Improving Lexical Choice in Neural Machine Translation

Toan Q. Nguyen & David Chiang

Overview

- Problem: Rare word mistranslation
- Model 1: softmax → cosine similarity (**fixnorm**)
- Model 2: direction connections (**fixnorm+lex**)
- Experiments & Results

Rare word mistranslation

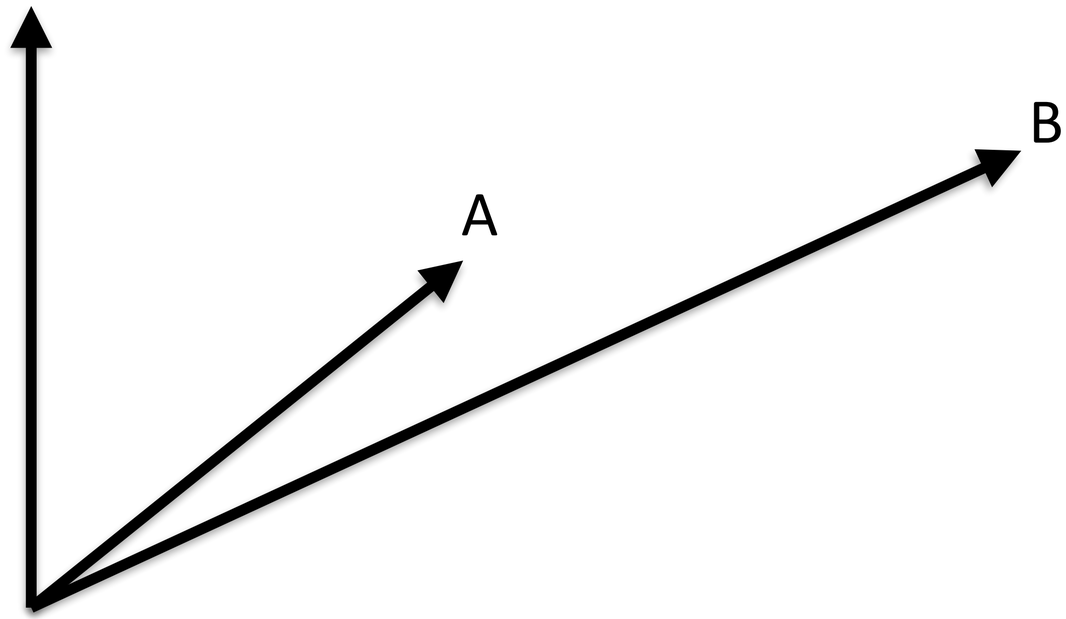
src	Ammo muammolar hali ko’p, deydi amerikalik olim **Entoni Fauchi**.
ref	But still there are many problems, says American scientist **Anthony Fauci**.
NMT	But there is still a lot of problems, says **James Chan**.

NMT tends to translate words that “*seem natural in the context, but do not reflect the content of the source sentence*”(Arthur et al., 2016).

Softmax

“query” vector

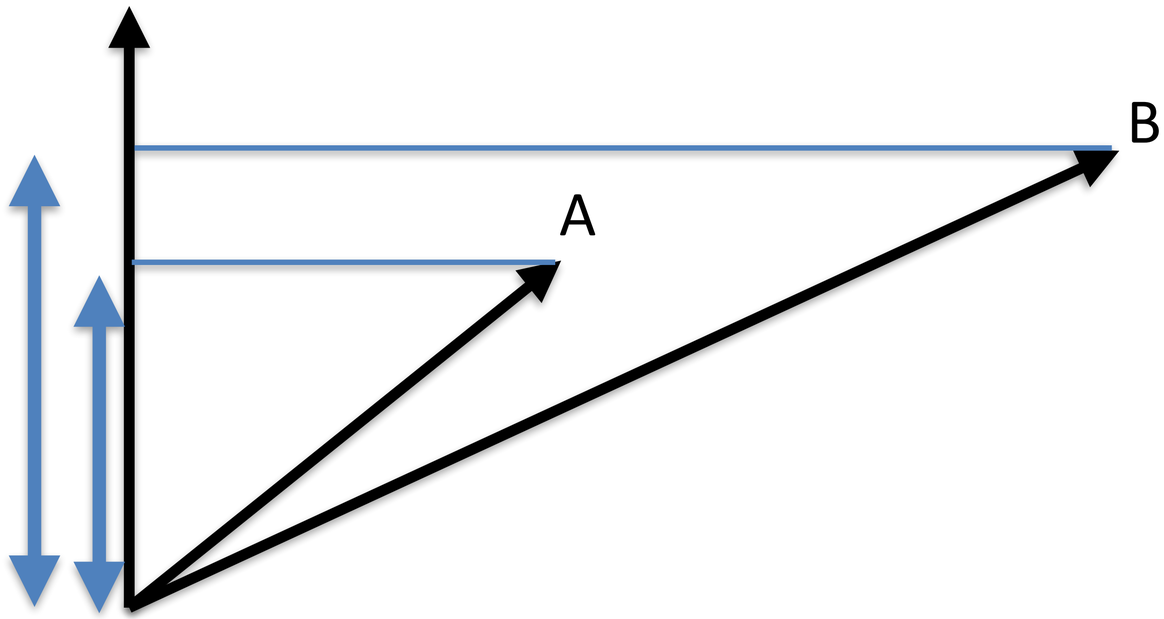
$\tilde{h}$



Softmax using dot product

“query” vector

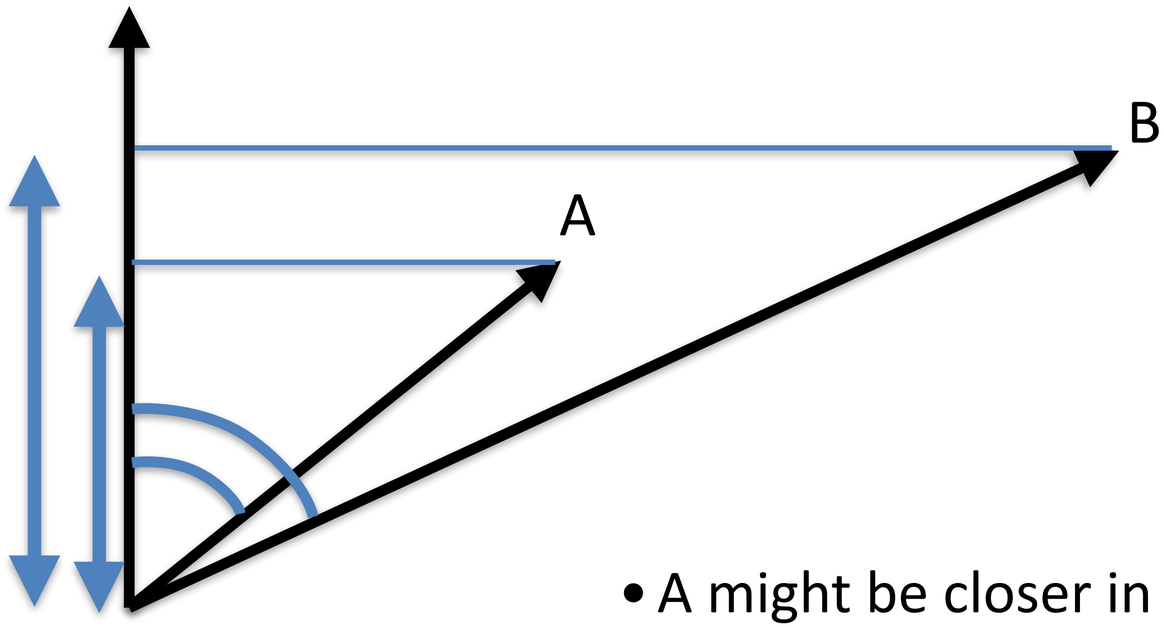
$\tilde{h}$



Softmax using dot product

"query" vector

$\tilde{h}$

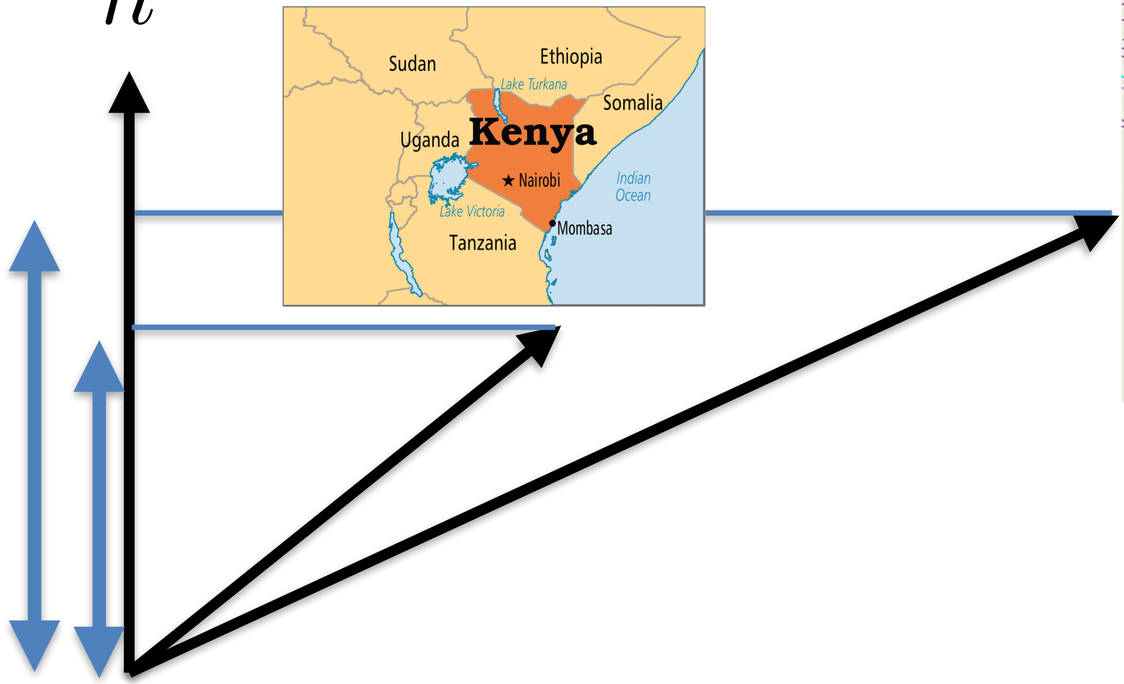


- A might be closer in direction
- still, B wins

Softmax using dot product

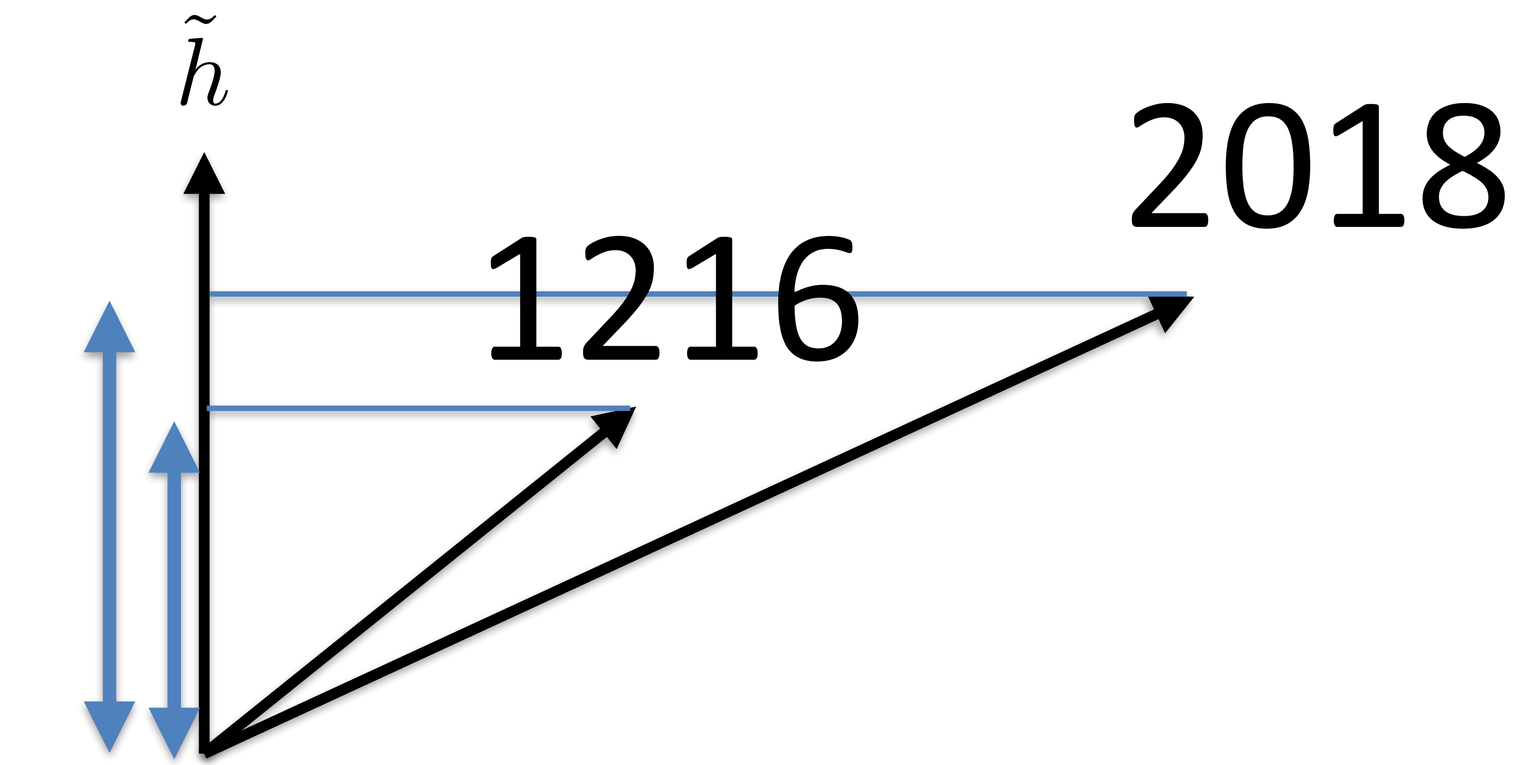
"query" vector

$\tilde{h}$



Softmax using dot product

“query” vector



Softmax using dot product

Source: Ammo muammolar hali ko’p, deydi amerikalik olim **Entoni Fauchi**.
Reference: But still there are many problems, says American scientist **Anthony Fauci**.
NMT: But there is still a lot of problems, says **James Chan**.

“query” vector

$\tilde{h}$

Anthony Fauci
Director NIAID

Margaret Chan
Former Director WHO

Softmax using cosine similarity (fixnorm)

Source: Ammo muammolar hali ko’p, deydi amerikalik olim **Entoni Fauchi**.
Reference: But still there are many problems, says American scientist **Anthony Fauci**.
NMT: But there is still a lot of problems, says **James Chan**.
Fixed-norm: But there is still a lot of problems , says American scientist **UNK UNK** .

“query” vector

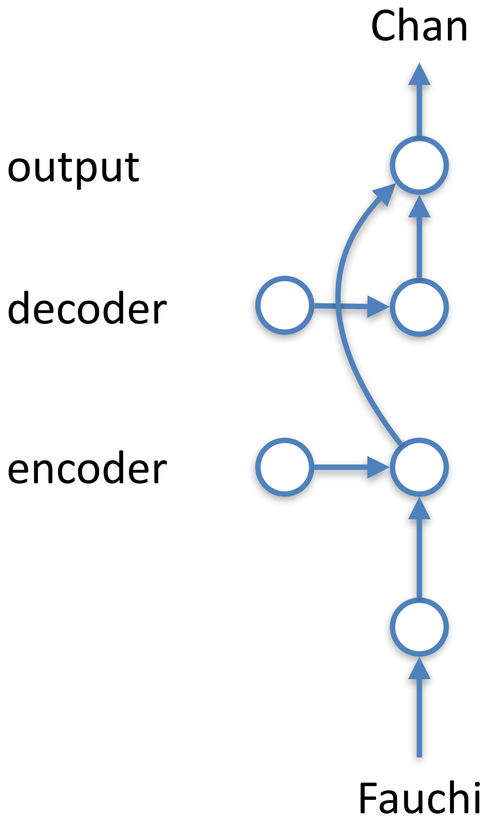
$\tilde{h}$

Anthony Fauci
Director NIAID

Margaret Chan
Former Director WHO

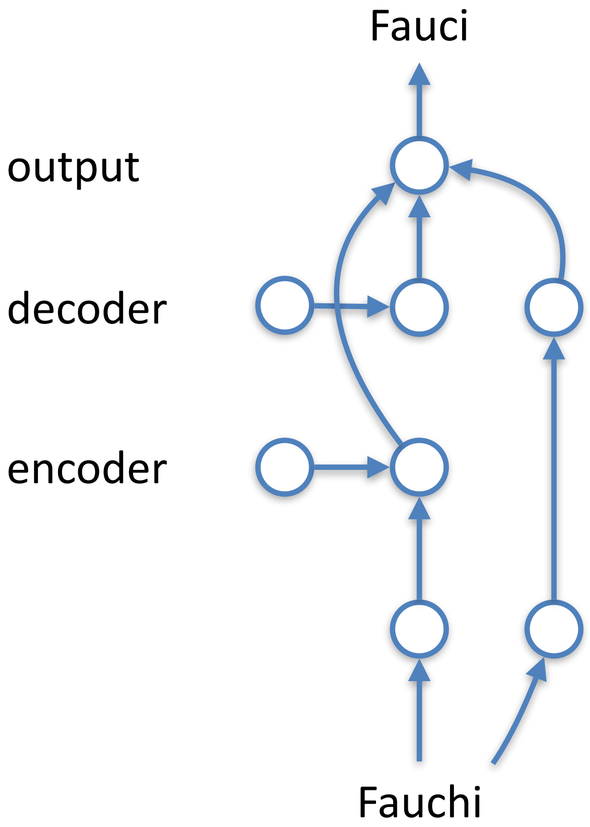
- Solution: **fix** magnitudes of vectors to constant with **weight normalization** Salimans and Kingma (2016)

Direct Connections (fixnorm+lex)



- Pro: Word choice depends on source and target context
- Con: Word choice depends on source and target context

Direct Connections (fixnorm+lex)

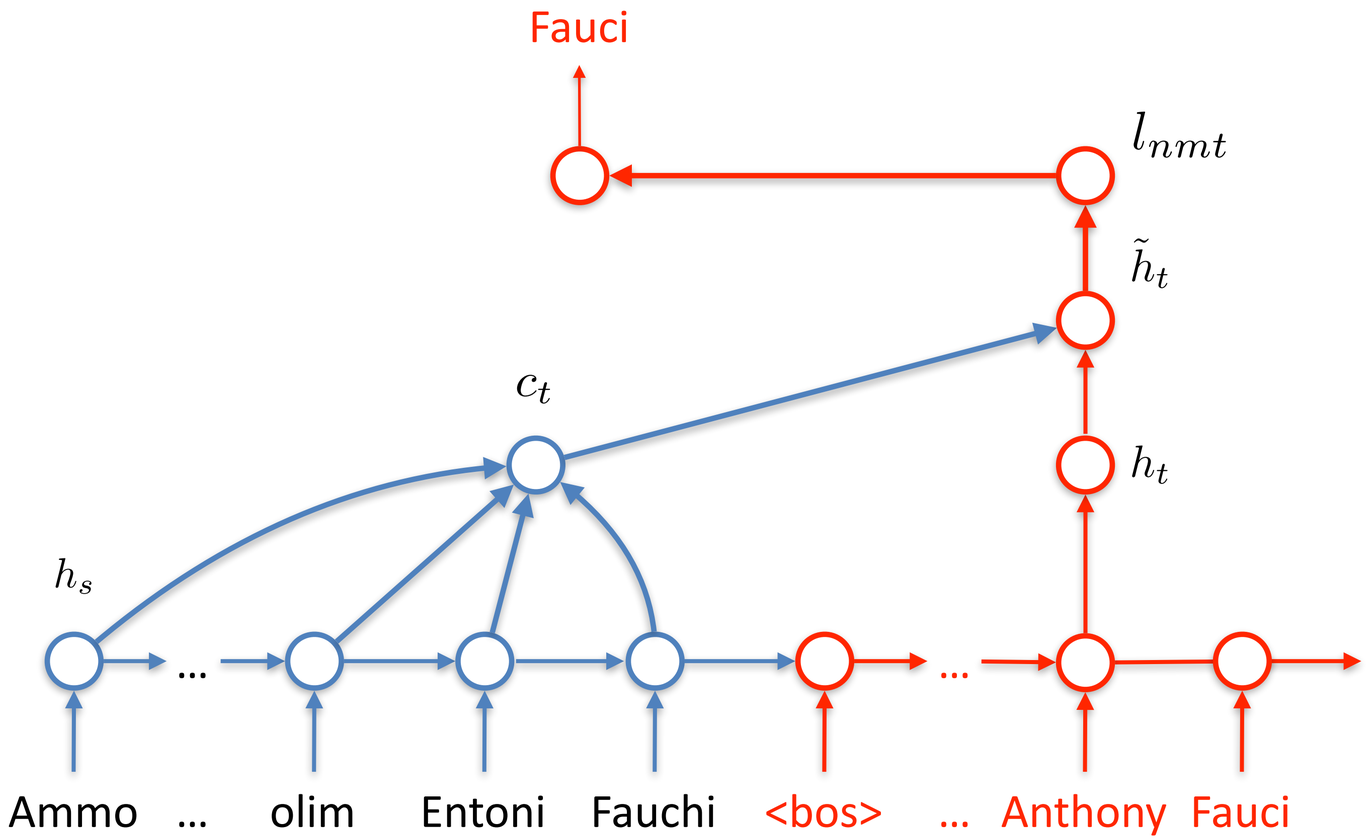


- Pro: Word choice depends on source and target context
- Con: Word choice depends on source and target context
- Solution: Add a more direct path that depends only on source word

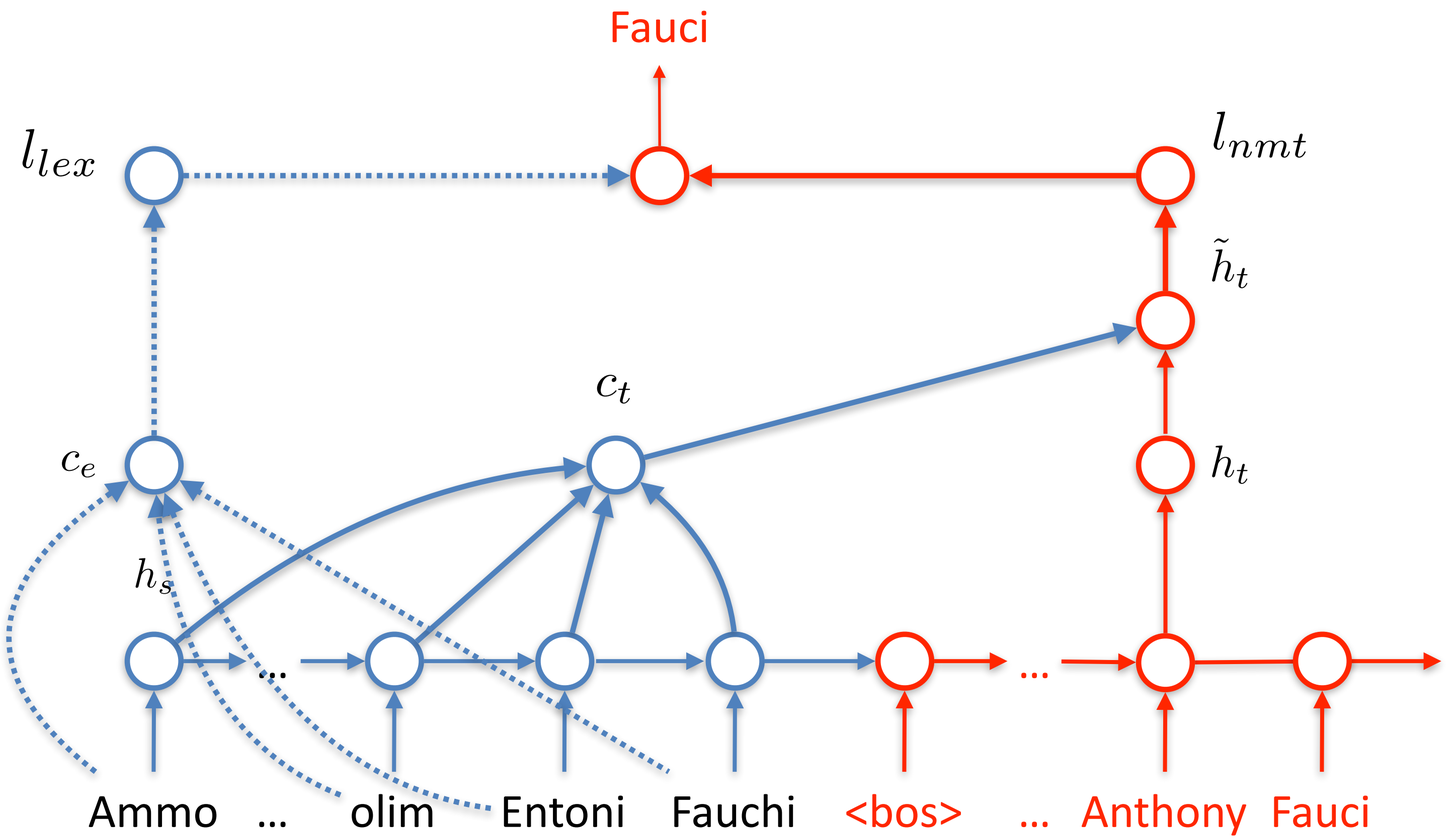
Reference: But still there are many problems, says American scientist **Anthony Fauci**.
Fixed-norm + lexical: But there are still problems, says American scientist **Anthony Fauci**.

Direct Connections (fixnorm+lex)

Direct Connections (fixnorm+lex)



Direct Connections (fixnorm+lex)



Experiments

We compare our models against the following baselines:

- Moses
- NMT + tied embedding (Inan et al., 2017; Press and Wolf, 2017)
- Arthur: NMT + tied embedding + discrete lexicon by Arthur et al. (2016)

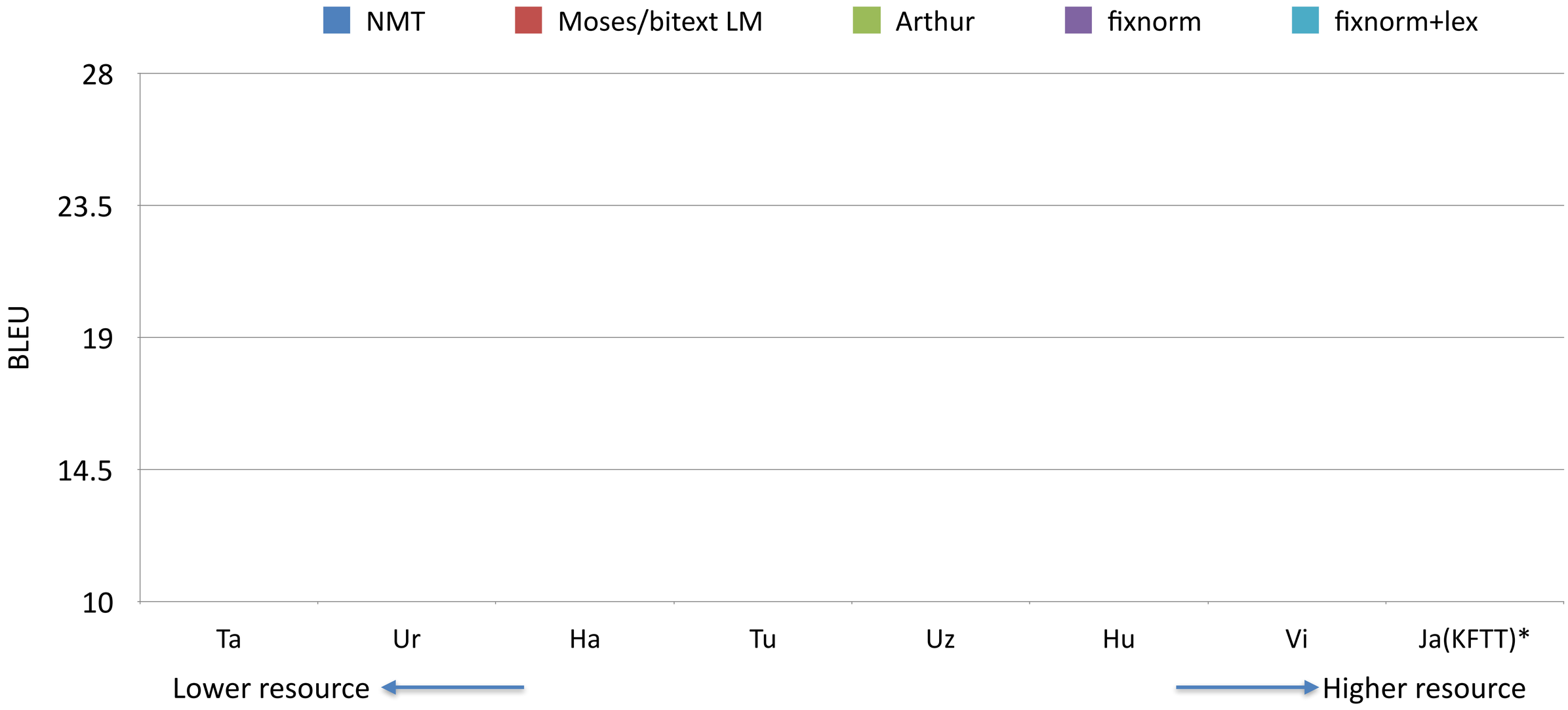
NMT systems: Global attention + general scoring + feed input (Luong et al., 2015a)

Datasets: 9 datasets, 8 language pairs, ranging from 0.1-8M words

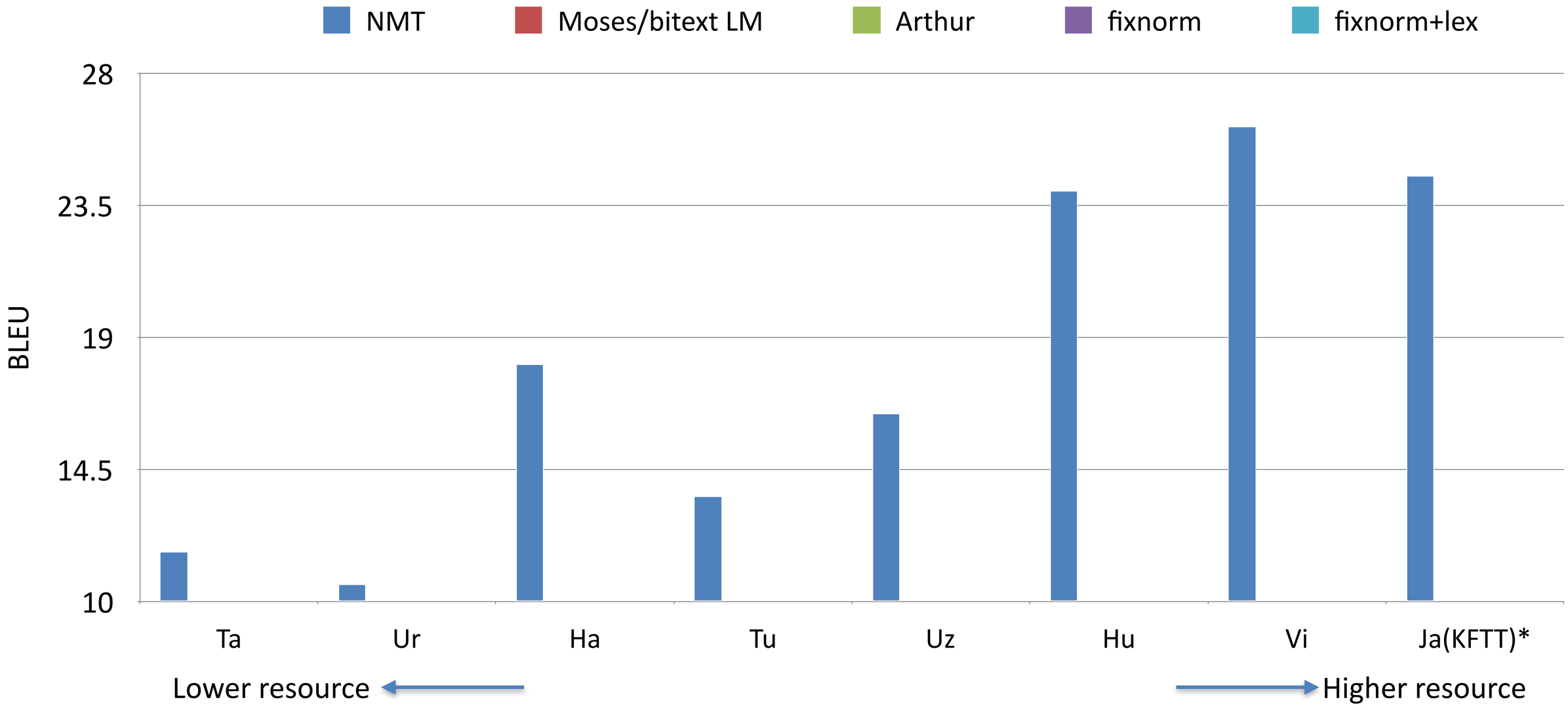
Training: Adadelta, dropout, select checkpoint based on dev BLEU…

Results (word-based)

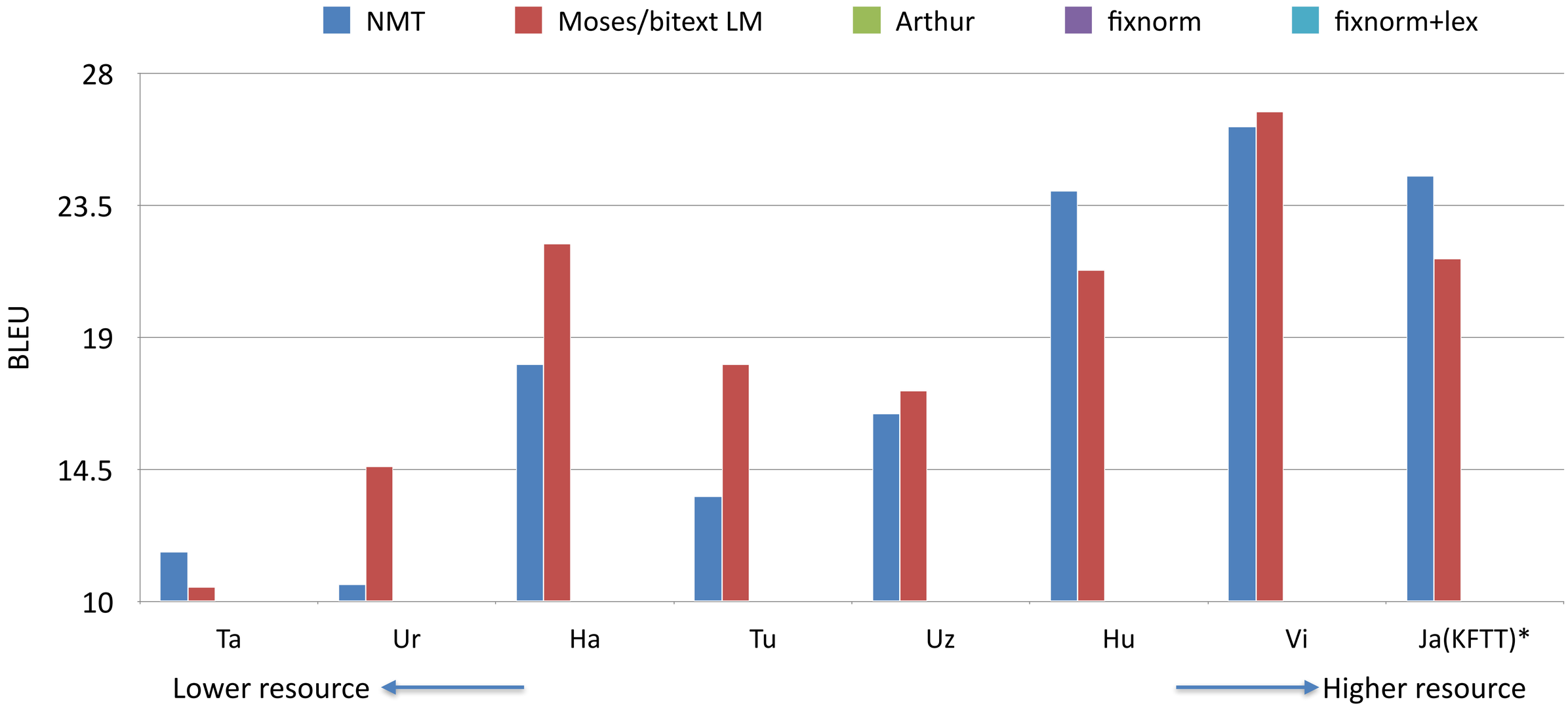
Results (word-based)



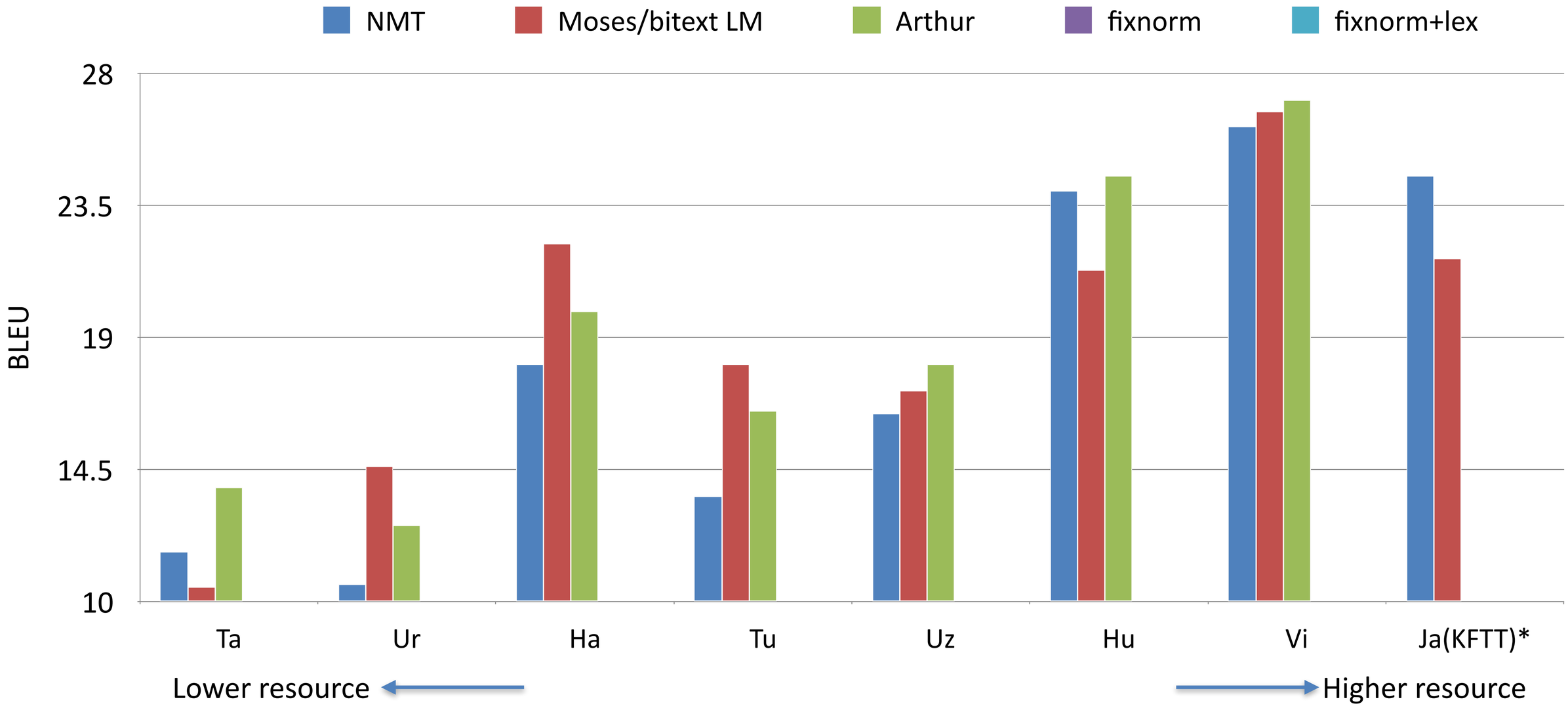
Results (word-based)



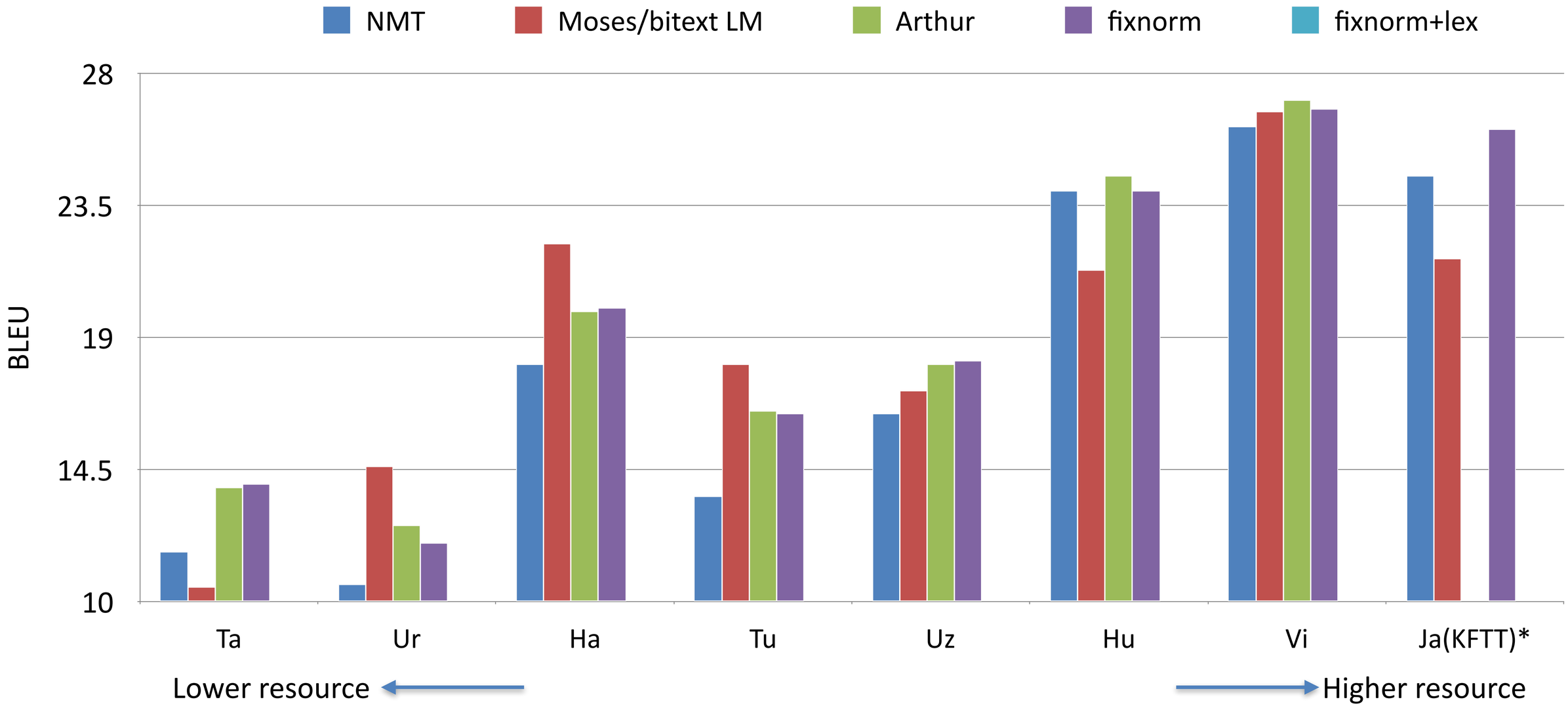
Results (word-based)



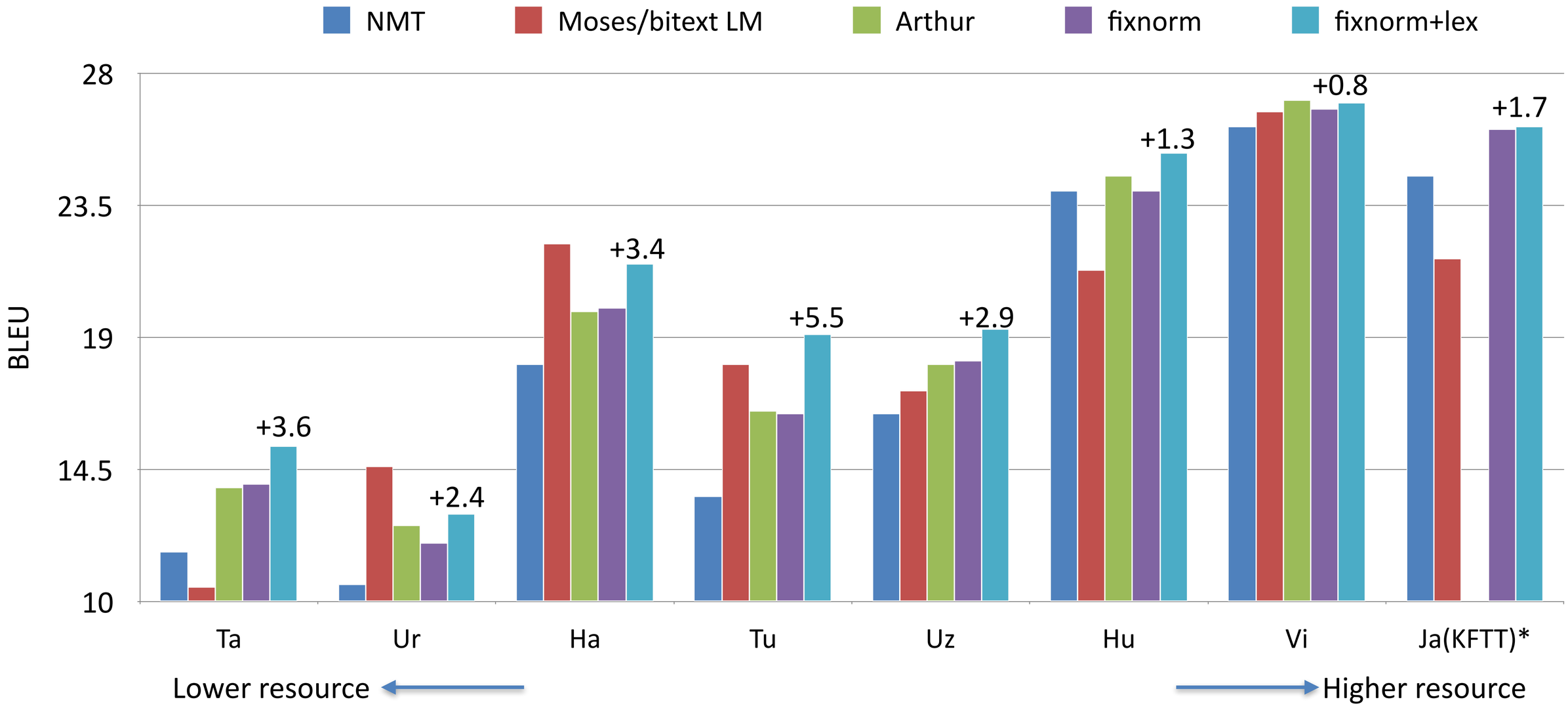
Results (word-based)



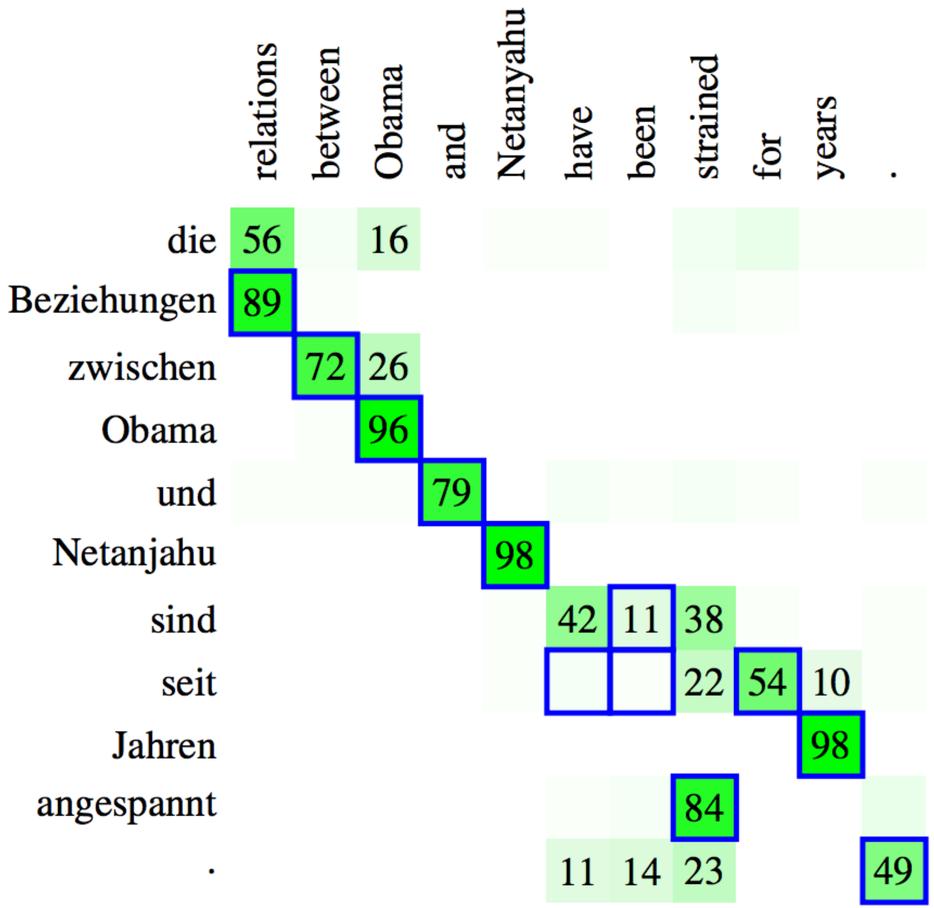
Results (word-based)



Results (word-based)

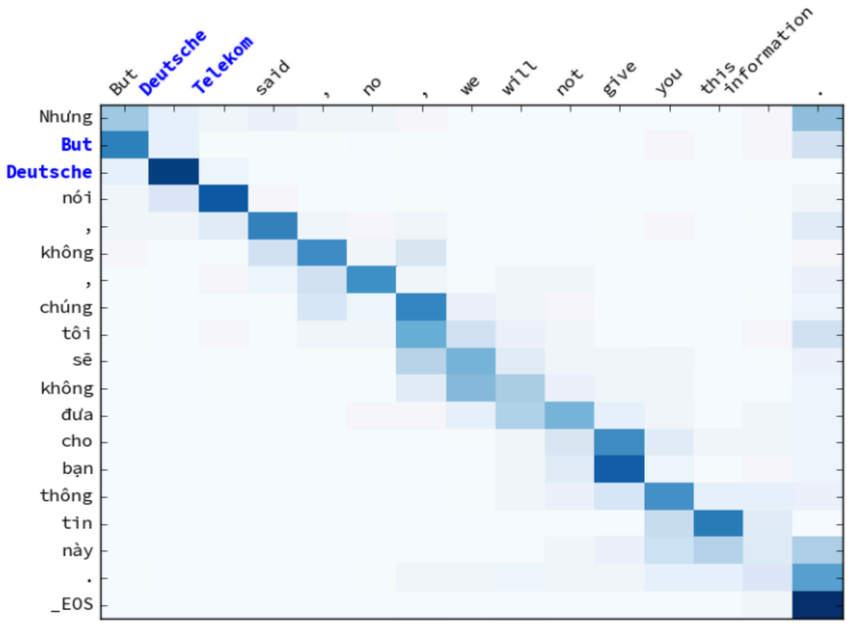


Alignments & UNK replacement

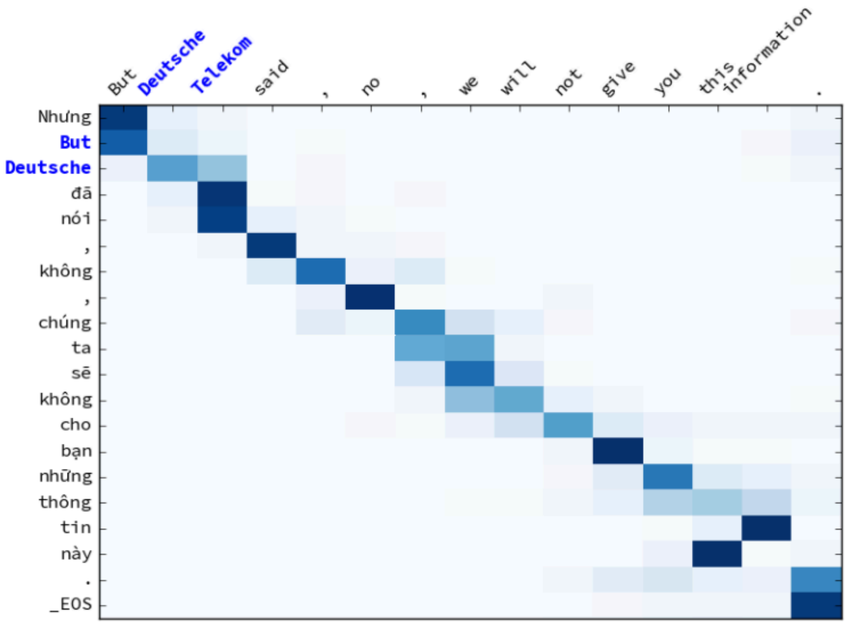


Koehn and Knowles, 2017

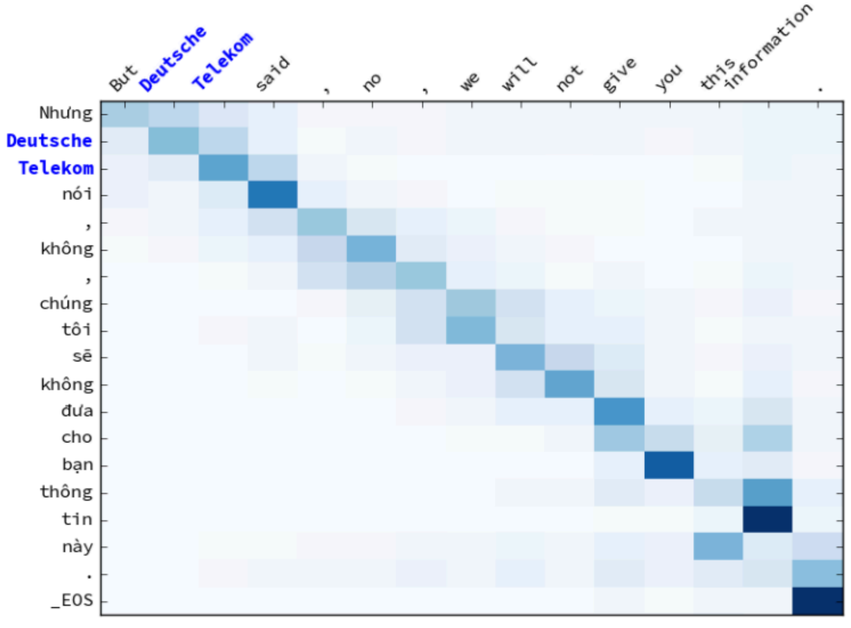
- Koehn and Knowles (2017): Alignments are sometimes shifted
- Could affect UNK replacement (Luong et al., 2015b)



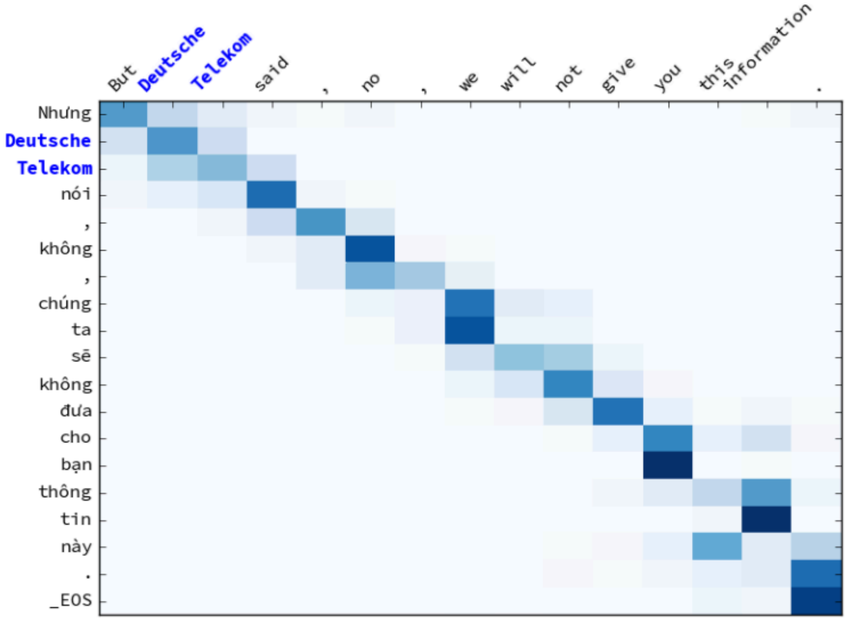
(a) **tied**



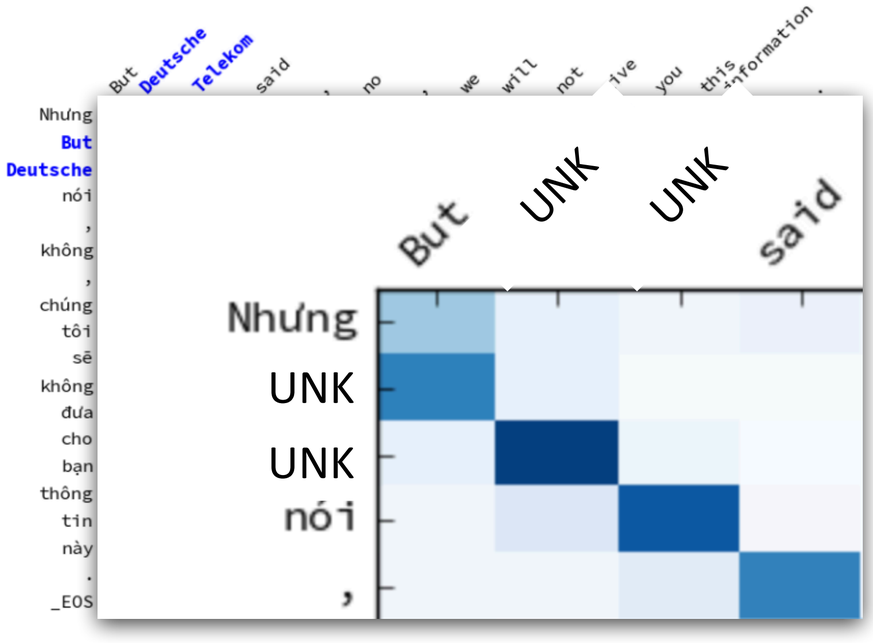
(b) **fixnorm**



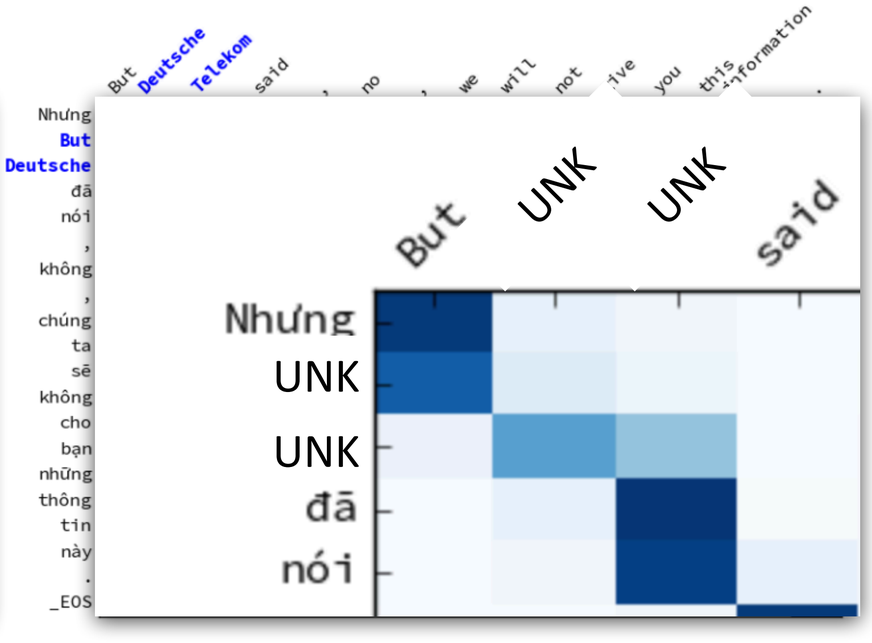
(c) **fixnorm+lex**



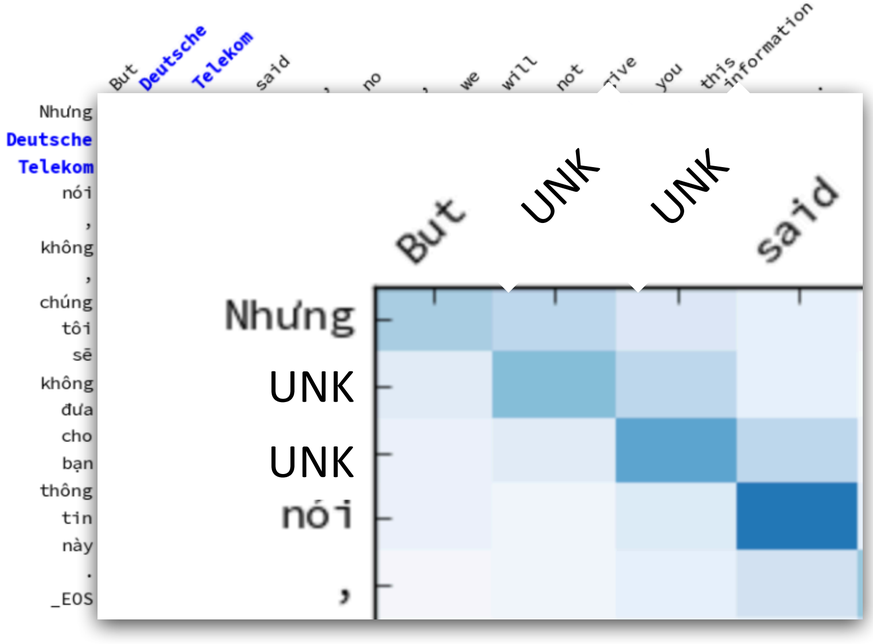
(d) Arthur et al. (2016)



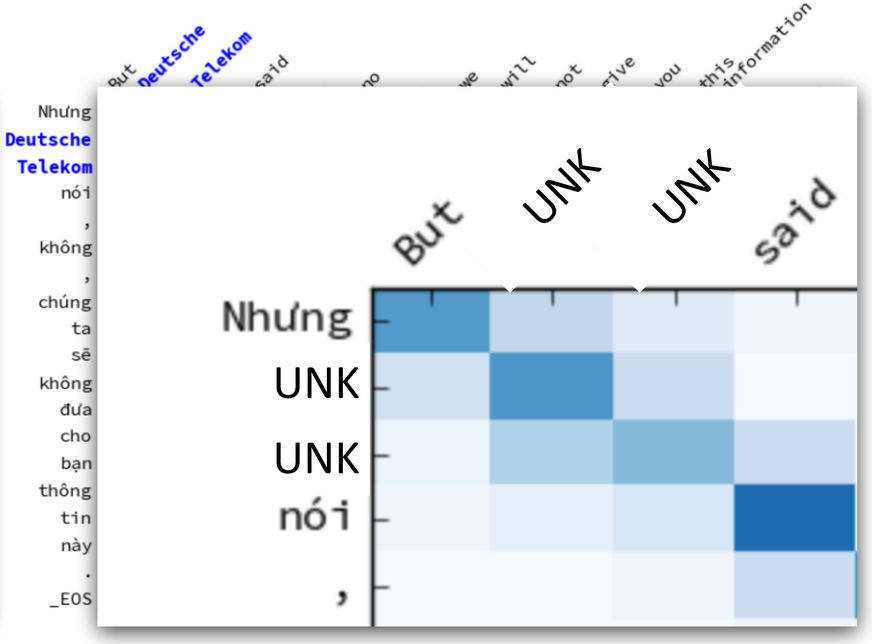
(a) **tied**



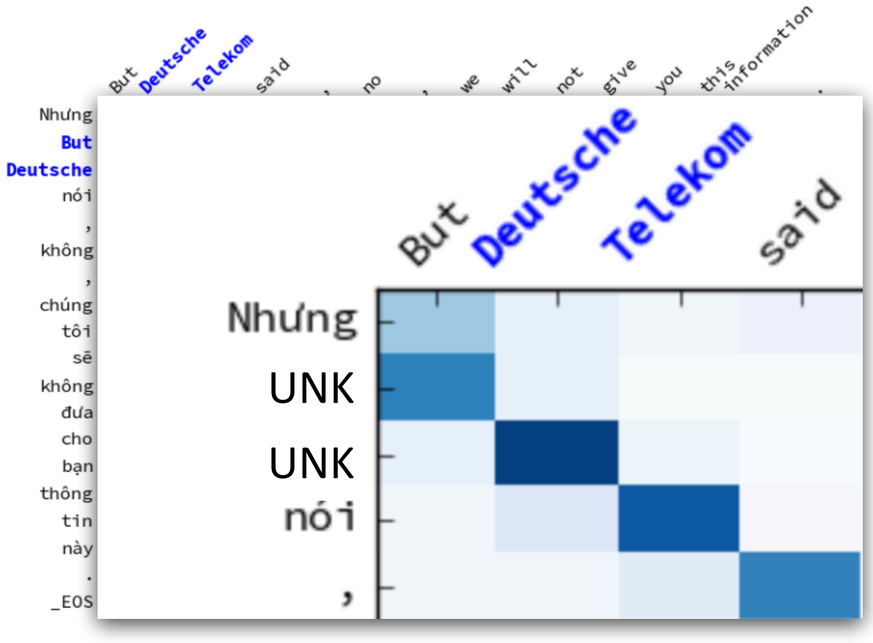
(b) **fixnorm**



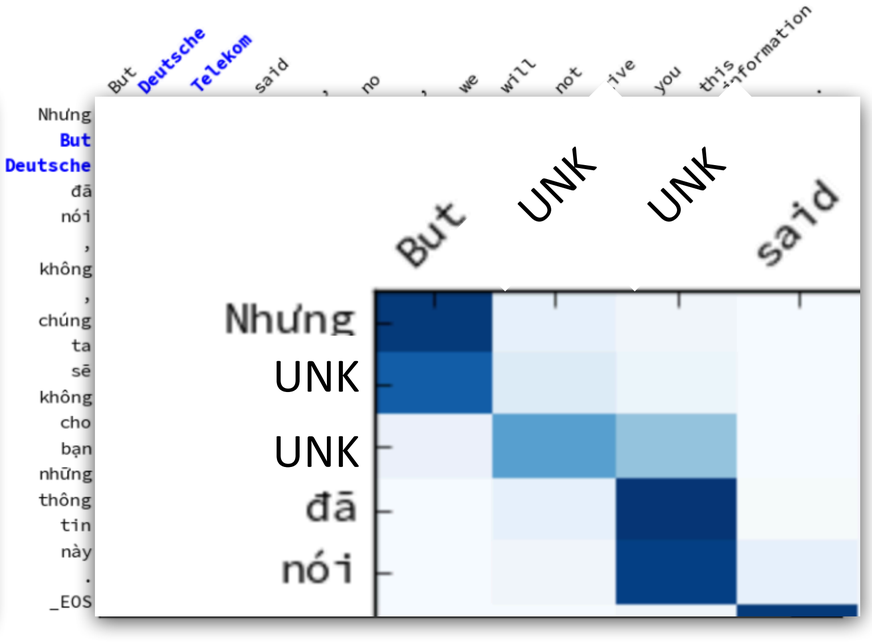
(c) **fixnorm+lex**



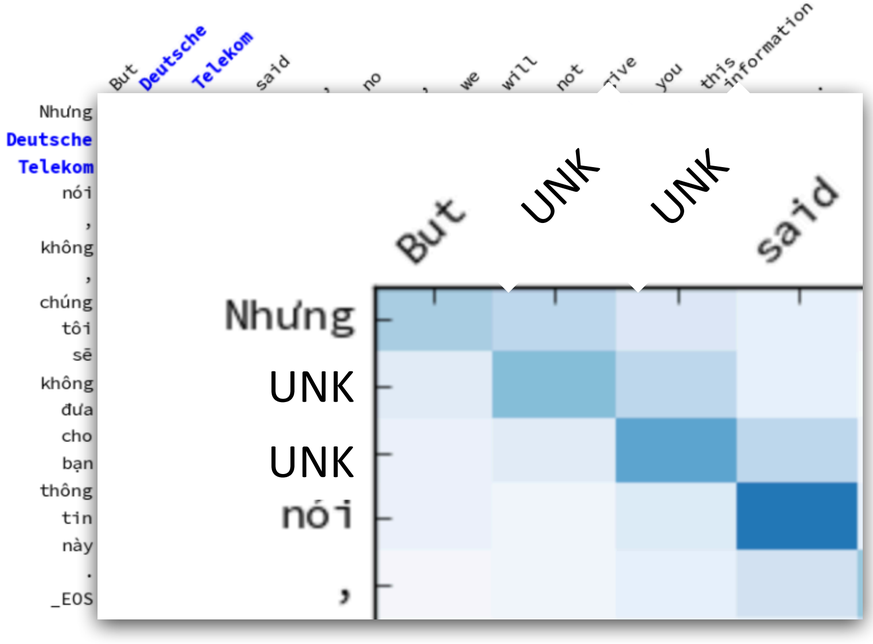
(d) Arthur et al. (2016)



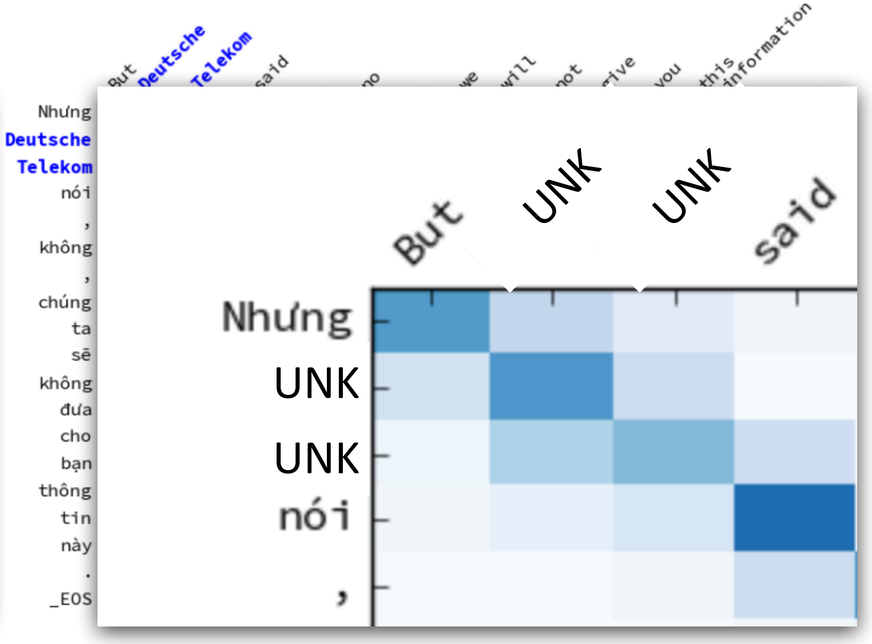
(a) **tied**



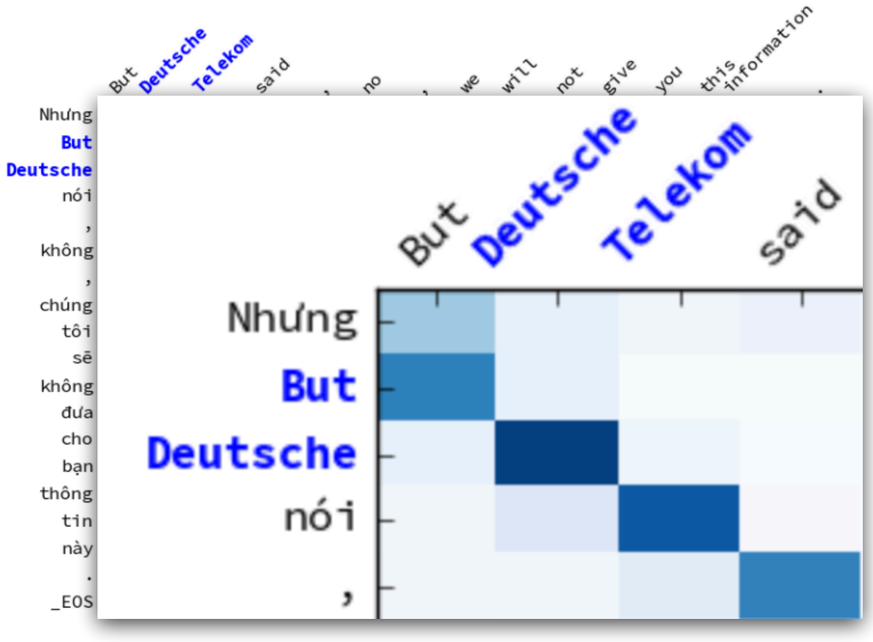
(b) **fixnorm**



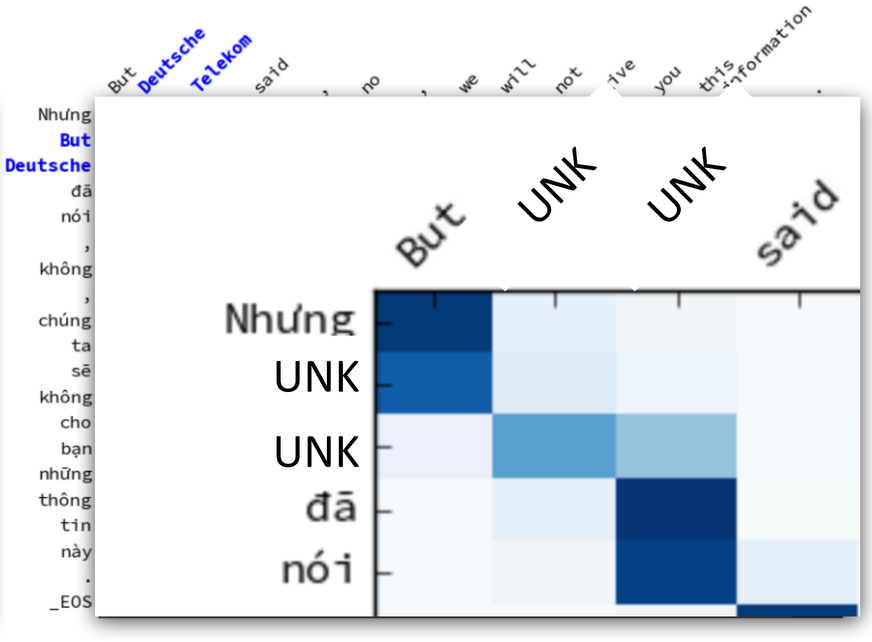
(c) **fixnorm+lex**



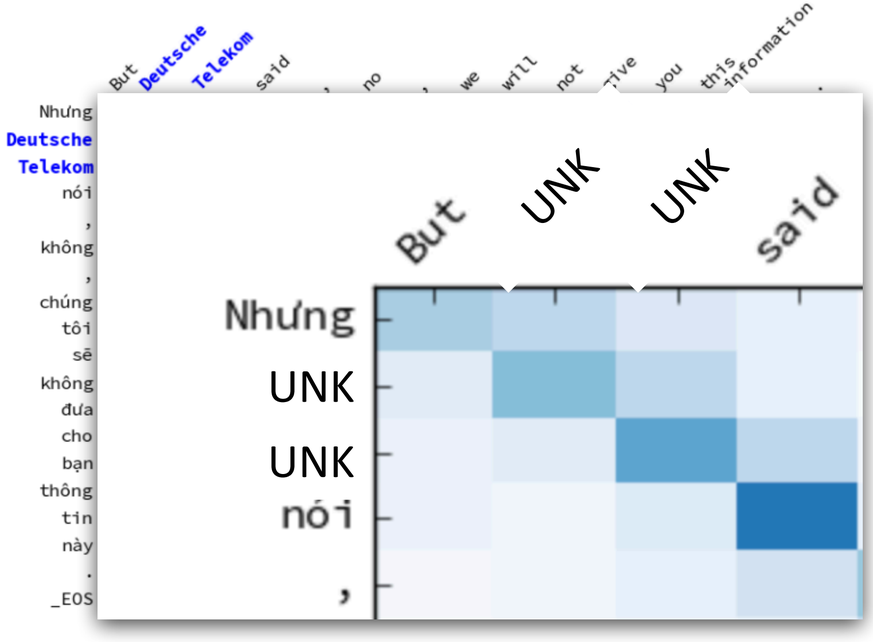
(d) Arthur et al. (2016)



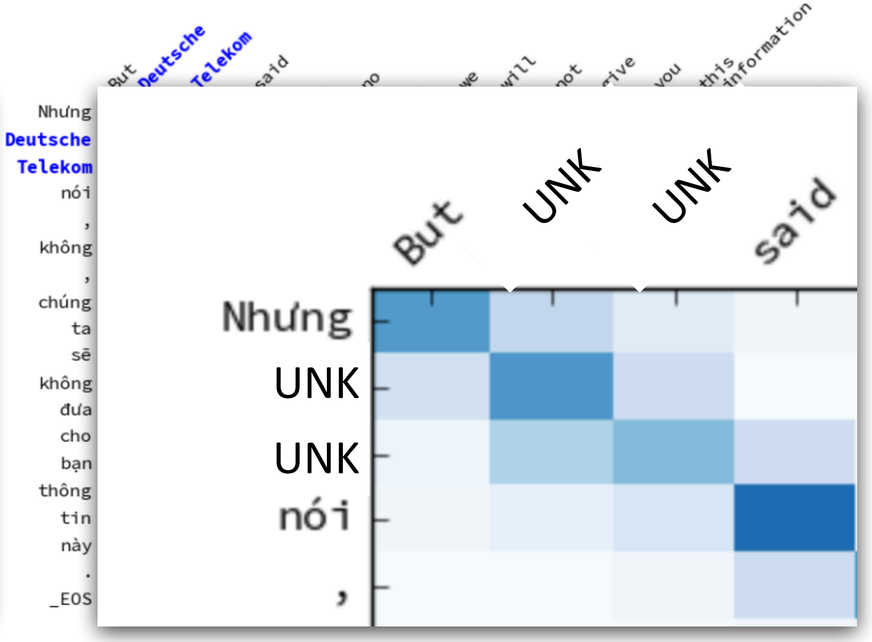
(a) **tied**



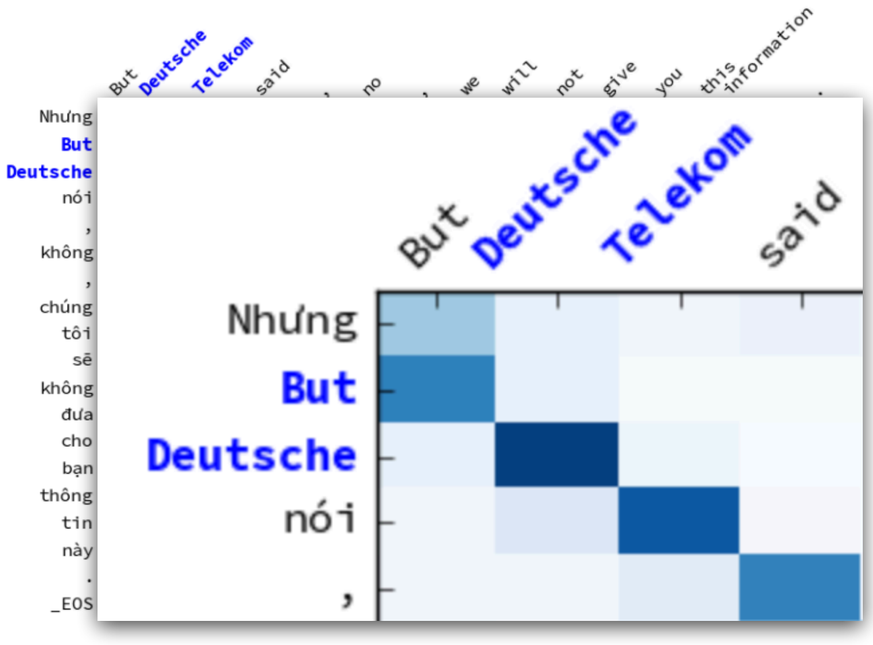
(b) **fixnorm**



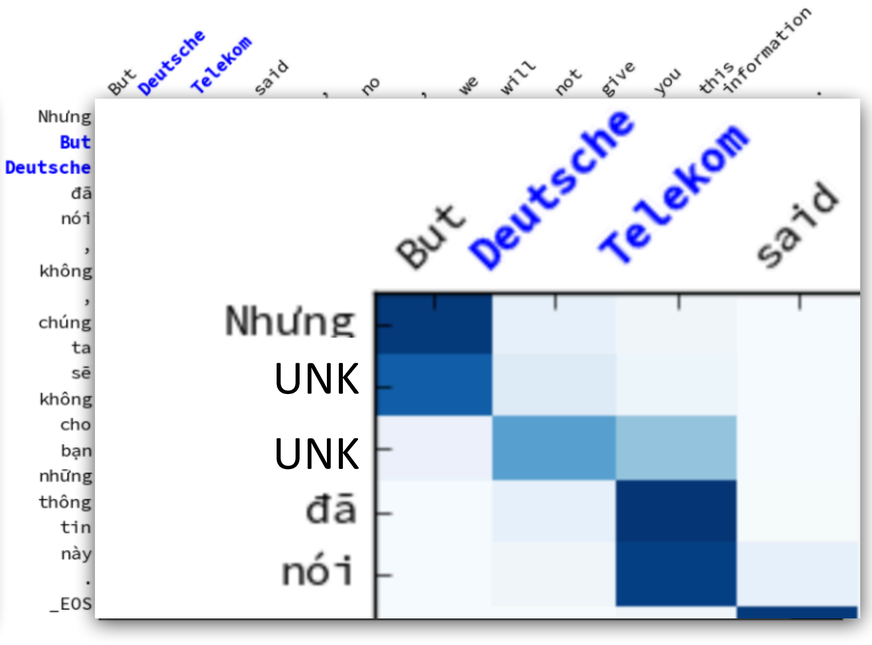
(c) **fixnorm+lex**



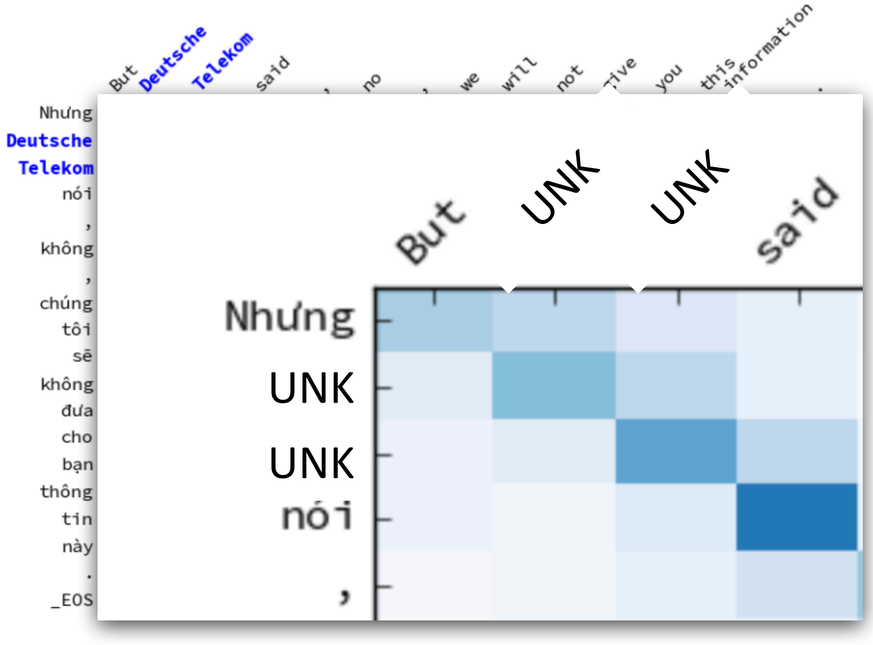
(d) Arthur et al. (2016)



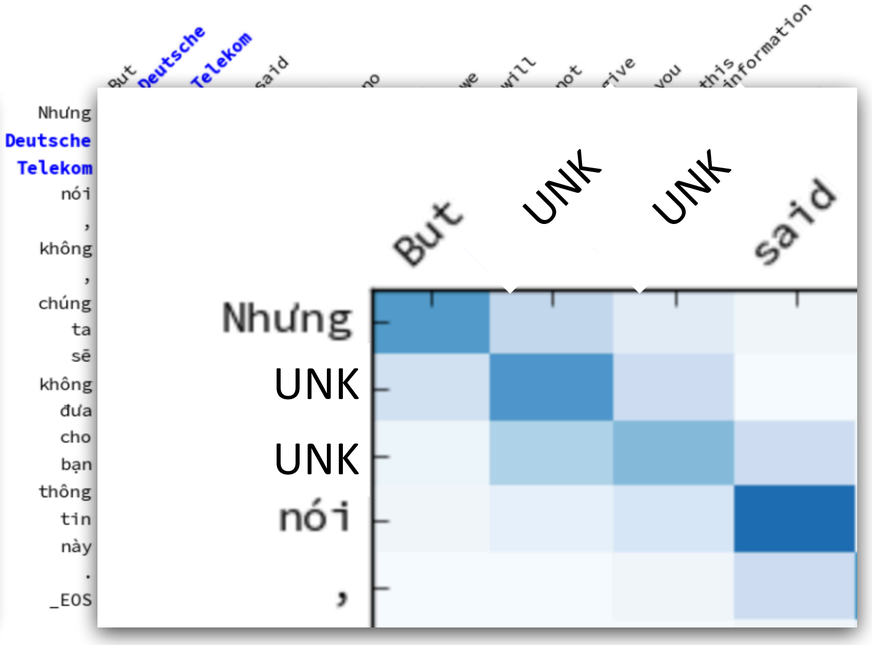
(a) **tied**



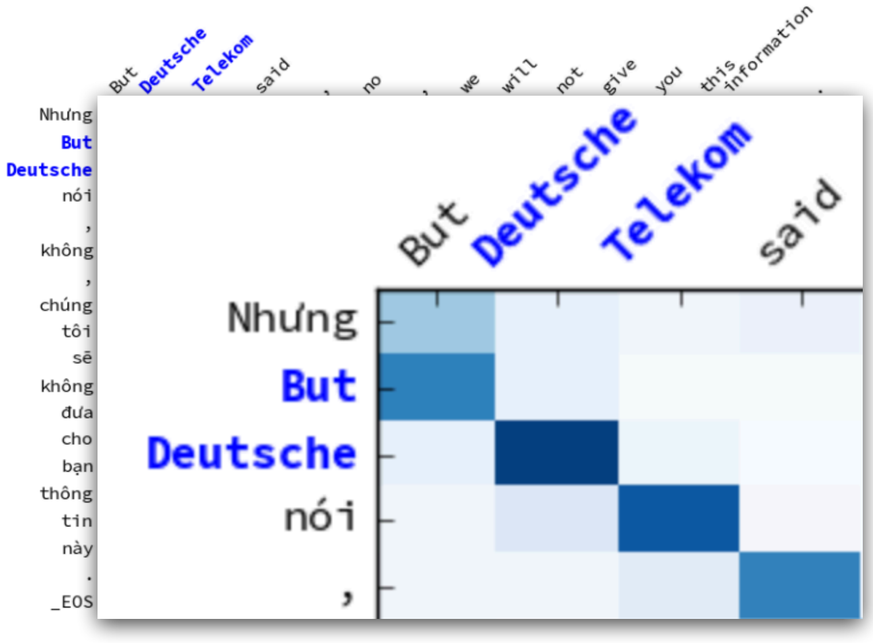
(b) **fixnorm**



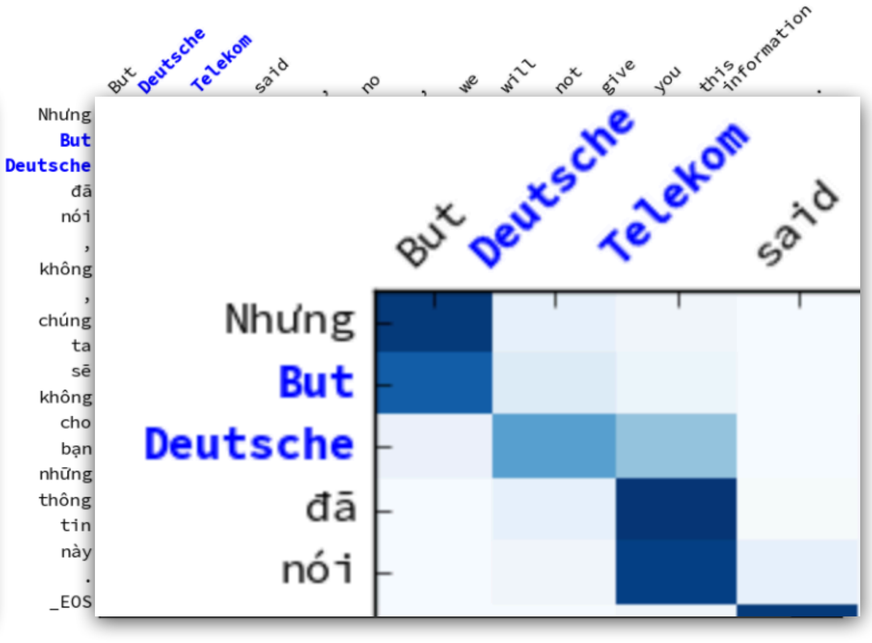
(c) **fixnorm+lex**



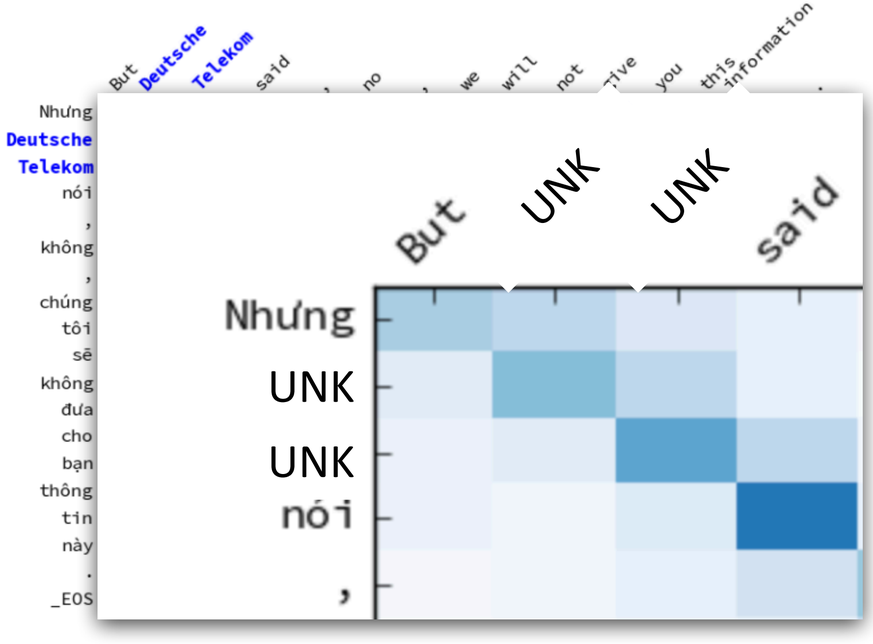
(d) Arthur et al. (2016)



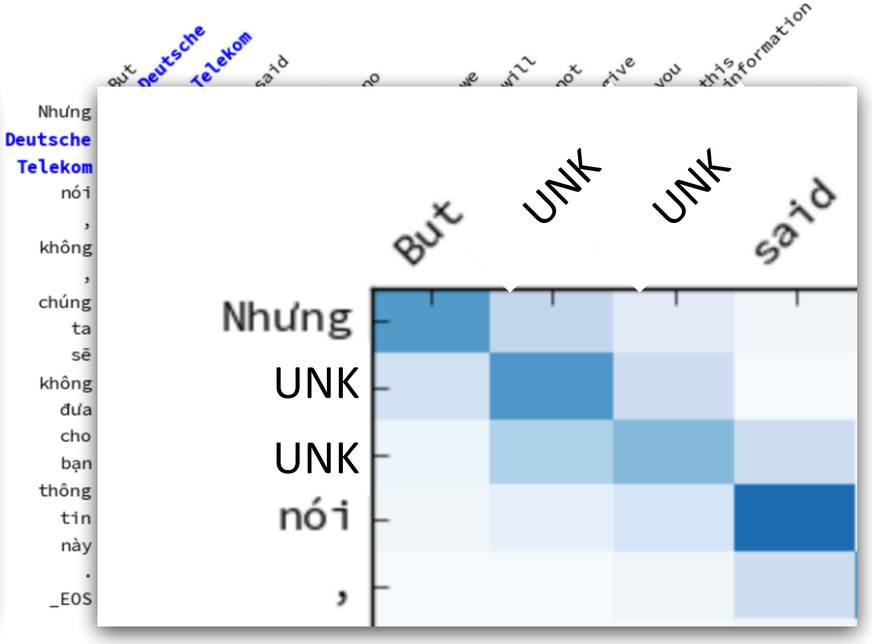
(a) **tied**



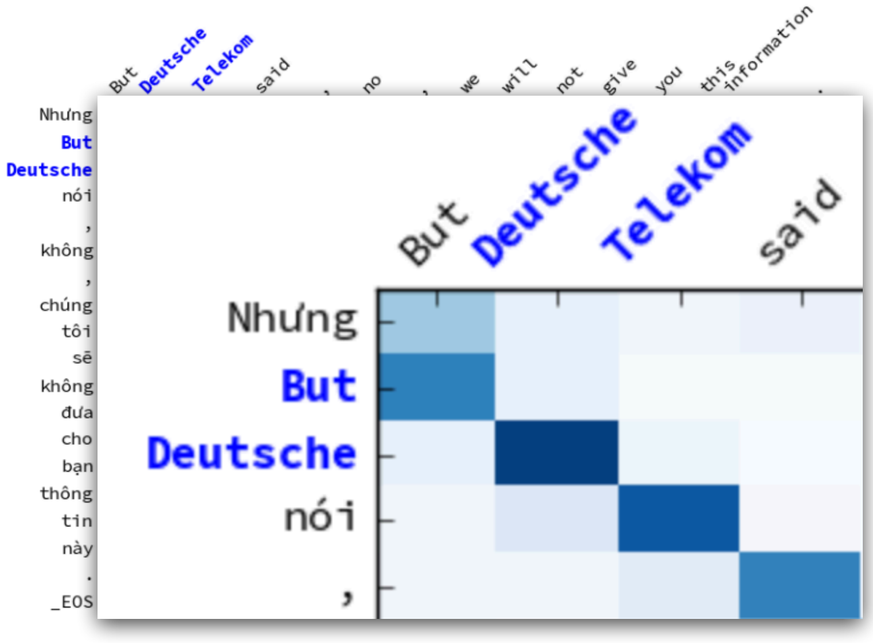
(b) **fixnorm**



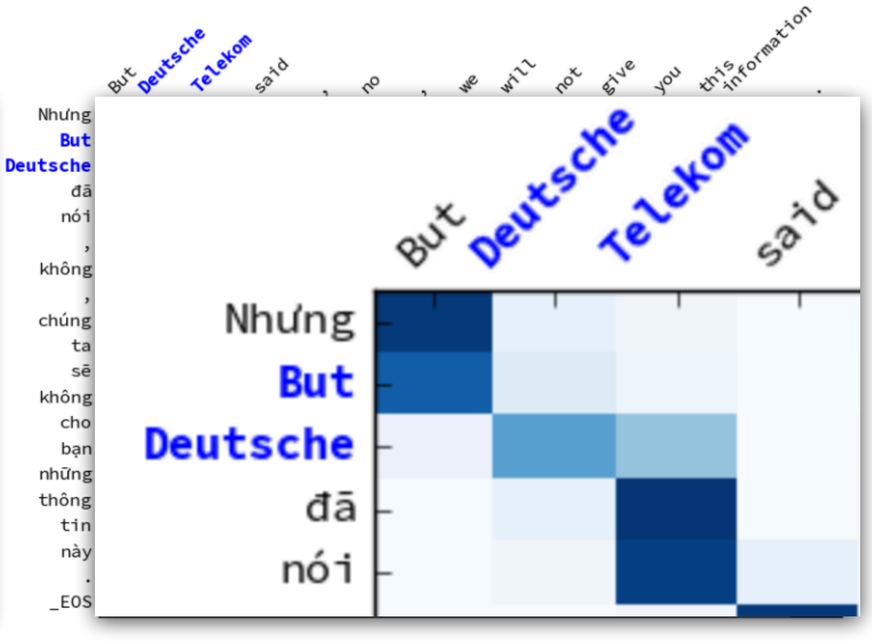
(c) **fixnorm+lex**



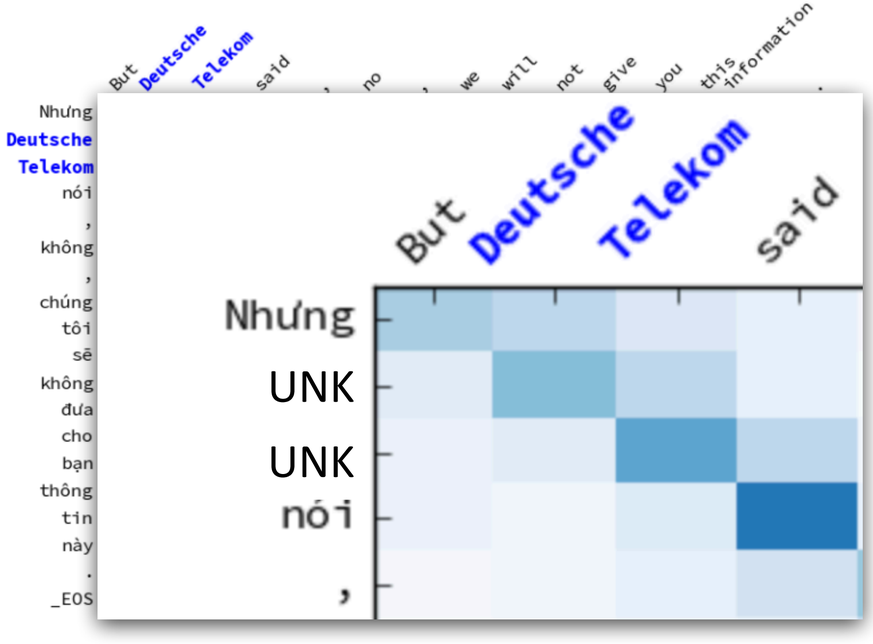
(d) Arthur et al. (2016)



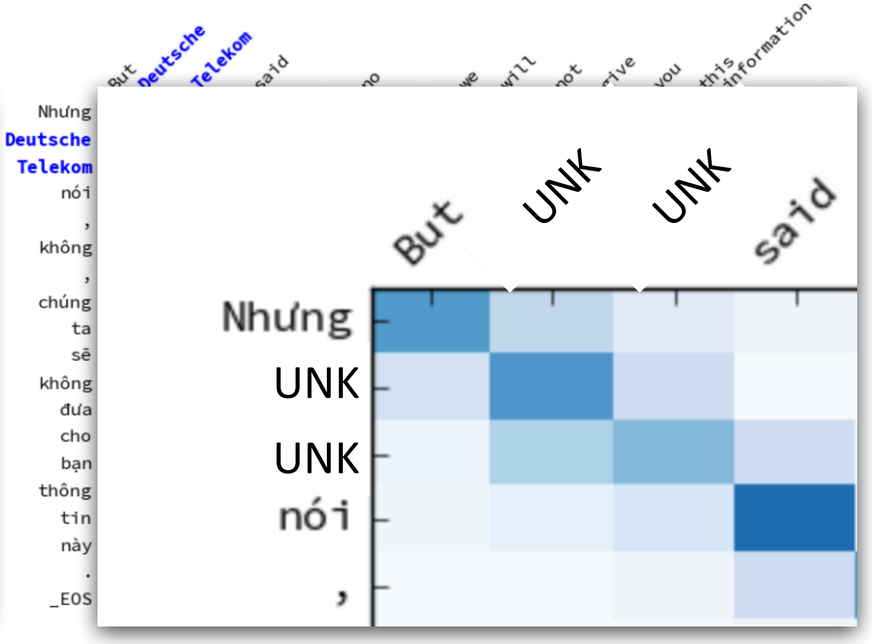
(a) **tied**



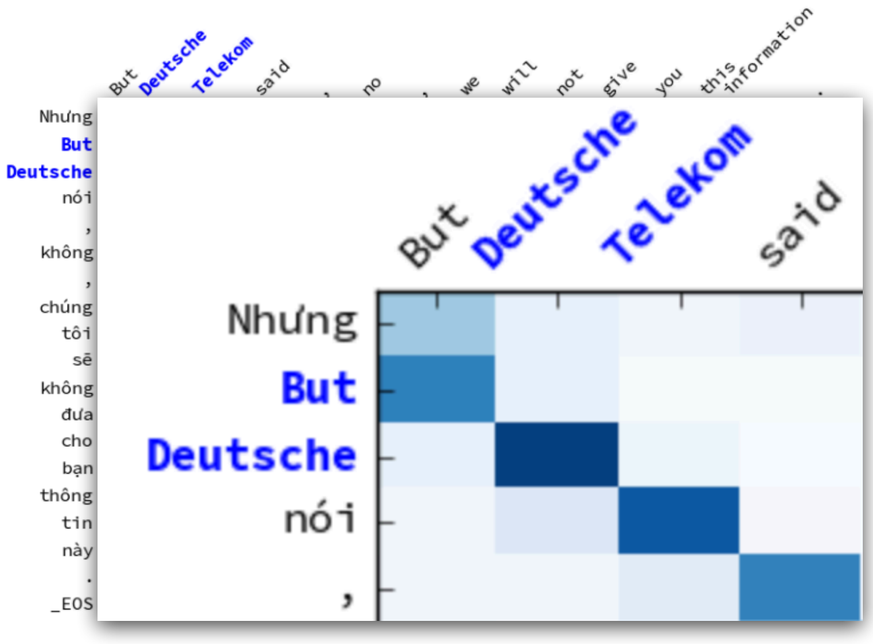
(b) **fixnorm**



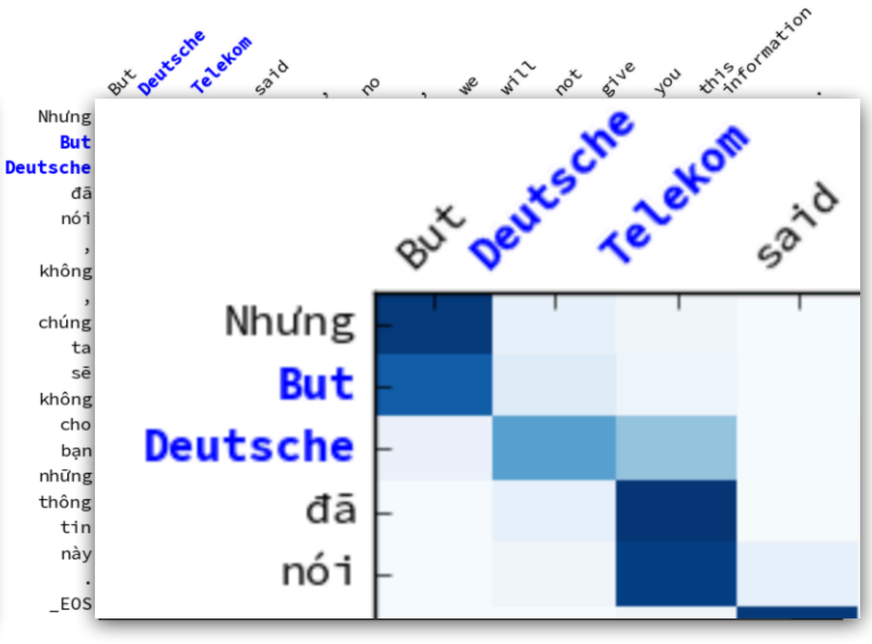
(c) **fixnorm+lex**



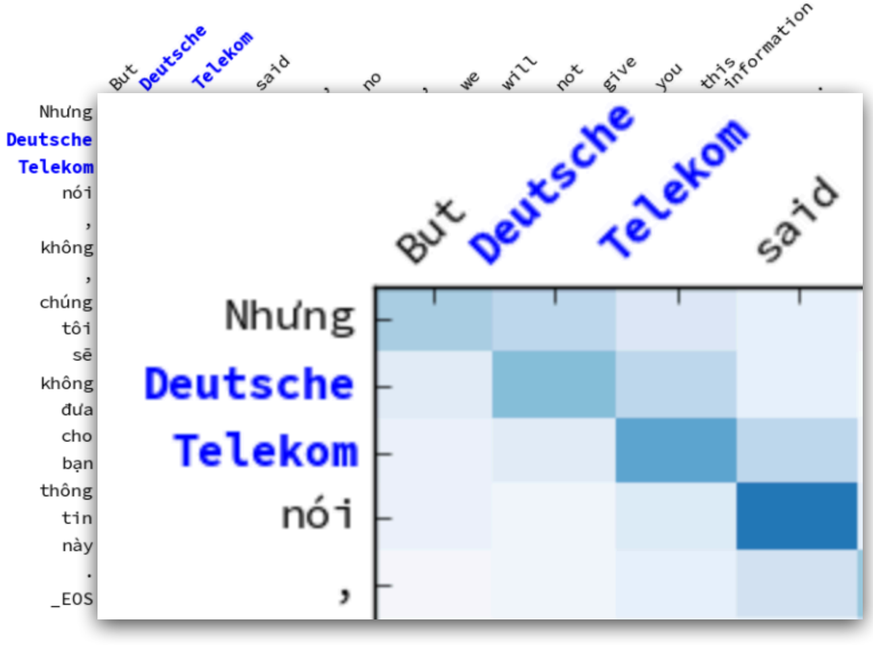
(d) Arthur et al. (2016)



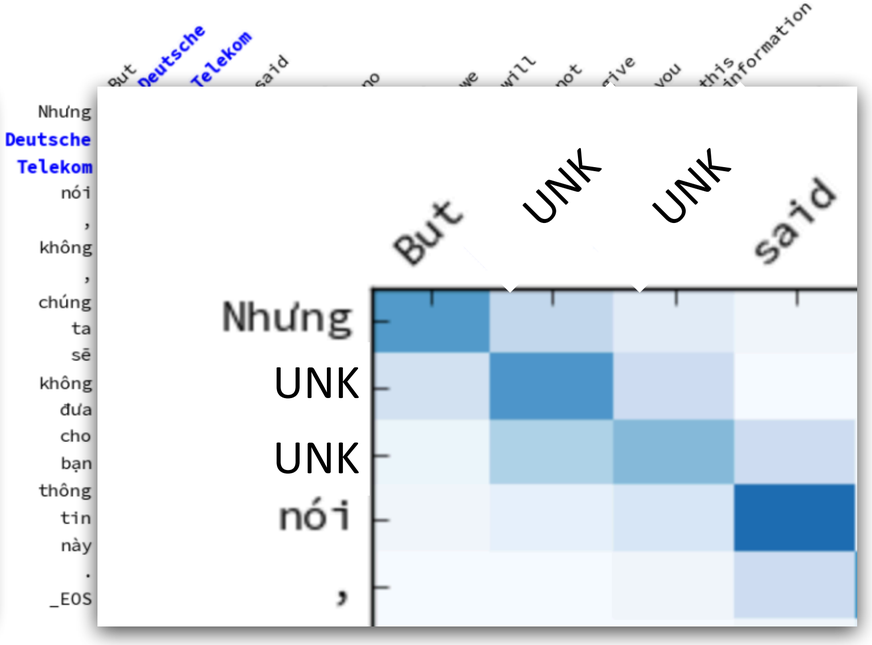
(a) **tied**



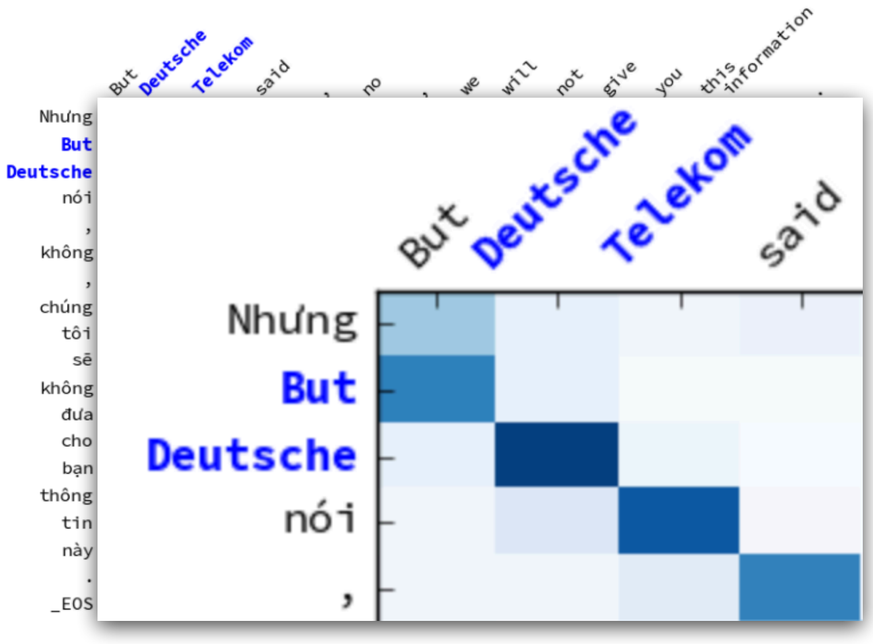
(b) **fixnorm**



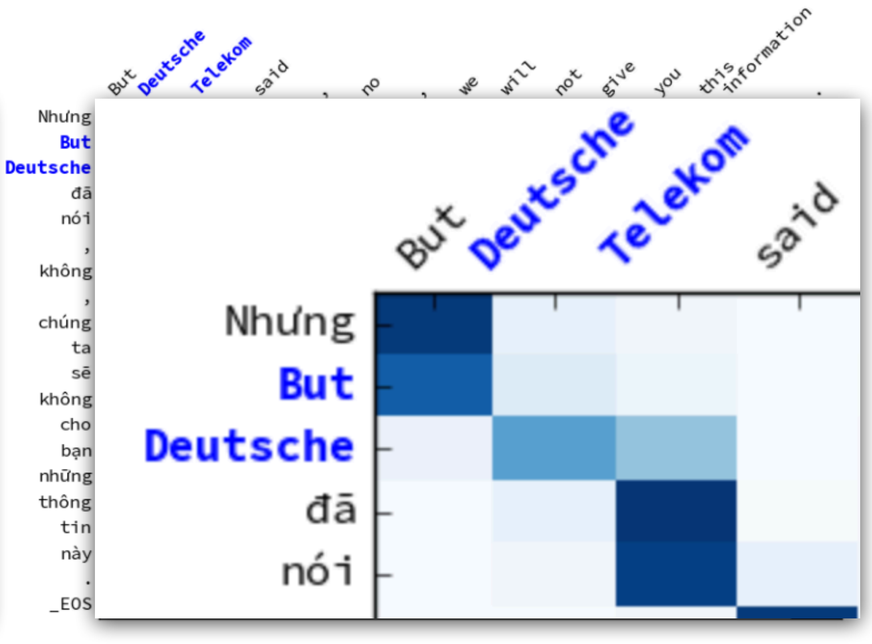
(c) **fixnorm+lex**



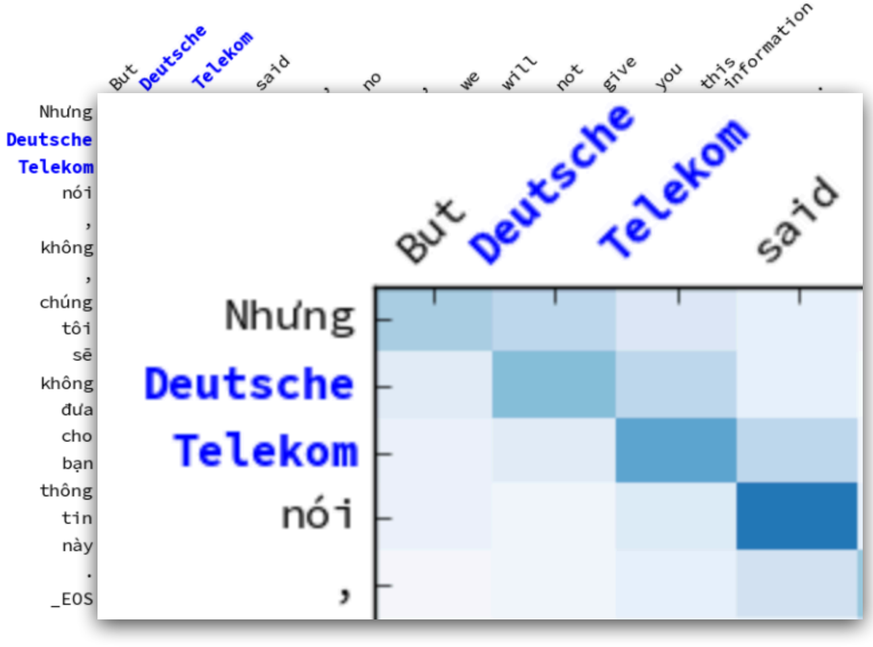
(d) Arthur et al. (2016)



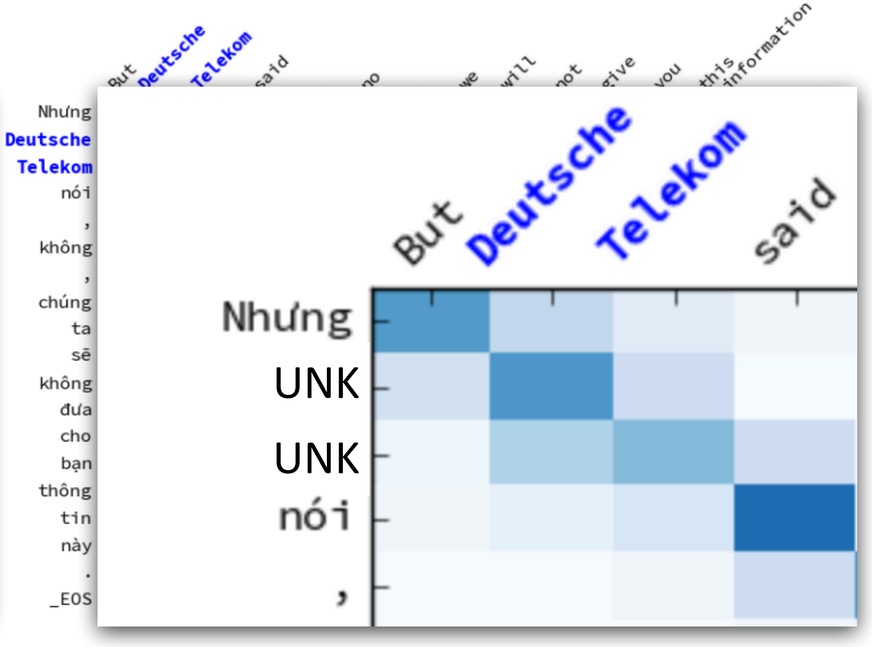
(a) **tied**



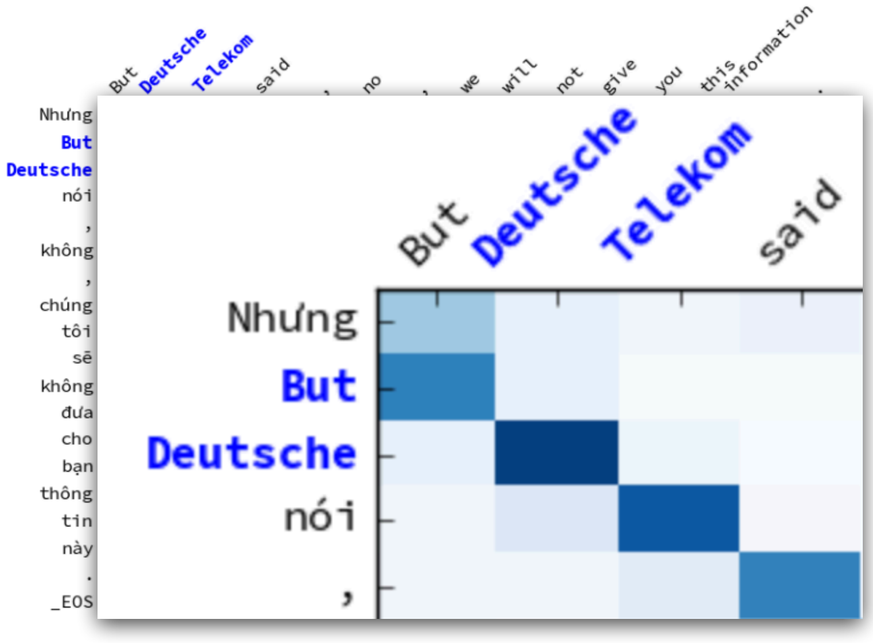
(b) **fixnorm**



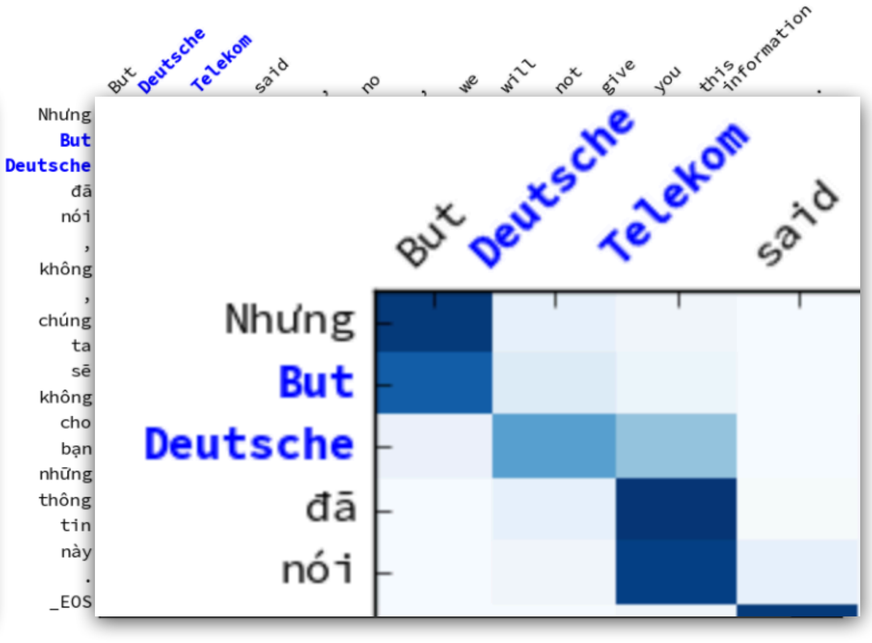
(c) **fixnorm+lex**



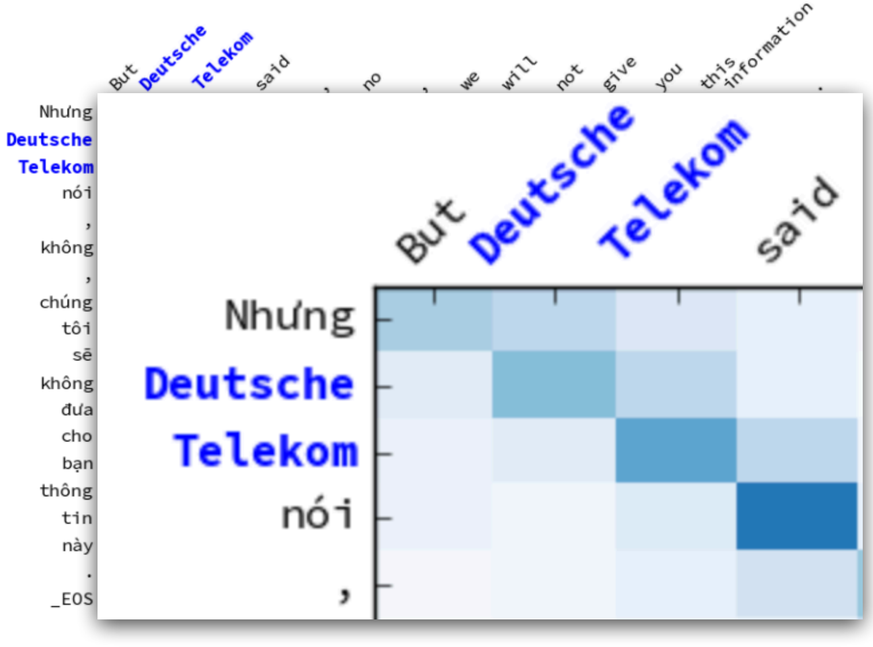
(d) Arthur et al. (2016)



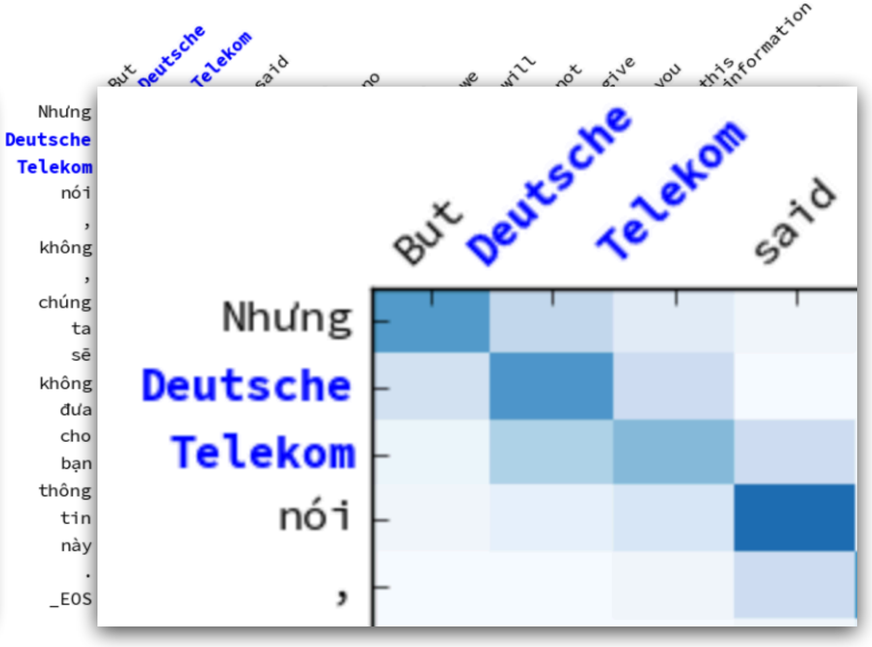
(a) **tied**



(b) **fixnorm**



(c) **fixnorm+lex**



(d) Arthur et al. (2016)

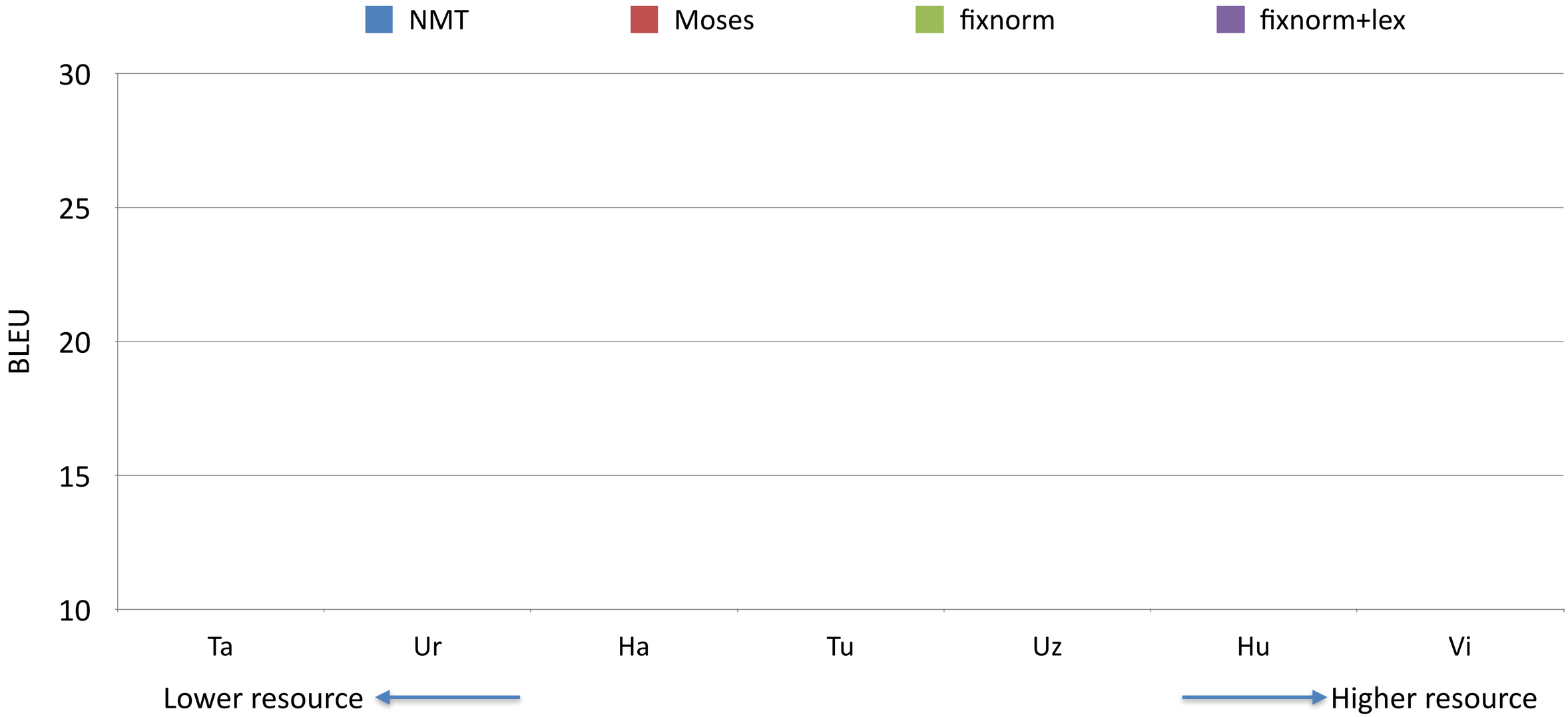
Byte Pair Encoding (BPE)

communists —> communi@@ + sts

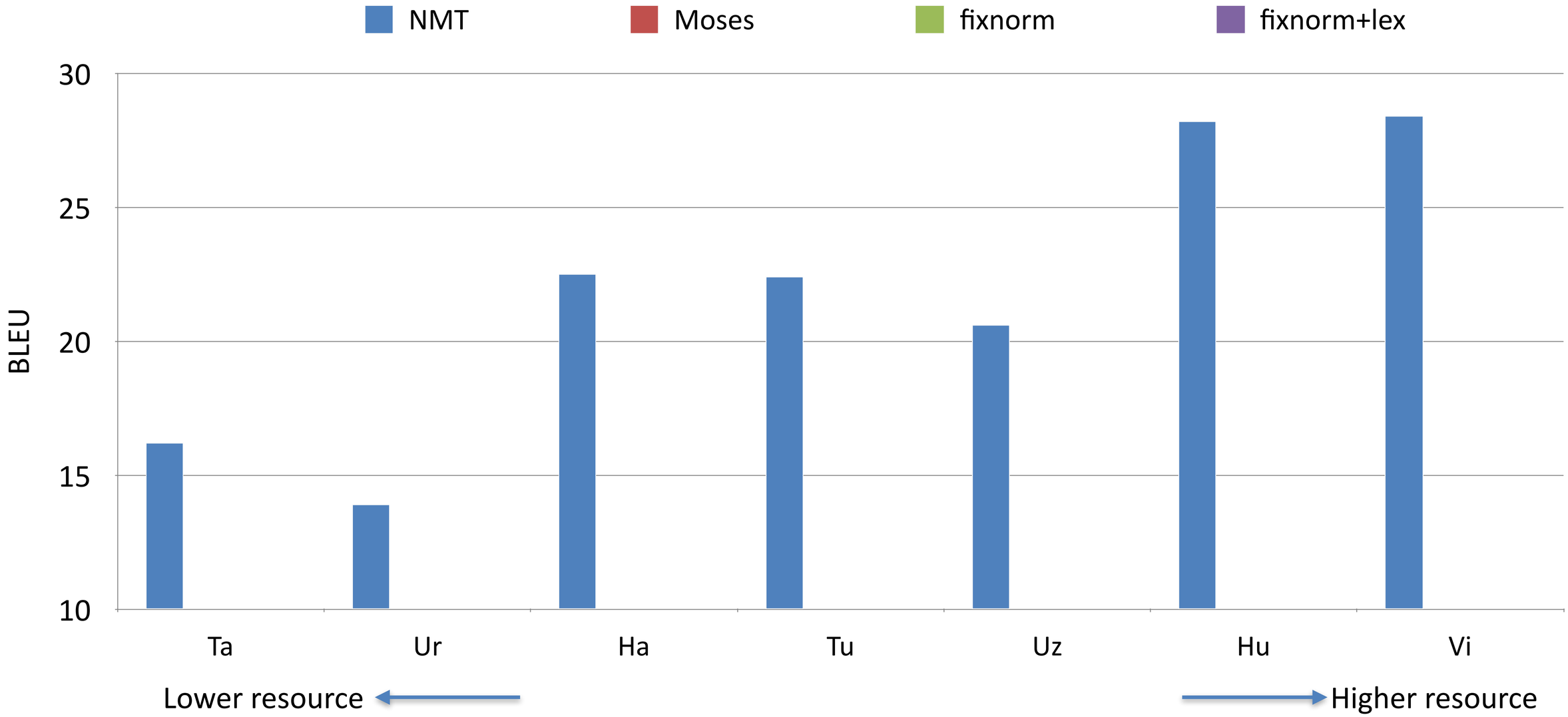
agglomerációs —> agg@@ + l@@ + om@@ + er@@ + ációs

Results (BPE)

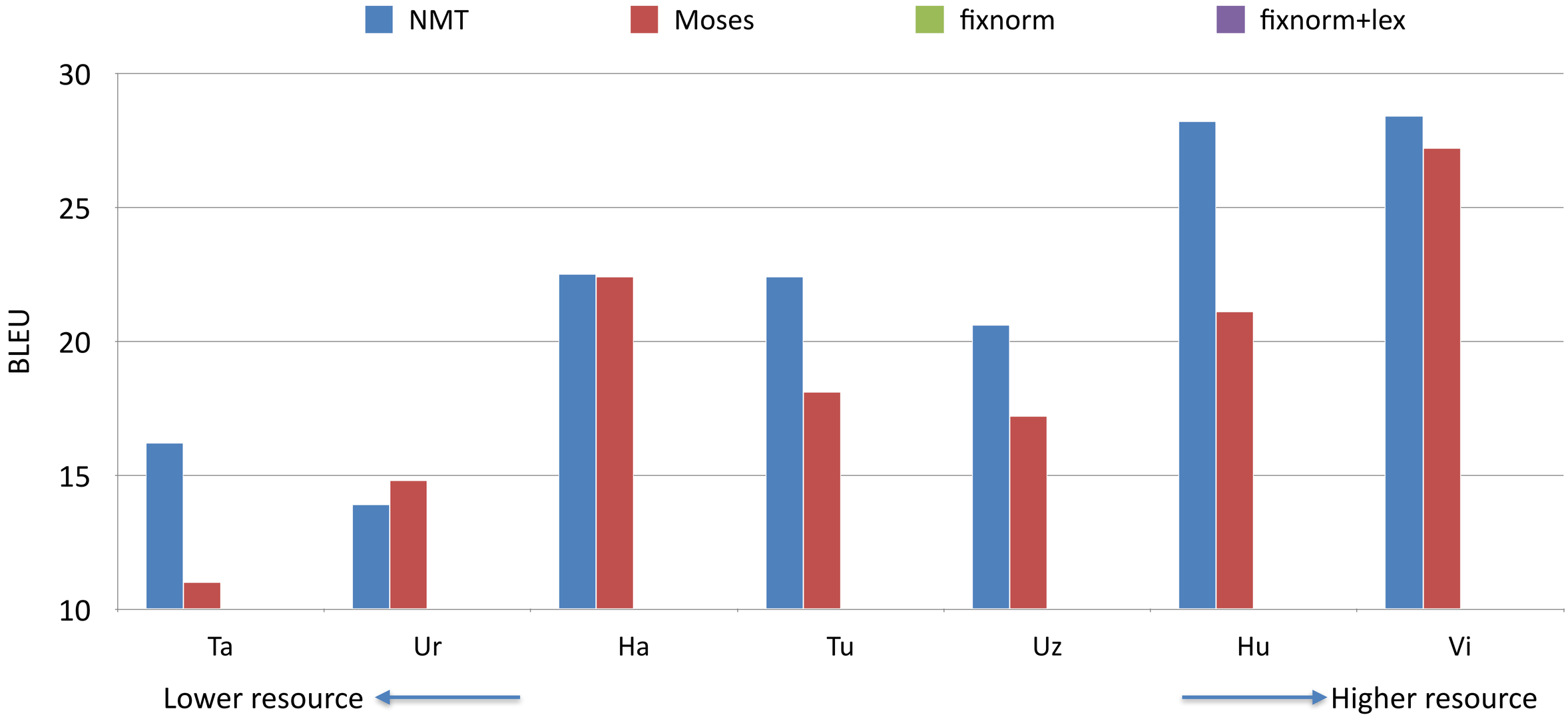
Results (BPE)


NMT
Moses
fixnorm
fixnorm+lex
BLEU
30
25
20
15
10
Ta
Ur
Ha
Tu
Uz
Hu
Vi
Lower resource
Higher resource

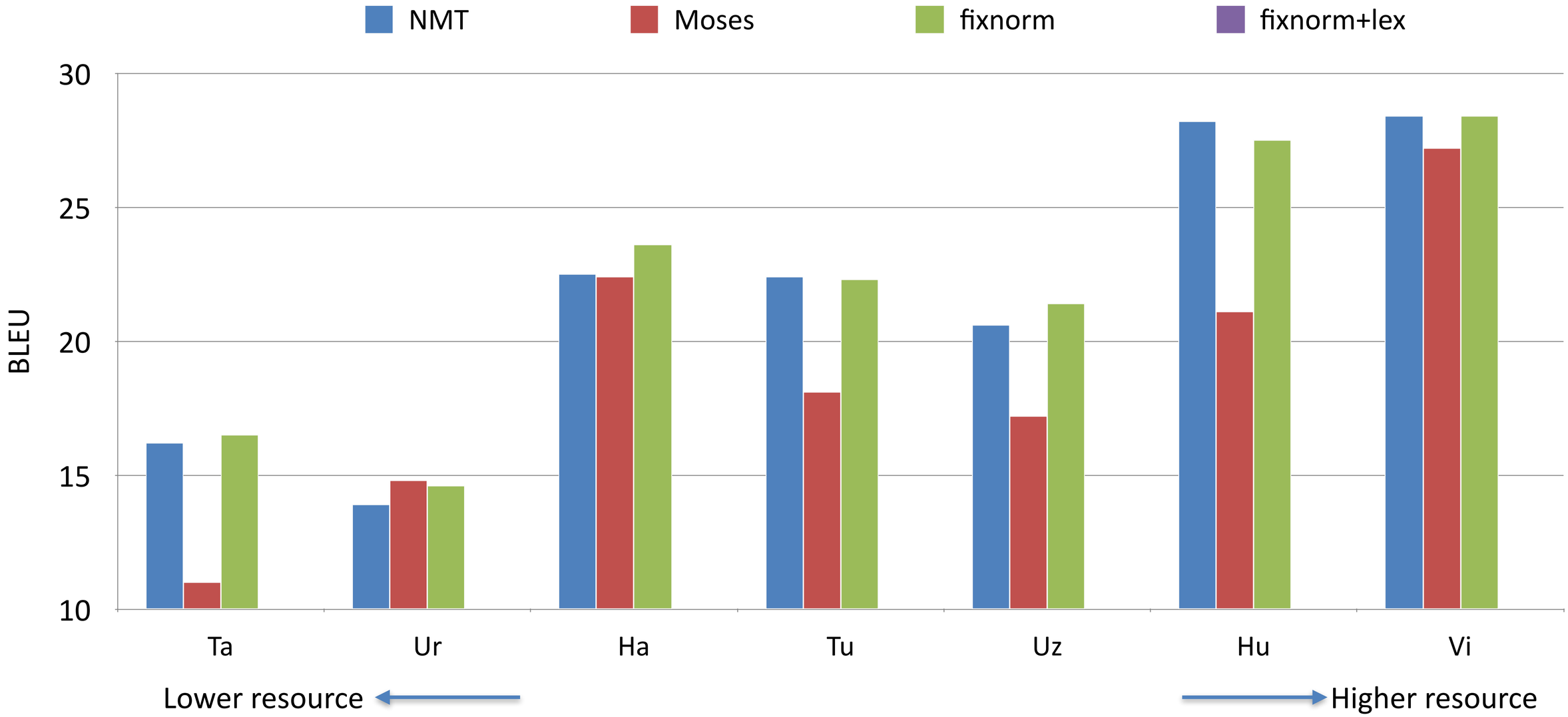
Results (BPE)


NMT
Moses
fixnorm
fixnorm+lex
BLEU
30
25
20
15
10
Ta
Ur
Ha
Tu
Uz
Hu
Vi
Lower resource
Higher resource

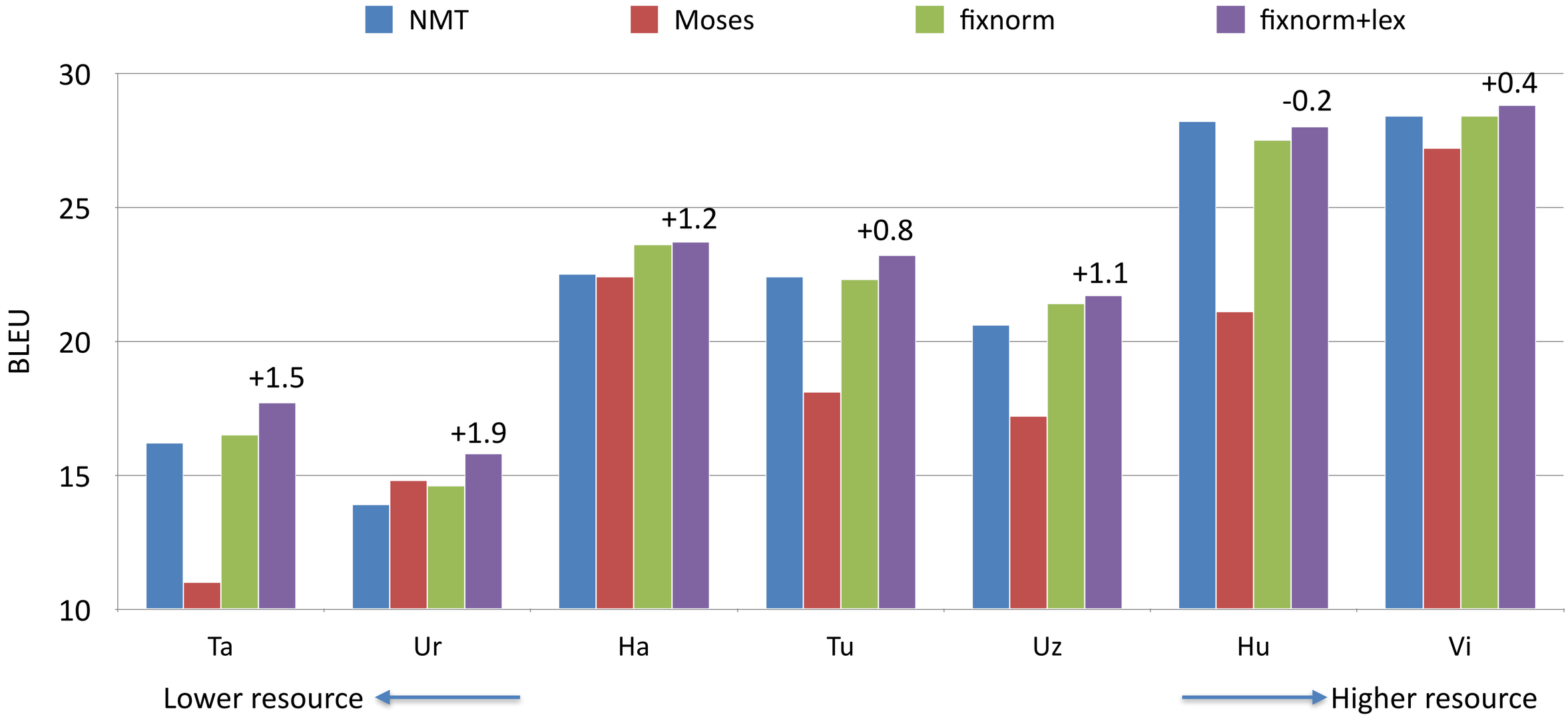
Results (BPE)


NMT
Moses
fixnorm
fixnorm+lex
BLEU
30
25
20
15
10
Ta
Ur
Ha
Tu
Uz
Hu
Vi
Lower resource
Higher resource

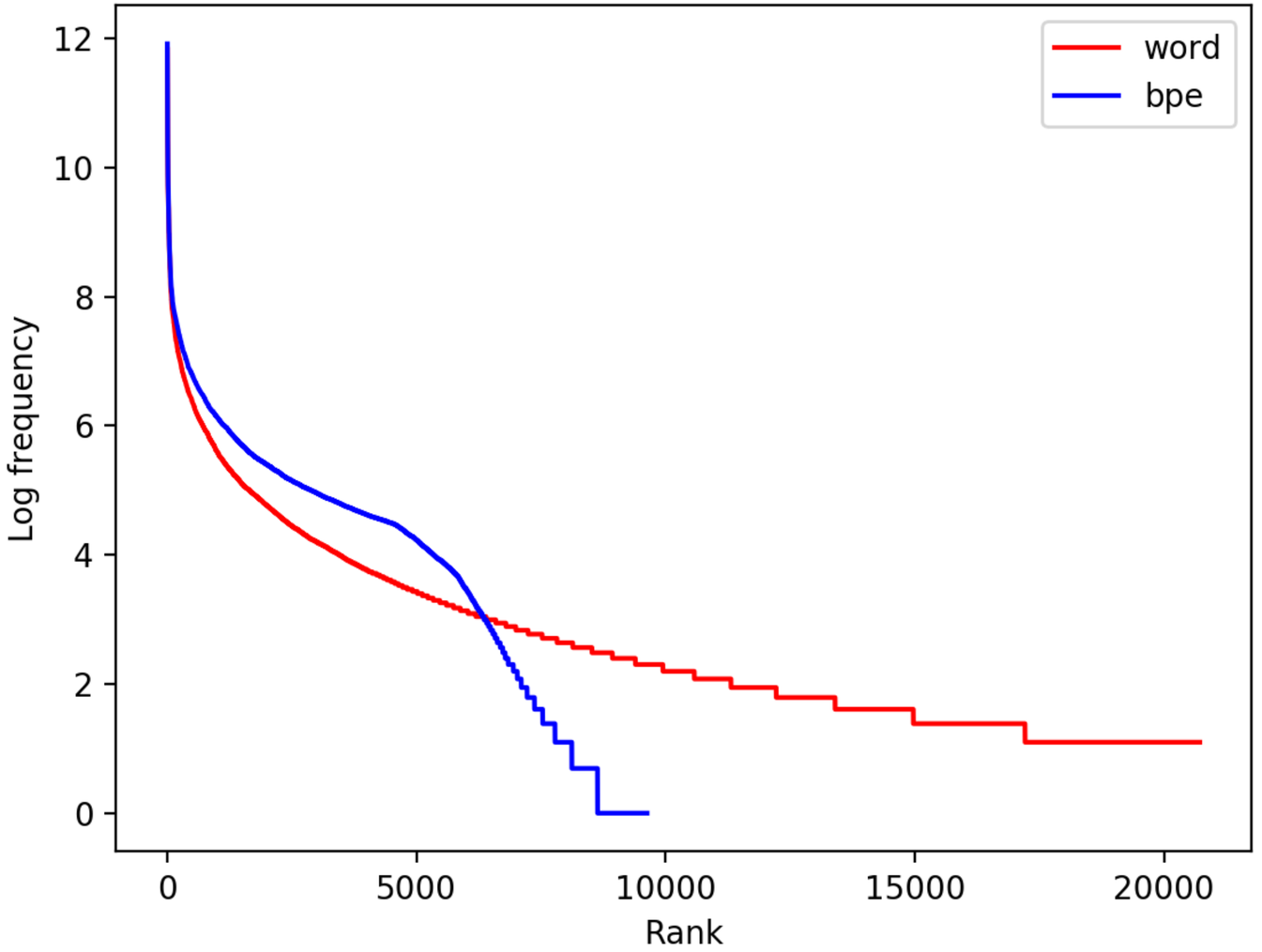
Results (BPE)


NMT
Moses
fixnorm
fixnorm+lex
BLEU
30
25
20
15
10
Ta
Ur
Ha
Tu
Uz
Hu
Vi
Lower resource
Higher resource

Results (BPE)


NMT
Moses
fixnorm
fixnorm+lex
BLEU
30
25
20
15
10
+1.5
+1.9
+1.2
+0.8
+1.1
-0.2
+0.4
Ta
Ur
Ha
Tu
Uz
Hu
Vi
Lower resource
Higher resource

What's different about BPE?



What’s different about BPE?

src	Here 's **V@@ la@@ di@@ mi@@ r T@@ sa@@ st@@ sin** form T@@ art@@ u in E@@ st@@ onia .
ref	Đây là **Vladimir Tsastsin** đến từ Tartu , Estonia
NMT	Đây là **V@@ la@@ di@@ mi@@ r T@@ sa@@ st@@ sin** ở E@@ st@@ onia . (Đây là **Vladimir Tsastsin** ở Estonia .)

Conclusion

- We present two simple and effective solutions for rare word mistranslation
- Word-based: improvements on all tested languages, up to +5.5 BLEU
- BPE-based: improvements on low-resource languages, up to +1.9 BLEU

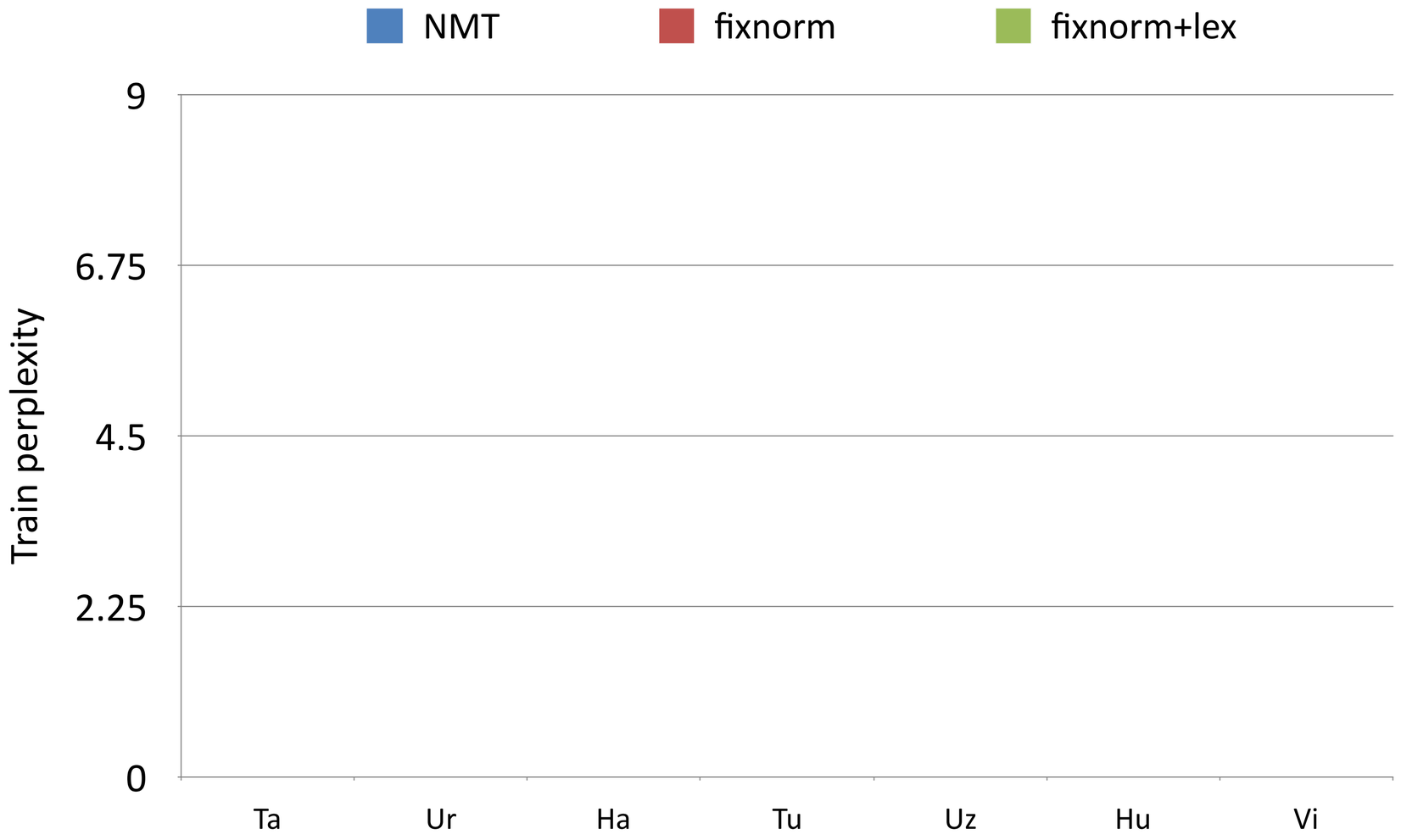
Code

https://github.com/tnq177/improving_lexical_choice_in_nmt

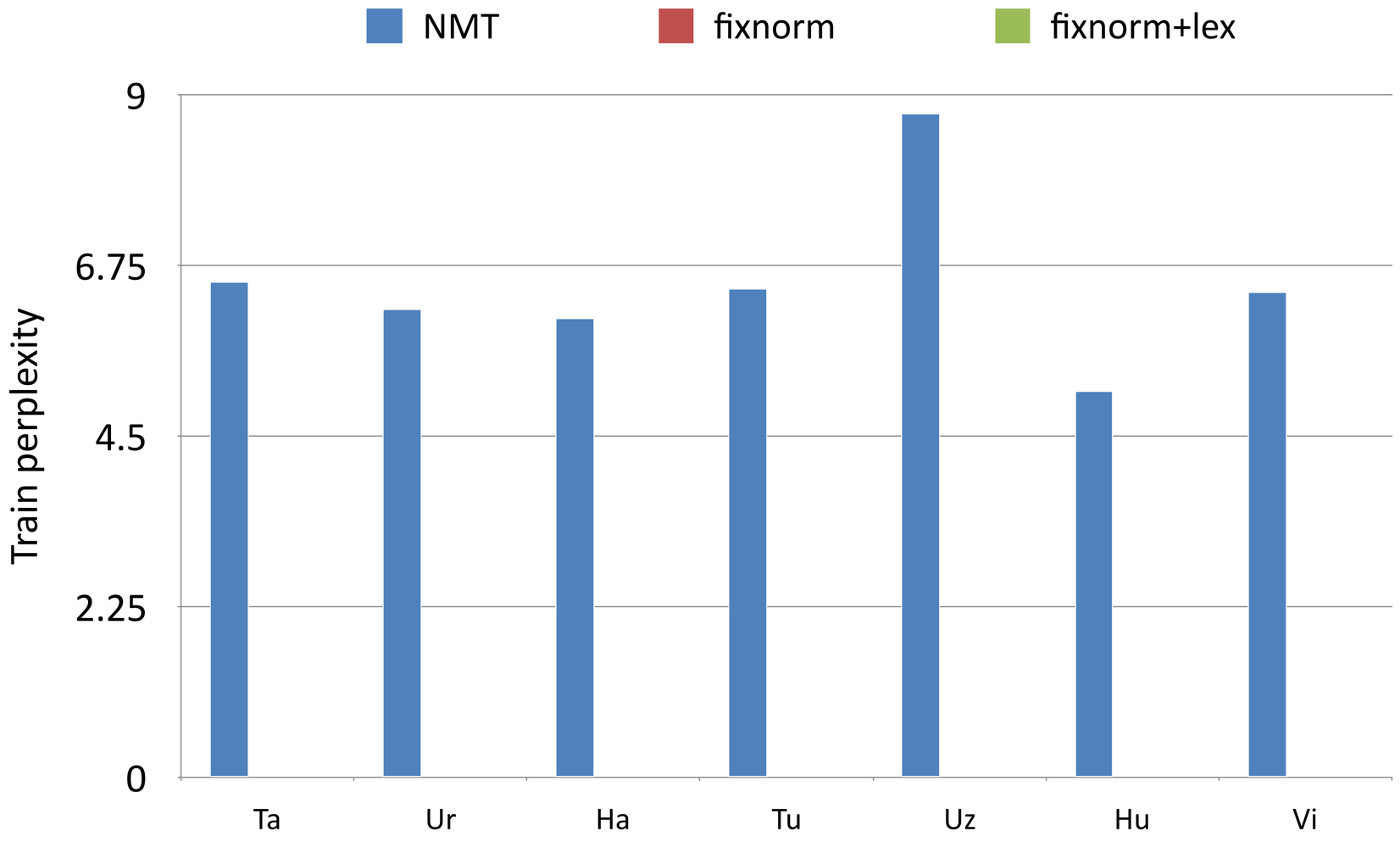
Thanks!

Train perplexity

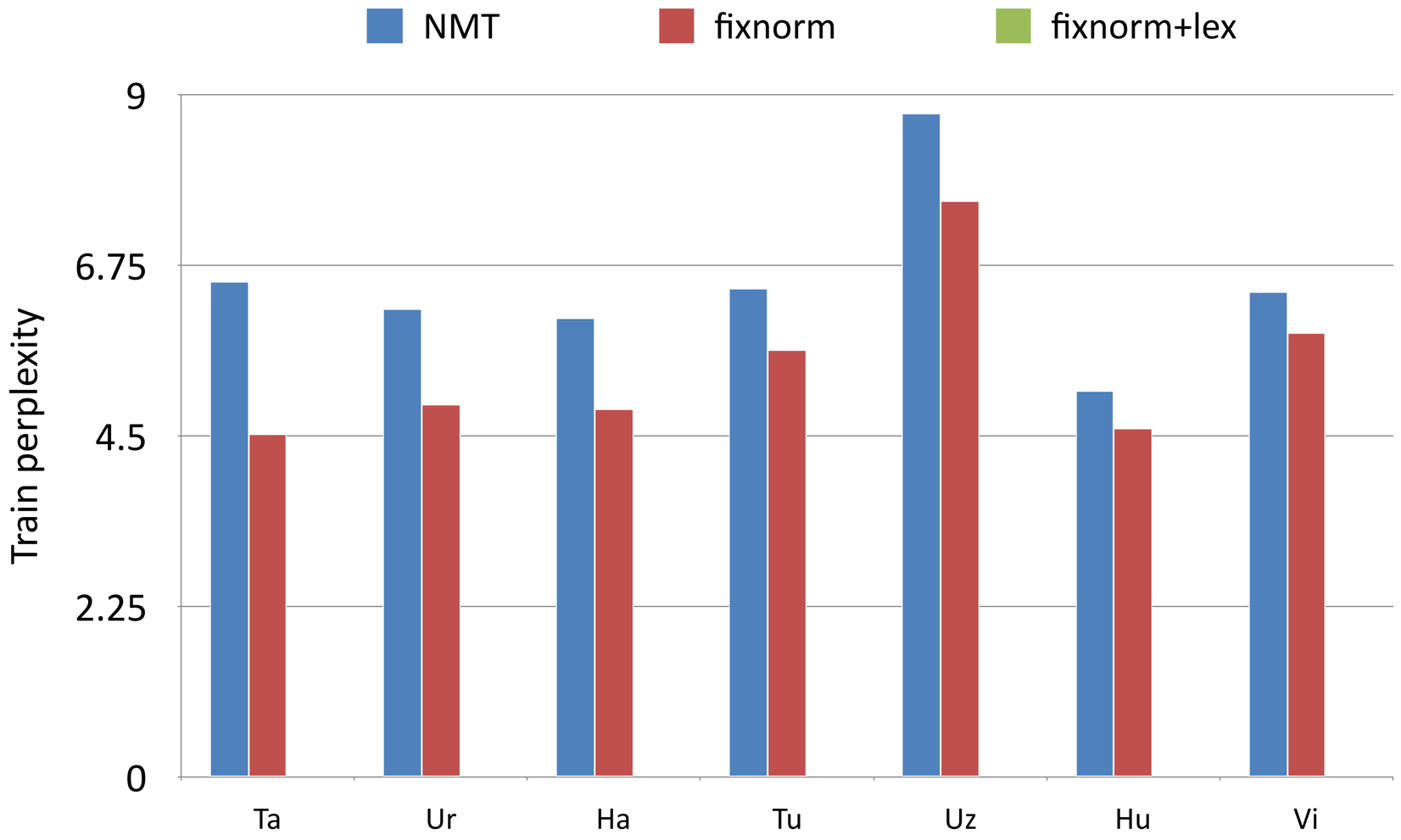
Train perplexity



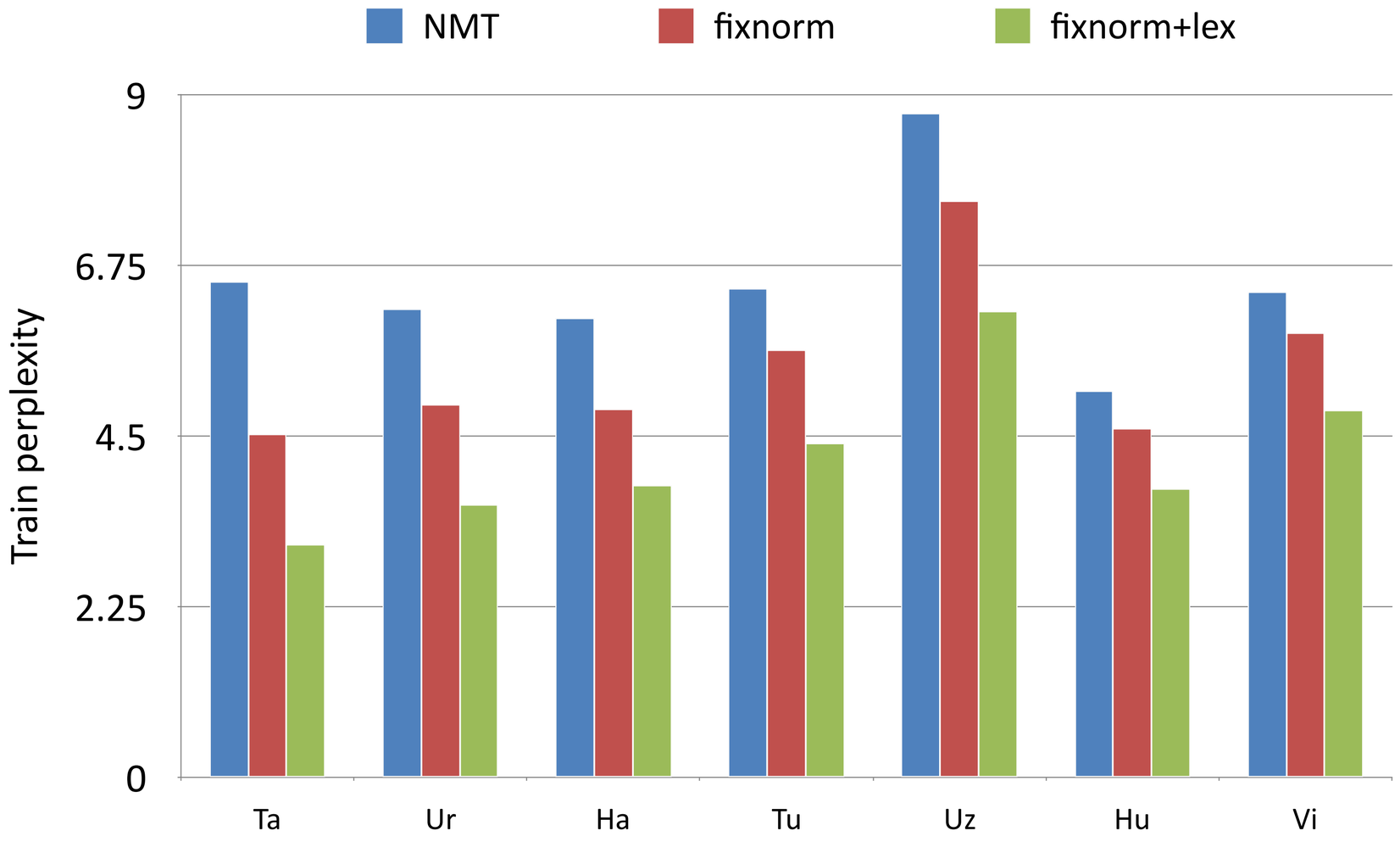
Train perplexity



Train perplexity

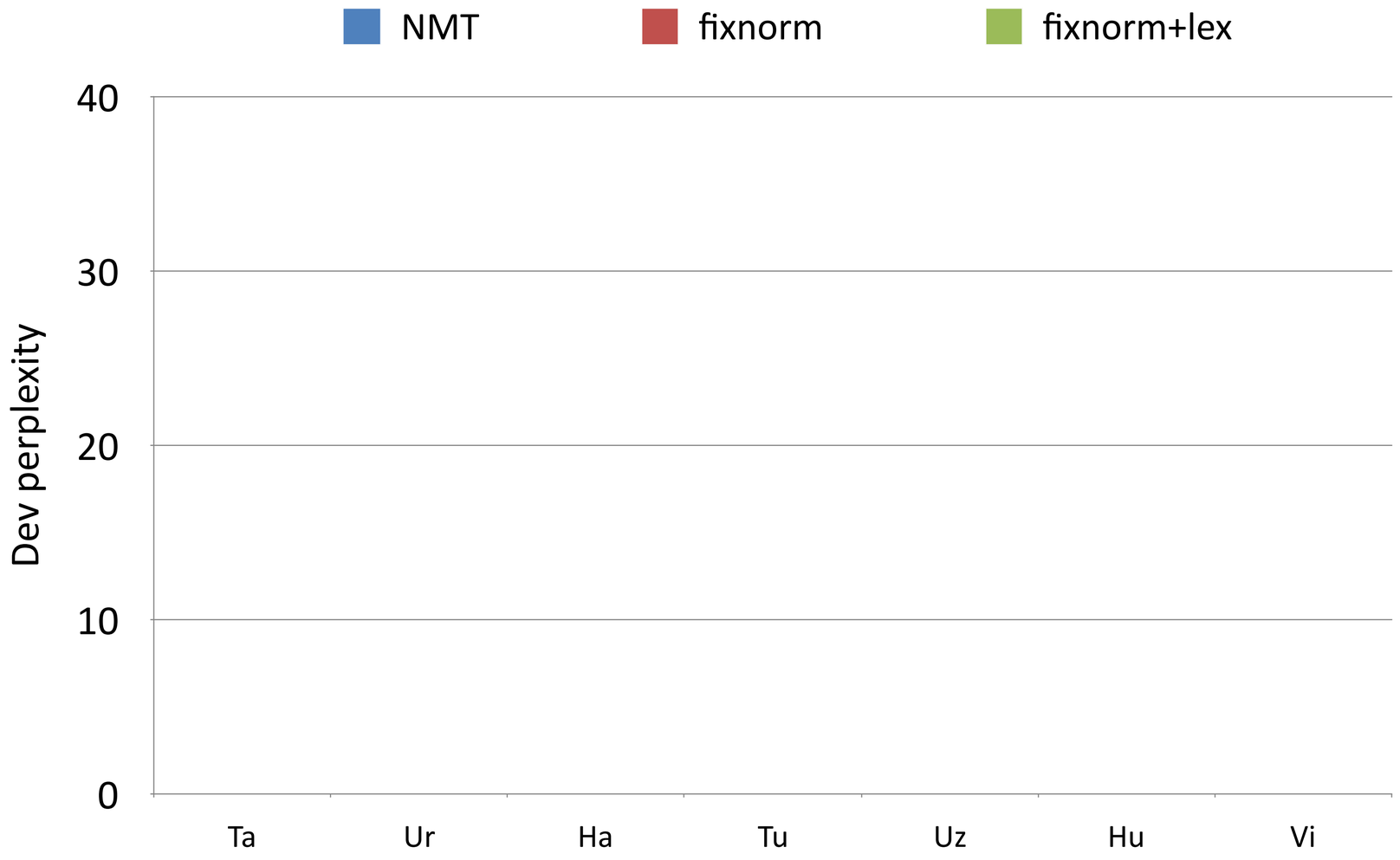


Train perplexity

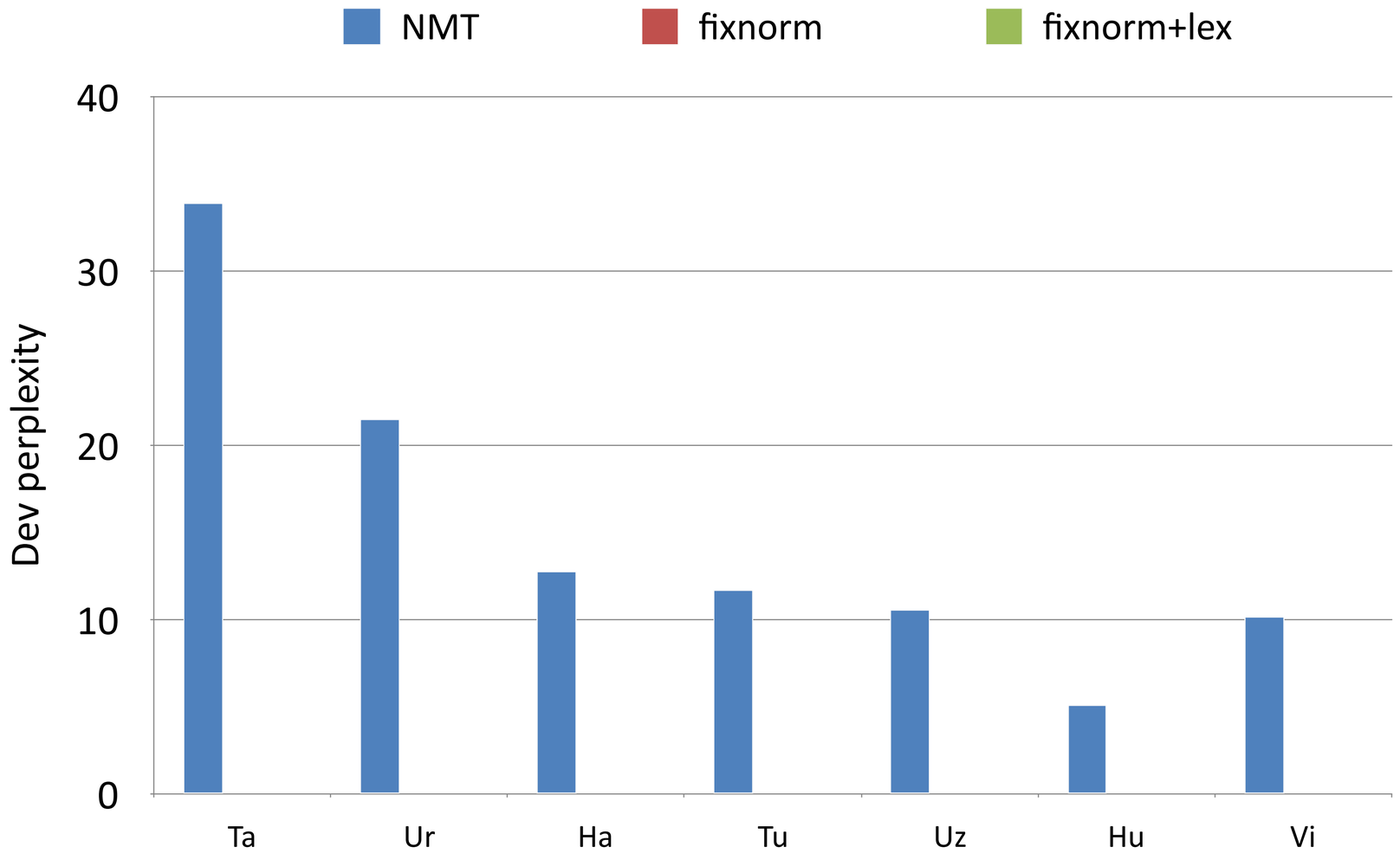


Dev perplexity

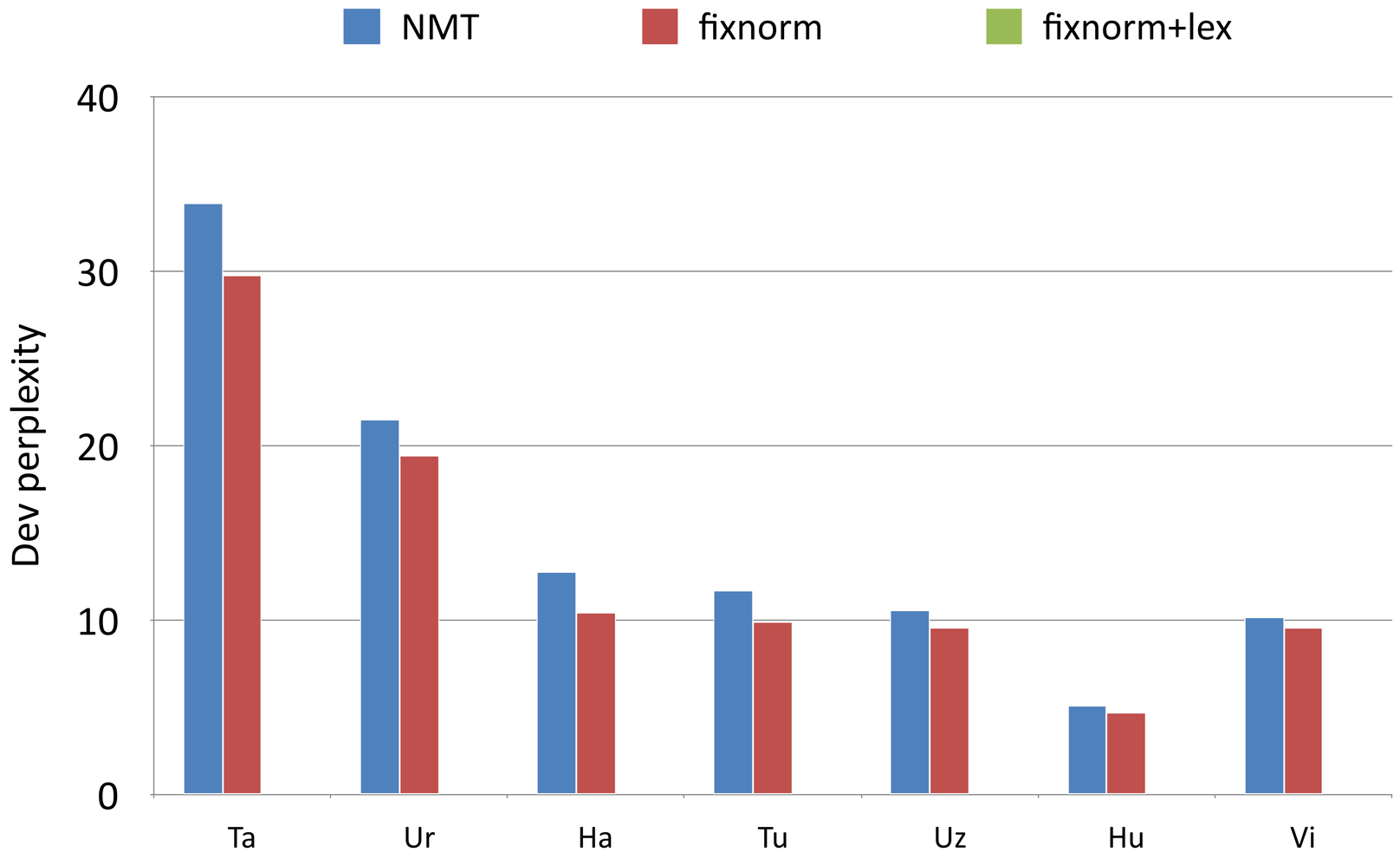
Dev perplexity



Dev perplexity



Dev perplexity



Dev perplexity