USING STATISTICAL MACHINE TRANSLATION TO IMPROVE STATISTICAL MACHINE TRANSLATION

NITIN MADNANI

RESEARCH SCIENTIST

TEXT, LANGUAGE & COMPUTATION

EDUCATIONAL TESTING SERVICE

PRINCETON, NJ

nmadnani@ets.org

HERE BE THREE PARTS ...

- Introduce statistical machine translation (SMT) using as little math as possible `(0 < |math| << boring)`
- Bring to light the dark magic of *parameter tuning* - without which SMT doesn't work - and its need for a special kind of data
- Show how I use SMT itself to "manufacture" this special data and significantly improve final translation performance

PART I

THE PIPELINE

MACHINE TRANSLATION

MACHINE TRANSLATION

“有机食品全程监管存在交叉或空当。”山东肥城市农业局副局长赵胜文等人表示，农业部门管生产，认监委管认证，工商部门管流通，卫生部门管餐桌，哪个部门都说得上话，哪个部门也不完全说了算。”

根据超市提供的信息，记者联系上一位常年在北京、山东、广西等地做“有机食品”收购、加工、包装的经纪人刘刚。他表示，有机食品行业认证环节随意性大。针对认证中出现的问题，记者联系到负有监管责任的国家认监委。

Documents in Source Language (SL)
(Chinese)

MACHINE TRANSLATION

“有机食品全程监管存在交叉或空当。”山东肥城市农业局副局长赵胜文等人表示，农业部门管生产，认监委管认证，工商部门管流通，卫生部门管餐桌，哪个部门都说得上话，哪个部门也不完全说了算。”

根据超市提供的信息，记者联系上一位常年在北京、山东、广西等地做“有机食品”收购、加工、包装的经纪人刘刚。他表示，有机食品行业认证环节随意性大。针对认证中出现的问题，记者联系到负有监管责任的国家认监委。

Documents in Source Language (SL)
(Chinese)

MT System

"Organic food full or empty when there is cross-regulation." Fat City, Shandong Zhaosheng Wen, Deputy Secretary of Agriculture, who said that the agricultural sector production management, commission management certification, the business sector pipe flow, the health sector management table, which the department put on the case, the department is not completely final say.

According to information provided by the supermarket, the reporter contacted a year-round in Beijing, Shandong, Guangxi and other places to do "organic food" acquisition, processing, packaging broker Gang. He said part of the organic food industry certification arbitrary. For certification problems, the reporter linked to the regulatory responsibility of the state commission bear.

Documents in Target Language (TL)
(English)

MACHINE TRANSLATION

“有机食品全程监管存在交叉或空当。”山东肥城市农业局副局长赵胜文等人表示，农业部门管生产，认监委管认证，工商部门管流通，卫生部门管餐桌，哪个部门都说得上话，哪个部门也不完全说了算。”

根据超市提供的信息，记者联系上一位常年在北京、山东、广西等地做“有机食品”收购、加工、包装的经纪人刘刚。他表示，有机食品行业认证环节随意性大。针对认证中出现的问题，记者联系到负有监管责任的国家认监委。

Documents in Source Language (SL)
(Chinese)

MT
System

"Organic food full or empty when there is cross-regulation." Fat City, Shandong Zhaosheng Wen, Deputy Secretary of Agriculture, who said that the agricultural sector production management, commission management certification, the business sector pipe flow, the health sector management table, which the department put on the case, the department is not completely final say.

According to information provided by the supermarket, the reporter contacted a year-round in Beijing, Shandong, Guangxi and other places to do "organic food" acquisition, processing, packaging broker Gang. He said part of the organic food industry certification arbitrary. For certification problems, the reporter linked to the regulatory responsibility of the state commission bear.

Documents in Target Language (TL)
(English)

Ideally, semantically <u>adequate</u> as well as linguistically <u>fluent</u>

MACHINE TRANSLATION

- First conceived by Warren Weaver in 1949†
- One of the most challenging (and popular) NLP tasks over the last two decades
- Three popular non-statistical approaches [1950s-1980s]
  - Rule-based. Manually construct rules that translate from SL to TL (with minimal analysis)
  - Interlingual. Reduce SL text to an abstract, language-independent base-form and then generate TL text
  - Transfer-based. Analyze SL text into syntactic components, transfer SL syntax to TL syntax and then generate TL text

† *Translation*. Warren Weaver. 1949. http://www.mt-archive.info/Weaver-1949.pdf

STATISTICAL MACHINE TRANSLATION

- Driven by statistical machine learning methods
  - Step 0: Find **LOTS** of example SL sentences and corresponding human translations into TL (*bilingual parallel corpora* or *bitext*)
  - Step 1: Apply a learning algorithm to parallel corpora and build an *approximate model* of human translation
  - Step 2: Apply learned model to new SL text and obtain translations in TL (notice that I didn't say *unseen* SL text)
- Represents current state-of-the-art and dominates MT research in both academia and industry
- Examples: Google Translate, Bing Translate

LEARNING A TRANSLATION MODEL

LEARNING A TRANSLATION MODEL

Parallel Corpus or Bitext

```
这是一个英文句子.
这是中国一句.
...
...
...
```

```
This is an English sentence.
That's a Chinese sentence.
...
...
...
```

Millions of Words
(Depends on SL)

LEARNING A TRANSLATION MODEL

Parallel Corpus or Bitext

```
这是一个英文句子.
这是中国一句.
...
...
...
```

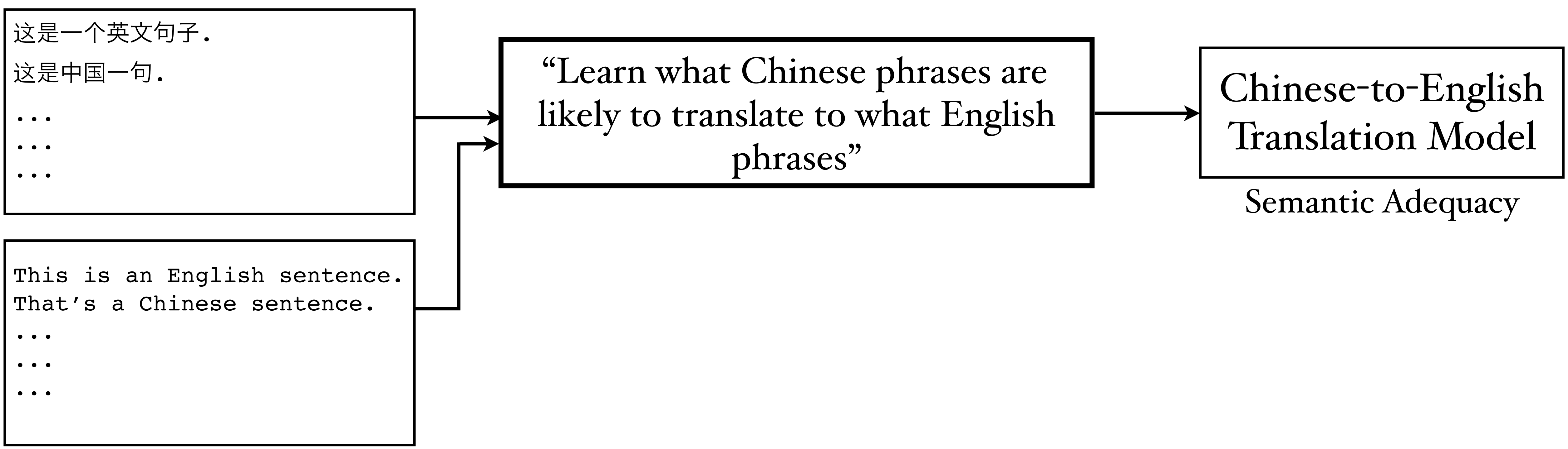
```
This is an English sentence.
That's a Chinese sentence.
...
...
...
```

"Learn what Chinese phrases are likely to translate to what English phrases"

Millions of Words
(Depends on SL)

LEARNING A TRANSLATION MODEL

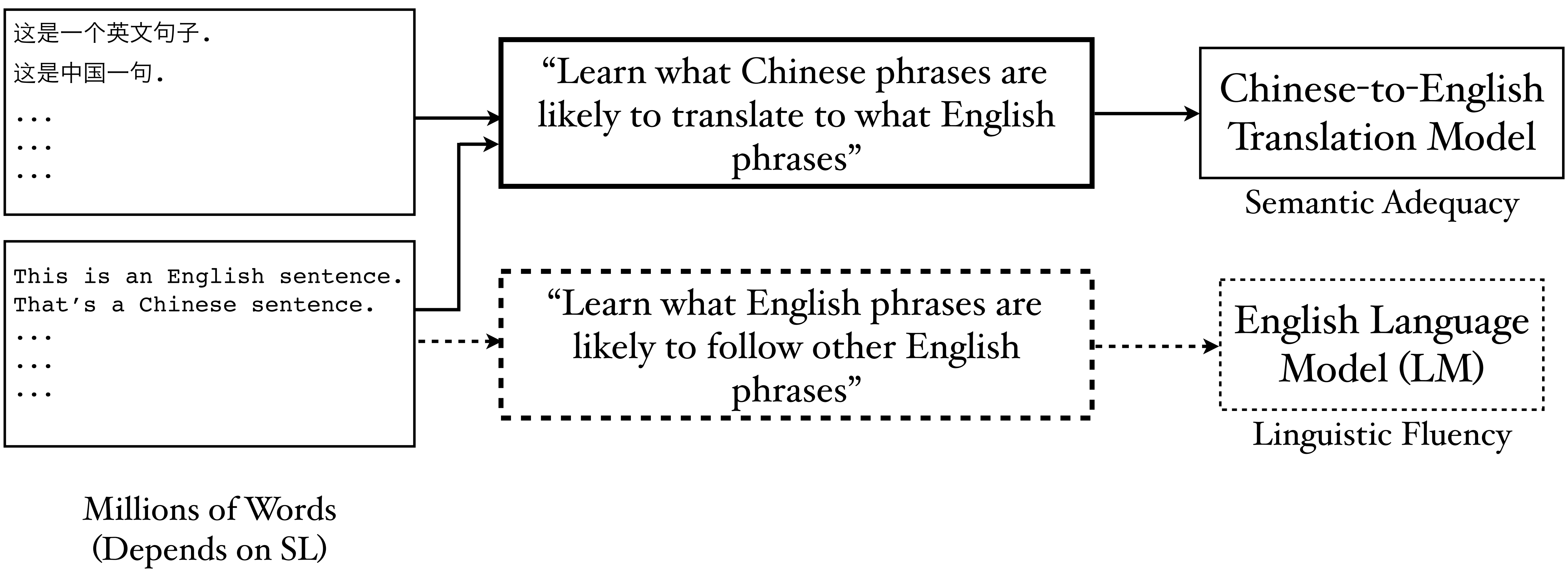
Parallel Corpus or Bitext



Millions of Words
(Depends on SL)

LEARNING A TRANSLATION MODEL

Parallel Corpus or Bitext

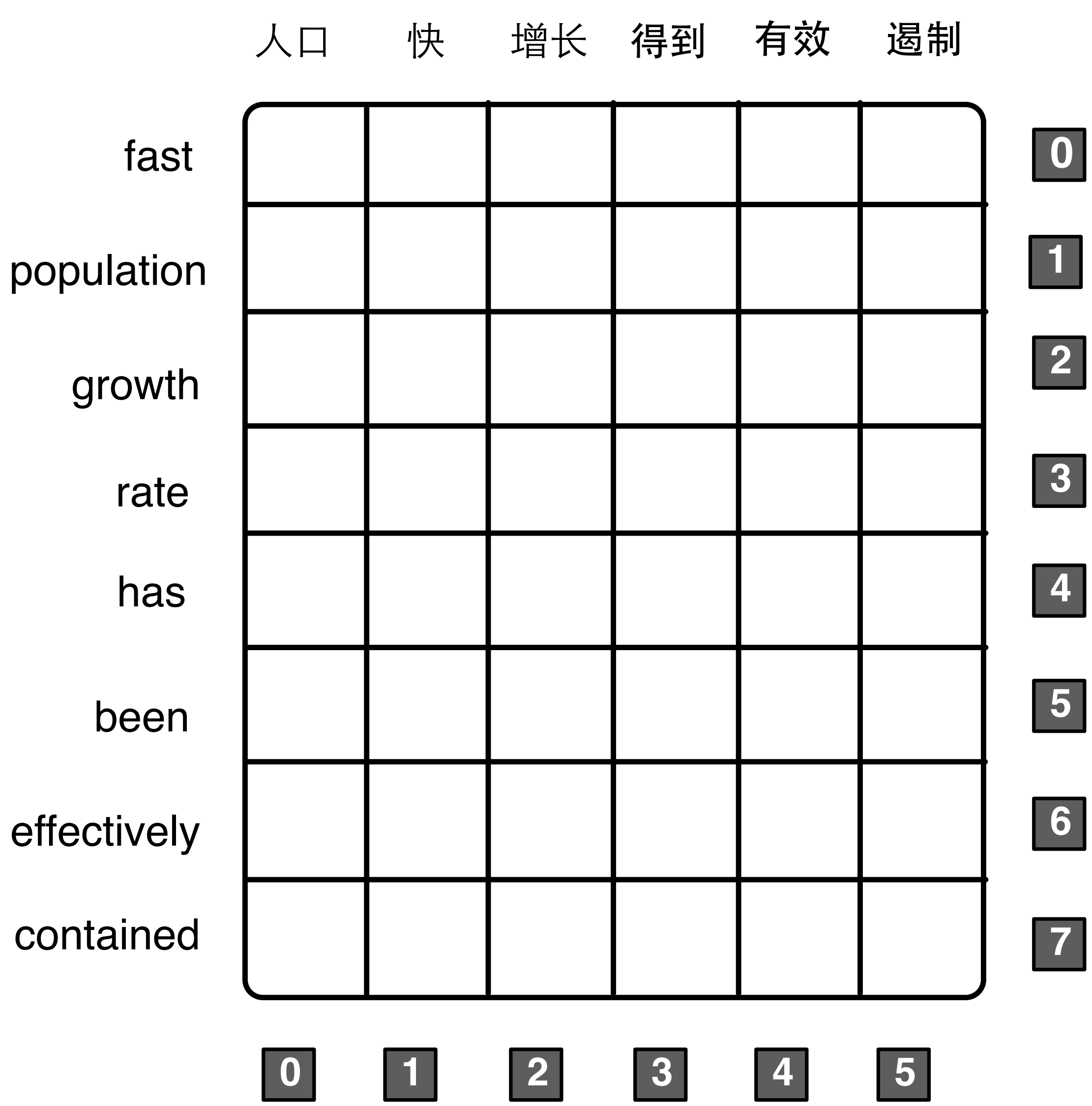


LEARNING A TRANSLATION MODEL

LEARNING A TRANSLATION MODEL

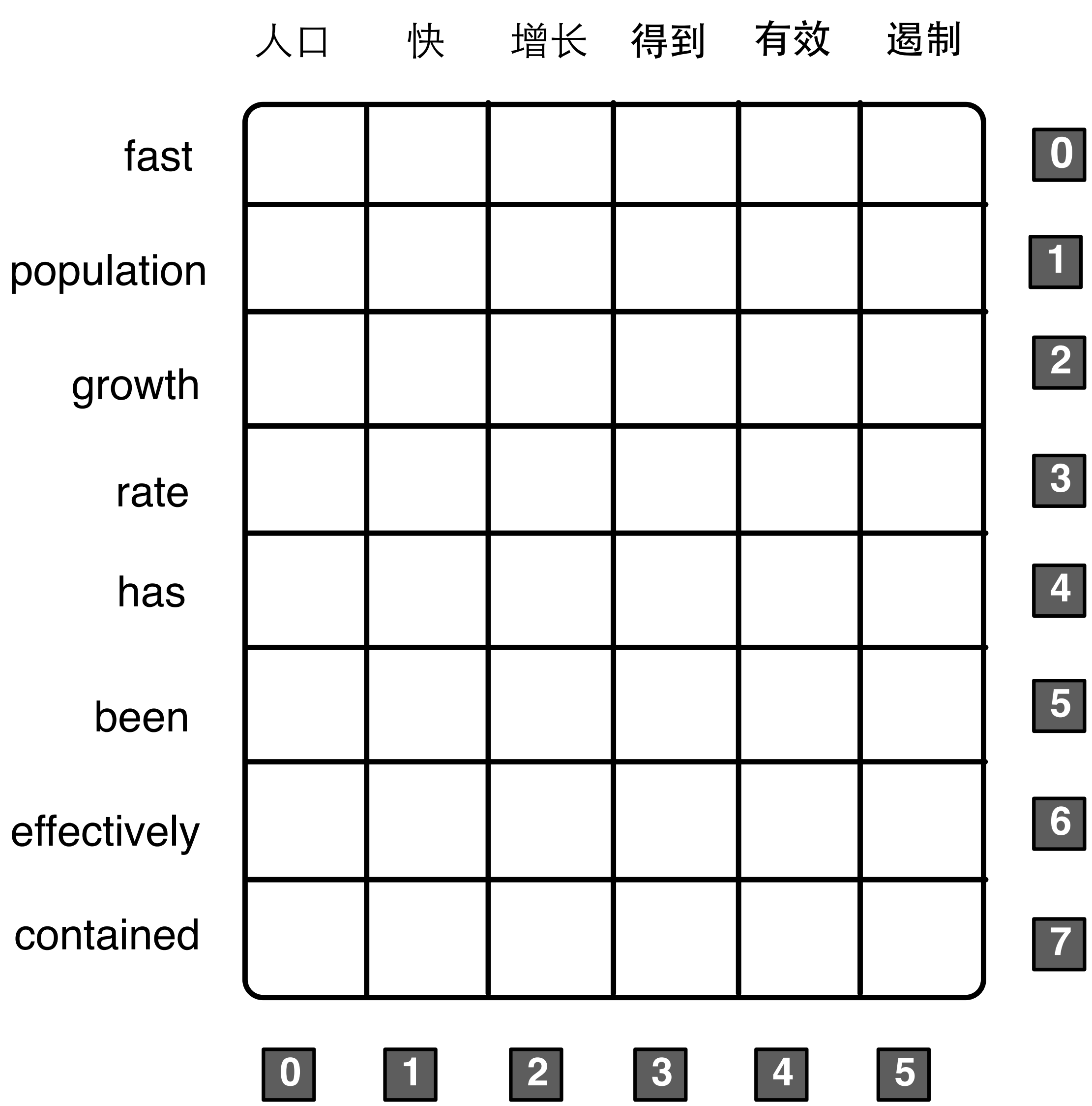
- Take each Chinese-English sentence pair in the bitext

LEARNING A TRANSLATION MODEL



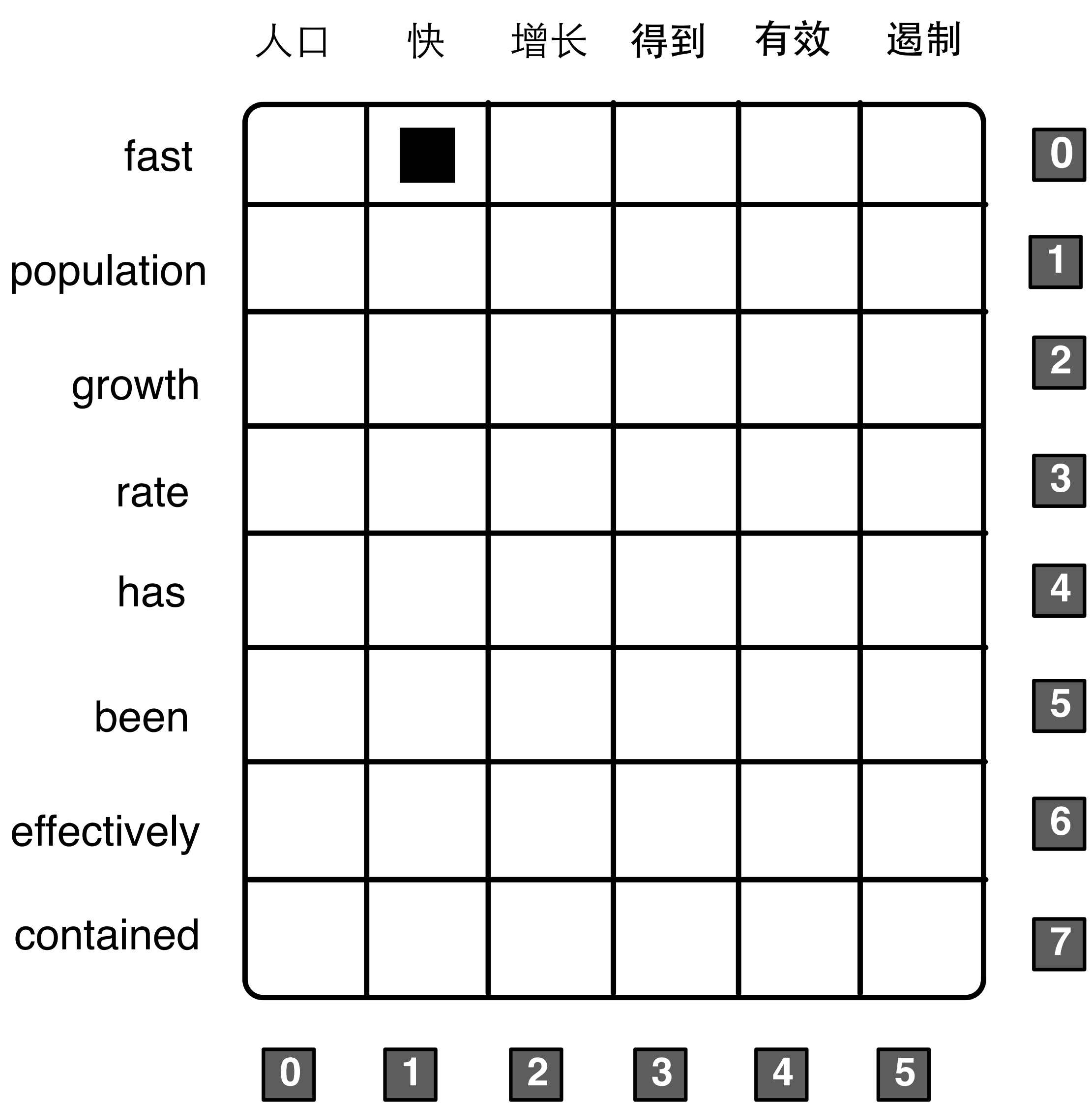
- Take each Chinese-English sentence pair in the bitext

LEARNING A TRANSLATION MODEL



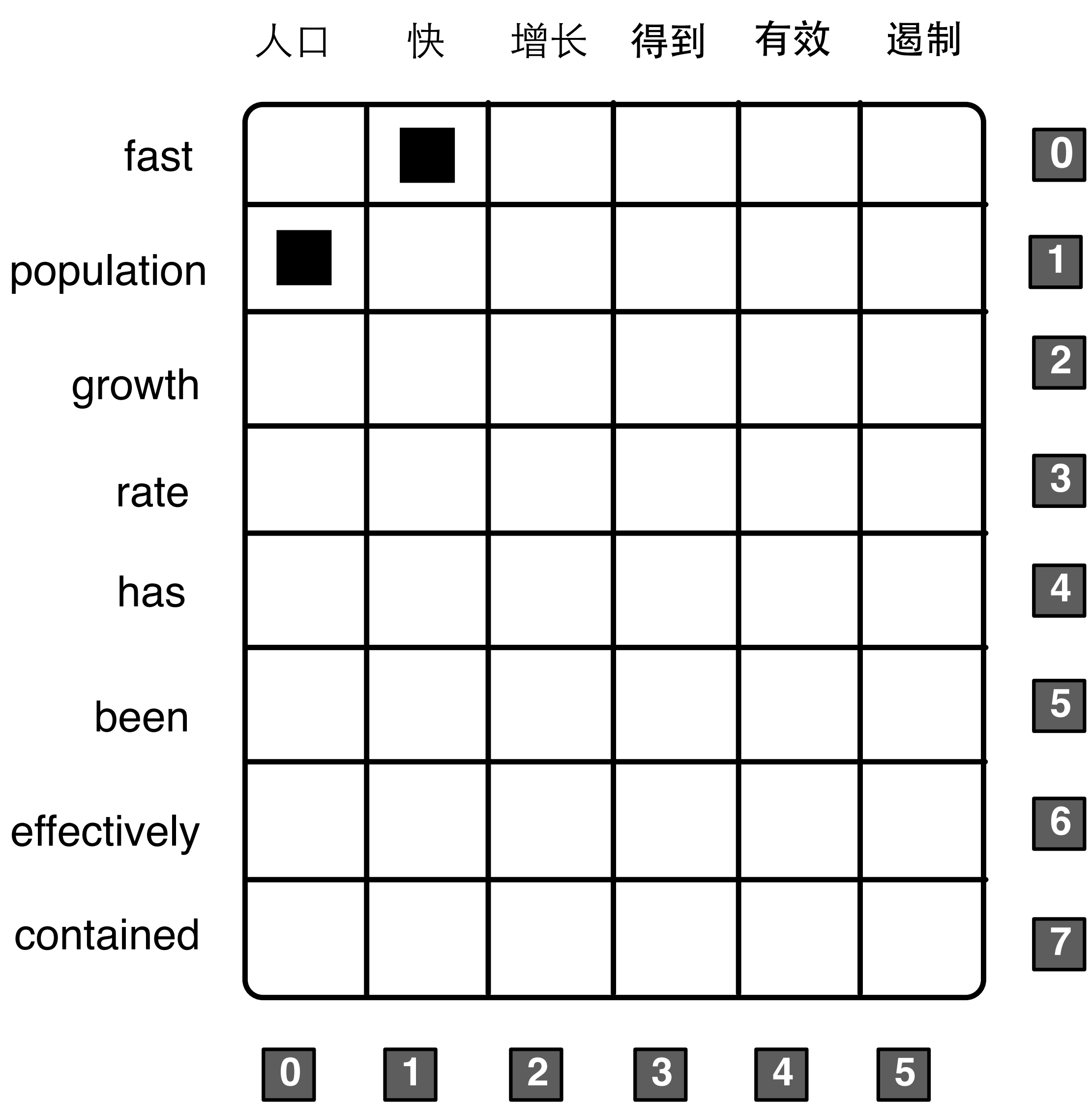
- Take each Chinese-English sentence pair in the bitext
- "Discover" what Chinese words correspond to what English words (*unsupervised* learning algorithm)

LEARNING A TRANSLATION MODEL



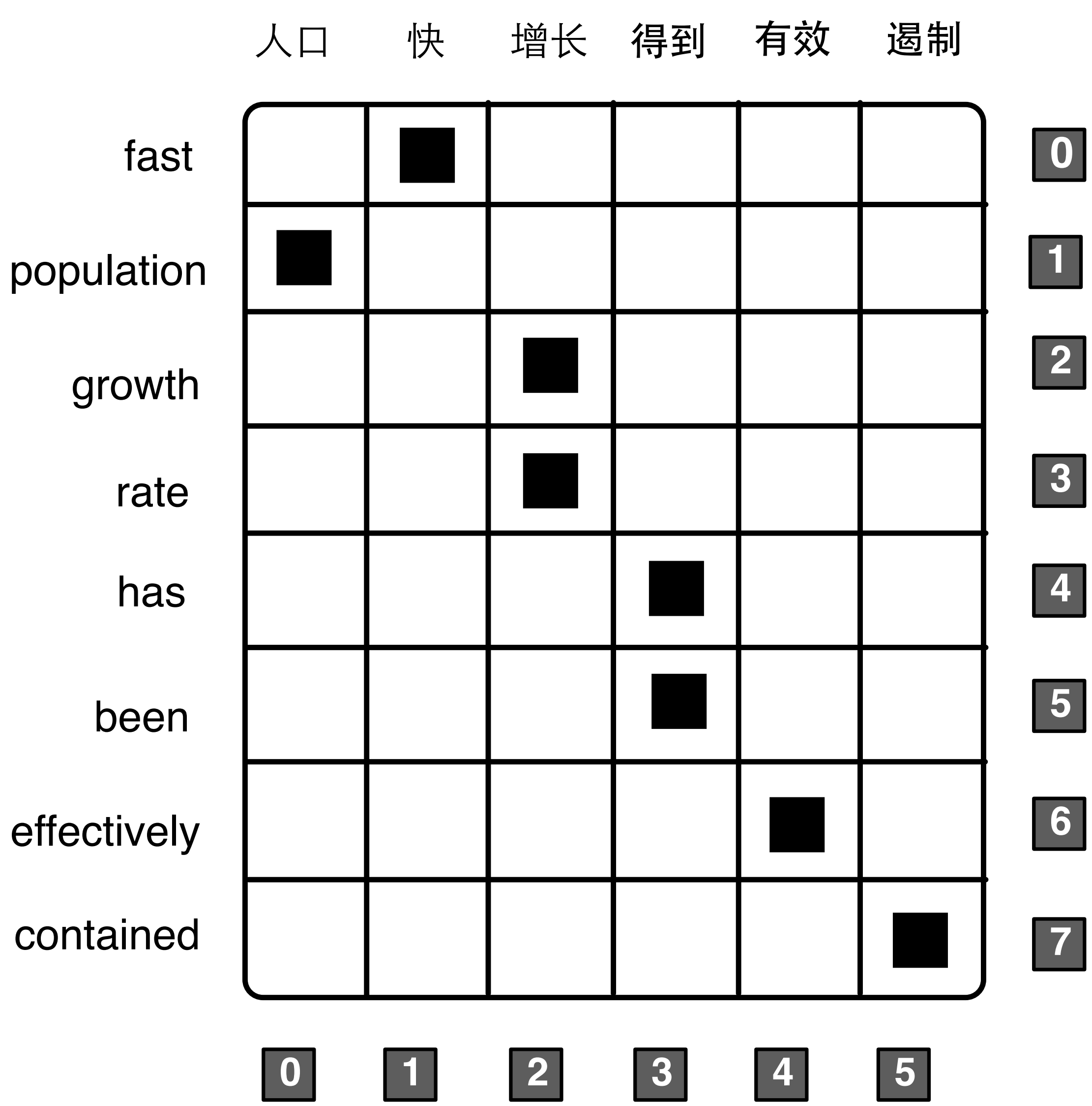
- Take each Chinese-English sentence pair in the bitext
- "Discover" what Chinese words correspond to what English words (*unsupervised* learning algorithm)

LEARNING A TRANSLATION MODEL



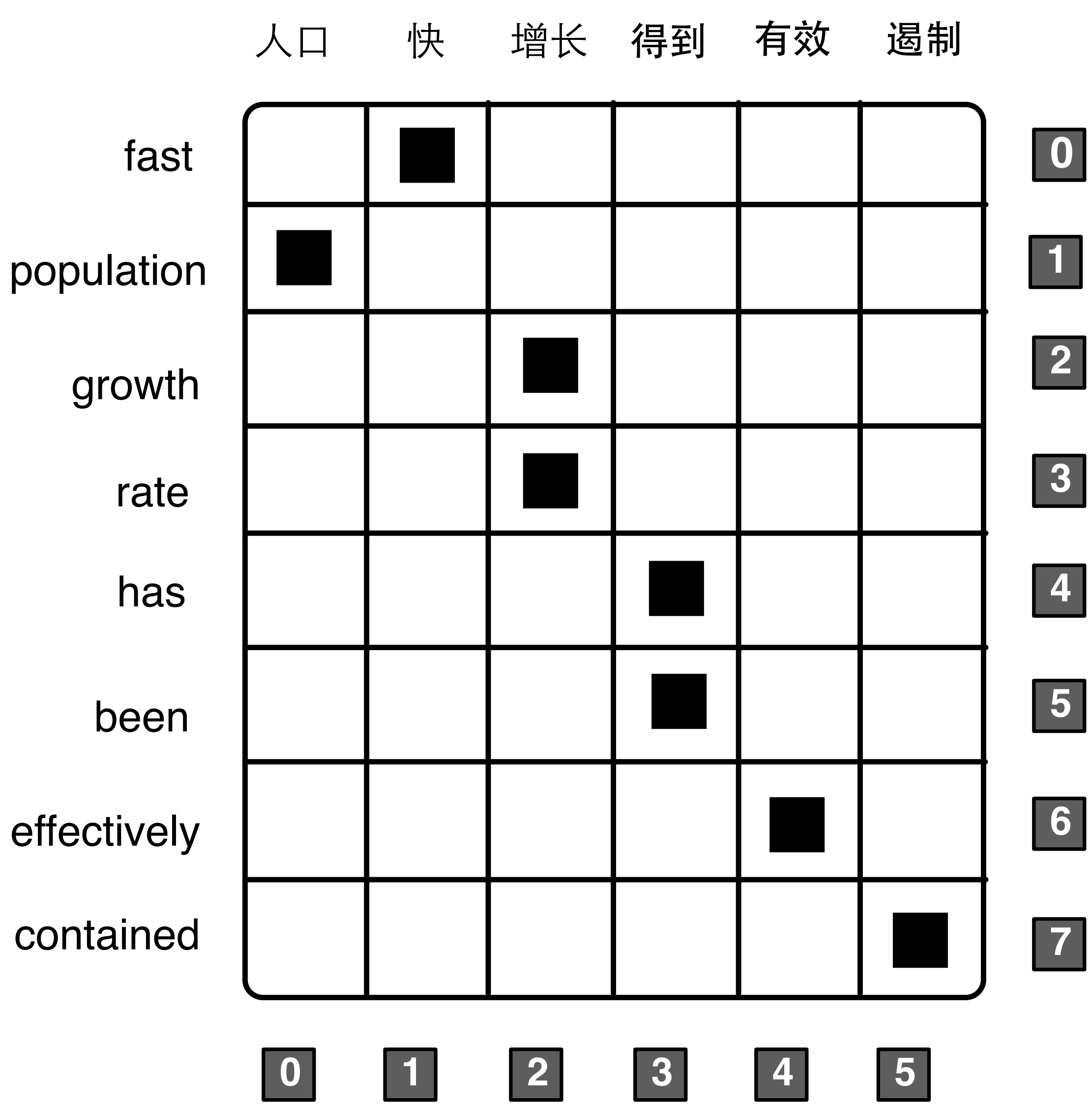
- Take each Chinese-English sentence pair in the bitext
- "Discover" what Chinese words correspond to what English words (*unsupervised* learning algorithm)

LEARNING A TRANSLATION MODEL



- Take each Chinese-English sentence pair in the bitext
- "Discover" what Chinese words correspond to what English words (*unsupervised* learning algorithm)

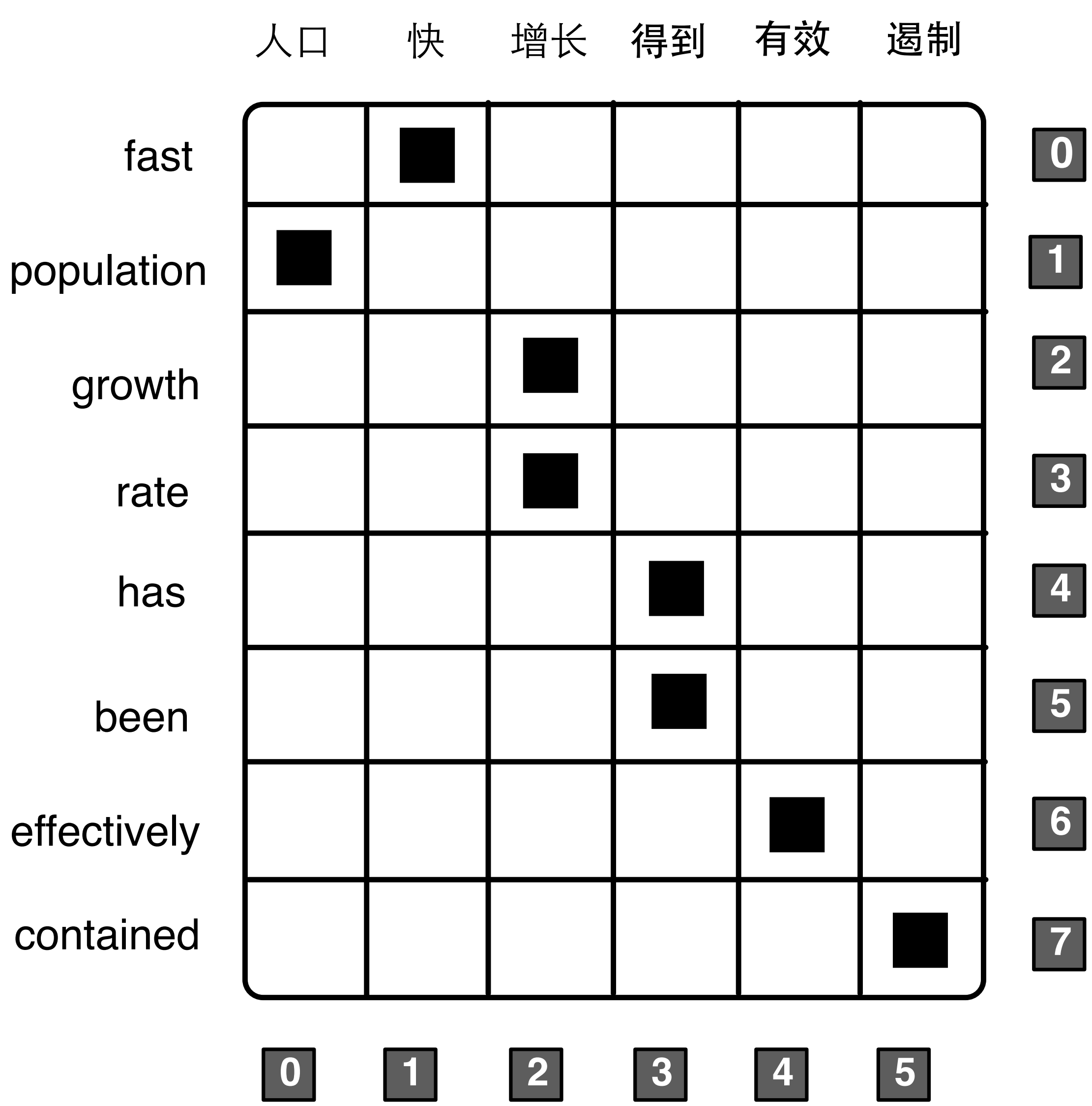
LEARNING A TRANSLATION MODEL



Alignment Matrix

- Take each Chinese-English sentence pair in the bitext
- "Discover" what Chinese words correspond to what English words (*unsupervised* learning algorithm)

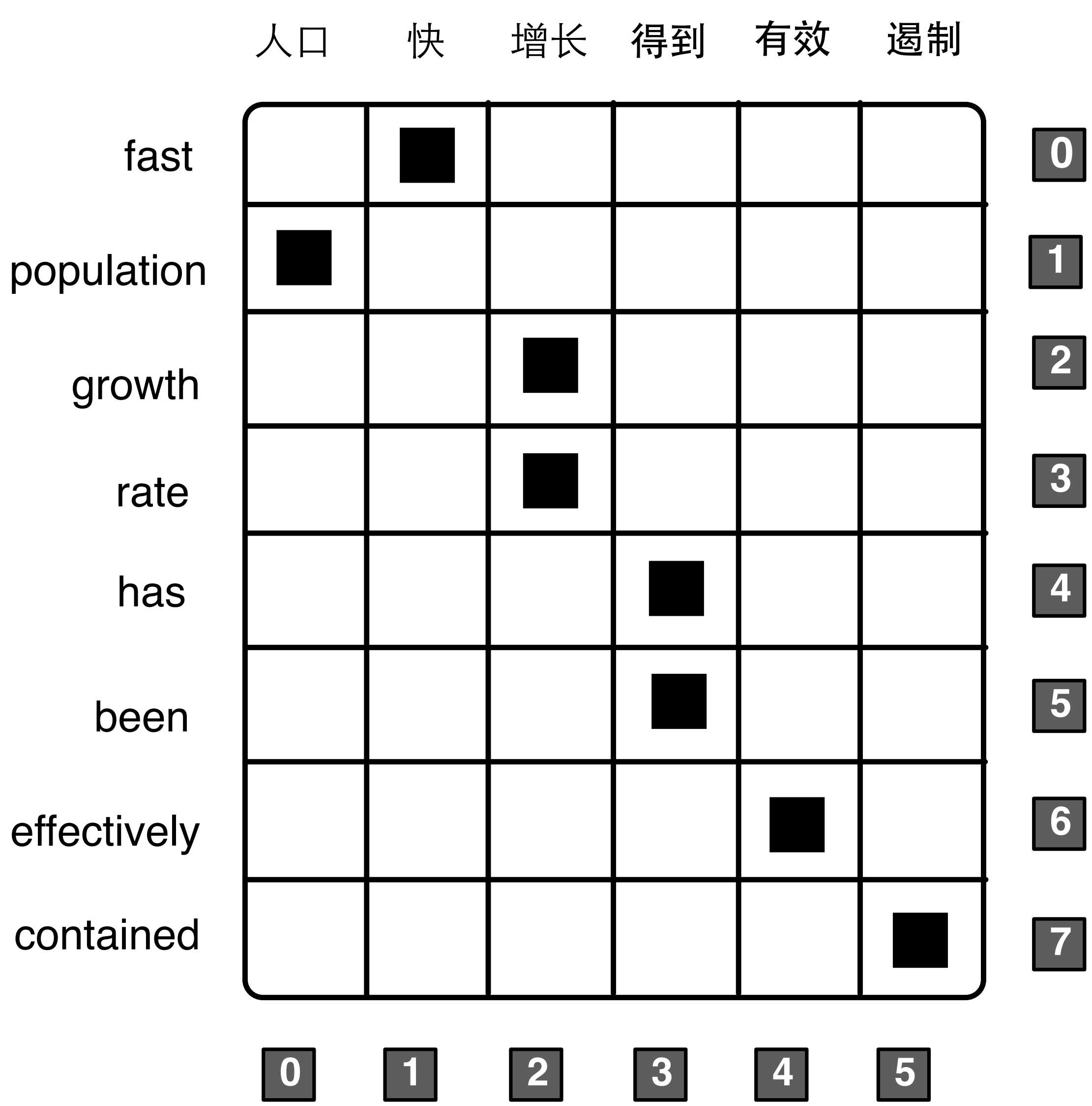
LEARNING A TRANSLATION MODEL



Alignment Matrix

- Take each Chinese-English sentence pair in the bitext
- "Discover" what Chinese words correspond to what English words (*unsupervised* learning algorithm)
- Now extract *phrasal* correspondences by drawing boxes around alignment points (each box should be self-contained)

LEARNING A TRANSLATION MODEL



Alignment Matrix

extracted bilingual phrase pairs

(0,0) x (1,1) ➔ <人口,population>
(1,1) x (0,0) ➔ <快,fast>
(2,2) x (2,3) ➔ <增长,growth rate>
...
...
(4,5) x (6,7) ➔ <有效 遏制,effectively contained>
...
...

LEARNING A TRANSLATION MODEL

LEARNING A TRANSLATION MODEL

- Compute *feature functions* $h(e_p, f_p)$ for each phrase pair $\langle e_p, f_p \rangle$

LEARNING A TRANSLATION MODEL

- Compute *feature functions* $h(e_p, f_p)$ for each phrase pair $<e_p, f_p>$
- Most features are computed via maximum likelihood estimation
- Examples:
  - How frequently was f_p extracted with e_p, relative to other e's?
  - How frequently was e_p extracted with f_p, relative to others f's?
  - How well do words in e_p align to those in f_p?
  - How well do words in f_p align to those in e_p?

LEARNING A TRANSLATION MODEL

LEARNING A TRANSLATION MODEL

- How to combine these various features (h_i) together into a probabilistic model?

LEARNING A TRANSLATION MODEL

- How to combine these various features (h_i) together into a probabilistic model?
- Use a discriminative model†

$$p(\mathbf{e}|\mathbf{f}) = \frac{\exp \sum_{k=1}^{N} \lambda_k h_k(\mathbf{e}, \mathbf{f})}{\sum_{e'} \exp \sum_{k=1}^{N} \lambda_k h_k(\mathbf{e}', \mathbf{f})}$$

  - Each λ_k is a weight for the corresponding feature h_k

†*Statistical Phrase-based Translation.* Philipp Koehn, Franz Josef Och, and Daniel Marcu. NAACL 2003

LEARNING A TRANSLATION MODEL

- How to combine these various features (h_i) together into a probabilistic model?
- Use a discriminative model†

$$p(\mathbf{e}|\mathbf{f}) = \frac{\exp \sum_{k=1}^{N} \lambda_k h_k(\mathbf{e}, \mathbf{f})}{\sum_{e'} \exp \sum_{k=1}^{N} \lambda_k h_k(\mathbf{e}', \mathbf{f})}$$

  - Each λ_k is a weight for the corresponding feature h_k
- This learned model represents the likelihood of generating TL sentence **e** given SL sentence **f**
- Now what?

†*Statistical Phrase-based Translation.* Philipp Koehn, Franz Josef Och, and Daniel Marcu. NAACL 2003

APPLYING A TRANSLATION MODEL

APPLYING A TRANSLATION MODEL

Math

APPLYING A TRANSLATION MODEL

Math

$$p(\mathbf{e}|\mathbf{f})$$

APPLYING A TRANSLATION MODEL

Math

$$\boxed{p(\mathbf{e}|\mathbf{f})}$$

Model Score

APPLYING A TRANSLATION MODEL

Math

$\mathbf{e}$ $\boxed{p(\mathbf{e}|\mathbf{f})}$

Model Score

APPLYING A TRANSLATION MODEL

Math

$\boxed{p(\mathbf{e}|\mathbf{f})}$

$\boxed{\mathbf{e}}$

All possible translations that can be constructed using the learned phrasal correspondences

Model Score

APPLYING A TRANSLATION MODEL

Math

$$\arg\max_{\boxed{\mathbf{e}}} \boxed{p(\mathbf{e}|\mathbf{f})}$$

All possible translations that can be constructed using the learned phrasal correspondences

Model Score

APPLYING A TRANSLATION MODEL

Math

Search operation

$$\underset{\mathbf{e}}{\arg\max}\; p(\mathbf{e}|\mathbf{f})$$

All possible translations that can be constructed using the learned phrasal correspondences

Model Score

APPLYING A TRANSLATION MODEL

Math

Search operation

$$\hat{\mathbf{e}} = \arg\max_{\mathbf{e}} p(\mathbf{e}|\mathbf{f})$$

All possible translations that can be constructed using the learned phrasal correspondences

Model Score

APPLYING A TRANSLATION MODEL

Math

"Best" TL translation for **f**

Search operation

$$\hat{\mathbf{e}} = \arg\max_{\mathbf{e}} p(\mathbf{e}|\mathbf{f})$$

All possible translations that can be constructed using the learned phrasal correspondences

Model Score

APPLYING A TRANSLATION MODEL

Math

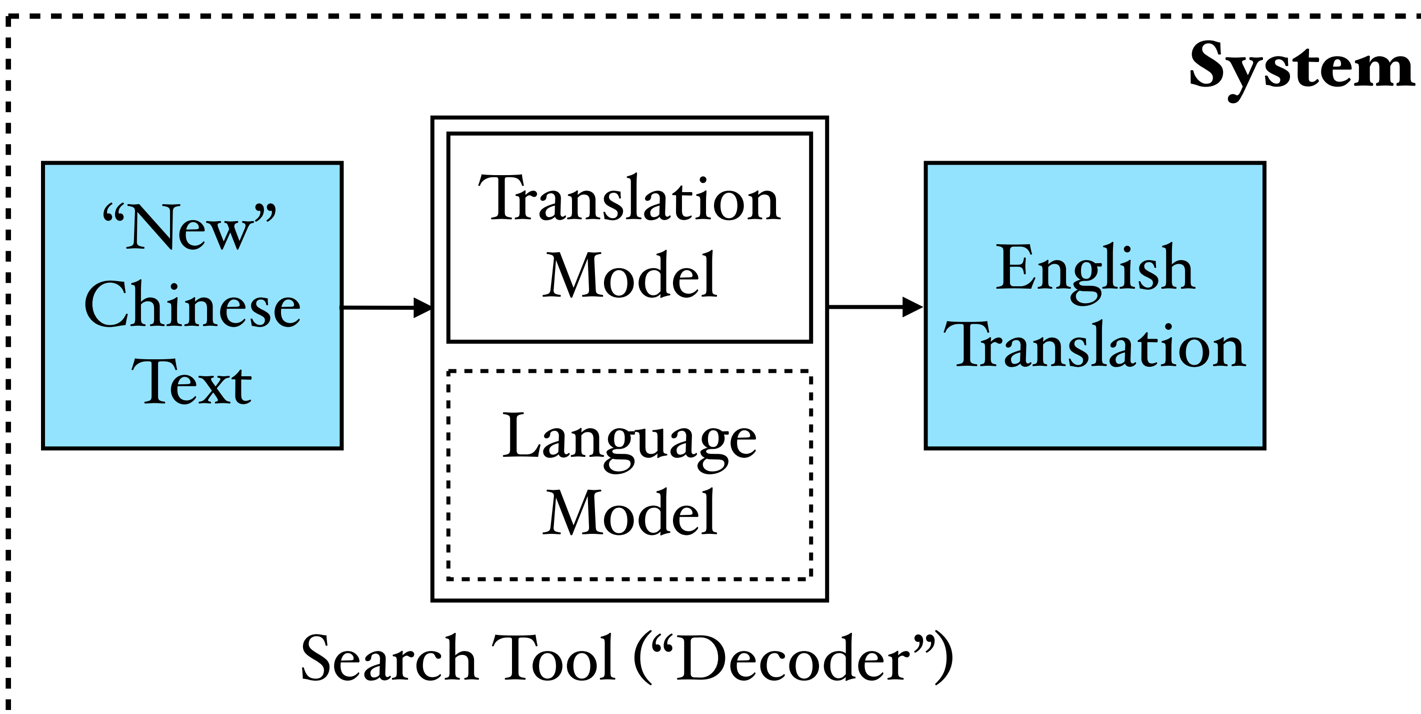
"Best" TL translation for **f**

Search operation

$$\hat{\mathbf{e}} = \arg\max_{\mathbf{e}} p(\mathbf{e}|\mathbf{f})$$

All possible translations that can be constructed using the learned phrasal correspondences

Model Score



APPLYING A TRANSLATION MODEL

Math

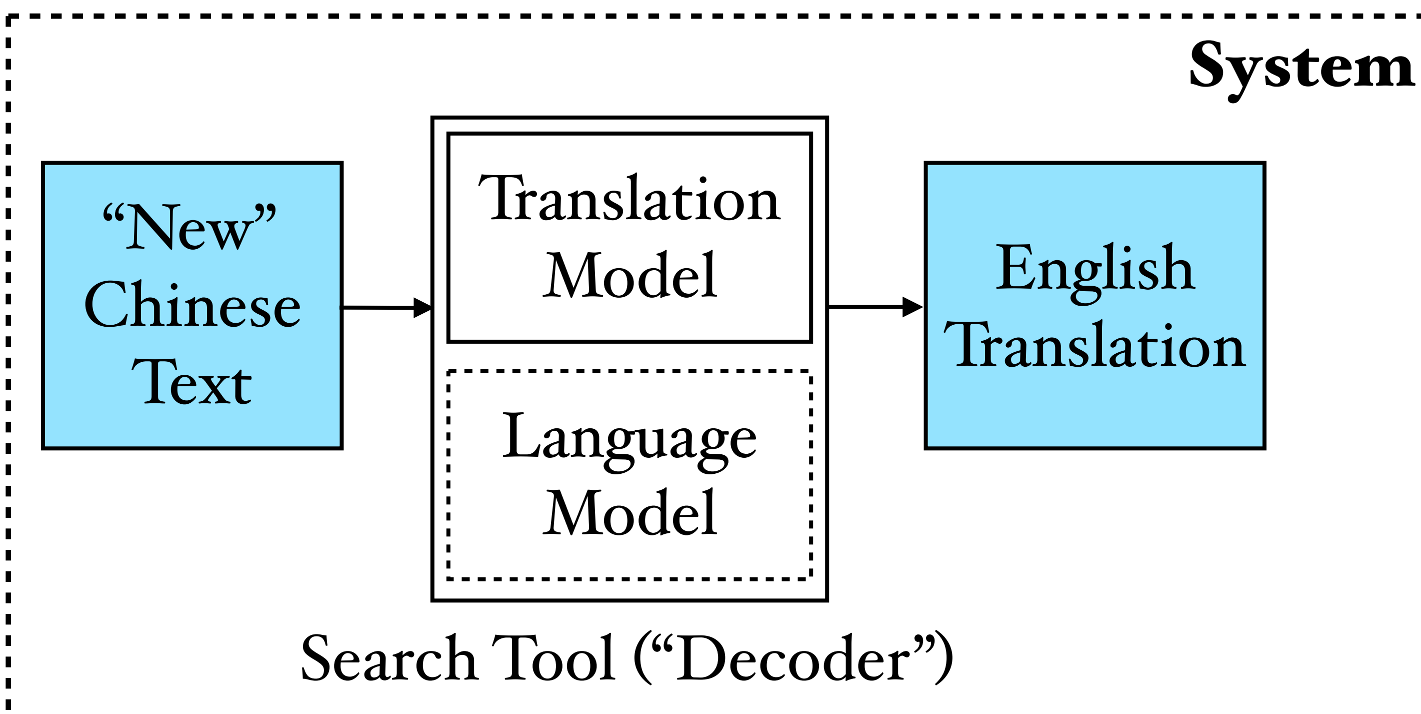
"Best" TL translation for **f**

Search operation

$$\hat{\mathbf{e}} = \arg\max_{\mathbf{e}} p(\mathbf{e}|\mathbf{f})$$

All possible translations that can be constructed using the learned phrasal correspondences

Model Score



- Search - "Decode" (Weaver thought of MT as "breaking a code")
- Brute-force decoding has been shown to be NP complete
- Writing an efficient decoder requires using heuristics e.g., beam search
- Phrasal reordering is a whole other problem
- Models/Decoders can both be imperfect (*model/search errors*)

NOT QUITE DONE YET ...

NOT QUITE DONE YET ...

- How do we tell that the SMT system is producing useful translations?

NOT QUITE DONE YET ...

- How do we tell that the SMT system is producing useful translations?
- Option 1: Ask bilingual Chinese-English speakers to rate the system output for adequacy and fluency
  - Informative but too slow to be useful as part of the system development cycle

NOT QUITE DONE YET ...

- How do we tell that the SMT system is producing useful translations?
- Option 1: Ask bilingual Chinese-English speakers to rate the system output for adequacy and fluency
  - Informative but too slow to be useful as part of the system development cycle
- Option 2: Test on datasets with already existing human-authored *reference translations;* use an **automated** metric to compare our system's translations to references

EVALUATING TRANSLATION

EVALUATING TRANSLATION

BLEU: MT metric that measures overlapping words sequences[†]

[†]*BLEU: A Method for Automatic Evaluation of Machine Translation*. Kishore Papineni et al. ACL 2002

EVALUATING TRANSLATION

BLEU: MT metric that measures overlapping words sequences†

`The issue of corruption has aroused strong resentment among the broad masses of people.`

System Output

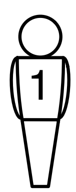
†*BLEU: A Method for Automatic Evaluation of Machine Translation*. Kishore Papineni et al. ACL 2002

EVALUATING TRANSLATION

BLEU: MT metric that measures overlapping words sequences†

`The issue of corruption has aroused strong resentment among the broad masses of people.`

System Output



`The problem of corruption has caused great dissatisfaction among the vast majority of people.`

Reference Translation

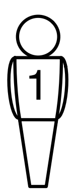
†*BLEU: A Method for Automatic Evaluation of Machine Translation*. Kishore Papineni et al. ACL 2002

EVALUATING TRANSLATION

BLEU: MT metric that measures overlapping words sequences†

The issue of corruption has aroused strong resentment among the broad masses of people.

System Output



The problem of corruption has caused great dissatisfaction among the vast majority of people.

Reference Translation

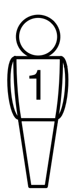
†*BLEU: A Method for Automatic Evaluation of Machine Translation*. Kishore Papineni et al. ACL 2002

EVALUATING TRANSLATION

BLEU: MT metric that measures overlapping words sequences†

The issue of corruption has aroused strong resentment among the broad masses of people.

System Output



The problem of corruption has caused great dissatisfaction among the vast majority of people.

Reference Translation

†*BLEU: A Method for Automatic Evaluation of Machine Translation*. Kishore Papineni et al. ACL 2002

EVALUATING TRANSLATION

BLEU: MT metric that measures overlapping words sequences†

The issue of corruption has aroused strong resentment among the broad masses of people.

System Output

The problem of corruption has caused great dissatisfaction among the vast majority of people.

Reference Translation

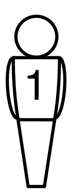
†*BLEU: A Method for Automatic Evaluation of Machine Translation*. Kishore Papineni et al. ACL 2002

EVALUATING TRANSLATION

BLEU: MT metric that measures overlapping words sequences†

The issue of corruption has aroused strong resentment among the broad masses of people.

System Output



The problem of corruption has caused great dissatisfaction among the vast majority of people.

Reference Translation

†*BLEU: A Method for Automatic Evaluation of Machine Translation*. Kishore Papineni et al. ACL 2002

EVALUATING TRANSLATION

BLEU: MT metric that measures overlapping words sequences†

The issue of corruption has aroused strong resentment among the broad masses of people.

System Output

1 *The problem of corruption has caused great dissatisfaction among the vast majority of people.*

2 *The issue of corruption has been causing immense dissatisfaction among the broad masses.*

3 *The issue of corruption has aroused great resentment among the vast majority of people.*

Reference Translation

†*BLEU: A Method for Automatic Evaluation of Machine Translation*. Kishore Papineni et al. ACL 2002

EVALUATING TRANSLATION

BLEU: MT metric that measures overlapping words sequences†

The issue of corruption has aroused strong resentment among the broad masses of people.

System Output

1 *The problem of corruption has caused great dissatisfaction among the vast majority of people.*

2 *The issue of corruption has been causing immense dissatisfaction among the broad masses.*

3 *The issue of corruption has aroused great resentment among the vast majority of people.*

Reference Translations

†*BLEU: A Method for Automatic Evaluation of Machine Translation*. Kishore Papineni et al. ACL 2002

EVALUATING TRANSLATION

BLEU: MT metric that measures overlapping words sequences†

The issue of corruption has aroused strong resentment among the broad masses of people.

System Output

1 *The problem of corruption has caused great dissatisfaction among the vast majority of people.*

2 *The issue of corruption has been causing immense dissatisfaction among the broad masses.*

3 *The issue of corruption has aroused great resentment among the vast majority of people.*

Reference Translations

†*BLEU: A Method for Automatic Evaluation of Machine Translation*. Kishore Papineni et al. ACL 2002

EVALUATING TRANSLATION

BLEU: MT metric that measures overlapping words sequences†

The issue of corruption has aroused strong resentment among the broad masses of people.

System Output

1

The problem of corruption has caused great dissatisfaction among the vast majority of people.

The issue of corruption has been causing immense dissatisfaction among the broad masses.

Too Expensive! Most datasets only have 1.

The issue of corruption has aroused great resentment among the vast majority of people.

Reference Translation

†*BLEU: A Method for Automatic Evaluation of Machine Translation*. Kishore Papineni et al. ACL 2002

THE SMT PIPELINE

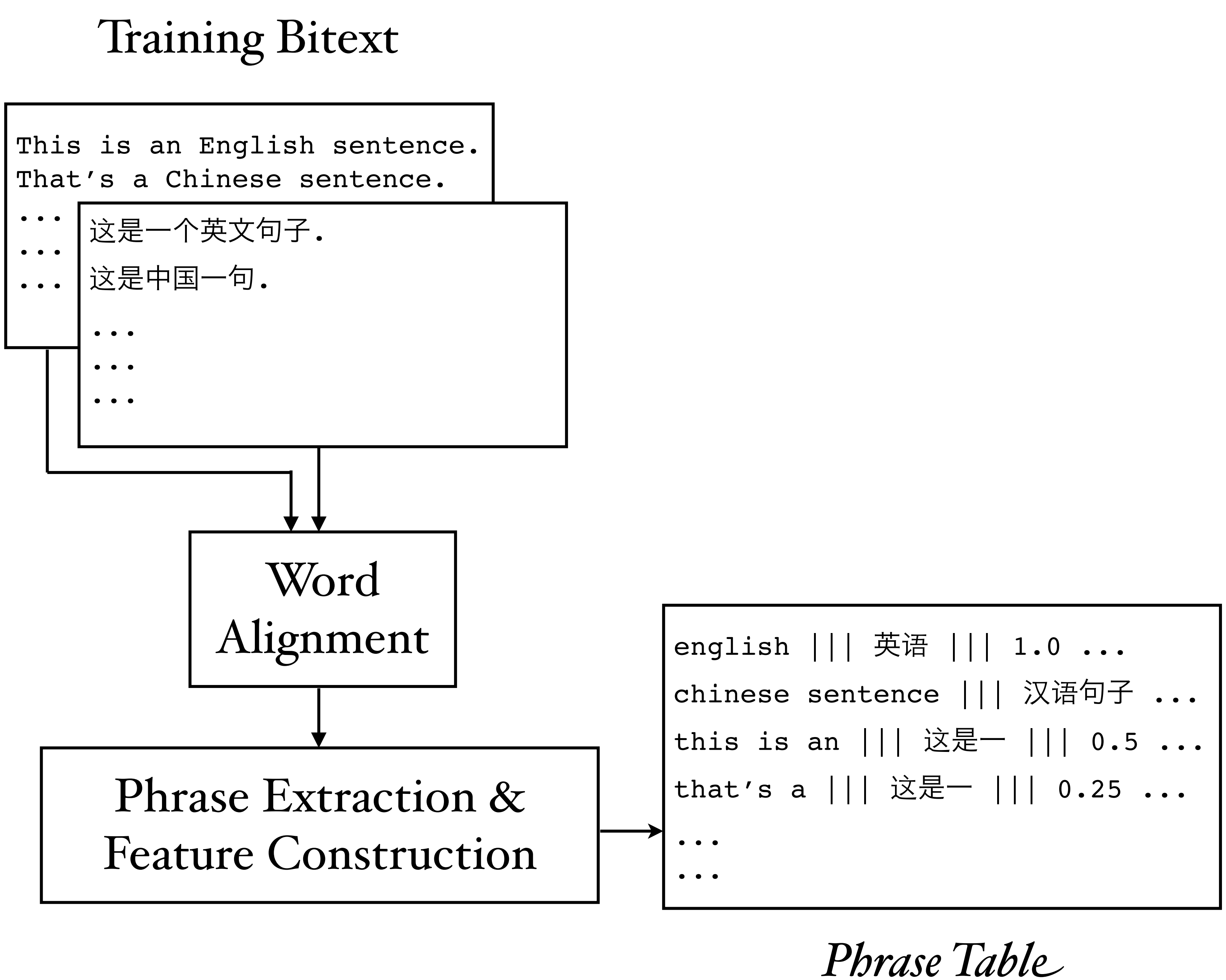
THE SMT PIPELINE

Training Bitext

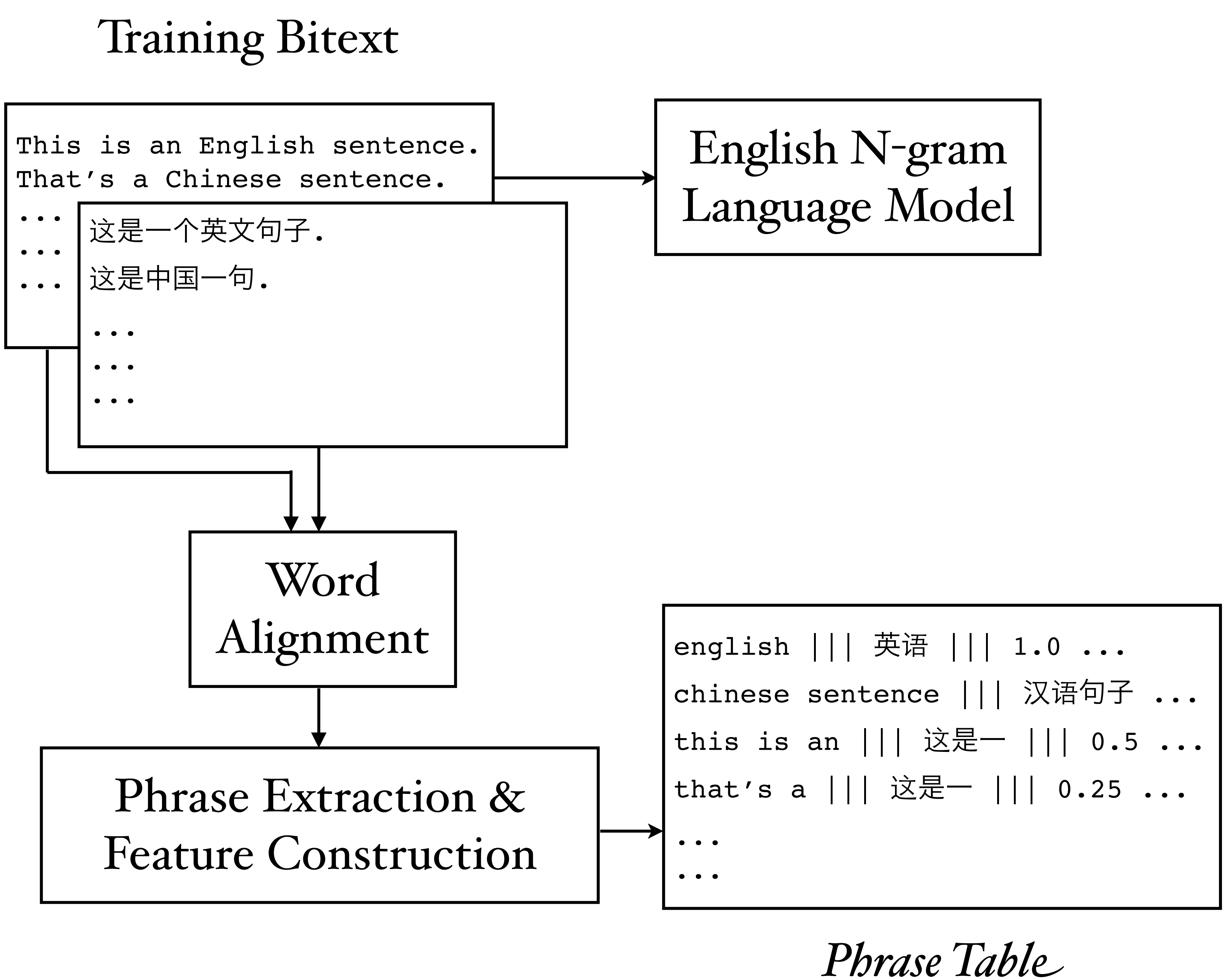
```
This is an English sentence.
That's a Chinese sentence.
...
...
...
```

```
这是一个英文句子.
这是中国一句.
...
...
...
```

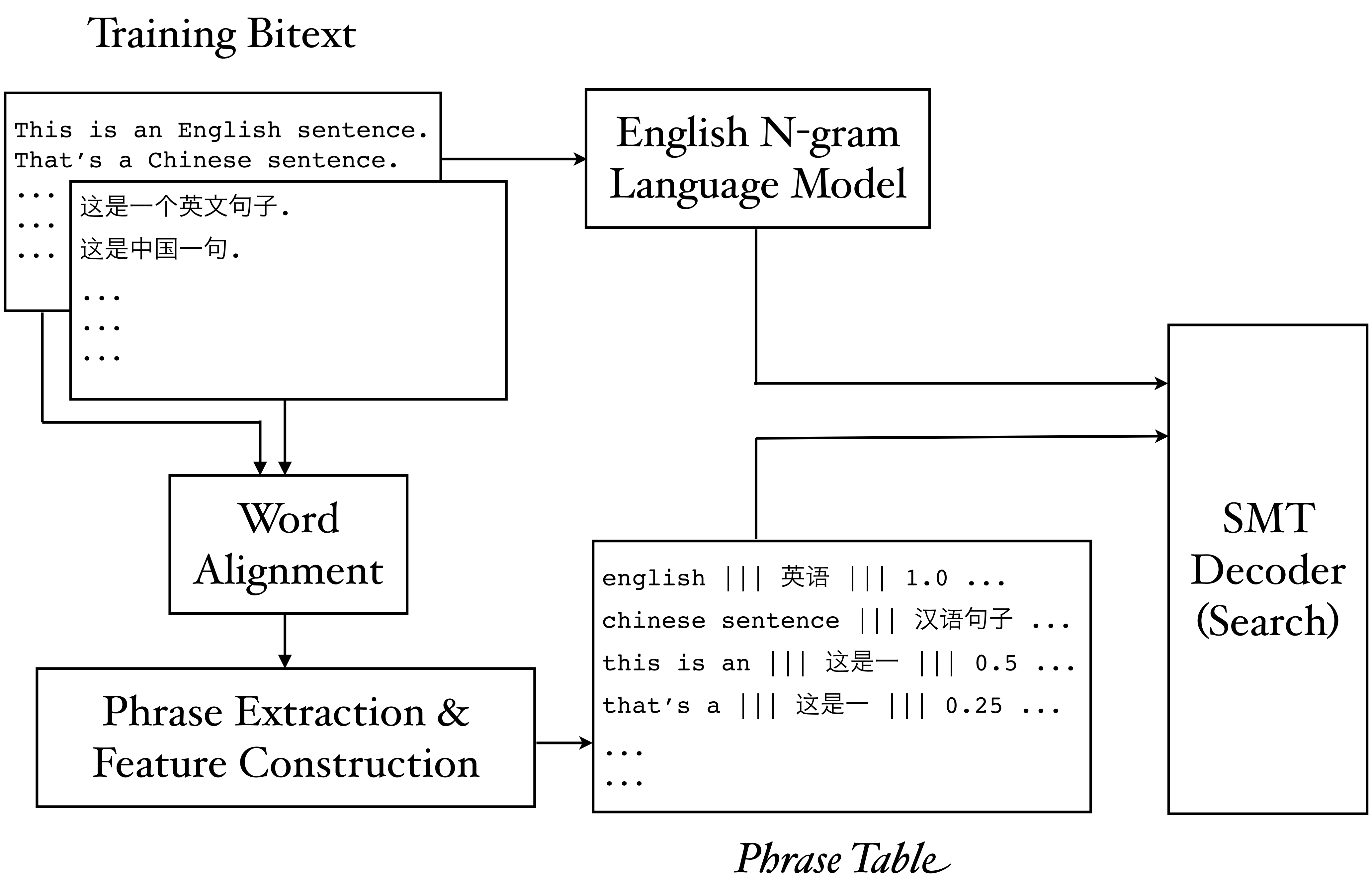
THE SMT PIPELINE



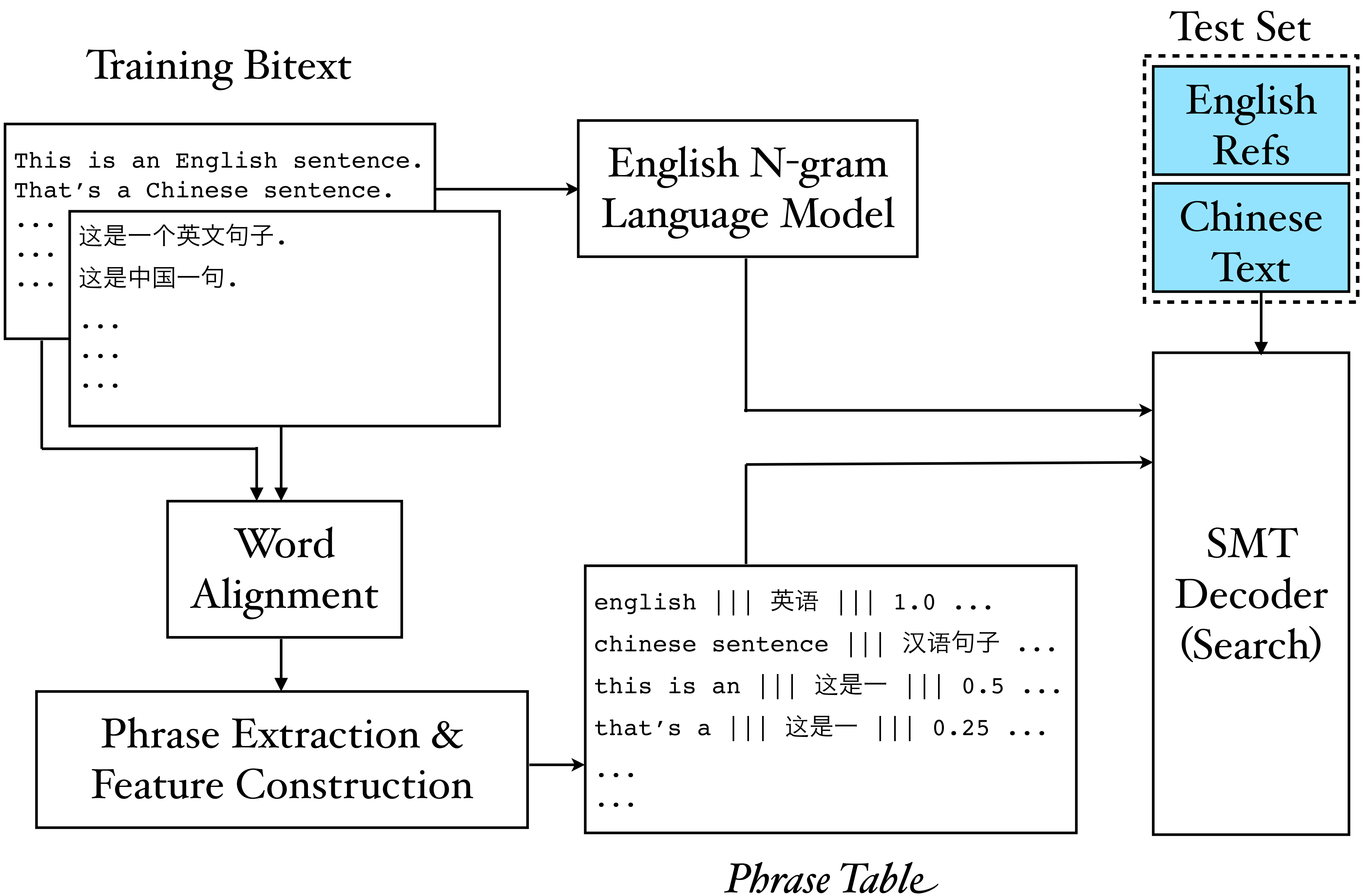
THE SMT PIPELINE



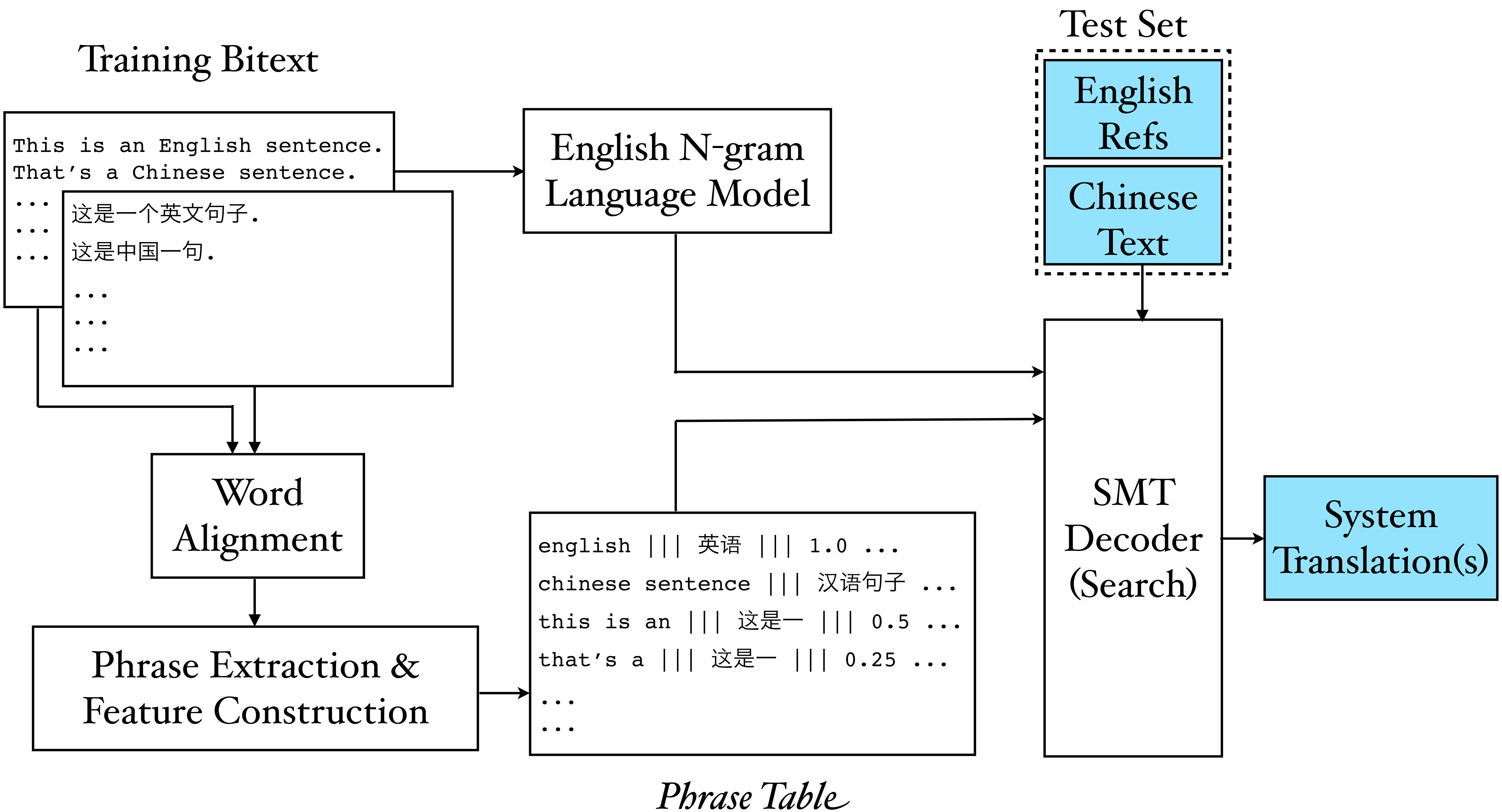
THE SMT PIPELINE



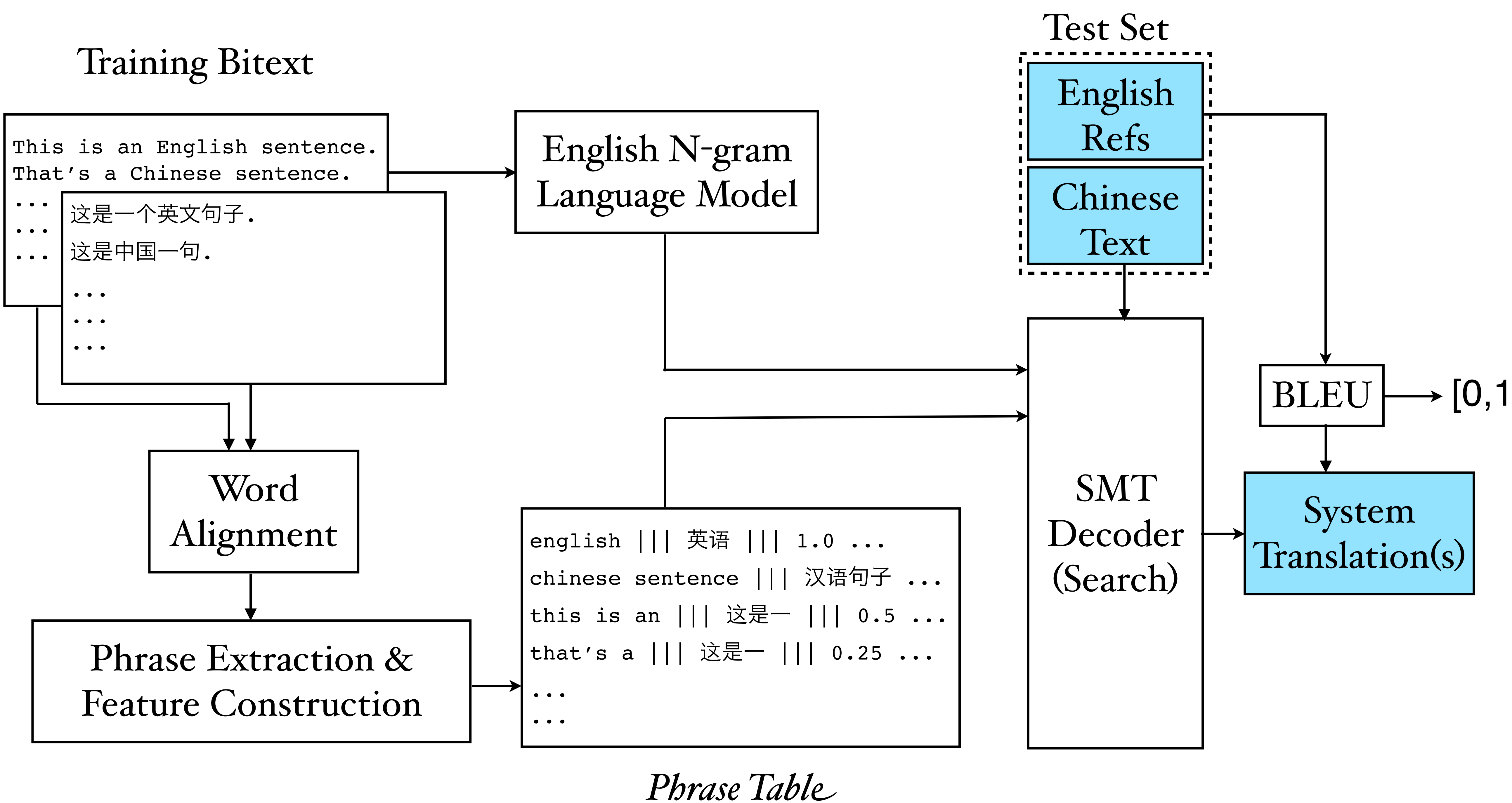
THE SMT PIPELINE



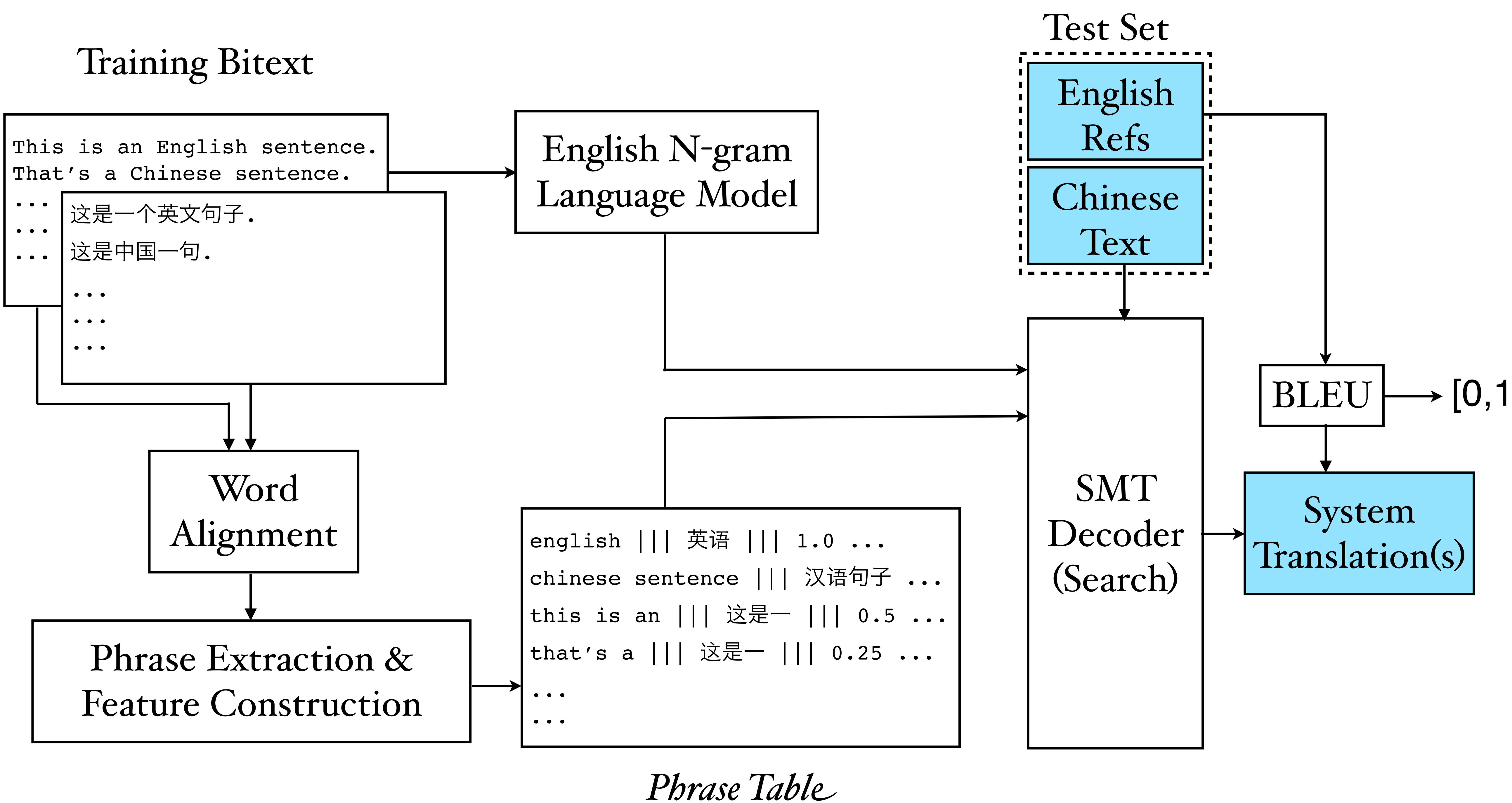
THE SMT PIPELINE



THE SMT PIPELINE



THE SMT PIPELINE



So, are we done?

PART II

THE MAGIC

PARAMETER TUNING

PARAMETER TUNING

- We still haven't talked about one very important step in the SMT pipeline

PARAMETER TUNING

- We still haven't talked about one very important step in the SMT pipeline

$$p(\mathbf{e}|\mathbf{f}) = \frac{\exp \sum_{k=1}^{N} \lambda_k h_k(\mathbf{e}, \mathbf{f})}{\sum_{e'} \exp \sum_{k=1}^{N} \lambda_k h_k(\mathbf{e}', \mathbf{f})}$$

PARAMETER TUNING

- We still haven't talked about one very important step in the SMT pipeline

$$p(\mathbf{e}|\mathbf{f}) = \frac{\exp \sum_{k=1}^{N} \overset{??}{\lambda_k} h_k(\mathbf{e}, \mathbf{f})}{\sum_{e'} \exp \sum_{k=1}^{N} \lambda_k h_k(\mathbf{e}', \mathbf{f})}$$

PARAMETER TUNING

- We still haven't talked about one very important step in the SMT pipeline

$$p(\mathbf{e}|\mathbf{f}) = \frac{\exp \sum_{k=1}^{N} \overset{??}{\lambda_k} h_k(\mathbf{e}, \mathbf{f})}{\sum_{e'} \exp \sum_{k=1}^{N} \lambda_k h_k(\mathbf{e}', \mathbf{f})}$$

PARAMETER TUNING

- We still haven't talked about one very important step in the SMT pipeline

$$p(\mathbf{e}|\mathbf{f}) = \frac{\exp \sum_{k=1}^{N} \overset{??}{\lambda_k} h_k(\mathbf{e},\mathbf{f})}{\sum_{e'} \exp \sum_{k=1}^{N} \lambda_k h_k(\mathbf{e}',\mathbf{f})}$$

- We need some held-out, development data (not training/test)

PARAMETER TUNING

- We still haven't talked about one very important step in the SMT pipeline

$$p(\mathbf{e}|\mathbf{f}) = \frac{\exp \sum_{k=1}^{N} \overset{??}{\lambda_k} h_k(\mathbf{e}, \mathbf{f})}{\sum_{e'} \exp \sum_{k=1}^{N} \lambda_k h_k(\mathbf{e}', \mathbf{f})}$$

- We need some held-out, development data (not training/test)

- Best estimates of parameters λ_k obtained by optimizing an objective related to translation quality (BLEU)

$$\lambda_1^k = \arg\max_{\hat{\lambda}_1^k} \sum_{(\mathbf{e},\mathbf{f})} \mathrm{BLEU}(\arg\max_{\mathbf{e}} p_{\hat{\lambda}}(\mathbf{e}|\mathbf{f}), \mathbf{e}_{\mathrm{ref}})$$

PARAMETER TUNING

- We still haven't talked about one very important step in the SMT pipeline

$$p(\mathbf{e}|\mathbf{f}) = \frac{\exp \sum_{k=1}^{N} \overset{??}{\lambda_k} h_k(\mathbf{e}, \mathbf{f})}{\sum_{e'} \exp \sum_{k=1}^{N} \lambda_k h_k(\mathbf{e}', \mathbf{f})}$$

- We need some held-out, development data (not training/test)

- Best estimates of parameters λ_k obtained by optimizing an objective related to translation quality (BLEU)

$$\lambda_1^k = \arg\max_{\hat{\lambda}_1^k} \sum_{(\mathbf{e},\mathbf{f})} \mathrm{BLEU}(\arg\max_{\mathbf{e}} p_{\hat{\lambda}}(\mathbf{e}|\mathbf{f}), \mathbf{e}_{\mathrm{ref}})$$

- The argmax inside BLEU( ) rules out gradient ascent

PARAMETER TUNING

PARAMETER TUNING

- The *log-linear* structure of our model allows us a way out

$$p(\mathbf{e}|\mathbf{f}) = \frac{\exp \sum_{k=1}^{N} \lambda_k h_k(\mathbf{e}, \mathbf{f})}{\sum_{e'} \exp \sum_{k=1}^{N} \lambda_k h_k(\mathbf{e}', \mathbf{f})}$$

PARAMETER TUNING

- The *log-linear* structure of our model allows us a way out

$$p(\mathbf{e}|\mathbf{f}) = \frac{\exp \sum_{k=1}^{N} \lambda_k h_k(\mathbf{e}, \mathbf{f})}{\sum_{e'} \exp \sum_{k=1}^{N} \lambda_k h_k(\mathbf{e}', \mathbf{f})}$$

- Notice that (denominator is a normalization constant)

$$\log p(\mathbf{e}|\mathbf{f}) \propto \sum_{k=1}^{N} \lambda_k h_k(\mathbf{e}, \mathbf{f})$$

PARAMETER TUNING

- The *log-linear* structure of our model allows us a way out

$$p(\mathbf{e}|\mathbf{f}) = \frac{\exp \sum_{k=1}^{N} \lambda_k h_k(\mathbf{e}, \mathbf{f})}{\sum_{e'} \exp \sum_{k=1}^{N} \lambda_k h_k(\mathbf{e}', \mathbf{f})}$$

- Notice that (denominator is a normalization constant)

$$\log p(\mathbf{e}|\mathbf{f}) \propto \sum_{k=1}^{N} \lambda_k h_k(\mathbf{e}, \mathbf{f})$$

- If we hold all λs except one constant

$$\lambda_k h_k(\mathbf{e}, \mathbf{f}) + \sum_{k' \neq k}^{N} \lambda_{k'} h_{k'}(\mathbf{e}, \mathbf{f}) \quad [\text{y = mx + C}]$$

PARAMETER TUNING

- The *log-linear* structure of our model allows us a way out

$$p(\mathbf{e}|\mathbf{f}) = \frac{\exp \sum_{k=1}^{N} \lambda_k h_k(\mathbf{e}, \mathbf{f})}{\sum_{e'} \exp \sum_{k=1}^{N} \lambda_k h_k(\mathbf{e}', \mathbf{f})}$$

- Notice that (denominator is a normalization constant)

$$\log p(\mathbf{e}|\mathbf{f}) \propto \sum_{k=1}^{N} \lambda_k h_k(\mathbf{e}, \mathbf{f})$$

- If we hold all λs except one constant

$$\lambda_k h_k(\mathbf{e}, \mathbf{f}) + \sum_{k' \neq k}^{N} \lambda_{k'} h_{k'}(\mathbf{e}, \mathbf{f}) \quad [y = mx + C]$$

- **Solution**: Use a variant of a line maximization algorithm

PARAMETER TUNING

PARAMETER TUNING

- Maximum BLEU Training Algorithm

PARAMETER TUNING

- Maximum BLEU Training Algorithm

```
Repeat
  - Initialize λ1..K
  - Generate 19 additional random values for λ1..K to avoid
    running into local maxima
  - Optimize each λ using line maximization, holding others
    constant
  - Values of λ1..K yielding greatest BLEU increase used as
    initial values for next iteration
Until no change in values of λ1..K
```

† *Minimum Error Rate Training in Statistical Machine Translation*. Franz Josef Och. ACL 2003.

PARAMETER TUNING

- Maximum BLEU Training Algorithm

```
Repeat
  - Initialize λ1..K
  - Generate 19 additional random values for λ1..K to avoid
    running into local maxima
  - Optimize each λ using line maximization, holding others
    constant
  - Values of λ1..K yielding greatest BLEU increase used as
    initial values for next iteration
Until no change in values of λ1..K
```

- Intelligently explore large multi-dimensional parameter space via translation quality feedback (BLEU) against reference translations

†*Minimum Error Rate Training in Statistical Machine Translation*. Franz Josef Och. ACL 2003.

PARAMETER TUNING

- Maximum BLEU Training Algorithm

```
Repeat
  - Initialize λ1..K
  - Generate 19 additional random values for λ1..K to avoid
    running into local maxima
  - Optimize each λ using line maximization, holding others
    constant
  - Values of λ1..K yielding greatest BLEU increase used as
    initial values for next iteration
Until no change in values of λ1..K
```

- Intelligently explore large multi-dimensional parameter space via translation quality feedback (BLEU) against reference translations

- Exploration is most useful when feedback is fair.

† *Minimum Error Rate Training in Statistical Machine Translation*. Franz Josef Och. ACL 2003.

PARAMETER TUNING

- Maximum BLEU Training Algorithm

```
Repeat
  - Initialize λ1..K
  - Generate 19 additional random values for λ1..K to avoid
    running into local maxima
  - Optimize each λ using line maximization, holding others
    constant
  - Values of λ1..K yielding greatest BLEU increase used as
    initial values for next iteration
Until no change in values of λ1..K
```

- Intelligently explore large multi-dimensional parameter space via translation quality feedback (BLEU) against reference translations
- Exploration is most useful when feedback is fair.
- What makes BLEU fair?

†*Minimum Error Rate Training in Statistical Machine Translation*. Franz Josef Och. ACL 2003.

PARAMETER TUNING

- Maximum BLEU Training Algorithm

```
Repeat
  - Initialize λ1..K
  - Generate 19 additional random values for λ1..K to avoid
    running into local maxima
  - Optimize each λ using line maximization, holding others
    constant
  - Values of λ1..K yielding greatest BLEU increase used as
    initial values for next iteration
Until no change in values of λ1..K
```

- Intelligently explore large multi-dimensional parameter space via translation quality feedback (BLEU) against reference translations

- Exploration is most useful when feedback is fair.

- What makes BLEU fair? Multiple (**Expensive**) Reference Translations.

†*Minimum Error Rate Training in Statistical Machine Translation*. Franz Josef Och. ACL 2003.

PART III

THE BOOTSTRAP

BITEXT TO THE RESCUE ...

BITEXT TO THE RESCUE …

- **Problem**: Given one human reference translation, can we *automatically* manufacture *more* that *mean the same thing*?

BITEXT TO THE RESCUE ...

- **Problem**: Given one human reference translation, can we *automatically* manufacture *more* that *mean the same thing*?
- Monolingual semantic knowledge has been shown to be latent in bitext†
- Can we exploit the *supposed* bilingual semantic adequacy?

†*Exploiting Hidden Meanings: Using Bilingual Text for Monolingual Annotation*. Philip Resnik. LNCS 2945 (2004)

BITEXT TO THE RESCUE ...

- **Problem**: Given one human reference translation, can we *automatically* manufacture *more* that *mean the same thing*?
- Monolingual semantic knowledge has been shown to be latent in bitext†
- Can we exploit the *supposed* bilingual semantic adequacy?
- "If a Chinese phrase C can translate into English as both E1 and E2, shouldn't E1 and E2 have the same meaning?"

†*Exploiting Hidden Meanings: Using Bilingual Text for Monolingual Annotation*. Philip Resnik. LNCS 2945 (2004)

BITEXT TO THE RESCUE ...

- **Problem**: Given one human reference translation, can we *automatically* manufacture *more* that *mean the same thing*?
- Monolingual semantic knowledge has been shown to be latent in bitext†
- Can we exploit the *supposed* bilingual semantic adequacy?
- "If a Chinese phrase C can translate into English as both E1 and E2, shouldn't E1 and E2 have the same meaning?"
- Theory aside, is there any empirical evidence that this works?

†*Exploiting Hidden Meanings: Using Bilingual Text for Monolingual Annotation*. Philip Resnik. LNCS 2945 (2004)

PRELIMINARY EVIDENCE

PRELIMINARY EVIDENCE

- Find all pairs of English phrases that have been extracted with the same Chinese phrase and posit them as *paraphrases* of each other†

†*Paraphrasing with Bilingual Parallel Corpora*. Colin Bannard & Chris Callison-Burch. ACL 2005.

PRELIMINARY EVIDENCE

- Find all pairs of English phrases that have been extracted with the same Chinese phrase and posit them as *paraphrases* of each other†

部長建大橋 ⇒ minister to build bridge

部長建大橋 ⇒ minister to construct overpass

minister to build bridge ⇒ minister to construct overpass

†*Paraphrasing with Bilingual Parallel Corpora*. Colin Bannard & Chris Callison-Burch. ACL 2005.

PRELIMINARY EVIDENCE

- Find all pairs of English phrases that have been extracted with the same Chinese phrase and posit them as *paraphrases* of each other[†]

部長建大橋 ⇒ minister to build bridge
部長建大橋 ⇒ minister to construct overpass

minister to build bridge ⇒ minister to construct overpass

總督建市 ⇒ governor to establish city
總督建市 ⇒ president to establish town

governor to establish city ⇒ president to establish town

[†] *Paraphrasing with Bilingual Parallel Corpora*. Colin Bannard & Chris Callison-Burch. ACL 2005.

PRELIMINARY EVIDENCE

- Find all pairs of English phrases that have been extracted with the same Chinese phrase and posit them as *paraphrases* of each other[†]

- Most *pivoted* paraphrase pairs found to be approximately paraphrastic

部長建大橋 ⇒ minister to build bridge
部長建大橋 ⇒ minister to construct overpass

minister to build bridge ⇒ minister to construct overpass

總督建市 ⇒ governor to establish city
總督建市 ⇒ president to establish town

governor to establish city ⇒ president to establish town

[†]*Paraphrasing with Bilingual Parallel Corpora*. Colin Bannard & Chris Callison-Burch. ACL 2005.

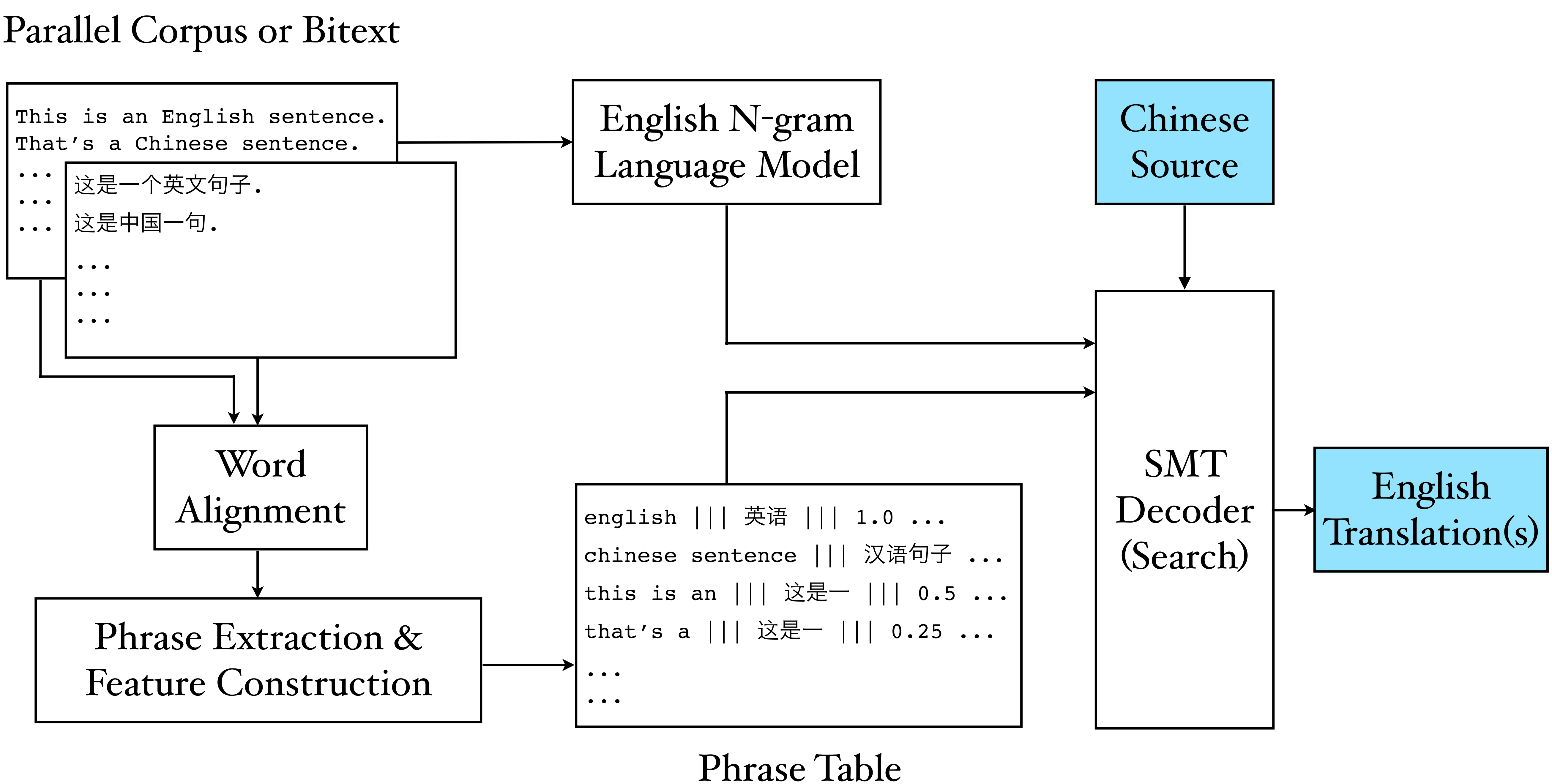
WHAT ABOUT SENTENCES?

WHAT ABOUT SENTENCES?

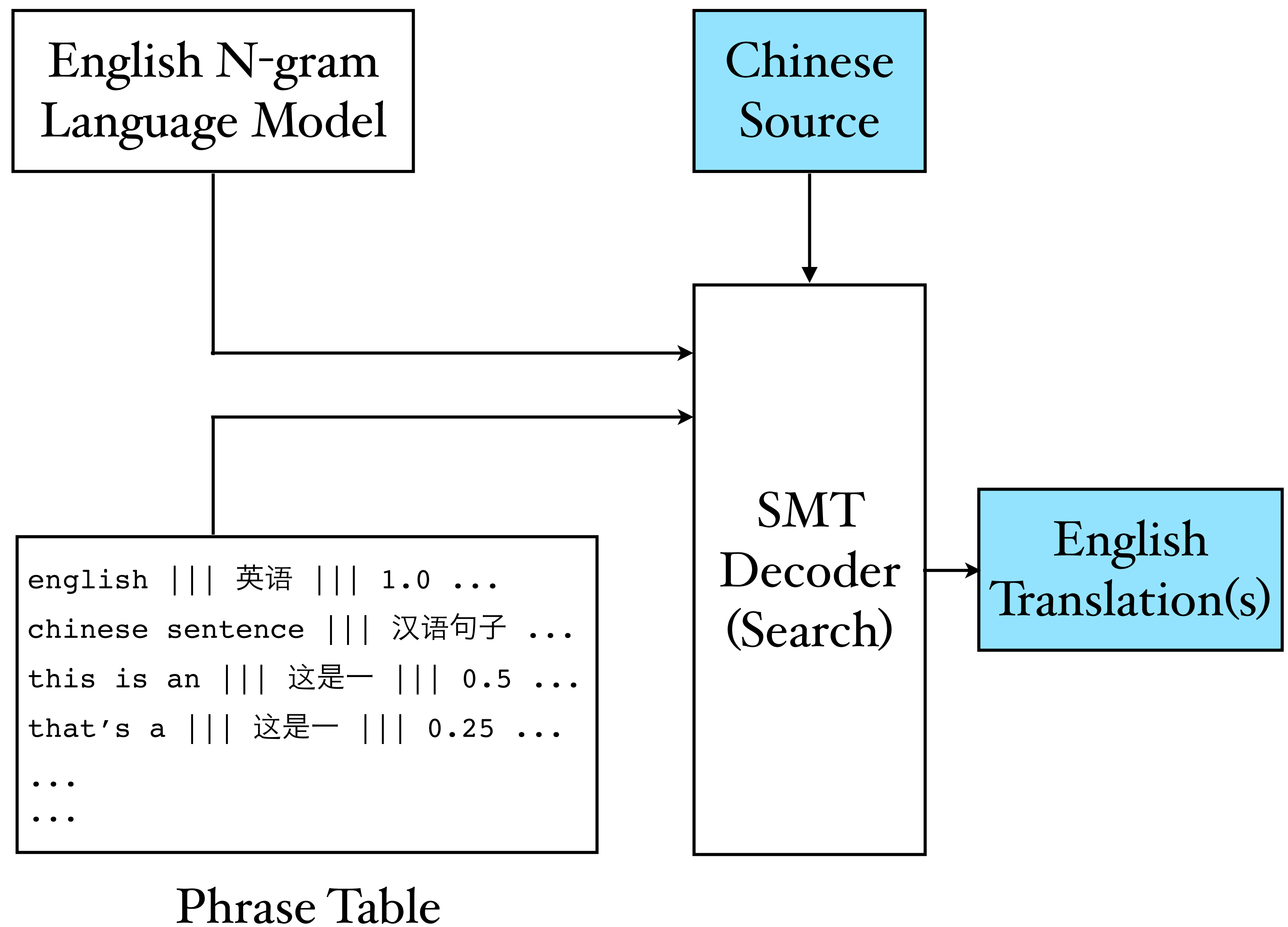
- Treat pivoted paraphrase pairs as English-to-English *translation* correspondences
- The English language model will still prove useful
- Combine (para)phrase table with language model inside a regular, unmodified SMT decoder
- Can now generate paraphrase(s) for *any* English sentence†
- Log-linear features in paraphrase space can also be computed via pivoting
  - # of times phrase e1 was "seen" with e2 = # of times e1 was extracted with pivot f * # of times e2 was extracted with pivot f, summed over *all* pivots

†*Using Paraphrases for Parameter Tuning in Statistical Machine Translation*. Nitin Madnani et al. WMT 2007

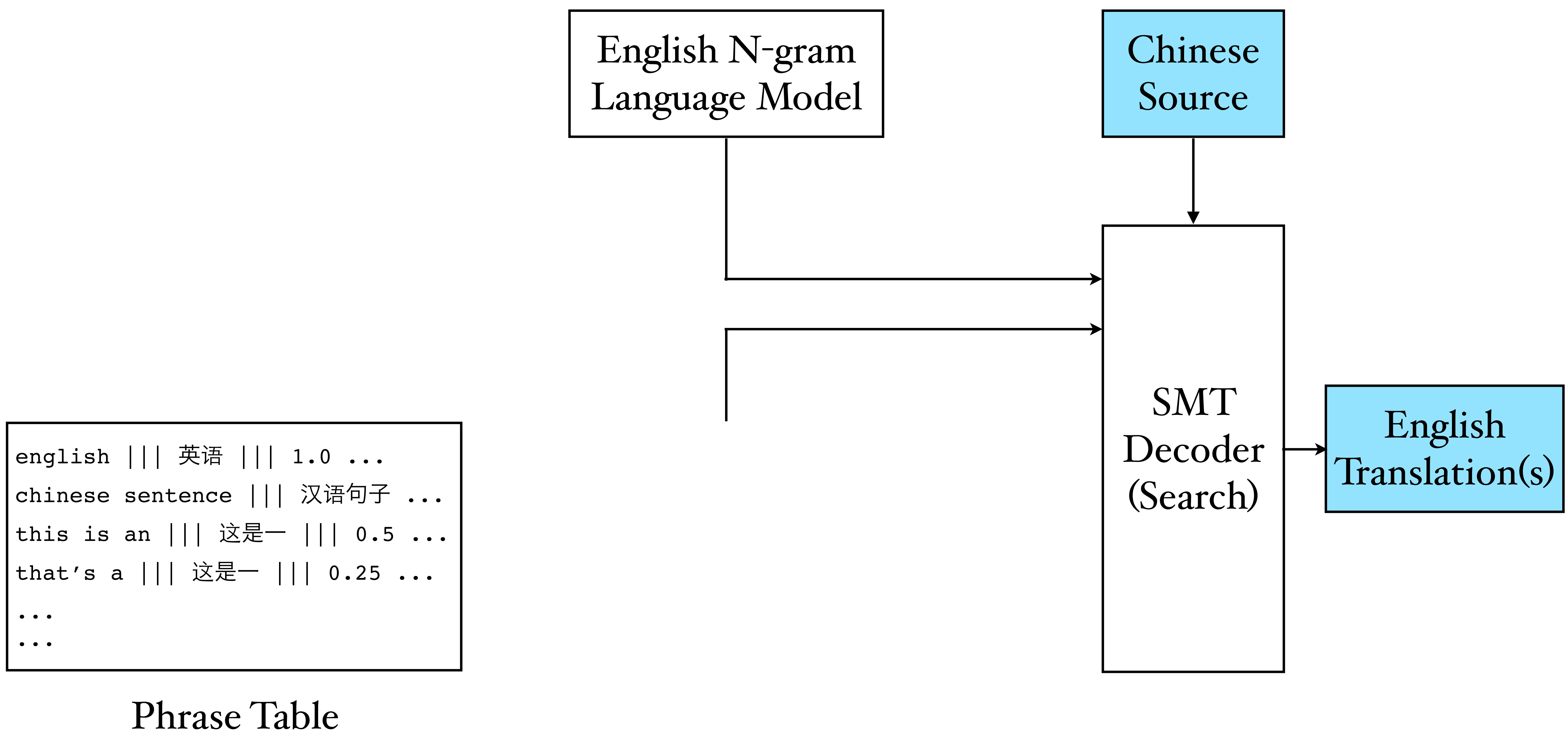
PARAPHRASE GENERATION



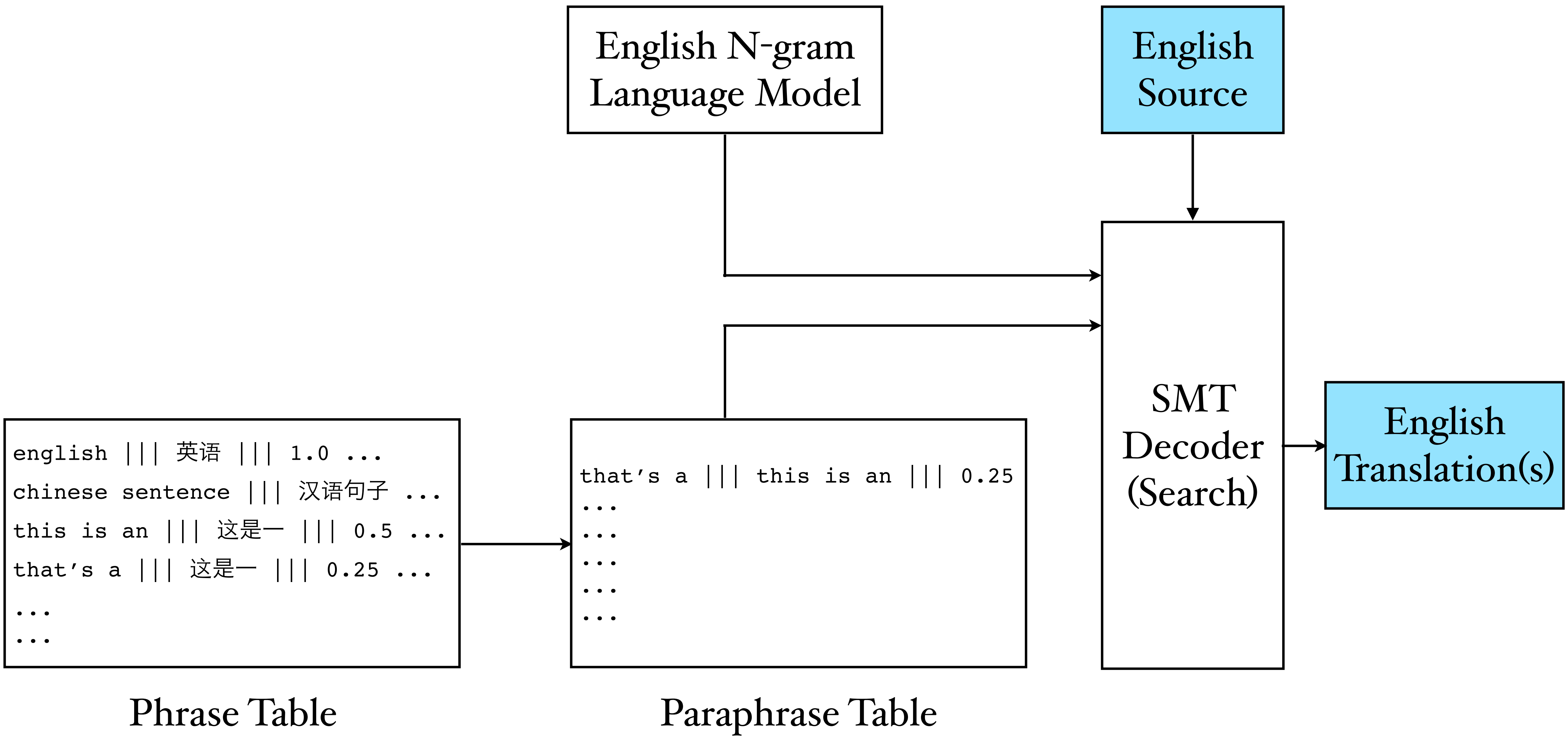
PARAPHRASE GENERATION



PARAPHRASE GENERATION



PARAPHRASE GENERATION



SENTENTIAL PARAPHRASES

Example paraphrases generated with Chinese as pivot language

SENTENTIAL PARAPHRASES

Alcatel added that the company's whole year earnings would be announced on February 4.
Alcatel said that the company's total annual revenues would be released on February 4.

He was now preparing a speech concerning the US policy for the upcoming World Economic Forum.
He was now ready to talk with regard to the US policies for the forthcoming International Economic Forum.

Tibet has entered an excellent phase of political stability, ethnic unity and people living in peace.
Tibetans have come to cordial political stability, national unity and lived in harmony.

Its ocean and blue-sky scenery and the mediterranean climate make it world's famous scenic spot.
Its harbour and blue-sky appearance and the border situation decided it world's renowned tourist attraction.

Paraphrase Quality ↓

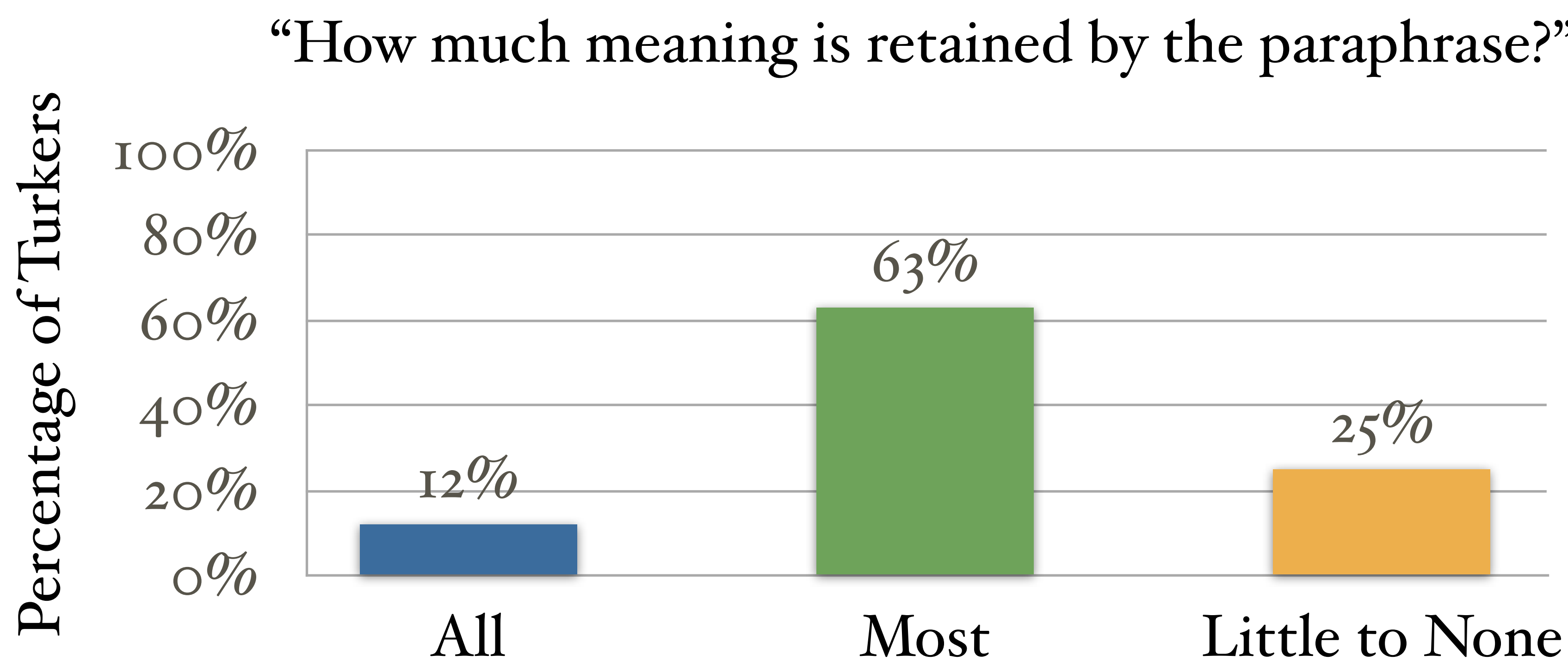
Example paraphrases generated with Chinese as pivot language

MTURK EVALUATION

MTURK EVALUATION

"How much meaning is retained by the paraphrase?"

MTURK EVALUATION



- Most "translations" are only *approximately* paraphrastic; Not surprising
- Paraphrases often not useful for direct human consumption
- Can they be used to solve our problem of reference sparsity for parameter tuning?

EXPERIMENTAL SETUP

$\boxed{R_1}$

$\boxed{S}$ +

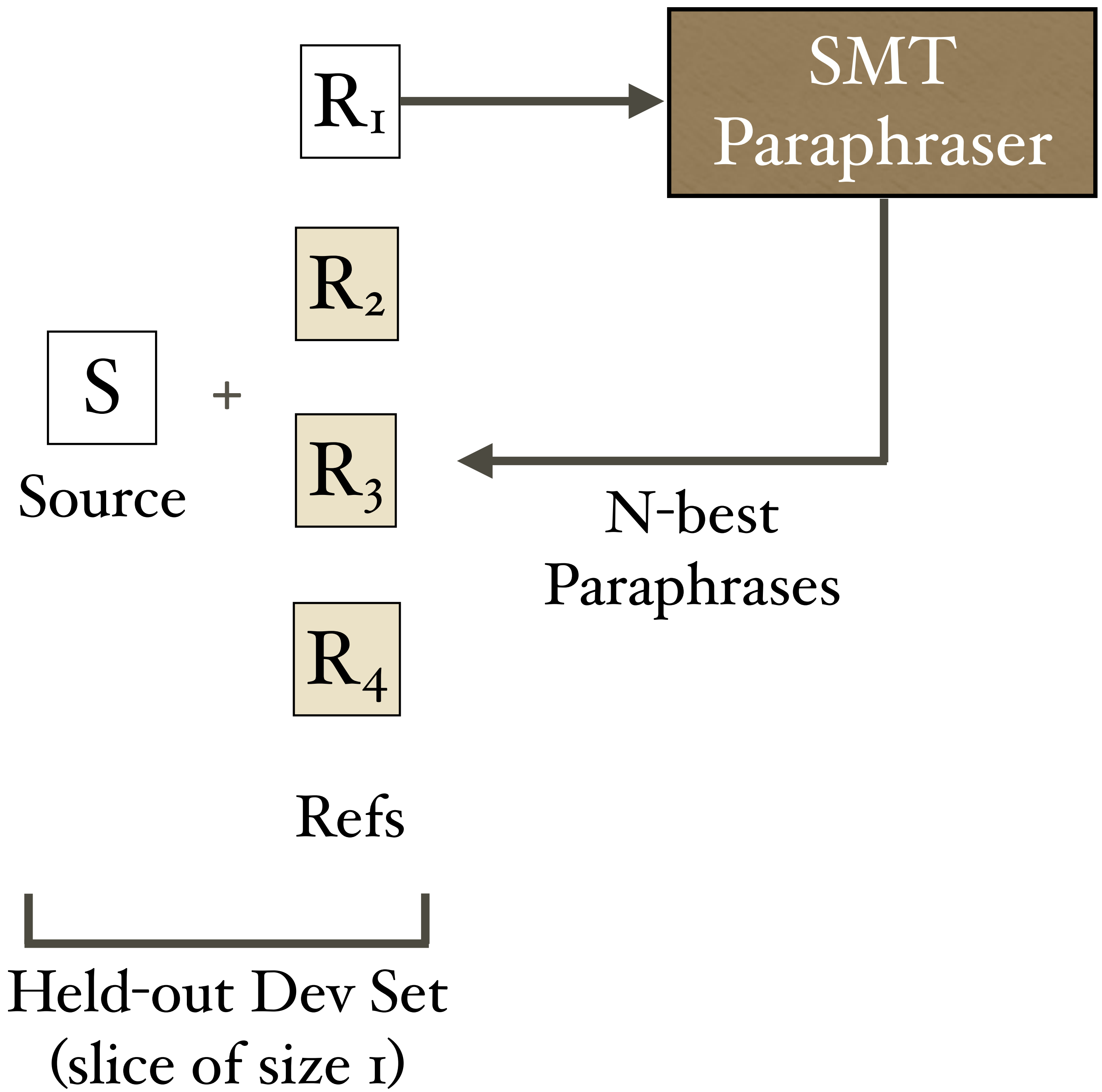
Source

Refs

Held-out Dev Set
(slice of size 1)

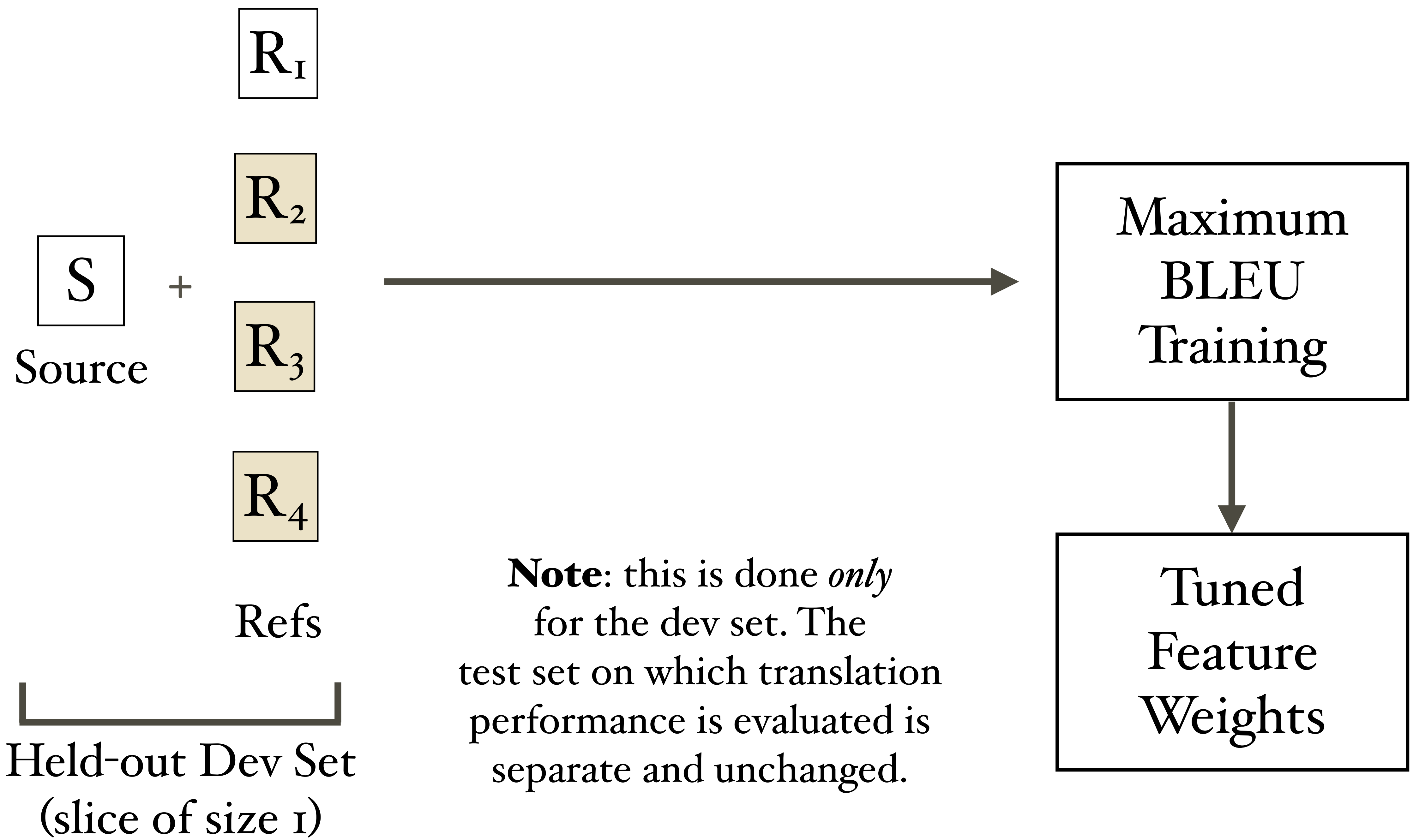
Investigating the Value of Paraphrased Reference Translations in Parameter Optimization. Nitin Madnani et al. AMTA 2008

EXPERIMENTAL SETUP



Investigating the Value of Paraphrased Reference Translations in Parameter Optimization. Nitin Madnani et al. AMTA 2008

EXPERIMENTAL SETUP



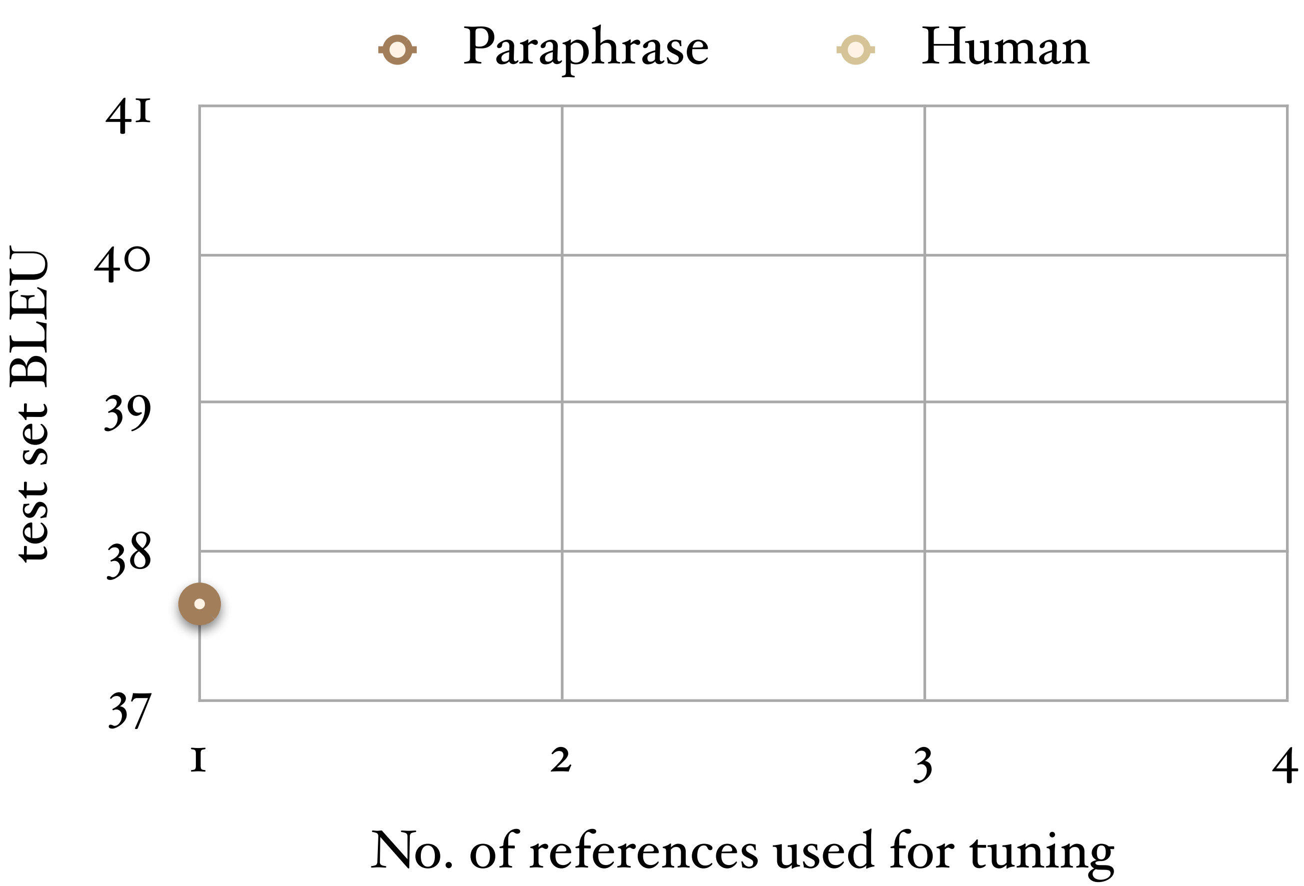
Note: this is done *only* for the dev set. The test set on which translation performance is evaluated is separate and unchanged.

Investigating the Value of Paraphrased Reference Translations in Parameter Optimization. Nitin Madnani et al. AMTA 2008

RESULTS: CHINESE TRANSLATION

Paraphrase Human

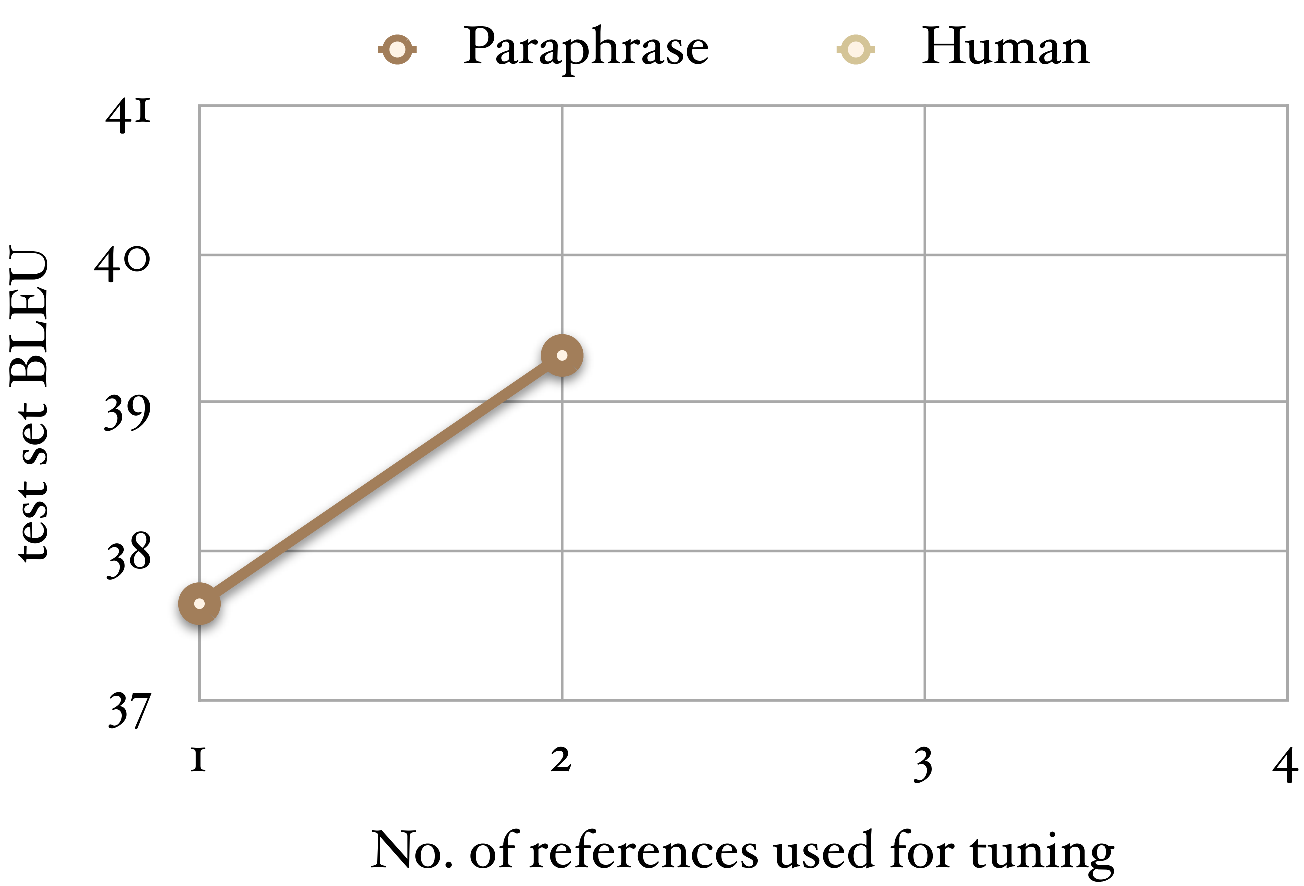
RESULTS: CHINESE TRANSLATION



# Tuning References	Paraphrase BLEU	Human BLEU
1 (1H+0)	37.65	37.65
2 (1H+1)	**39.32**	**39.20**
3 (1H+2)	**39.58**	**40.21**
4 (1H+3)	**39.21**	**40.69**

Higher BLEU is better
Bold denotes statistical significance for BLEU

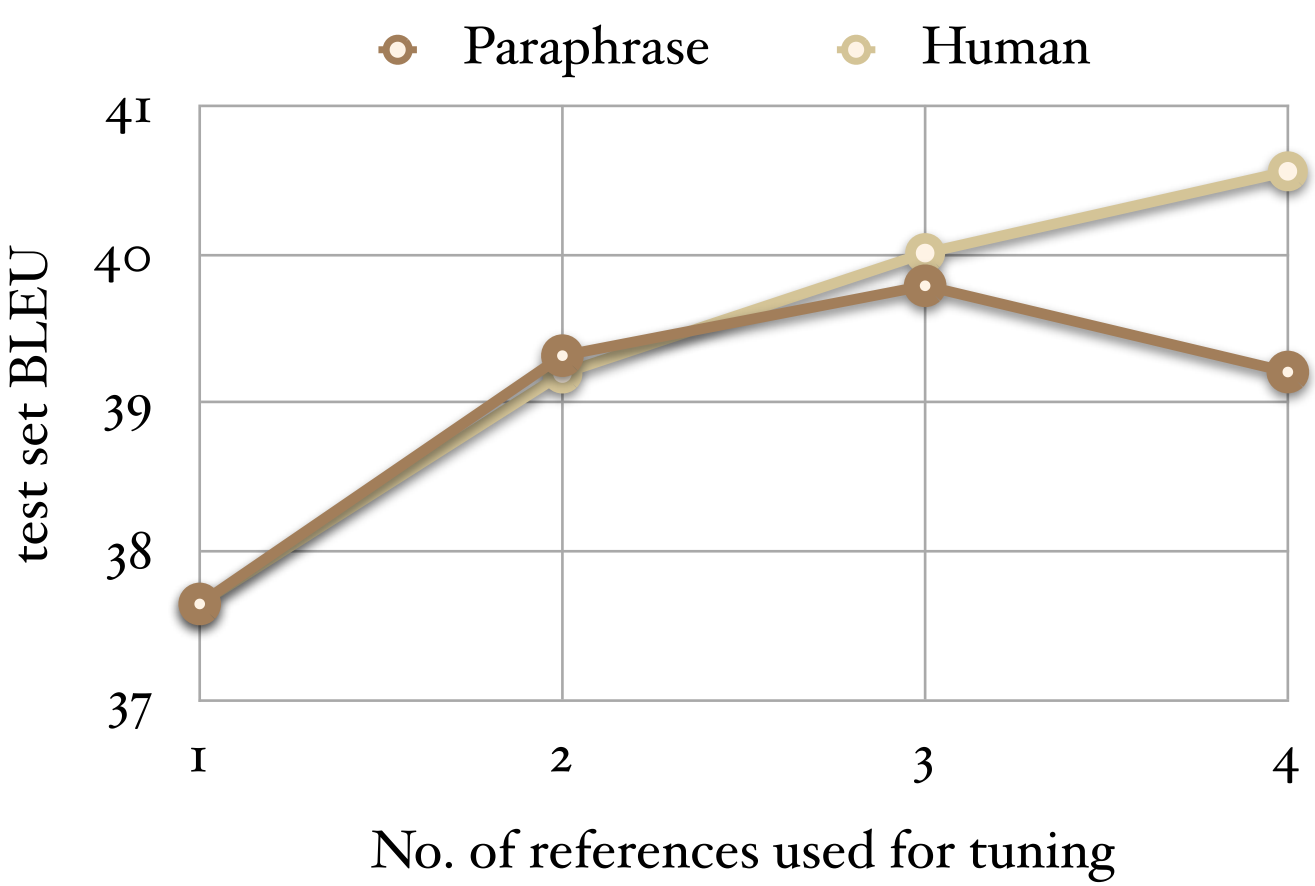
RESULTS: CHINESE TRANSLATION



# Tuning References	Paraphrase BLEU	Human BLEU
1 (1H+0)	37.65	37.65
2 (1H+1)	**39.32**	**39.20**
3 (1H+2)	**39.58**	**40.21**
4 (1H+3)	**39.21**	**40.69**

Higher BLEU is better
Bold denotes statistical significance for BLEU

RESULTS: CHINESE TRANSLATION

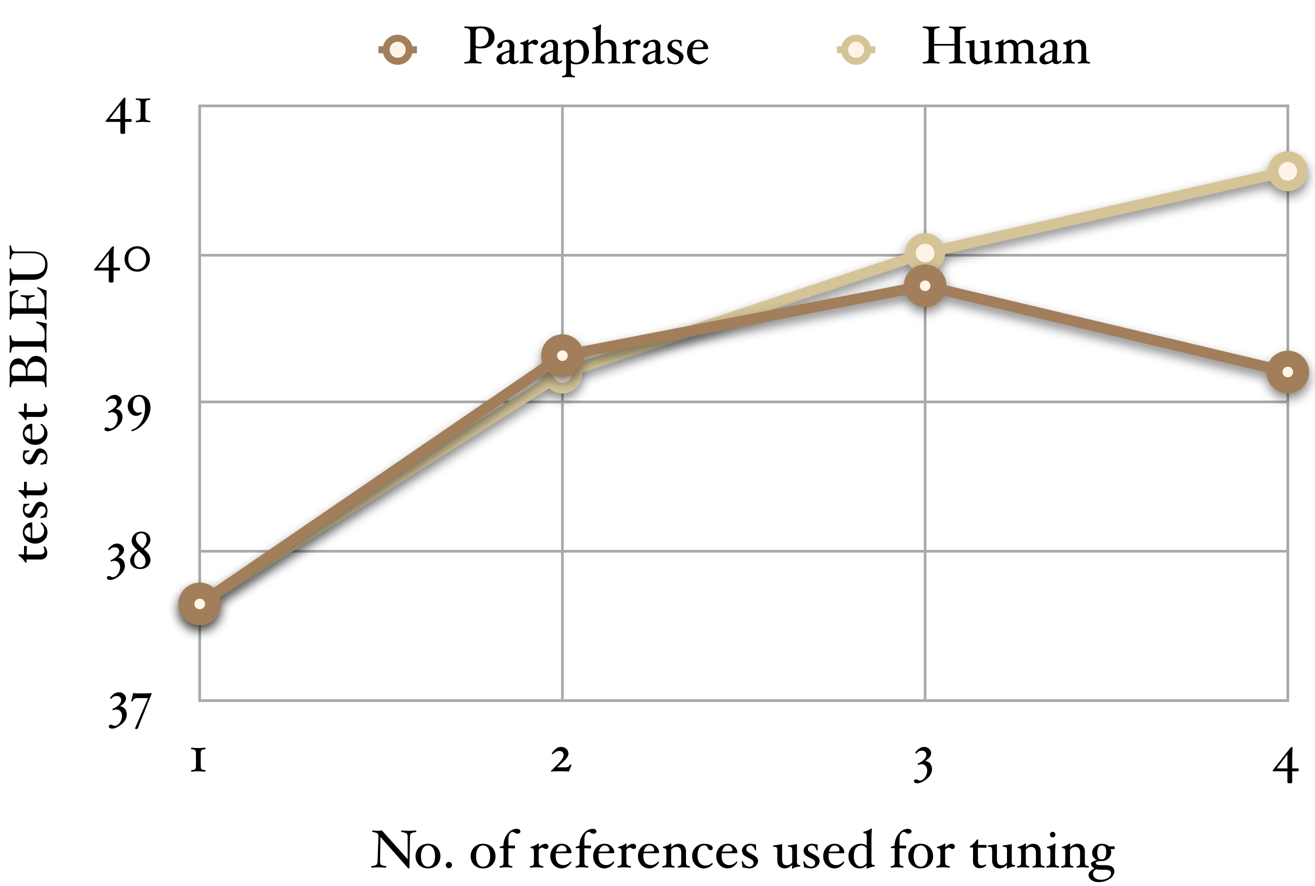


# Tuning References	Paraphrase BLEU	Human BLEU
1 (1H+0)	37.65	37.65
2 (1H+1)	**39.32**	**39.20**
3 (1H+2)	**39.58**	**40.21**
4 (1H+3)	**39.21**	**40.69**

Higher BLEU is better
Bold denotes statistical significance for BLEU

- Significant improvements in BLEU and TER on test set (**note:** not tuning/dev set)
- Adding 2-best or 3-best paraphrased references gives smaller improvements
- Effect of adding more than 1 human reference is better

RESULTS: CHINESE TRANSLATION



# Tuning References	Paraphrase BLEU	Human BLEU
1 (1H+0)	37.65	37.65
2 (1H+1)	**39.32**	**39.20**
3 (1H+2)	**39.58**	**40.21**
4 (1H+3)	**39.21**	**40.69**

Higher BLEU is better
Bold denotes statistical significance for BLEU

- Significant improvements in BLEU and TER on test set (**note:** not tuning/dev set)
- Adding 2-best or 3-best paraphrased references gives smaller improvements
- Effect of adding more than 1 human reference is better
- Similar results for French, Spanish and German translation (to English)

MORE != BETTER?

MORE != BETTER?

- The current SMT paraphraser changes *everything it can*
- Basically a crap-shoot; change everything and hope that some changes will turn out to be useful during parameter tuning
- How about only making changes that are likely to be *useful*?
- Useful: paraphrases that are *a priori* more likely to match the system translation output
- One way to do this is to create a "targeted" version of the paraphraser

TARGETED PARAPHRASER

TARGETED PARAPHRASER

O - AWB was severely hit after the relevant inquiry report into the matter was made public on the 27th.

T - After the release of the investigation report on the 27th, the company suffered a serious blow.

P_u - AWB was significantly impacted after the concerning review report about the issue was made release on the 27th.

P_t - AWB suffered a serious blow after the relevant inquiry report into the matter was made public on the 27th.

Actual Examples

T: MT output, **O**: Original Reference, **P_u**: “Untargeted” paraphrase, **P_t**: Targeted Paraphrase

TARGETED PARAPHRASER

O - AWB was severely hit after the relevant inquiry report into the matter was made public on the 27th.

T - After the release of the investigation report on the 27th, the company suffered a serious blow.

P_u - AWB was significantly impacted after the concerning review report about the issue was made release on the 27th.

P_t - AWB suffered a serious blow after the relevant inquiry report into the matter was made public on the 27th.

O - Singapore economic review committee: economy expected to see complete recovery in 2004

T - Singapore : the economy in 2004 is thought to recover fully

P_u - New economy: economic review board thought possible recovery in 2004

P_t - Singapore economic review committee: economy expected to recover fully in 2004

Actual Examples

T: MT output, **O**: Original Reference, **P_u**: “Untargeted” paraphrase, **P_t**: Targeted Paraphrase

TARGETED PARAPHRASER

- Tune SMT system with single human reference and define a new *targeting* feature for paraphrase decoder
    - # of words in **paraphrase** hypothesis NOT in the **translation** system translation output
- By negatively weighting this feature, paraphrases can be made to look more like the translation output
- This could lead to a nasty feedback loop that didn't exist before!
    - Bad translation ==> Bad targeted paraphrase ==> Bad translation ...
- Need a counter-balance feature that prevents such a loop
    - *Self-paraphrase bias*: reserve fixed amount of prob. mass for identity paraphrases
- Need some fancy math to find an operating point that balances the two

The Circle of Meaning: From Translation to Paraphrasing & Back. Nitin Madnani. PhD Dissertation. 2010

TARGETED PARAPHRASER

$\boxed{R_1}$

$\boxed{S}$ +

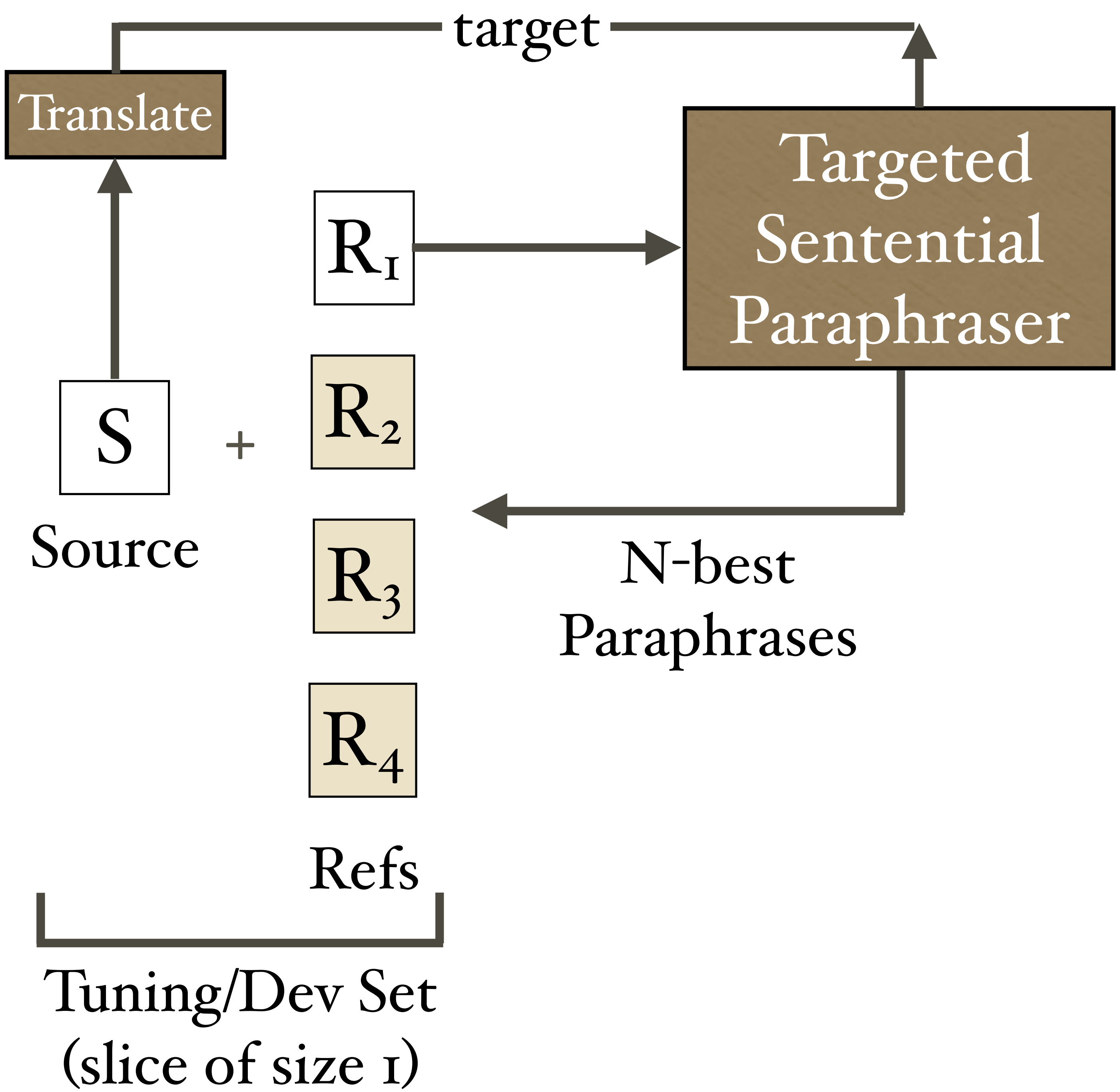
Source

Refs

Tuning/Dev Set
(slice of size 1)

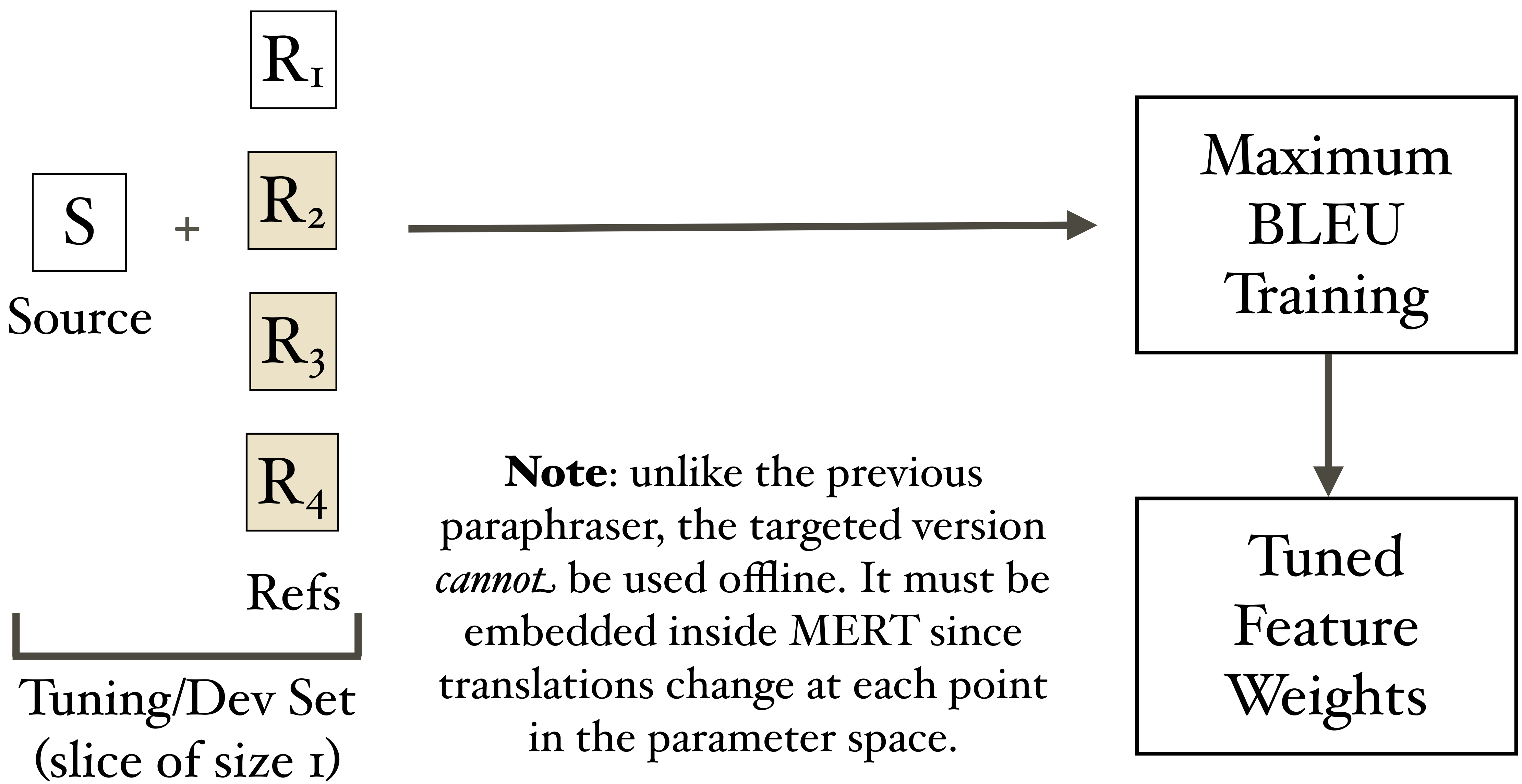
Generating Targeted Paraphrases for Improved Translation. Nitin Madnani & Bonnie Dorr. ACM TIST (To Appear)

TARGETED PARAPHRASER



Generating Targeted Paraphrases for Improved Translation. Nitin Madnani & Bonnie Dorr. ACM TIST (To Appear)

TARGETED PARAPHRASER



Note: unlike the previous paraphraser, the targeted version *cannot* be used offline. It must be embedded inside MERT since translations change at each point in the parameter space.

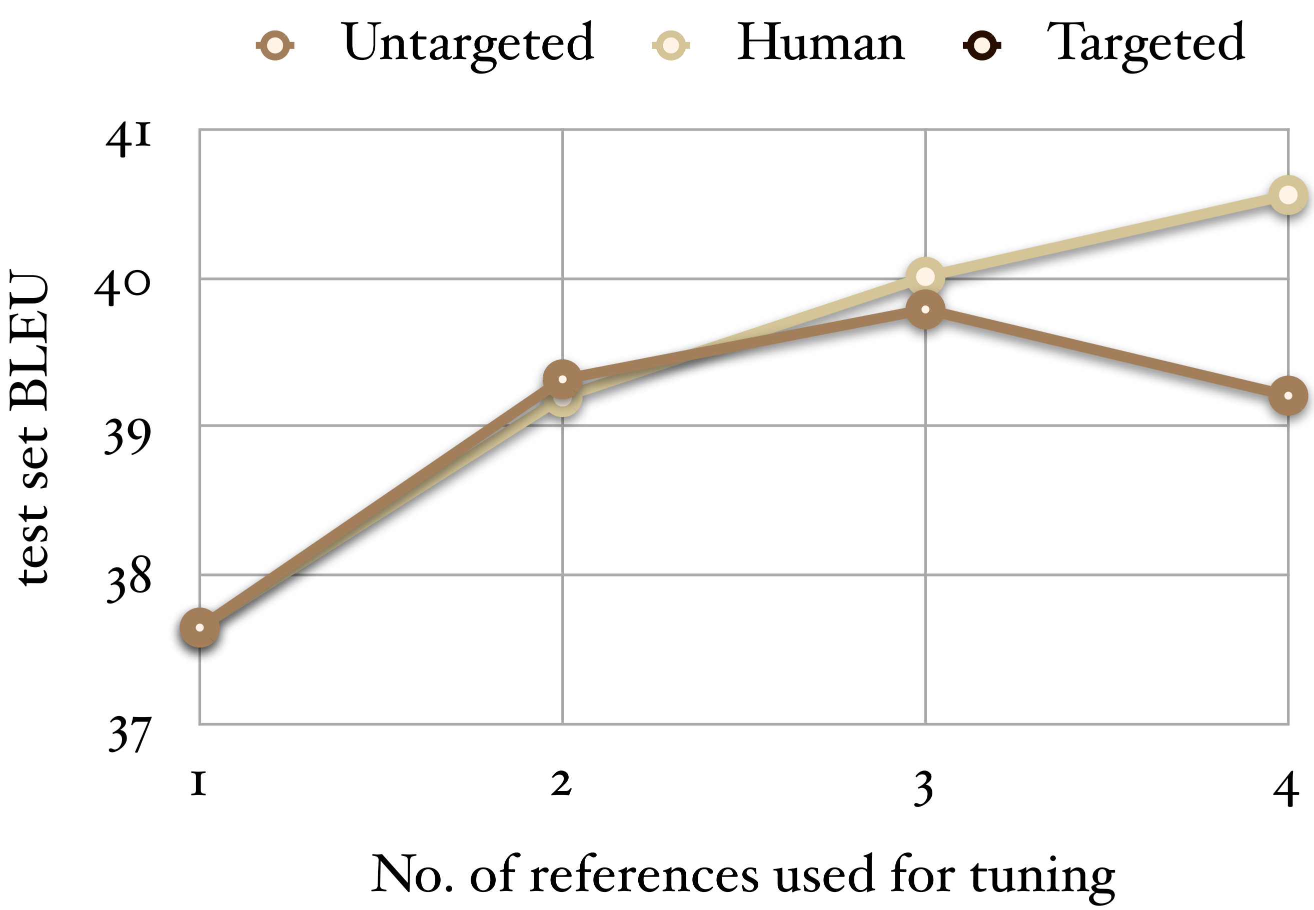
Generating Targeted Paraphrases for Improved Translation. Nitin Madnani & Bonnie Dorr. ACM TIST (To Appear)

RESULTS: CHINESE TRANSLATION

Untargeted   Human   Targeted

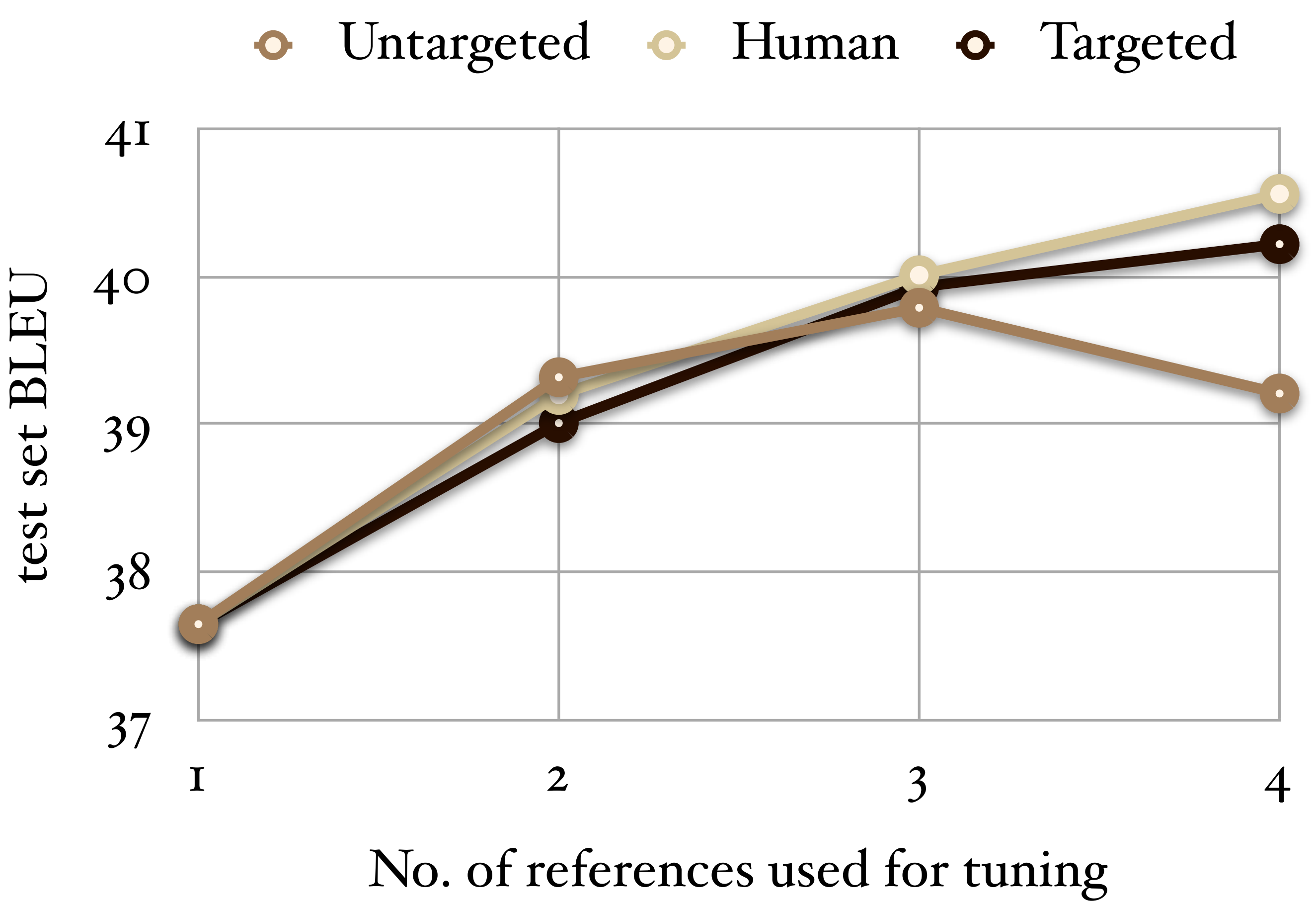
Refs	Untargeted	Targeted	Human
	BLEU	BLEU	BLEU
1H+0	37.65	37.65	37.65
1H+1	**39.32**	**39.01**	**39.20**
1H+2	**39.58**	**39.93**	**40.21**
1H+3	**39.21**	**40.22**	**40.69**

RESULTS: CHINESE TRANSLATION



Refs	Untargeted BLEU	Targeted BLEU	Human BLEU
1H+0	37.65	37.65	37.65
1H+1	**39.32**	**39.01**	**39.20**
1H+2	**39.58**	**39.93**	**40.21**
1H+3	**39.21**	**40.22**	**40.69**

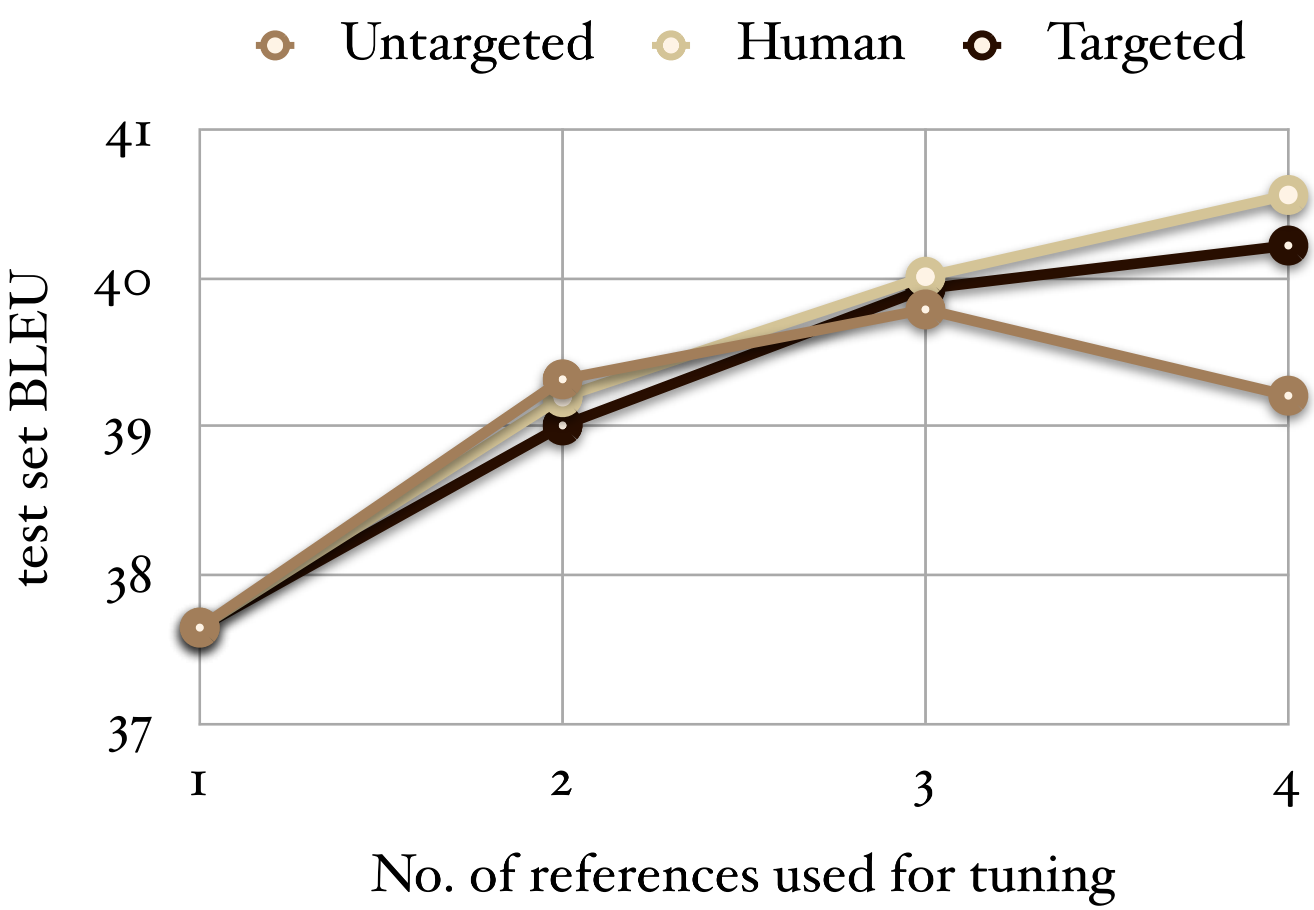
RESULTS: CHINESE TRANSLATION



Refs	Untargeted	Targeted	Human
	BLEU	BLEU	BLEU
1H+0	37.65	37.65	37.65
1H+1	**39.32**	**39.01**	**39.20**
1H+2	**39.58**	**39.93**	**40.21**
1H+3	**39.21**	**40.22**	**40.69**

- *k*-best targeted paraphrases behave much better than *k*-best *untargeted* paraphrases
- Significant improvements in translation performance compared to baseline (1H)

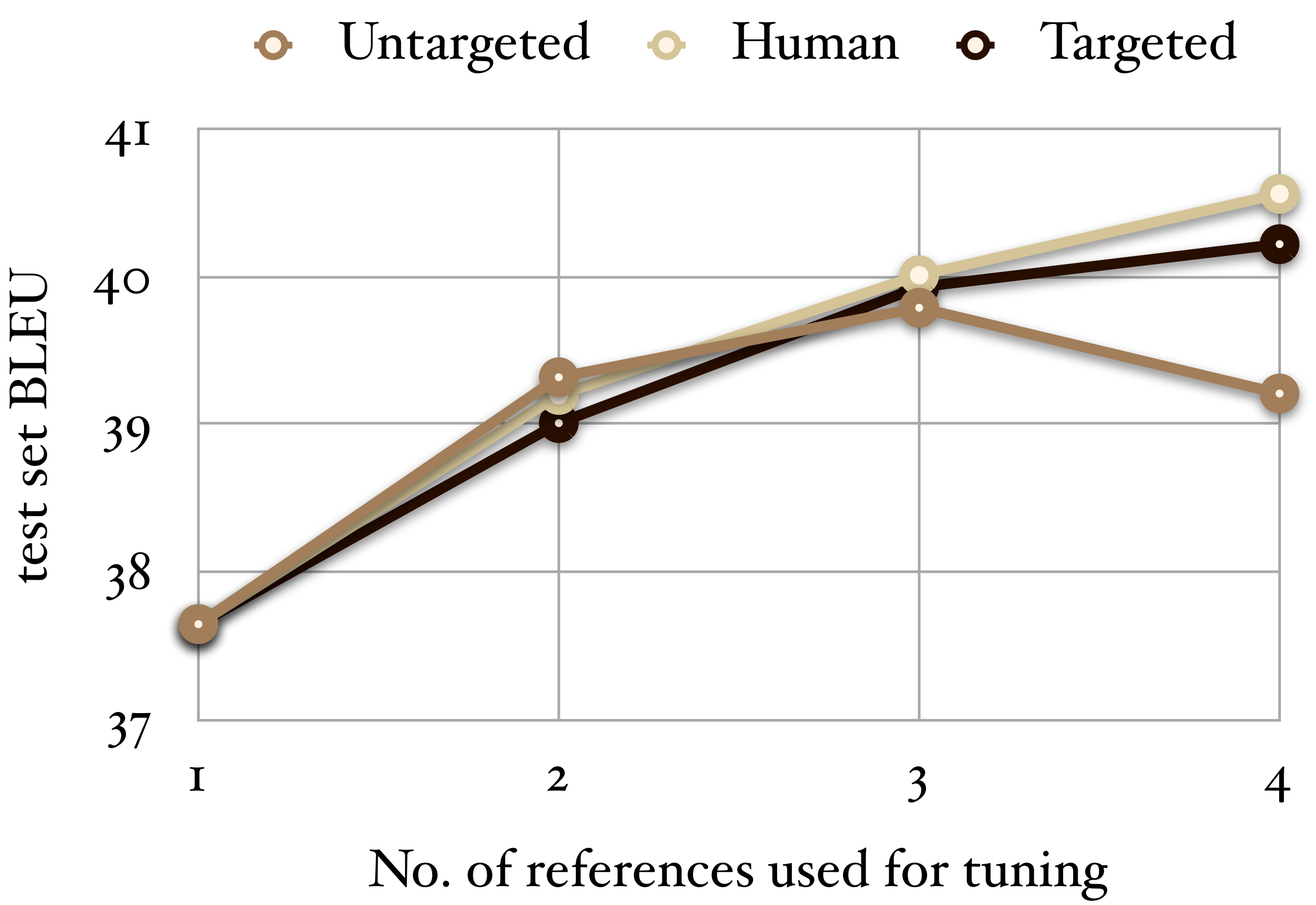
RESULTS: CHINESE TRANSLATION



Refs	Untargeted BLEU	Targeted BLEU	Human BLEU
1H+0	37.65	37.65	37.65
1H+1	**39.32**	**39.01**	**39.20**
1H+2	**39.58**	**39.93**	**40.21**
1H+3	**39.21**	**40.22**	**40.69**

- *k*-best targeted paraphrases behave much better than *k*-best *untargeted* paraphrases
- Significant improvements in translation performance compared to baseline (1H)
- Similar results obtained for French, Spanish and German translation

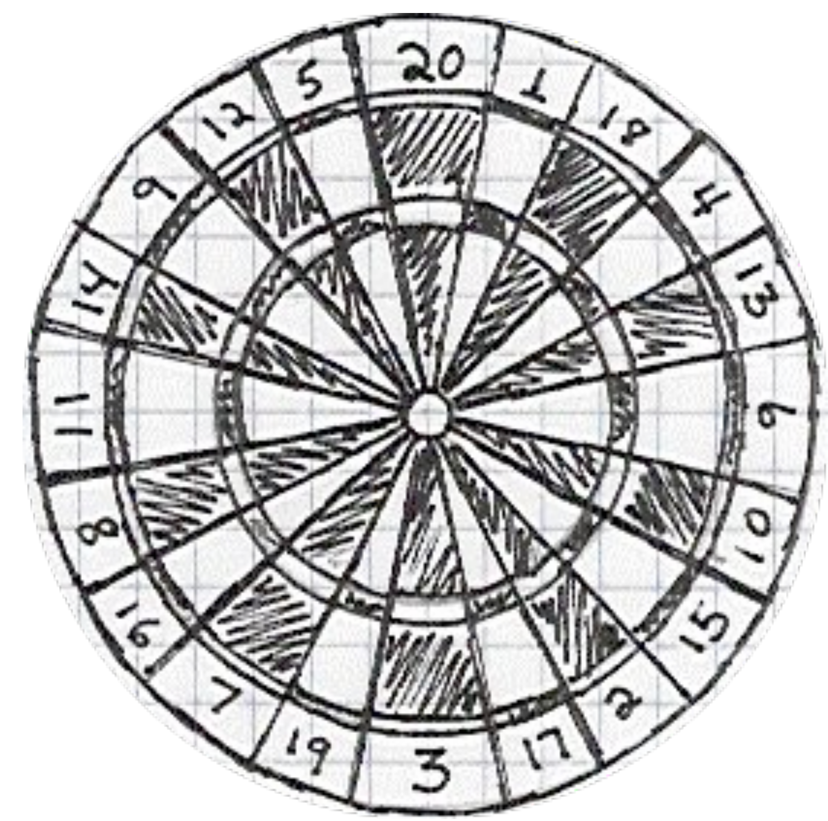
RESULTS: CHINESE TRANSLATION



Refs	Untargeted BLEU	Targeted BLEU	Human BLEU
1H+0	37.65	37.65	37.65
1H+1	**39.32**	**39.01**	**39.20**
1H+2	**39.58**	**39.93**	**40.21**
1H+3	**39.21**	**40.22**	**40.69**

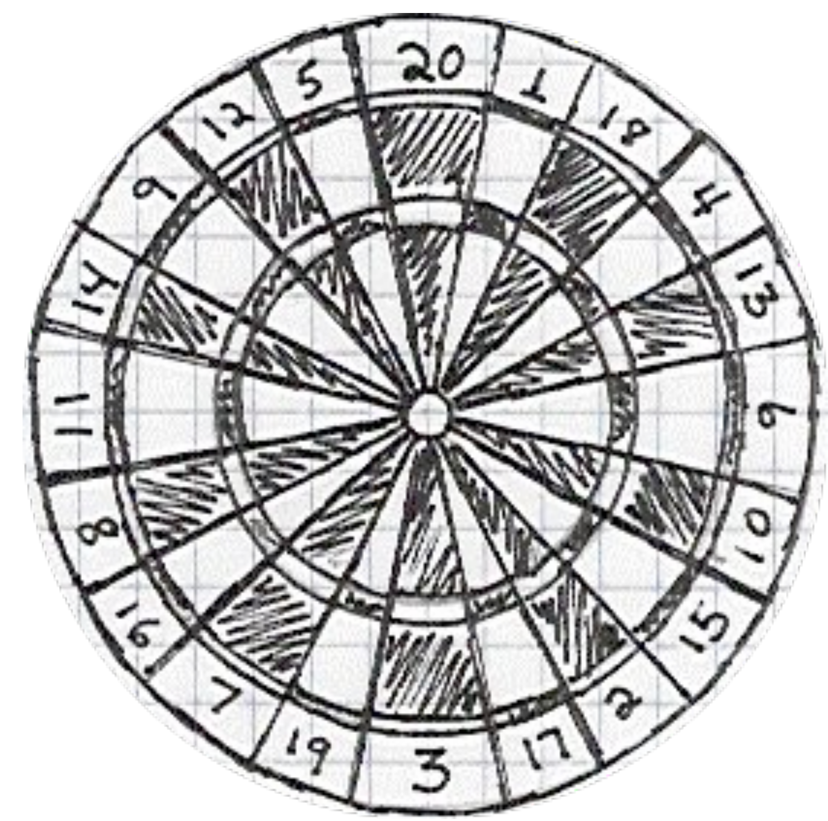
- *k*-best targeted paraphrases behave much better than *k*-best *untargeted* paraphrases
- Significant improvements in translation performance compared to baseline (1H)
- Similar results obtained for French, Spanish and German translation
- All results also validated using human judgments of translation via Mechanical Turk

A DARTBOARD ANALOGY



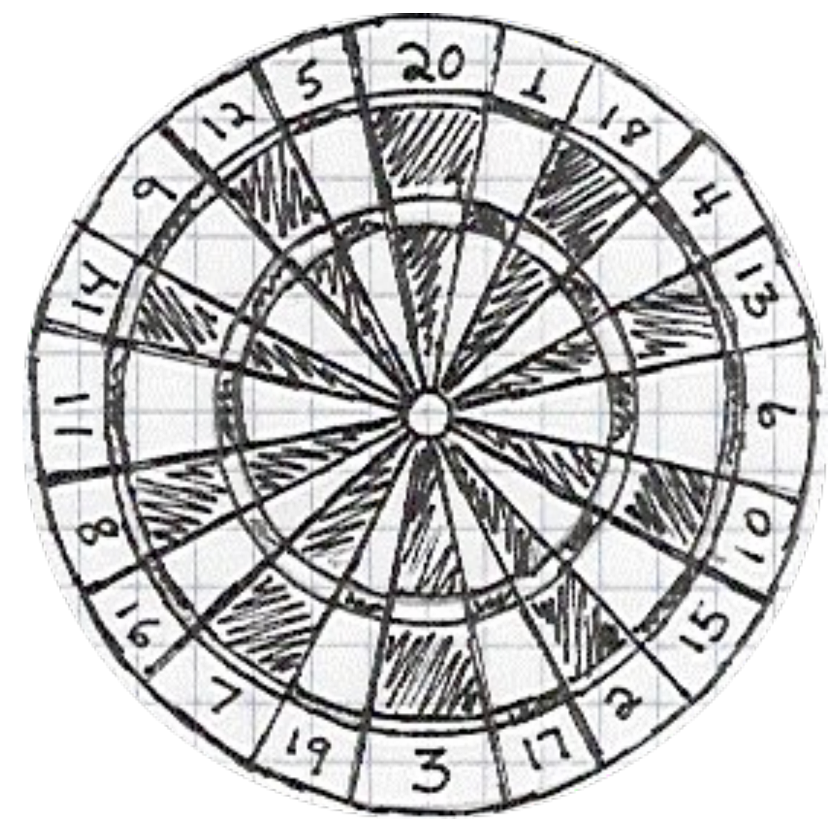
A DARTBOARD ANALOGY

- Imagine matching an a word sequence as hitting the bullseye on a dartboard (BLEU)



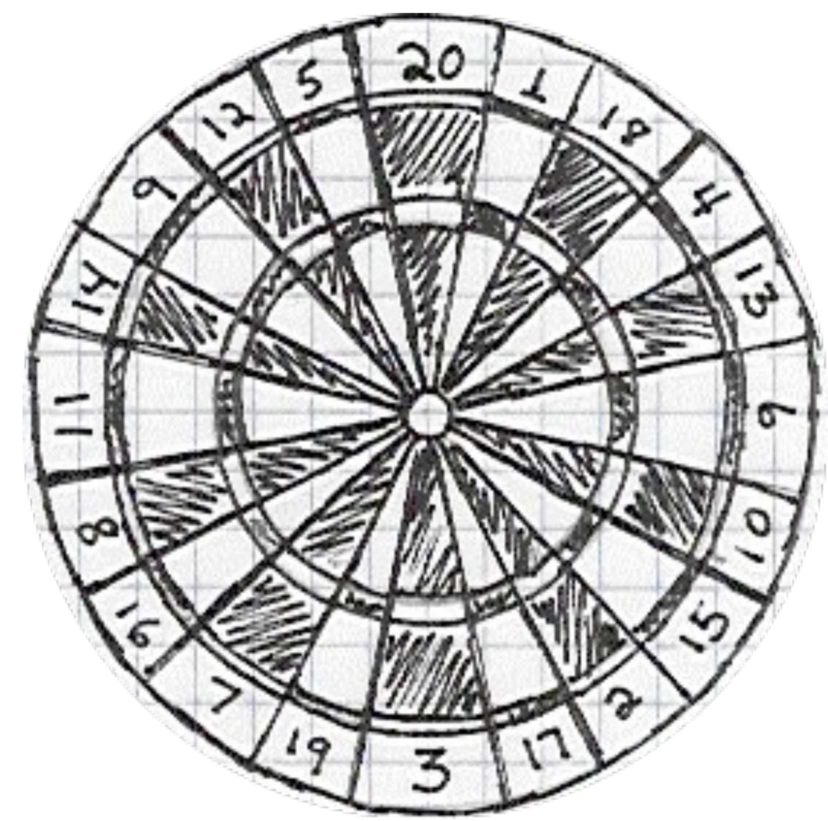
A DARTBOARD ANALOGY

- Imagine matching an a word sequence as hitting the bullseye on a dartboard (BLEU)
- Using 4 human references is like scaling the dartboard 4x (the bullseye is 4 times bigger)



A DARTBOARD ANALOGY

- Imagine matching an a word sequence as hitting the bullseye on a dartboard (BLEU)
- Using 4 human references is like scaling the dartboard 4x (the bullseye is 4 times bigger)
- Using untargeted paraphrases is like scaling the board but with the bullseye scrambled all over the board



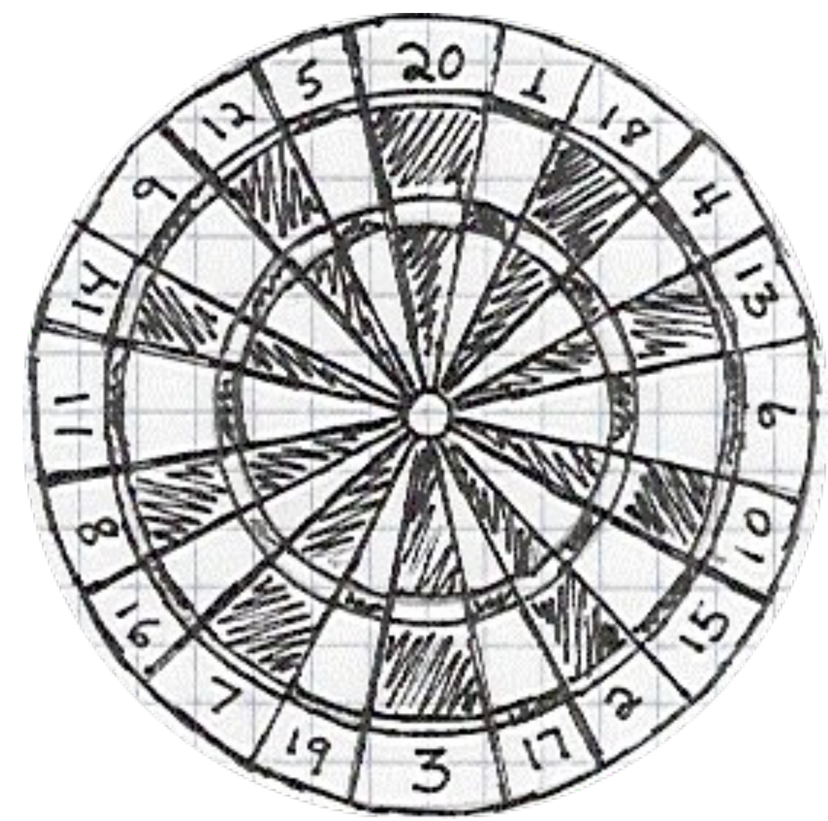
A DARTBOARD ANALOGY

- Imagine matching an a word sequence as hitting the bullseye on a dartboard (BLEU)

- Using 4 human references is like scaling the dartboard 4x (the bullseye is 4 times bigger)

- Using untargeted paraphrases is like scaling the board but with the bullseye scrambled all over the board

- With targeted paraphrases, the bullseye is still somewhat scrambled but we get to shoot the dart out of a rifle with a scope!



SUMMARY

- SMT represents the current state of the art in MT
- Besides bitext, SMT systems require multiple **reference** translations that aren't cheap
- We can use the SMT system itself to manufacture additional references from a single, good quality reference
- No reason for the paraphraser to be restricted to SMT
  - Generate new reference answers for short-answer tests
  - Generate multiple choice items for "paraphrase" questions
  - Expanding sentiment lexicon for essay opinion mining

QUESTIONS?

BACKUP SLIDES

SENTENTIAL PARAPHRASES

We must bear in mind the community as a whole.
We must remember the wider community.

They should be better coordinated and more effective.
They should improve the coordination and efficacy.

Women are still one of the most vulnerable sections of society, whose rights are rudely trampled underfoot by the current social and economic system.
They remain one of the weakest in society, whose duties are abruptly scorned by the present social and economic order.

That is what we are waiting to hear from the European Commission.
That is what we expected from the meeting.

This occurred not far away and not very long ago.
This substances not far behind and very recently.

Original Sentence, *Generated Paraphrase* (via **French)**

PHRASAL PARAPHRASES

PHRASAL PARAPHRASES

- Analyzed phrasal paraphrases with Arabic as pivot language
- Only those with $p(e_p|e_q) > 0.9$ to concentrate on pairs more likely to be paraphrases
- Roughly five types of paraphrases

PHRASAL PARAPHRASES

```
polish troops ||| polish soldiers
accounting firms ||| auditing firms
armed source ||| military source
...
```

Lexical

- Analyzed phrasal paraphrases with Arabic as pivot language
- Only those with $p(e_p|e_q) > 0.9$ to concentrate on pairs <u>more likely</u> to be paraphrases
- Roughly five types of paraphrases

PHRASAL PARAPHRASES

```
polish troops ||| polish soldiers
accounting firms ||| auditing firms
armed source ||| military source
...
```

Lexical

```
50 ton ||| 50 tons
caused clouds ||| causing clouds
syria deny ||| syria denies
...
```

Morphological variants

- Analyzed phrasal paraphrases with Arabic as pivot language
- Only those with $p(e_p|e_q) > 0.9$ to concentrate on pairs <u>more likely</u> to be paraphrases
- Roughly five types of paraphrases

PHRASAL PARAPHRASES

```
polish troops ||| polish soldiers
accounting firms ||| auditing firms
armed source ||| military source
...
```

Lexical

```
50 ton ||| 50 tons
caused clouds ||| causing clouds
syria deny ||| syria denies
...
```

Morphological variants

```
mutual proposal ||| suggest
them were exiled ||| them abroad
my parents ||| my father
...
```

Approximate

- Analyzed phrasal paraphrases with Arabic as pivot language
- Only those with $p(e_p|e_q) > 0.9$ to concentrate on pairs <u>more likely</u> to be paraphrases
- Roughly five types of paraphrases

PHRASAL PARAPHRASES

```
polish troops ||| polish soldiers
accounting firms ||| auditing firms
armed source ||| military source
...
```

Lexical

```
50 ton ||| 50 tons
caused clouds ||| causing clouds
syria deny ||| syria denies
...
```

Morphological variants

```
mutual proposal ||| suggest
them were exiled ||| them abroad
my parents ||| my father
...
```

Approximate

```
agence presse ||| news agency
army roadblock ||| military barrier
staff walked out ||| team withdrew
controversy over ||| polemic about
...
```

Exact

- Analyzed phrasal paraphrases with Arabic as pivot language
- Only those with $p(e_p|e_q) > 0.9$ to concentrate on pairs more likely to be paraphrases
- Roughly five types of paraphrases

PHRASAL PARAPHRASES

```
polish troops ||| polish soldiers
accounting firms ||| auditing firms
armed source ||| military source
...
```

Lexical

```
50 ton ||| 50 tons
caused clouds ||| causing clouds
syria deny ||| syria denies
...
```

Morphological variants

```
mutual proposal ||| suggest
them were exiled ||| them abroad
my parents ||| my father
...
```

Approximate

```
counterpart salam ||| peace
regulation dealing ||| list
recall one ||| deported
...
```

Useless (Noise)

```
agence presse ||| news agency
army roadblock ||| military barrier
staff walked out ||| team withdrew
controversy over ||| polemic about
...
```

Exact

- Analyzed phrasal paraphrases with Arabic as pivot language
- Only those with $p(e_p|e_q) > 0.9$ to concentrate on pairs <u>more likely</u> to be paraphrases
- Roughly five types of paraphrases

PHRASAL PARAPHRASES

```
polish troops ||| polish soldiers
accounting firms ||| auditing firms
armed source ||| military source
...
```

Lexical

```
50 ton ||| 50 tons
caused clouds ||| causing clouds
syria deny ||| syria denies
...
```

Morphological variants

```
mutual proposal ||| suggest
them were exiled ||| them abroad
my parents ||| my father
...
```

Approximate

```
counterpart salam ||| peace
regulation dealing ||| list
recall one ||| deported
...
```

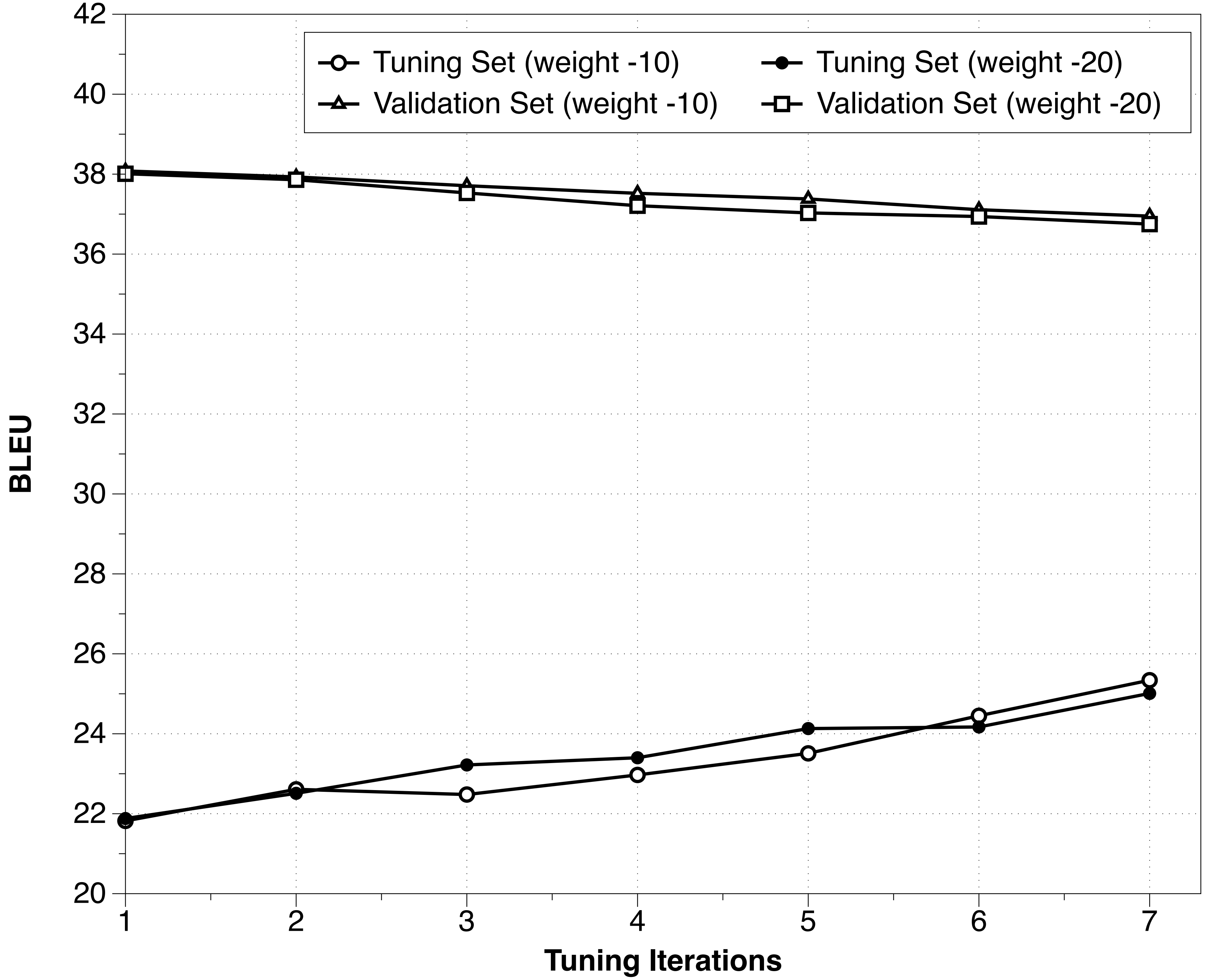
Useless (Noise)

```
agence presse ||| news agency
army roadblock ||| military barrier
staff walked out ||| team withdrew
controversy over ||| polemic about
...
```

Exact

- Analyzed phrasal paraphrases with Arabic as pivot language
- Only those with $p(e_p|e_q) > 0.9$ to concentrate on pairs more likely to be paraphrases
- Roughly five types of paraphrases
- #Approximate + #Exact >> #Useless

NEED FOR SELF-PARAPHRASE BIAS



EXPERIMENTAL DETAILS

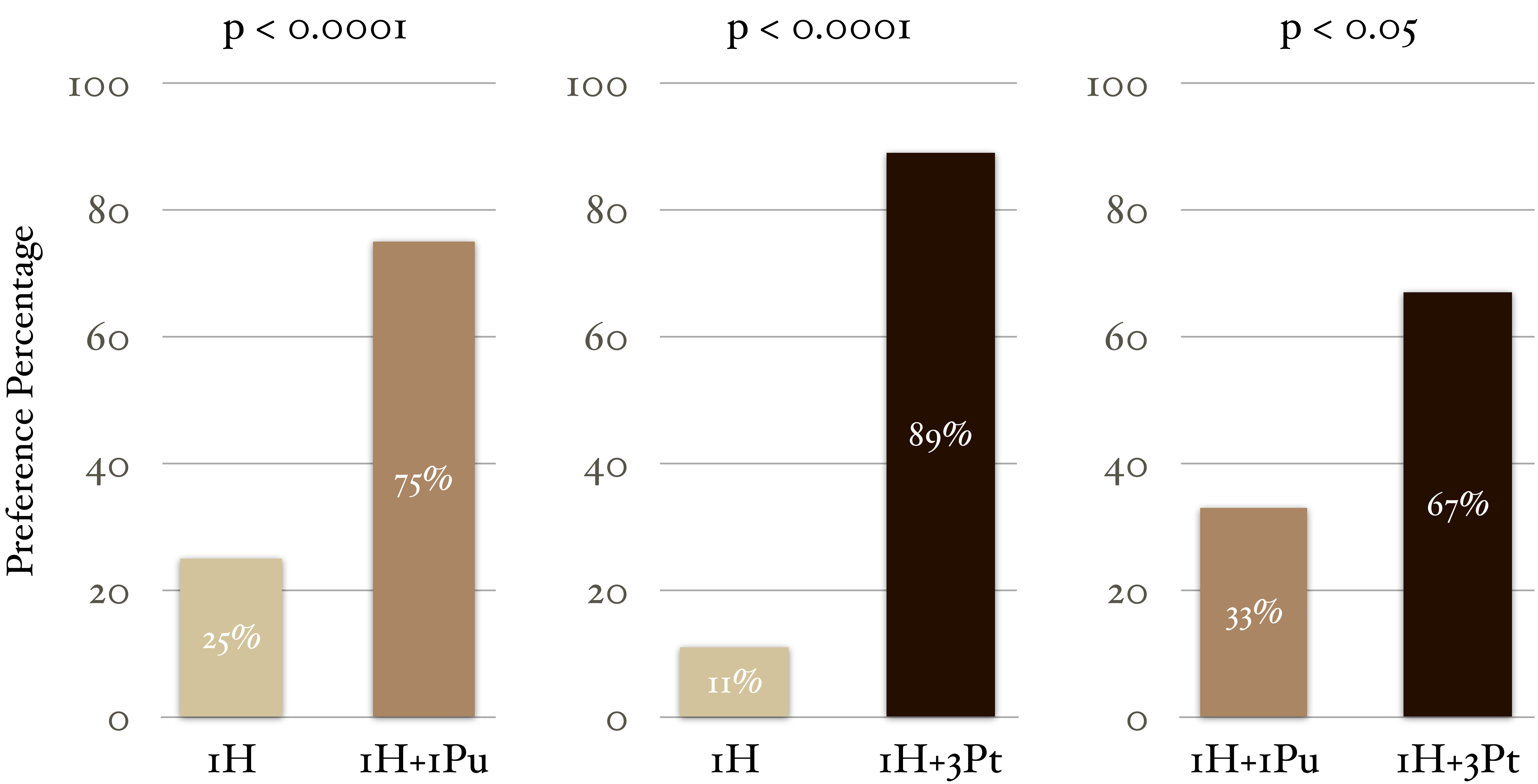
EXPERIMENTAL DETAILS

	Bitext	LM data	Tuning Set	Validation Set
Zh-En	2.5 million sentences (newswire)	8 billion words (Trigram, 5-gram)	919 sentences 4 references	2870 sentences 4 references
Fr-En	1.7 million sentences (Europarl)	8 billion words (Trigram, 5-gram)	2051 sentences 1 reference	2525 sentences 1 reference
De-En	1.6 million sentences (Europarl)	8 billion words (Trigram, 5-gram)	2051 sentences 1 reference	2525 sentences 1 reference
Es-En	1.7 million sentences (Europarl)	8 billion words (Trigram, 5-gram)	2051 sentences 1 reference	2525 sentences 1 reference

HUMAN JUDGMENTS: CHINESE

Pu: Untargeted, Pt: Targeted

HUMAN JUDGMENTS: CHINESE

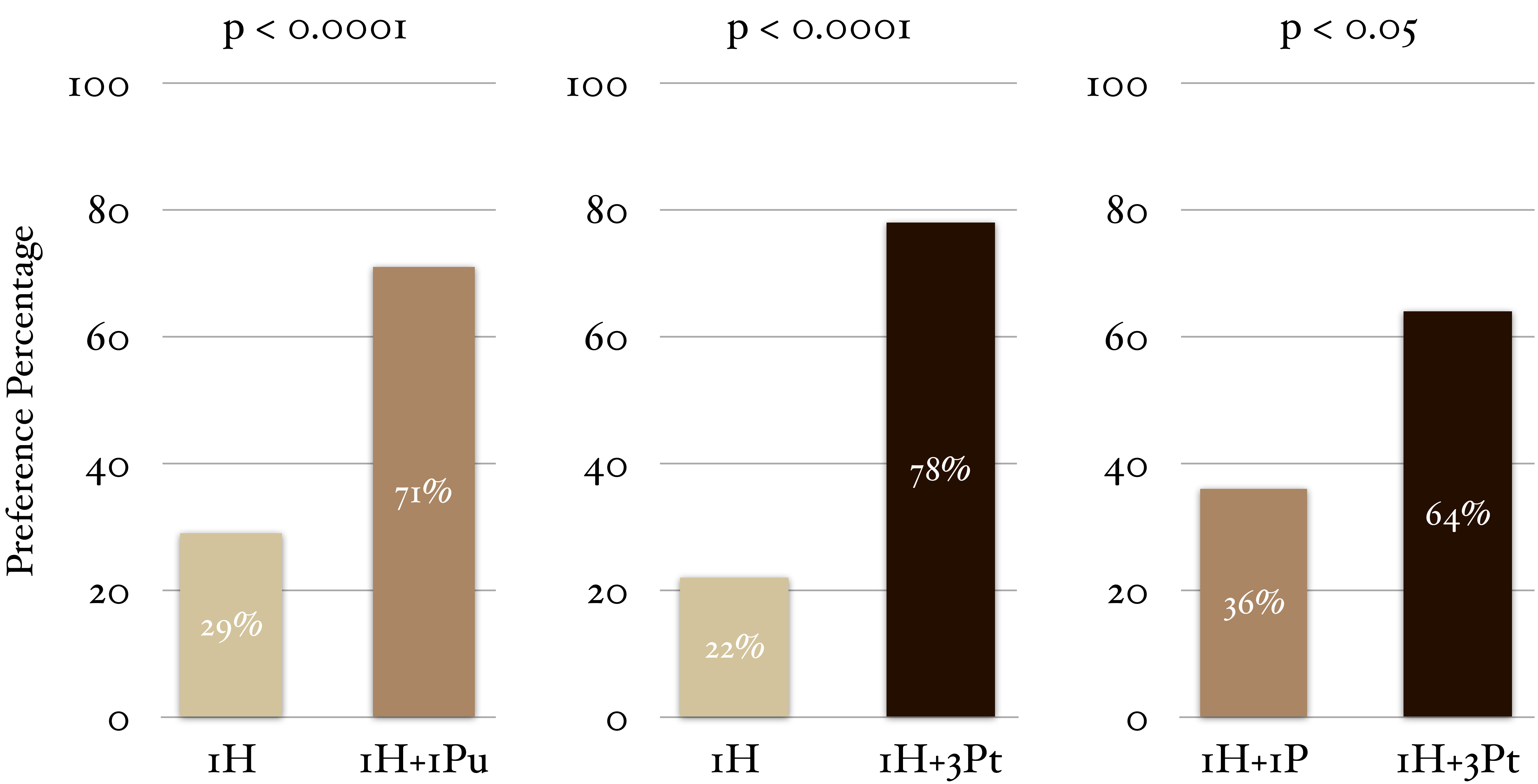


Pu: Untargeted, Pt: Targeted

HUMAN JUDGMENTS: FRENCH

Pu: Untargeted, Pt: Targeted

HUMAN JUDGMENTS: FRENCH

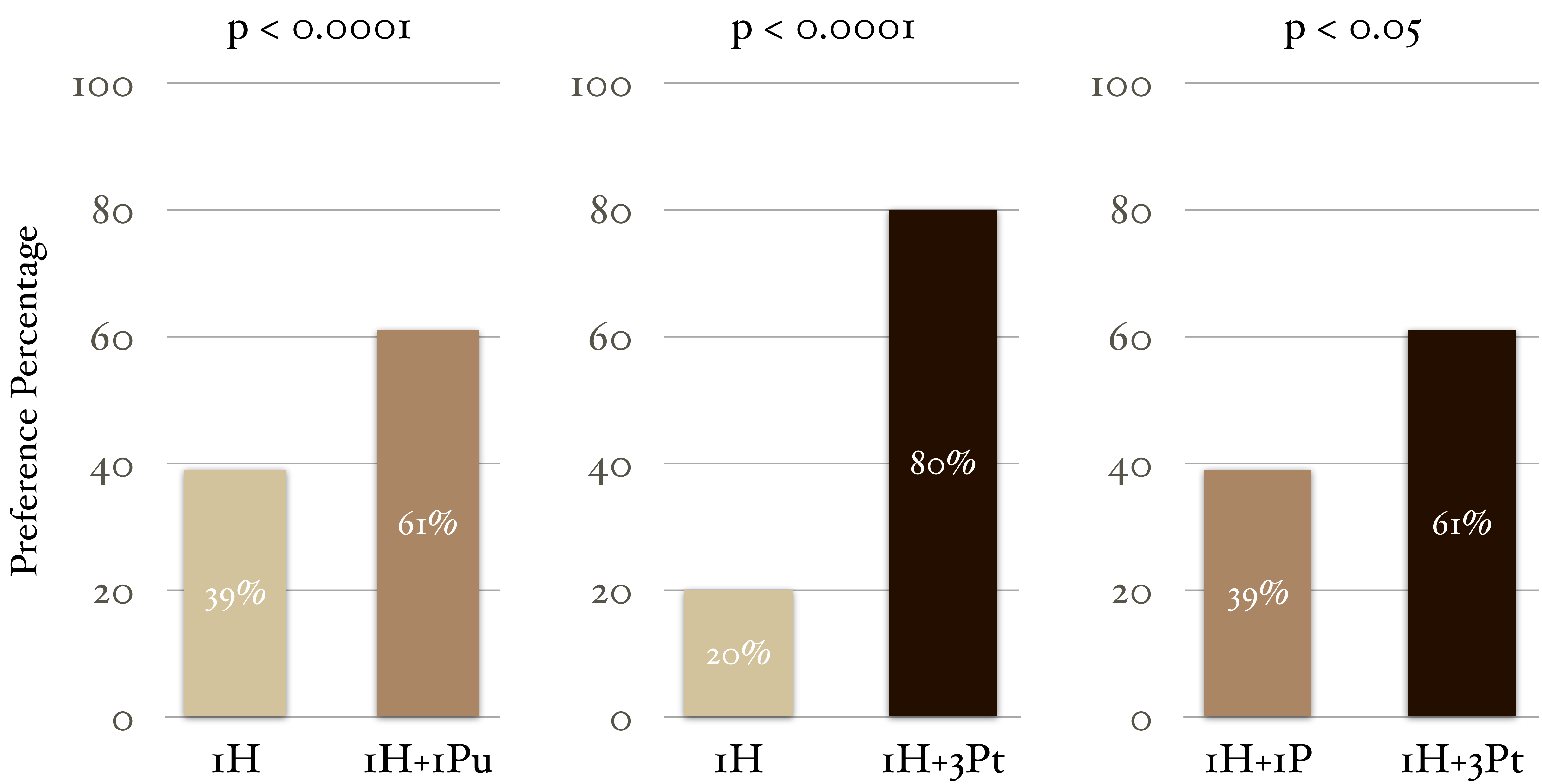


Pu: Untargeted, Pt: Targeted

HUMAN JUDGMENTS: GERMAN

Pu: Untargeted, Pt: Targeted

HUMAN JUDGMENTS: GERMAN

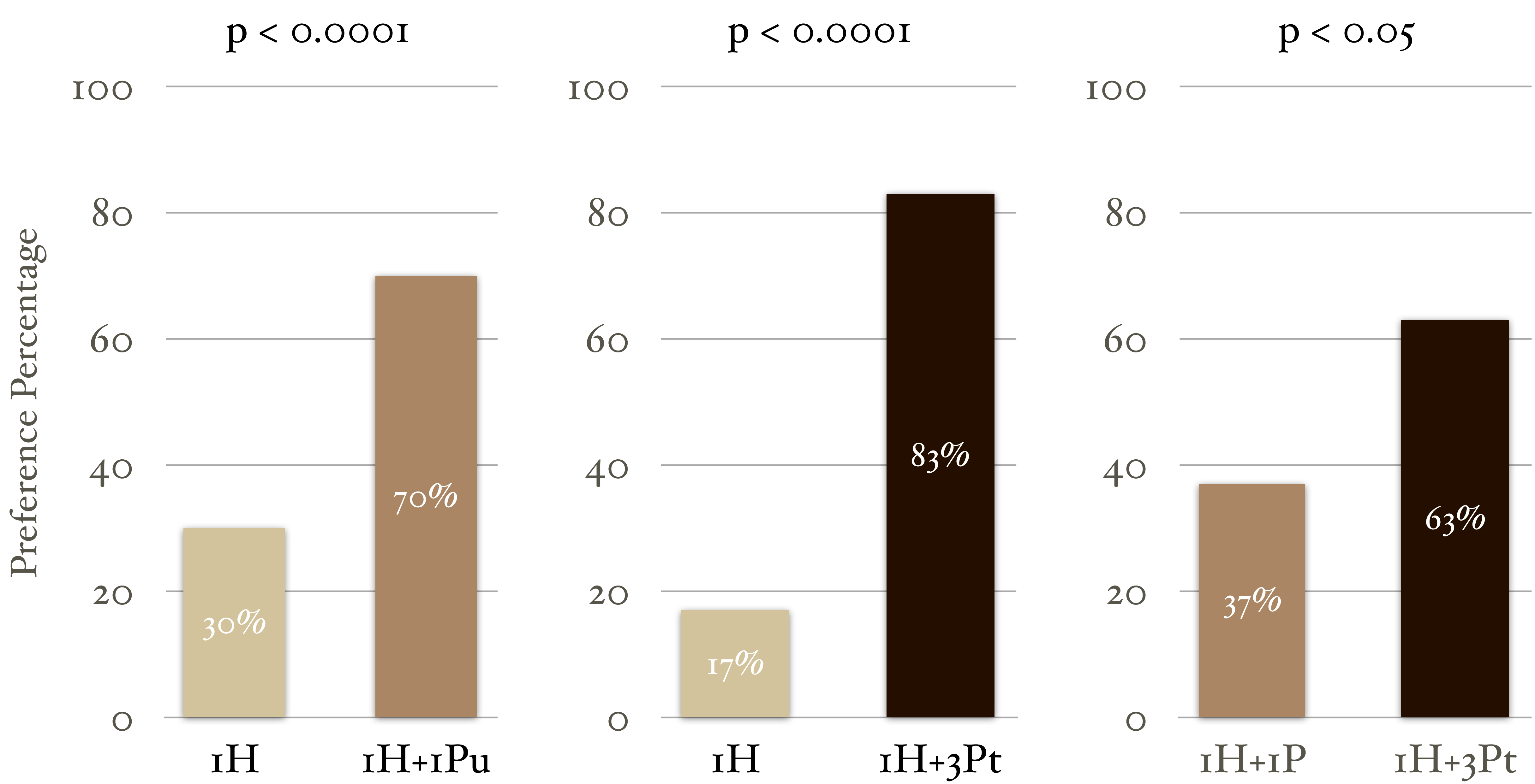


Pu: Untargeted, Pt: Targeted

HUMAN JUDGMENTS: SPANISH

Pu: Untargeted, Pt: Targeted

HUMAN JUDGMENTS: SPANISH



Pu: Untargeted, Pt: Targeted

RELATED MT-PARAPHRASING WORK

- Kauchak & Barzilay used MT output to change the reference†
  - Goal: Create a paraphrased reference more useful for *evaluation*
  - Only a *single* paraphrase instead of *k*-best
  - Paraphrasing effected via machinery completely unrelated to SMT
  - Only lexical paraphrasing
  - Required WordNet for synonyms

†David Kauchak & Regina Barzilay. *Paraphrasing for Automatic Evaluation.* HLT/NAACL 2006.

UNTARGETED PARAPHRASES

We must bear in mind the community as a whole.
We must remember the wider community.

They should be better coordinated and more effective.
They should improve the coordination and efficacy.

Women are still one of the most vulnerable sections of society, whose rights are rudely trampled underfoot by the current social and economic system.
They remain one of the weakest in society, whose duties are abruptly scorned by the present social and economic order.

That is what we are waiting to hear from the European Commission.
That is what we expected from the meeting.

This occurred not far away and not very long ago.
This substances not far behind and very recently.

Pivot Language: French

TRANSLATION EXAMPLES: FRENCH

S - N'empêche qu'il existe suffisamment de raisons de se procurer un lecteur indépendant.

O - In spite of this, there are many reasons to get a separate MP3 player.

T_b - Despite that it sufficiently exists of reason for providing an independent player.

T_u - But there are plenty of reasons to get an independent player.

S - Celui qui croît en Dieu ressent-il moins la douleur ?

O - Does it hurt less if you believe in God?

T_b - Anyone believes in God has less pain?

T_t - Whoever believes in God, does he feel less pain?

S: Source, **O**: Original Reference, **T_b**: Baseline translation, **$T_{u|t}$**: Translation with untargeted|targeted paraphrase

TRANSLATION EXAMPLES: GERMAN

S - Eine Ratte oder eine Schabe flieht bei Gefahr heißt das, dass sie auch Furcht empfindet?

O - When in danger, a rat or roach will run away. Does it mean they experience fear, too?

$\mathbf{T_b}$ - A rat or a Schabe flees by danger that means that they also feel fears?

$\mathbf{T_u}$ - A rat or a cockroach is fleeing when in danger, that means that they felt fear?

S - Nach dem steilen Abfall am Morgen konnte die Prager Börse die Verluste korrigieren.

O - After a sharp drop in the morning, the Prague Stock Market corrected its losses.

$\mathbf{T_b}$ - After the steep waste at tomorrow the Prague stock exchange cannot correct the losses.

$\mathbf{T_t}$ - After the steep waste in the morning, the Prague Stock Exchange losses corrected.

S: Source, **O**: Original Reference, $\mathbf{T_b}$: Baseline translation, $\mathbf{T_{u|t}}$: Translation with untargeted|targeted paraphrase

PARAPHRASING BEYOND TRANSLATION

PARAPHRASING BEYOND TRANSLATION

- I presented a novel and general sentential paraphrasing architecture that is entirely data-driven and built using existing machinery

PARAPHRASING BEYOND TRANSLATION

- I presented a novel and general sentential paraphrasing architecture that is entirely data-driven and built using existing machinery

- Can be used for applications that require higher quality paraphrases: use better features or trade off coverage by using more pivot languages

PARAPHRASING BEYOND TRANSLATION

- I presented a novel and general sentential paraphrasing architecture that is entirely data-driven and built using existing machinery

- Can be used for applications that require higher quality paraphrases: use better features or trade off coverage by using more pivot languages

- I have also worked on a paraphrase-enhanced MT evaluation metric (TERp) which can also be employed for paraphrase recognition †

†M. Snover, N. Madnani, B, Dorr and R. Schwartz. *TER-plus: Paraphrase, Semantic, and Alignment Enhancements to Translation Edit Rate*. *Machine Translation*. 23(2-3), 2009

PARAPHRASING BEYOND TRANSLATION

- I presented a novel and general sentential paraphrasing architecture that is entirely data-driven and built using existing machinery

- Can be used for applications that require higher quality paraphrases: use better features or trade off coverage by using more pivot languages

- I have also worked on a paraphrase-enhanced MT evaluation metric (TERp) which can also be employed for paraphrase recognition †

- Note that the paraphraser is essentially an SCFG parser

†M. Snover, N. Madnani, B, Dorr and R. Schwartz. *TER-plus: Paraphrase, Semantic, and Alignment Enhancements to Translation Edit Rate*. *Machine Translation*. 23(2-3), 2009

PARAPHRASING BEYOND TRANSLATION

- I presented a novel and general sentential paraphrasing architecture that is entirely data-driven and built using existing machinery
- Can be used for applications that require higher quality paraphrases: use better features or trade off coverage by using more pivot languages
- I have also worked on a paraphrase-enhanced MT evaluation metric (TERp) which can also be employed for paraphrase recognition [†]
- Note that the paraphraser is essentially an SCFG parser
  - So far, I have talked about generation: parse source sentence using one side of monolingual grammar and read off target tree

[†]M. Snover, N. Madnani, B, Dorr and R. Schwartz. *TER-plus: Paraphrase, Semantic, and Alignment Enhancements to Translation Edit Rate*. *Machine Translation*. 23(2-3), 2009

PARAPHRASING BEYOND TRANSLATION

- I presented a novel and general sentential paraphrasing architecture that is entirely data-driven and built using existing machinery
- Can be used for applications that require higher quality paraphrases: use better features or trade off coverage by using more pivot languages
- I have also worked on a paraphrase-enhanced MT evaluation metric (TERp) which can also be employed for paraphrase recognition †
- Note that the paraphraser is essentially an SCFG parser
  - So far, I have talked about generation: parse source sentence using one side of monolingual grammar and read off target tree
  - With additional work, we can do recognition: synchronously parse two sentences with induced monolingual grammar

†M. Snover, N. Madnani, B, Dorr and R. Schwartz. *TER-plus: Paraphrase, Semantic, and Alignment Enhancements to Translation Edit Rate*. *Machine Translation*. 23(2-3), 2009