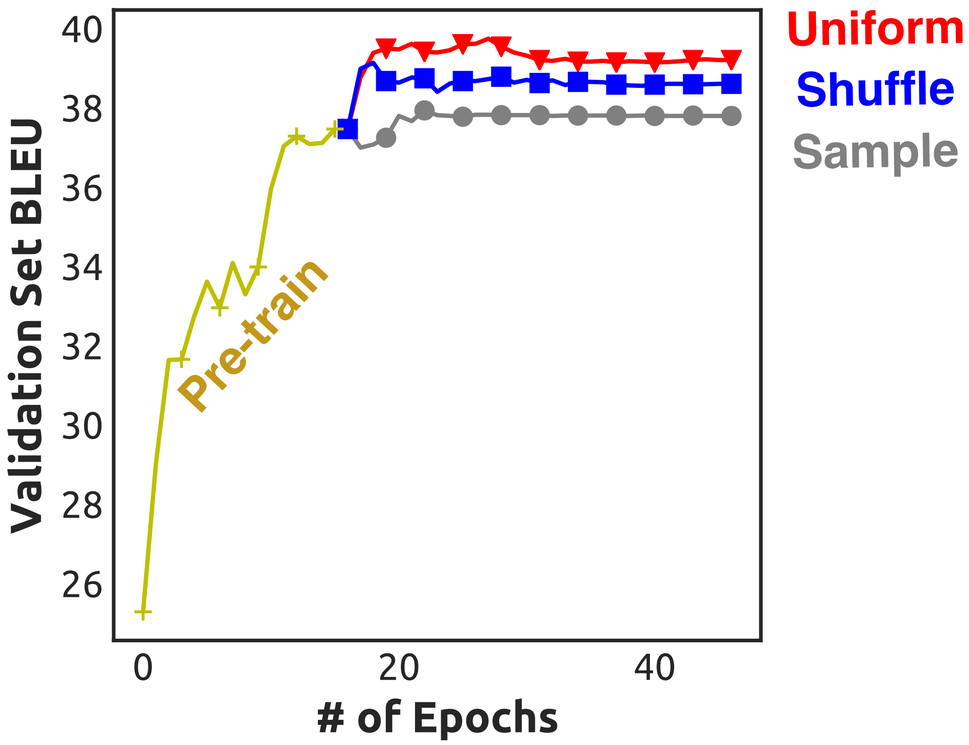
NIST Machine Translation

# of Refs		Method	BLEU
pre-train	fine-tune		
1		Pre-train	33.6
1	1	First	34.5
1	4	Shuffle	35.2
1	50	Uniform	36.0

+0.7

+1.5

pre-train on 1M NIST
fine-tune on 4,667 NIST02-05
test results on 691 NIST08



learning curve of different methods
with 50 references

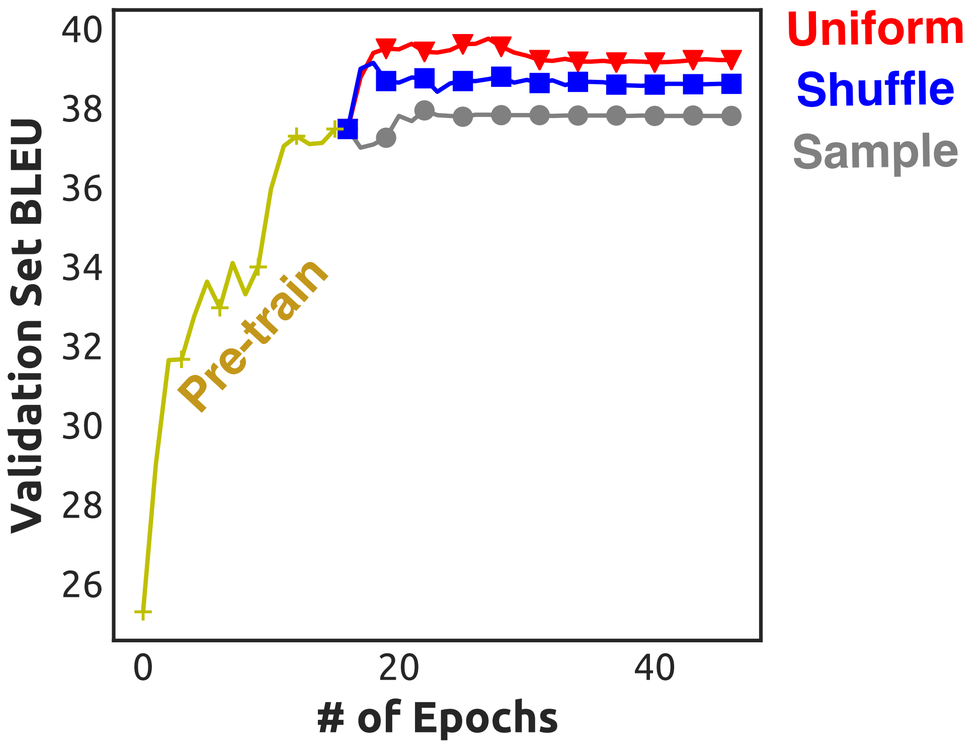
Results on Test set

NIST Machine Translation

# of Refs		Method	BLEU
pre-train	fine-tune		
1		Pre-train	33.6
1	1	First	34.5
1	4	Shuffle	35.2
1	50	Uniform	36.0

+2.4

pre-train on 1M NIST
fine-tune on 4,667 NIST02-05
test results on 691 NIST08



learning curve of different methods
with 50 references

Results on Test set

NIST Machine Translation

# of Refs		Method	BLEU
pre-train	fine-tune		
1		Pre-train	33.6
1	1	First	34.5
1	4	Shuffle	35.2
1	50	Uniform	36.0

+2.4

pre-train on 1M NIST
fine-tune on 4,667 NIST02-05
test results on 691 NIST08

MSCOCO Image Captioning

# of Refs	Method	BLEU	CIDEr
1	First	26.70	80.70
5	Sample One	28.67	85.41
5	Shuffle	30.94 (+2.3)	94.10 (+8.7)
20	Shuffle	31.79 (+3.1)	97.10 (+11.7)

MSCOCO Image Captioning test results

Conclusions

- We use multiple-reference for training MT (image captioning)
- We introduce a soft alignment based lattice compression framework
  - which can generate more training references based on existing ones
  - our methods outperforms baselines on both MT and image captioning

Conclusions

- We use multiple-reference for training MT (image captioning)
- We introduce a soft alignment based lattice compression framework
  - which can generate more training references based on existing ones
  - our methods outperforms baselines on both MT and image captioning

Oregon State University

We're hiring PhDs

Baidu Research

We're hiring researchers

Thanks!

Codes available at: https://github.com/renj/pseudo-ref