Compositional Morphology for Word Representations and Language Modelling

Jan Botha, Phil Blunsom

ICML 2014, Beijing

UNIVERSITY OF OXFORD

Motivating Example

What we see

The king finally abdicated after years of unkingly conduct .

Motivating Example

What we see

The king finally abdicated after years of unkingly conduct .

Wait what – unkingly?

Motivating Example

What we see

The king finally abdicated after years of unkingly conduct .

Wait what – unkingly?

unkingly ʌnˈkɪŋli

a word you have probably never seen, but still understand

MOTIVATING EXAMPLE

WHAT WE SEE

The king finally abdicated after years of unkingly conduct .

Wait what – unkingly?

unkingly ʌnˈkɪŋli

a word you have probably never seen, but still understand

⇒ compositional morphology in action

MOTIVATING EXAMPLE

WHAT WE SEE

The king finally abdicated after years of unkingly conduct .

Wait what – unkingly?

unkingly ʌnˈkɪŋli
a word you have probably never seen, but still understand

⇒ compositional morphology in action

WHAT OUR MODELS SEE (MOSTLY)

10 2 95 529 11 88 21 50 74 239

MOTIVATING EXAMPLE

WHAT WE SEE

The king finally abdicated after years of unkingly conduct .

Wait what – unkingly?

unkingly ʌnˈkɪŋli

a word you have probably never seen, but still understand

⇒ compositional morphology in action

WHAT OUR MODELS SEE (MOSTLY)

10 2 95 529 11 88 21 50 74 239

MOTIVATING EXAMPLE 2

Other languages display still more variation

CZECH CONJUGATION

čistit *(to clean)*
čistím
čistíš
čistí
čistíme
čistíte
čistil
čištěn
čisti
čistěte
čistěme

TURKISH PRODUCTIVE DERIVATION

Avrupa	*(Europe)*
Avrupalı	*(of Europe)*
Avrupalılaş	*(become of Europe)*
Avrupalılaştır	*(to Europeanise)*
Avrupalılaştırama	*(be unable to Europeanise)*
Avrupalılaştıramadık	*(we were unable to Europeanise)*
...	

MOTIVATING EXAMPLE 2

Other languages display still more variation

CZECH CONJUGATION

čistit *(to clean)*
čistím
čistíš
čistí
čistíme
čistíte
čistil
čištěn
čisti
čistěte
čistěme

TURKISH PRODUCTIVE DERIVATION

Avrupa	*(Europe)*
Avrupalı	*(of Europe)*
Avrupalılaş	*(become of Europe)*
Avrupalılaştır	*(to Europeanise)*
Avrupalılaştırama	*(be unable to Europeanise)*
Avrupalılaştıramadık	*(we were unable to Europeanise)*
...	

⇒ we should model morphemes!

REPRESENTING WORDS

- Discrete set?

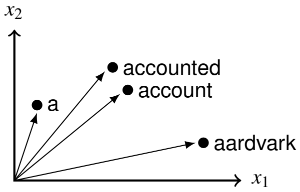
  {a, aardvark, ..., account, accounted, accounting, ...}

REPRESENTING WORDS

- Discrete set?

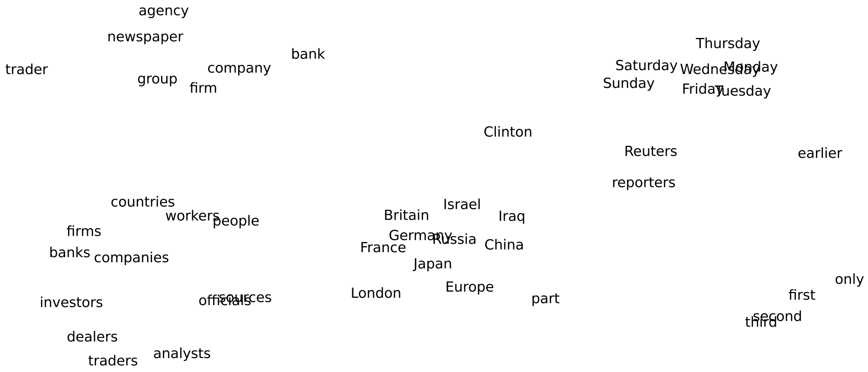
  {a, aardvark, . . . , account, accounted, accounting, . . . }

- Vector space?

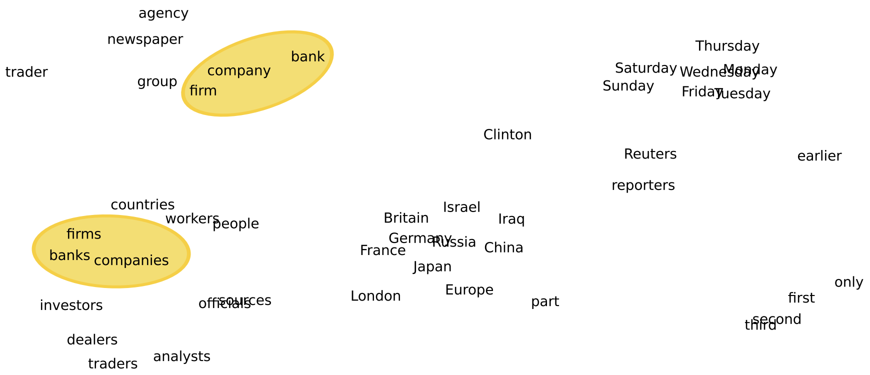


Extract from Collobert & Weston Embeddings

EXTRACT FROM COLLOBERT & WESTON EMBEDDINGS



Extract from Collobert & Weston Embeddings



Morpheme vectors

Existing word vectors already capture some morphology.

- $\overrightarrow{\text{banks}} - \overrightarrow{\text{bank}} \approx \overrightarrow{\text{kings}} - \overrightarrow{\text{king}} \approx \overrightarrow{\text{queens}} - \overrightarrow{\text{queen}}$

(Mikolov et al. 2013)

MORPHEME VECTORS

Existing word vectors already capture some morphology.

- $\overrightarrow{\text{banks}} - \overrightarrow{\text{bank}} \approx \overrightarrow{\text{kings}} - \overrightarrow{\text{king}} \approx \overrightarrow{\text{queens}} - \overrightarrow{\text{queen}}$

(Mikolov et al. 2013)

Logical extension:

- $\overrightarrow{\text{kings}} \approx \overrightarrow{\text{king}} + \overrightarrow{\text{-s}}$
- $\overrightarrow{\text{unkingly}} \approx \overrightarrow{\text{un-}} + \overrightarrow{\text{king}} + \overrightarrow{\text{-ly}}$

MORPHEME VECTORS

Existing word vectors already capture some morphology.

- $\overrightarrow{\text{banks}} - \overrightarrow{\text{bank}} \approx \overrightarrow{\text{kings}} - \overrightarrow{\text{king}} \approx \overrightarrow{\text{queens}} - \overrightarrow{\text{queen}}$

(Mikolov et al. 2013)

Logical extension:

- $\overrightarrow{\text{kings}} \approx \overrightarrow{\text{king}} + \overrightarrow{\text{-s}}$
- $\overrightarrow{\text{unkingly}} \approx \overrightarrow{\text{un-}} + \overrightarrow{\text{king}} + \overrightarrow{\text{-ly}}$

HOW TO...

- obtain morpheme vectors
- compose morpheme vectors
- do it all *within* a language model usable in an MT decoder

Morphological composition as addition

Literally, word = sum of its parts?

MORPHOLOGICAL COMPOSITION AS ADDITION

Literally, word = sum of its parts?

Problems:

- bag of morphemes: $\overrightarrow{\text{hang}} + \overrightarrow{\text{over}} \neq \overrightarrow{\text{over}} + \overrightarrow{\text{hang}}$
- non-compositionality: $\overrightarrow{\text{greenhouse}} \neq \overrightarrow{\text{green}} + \overrightarrow{\text{house}}$

MORPHOLOGICAL COMPOSITION AS ADDITION

Literally, word = sum of its parts?

Problems:

- bag of morphemes: $\overrightarrow{\text{hang}} + \overrightarrow{\text{over}} \neq \overrightarrow{\text{over}} + \overrightarrow{\text{hang}}$
- non-compositionality: $\overrightarrow{\text{greenhouse}} \neq \overrightarrow{\text{green}} + \overrightarrow{\text{house}}$

PRAGMATIC SOLUTION

include word identity as component too:

$$\overrightarrow{\text{greenhouse}} \equiv \overrightarrow{\text{greenhouse}} + \overrightarrow{green}_{stem} + \overrightarrow{house}_{stem}$$

$$\overrightarrow{\text{unkingly}} \equiv \overrightarrow{\text{unkingly}} + \overrightarrow{un}_{pre} + \overrightarrow{king}_{stem} + \overrightarrow{ly}_{suf}$$

Morphological composition as addition

Literally, word = sum of its parts?

Problems:

- bag of morphemes: $\overrightarrow{\text{hang}} + \overrightarrow{\text{over}} \neq \overrightarrow{\text{over}} + \overrightarrow{\text{hang}}$
- non-compositionality: $\overrightarrow{\text{greenhouse}} \neq \overrightarrow{\text{green}} + \overrightarrow{\text{house}}$

Pragmatic Solution

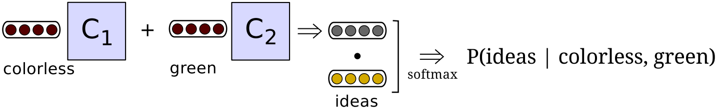
include word identity as component too:

$$\overrightarrow{\text{greenhouse}} \equiv \overrightarrow{greenhouse}_{id} + \overrightarrow{green}_{stem} + \overrightarrow{house}_{stem}$$

$$\overrightarrow{\text{unkingly}} \equiv \overrightarrow{unkingly}_{id} + \overrightarrow{un}_{pre} + \overrightarrow{king}_{stem} + \overrightarrow{ly}_{suf}$$

SIMPLEST VECTOR-BASED PROBABILISTIC LM

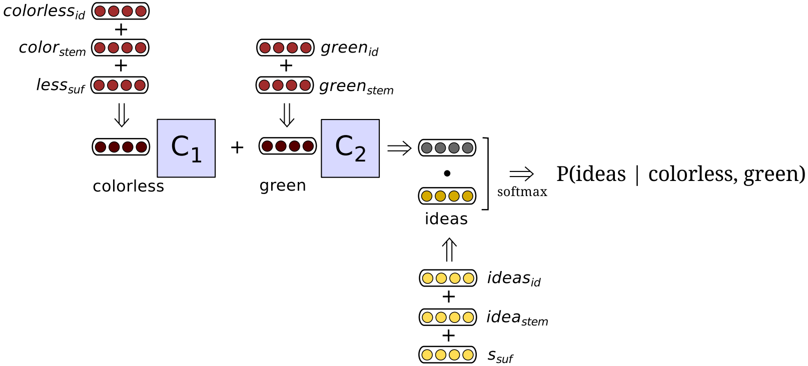
LBL (Log-bilinear model) (Mnih & Hinton, 2007; Mnih & Teh, 2012)



"colorless green ideas sleep furiously ."

ADD MORPHEME VECTORS INSIDE LM

LBL++



"colorless green ideas sleep furiously ."

COMPUTATIONAL EFFICIENCY

Problem:

Each probability query requires normalisation over vocabulary.

- $\mathcal{O}(\text{vocab size})$
- rich morphology $\Rightarrow$ large vocabulary

COMPUTATIONAL EFFICIENCY

Problem:
Each probability query requires normalisation over vocabulary.

- $\mathcal{O}(\text{vocab size})$
- rich morphology $\Rightarrow$ large vocabulary

SOLUTION: DECOMPOSE MODEL USING WORD CLASSES

$$\begin{aligned} P\big(\text{word} \mid \text{history}\big) = \; & P\big(\text{class}(\text{word}) \mid \text{history}\big) \\ & \times P\big(\text{word} \mid \text{class}(\text{word}), \text{history}\big) \end{aligned}$$

- use unsupervised Brown-clustering
- each LM query becomes $2 \times \mathcal{O}(\sqrt{\text{vocab size}})$
  $\Rightarrow$ fast enough for MT-decoding

EVALUATION OVERVIEW

Setup

- 4-gram models
- Czech, English, French, German, Spanish, Russian
- train on 20–50m tokens
- large vocabularies (exclude 5% of singletons)

Evaluation Overview

Setup

- 4-gram models
- Czech, English, French, German, Spanish, Russian
- train on 20–50m tokens
- large vocabularies (exclude 5% of singletons)

Three evaluation contexts:

- Perplexity on test data
- Word similarity rating
- Machine translation

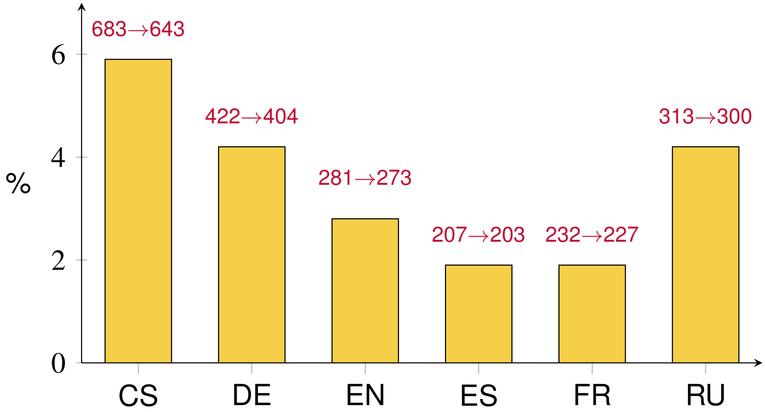
EVALUATION OVERVIEW

Three evaluation contexts:

- **Perplexity on test data**
- Word similarity rating
- Machine translation

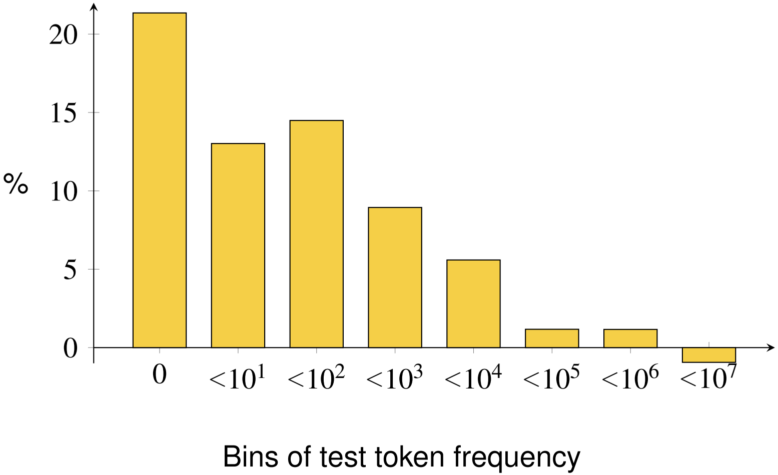
PERPLEXITY IMPROVEMENTS BY LANGUAGE

CLBL→CLBL++



PERPLEXITY IMPROVEMENTS ON GERMAN

CLBL→CLBL++ (BREAK-DOWN BY TOKEN FREQUENCY)



EVALUATION OVERVIEW

Three evaluation contexts:

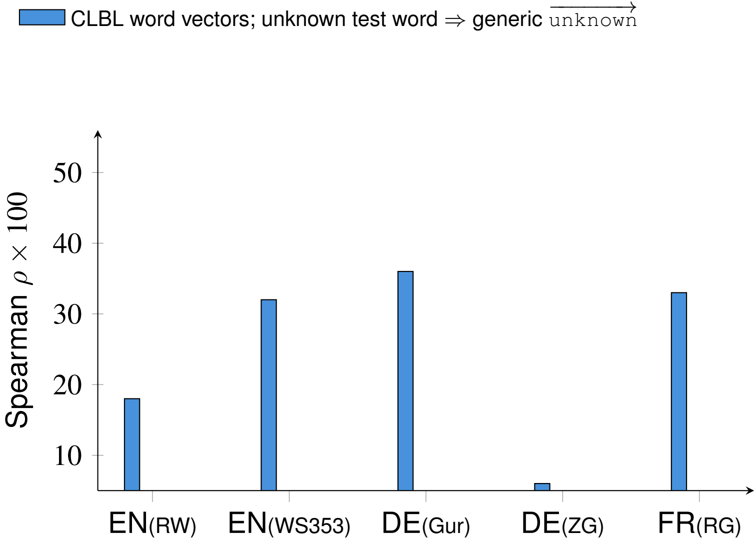
- **Perplexity on test data**
- Word similarity rating
- Machine translation

EVALUATION OVERVIEW

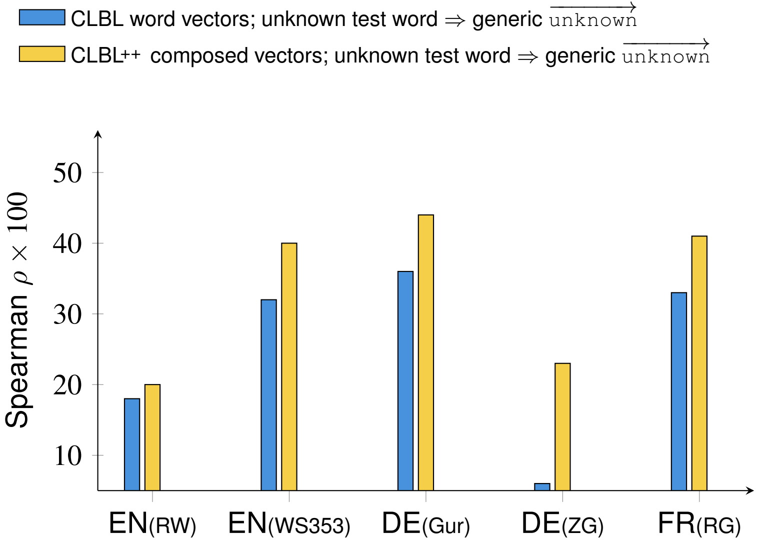
Three evaluation contexts:

- Perplexity on test data
- **Word similarity rating**
- Machine translation

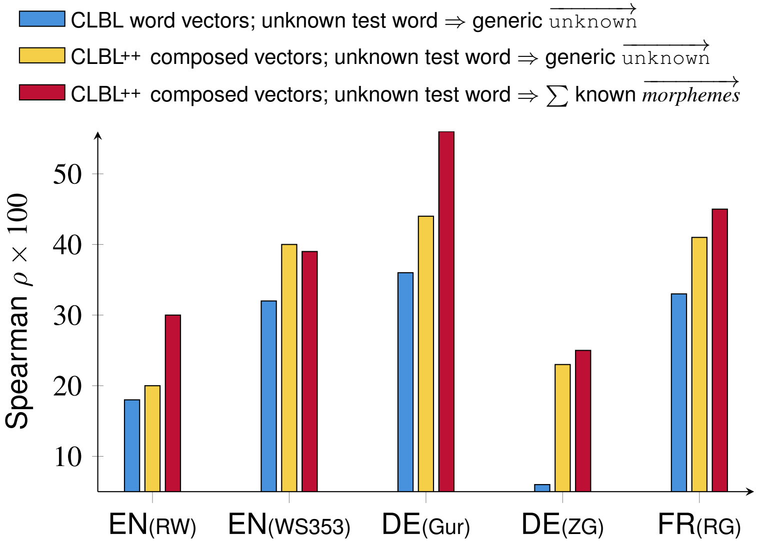
WORD SIMILARITY RATING



WORD SIMILARITY RATING



Word similarity rating



EVALUATION OVERVIEW

Three evaluation contexts:

- Perplexity on test data
- **Word similarity rating**
- Machine translation

EVALUATION OVERVIEW

Three evaluation contexts:

- Perplexity on test data
- Word similarity rating
- **Machine translation**

Machine Translation Evaluation

How to use the LM?

- rescore n-best list $<$ rescore lattice $<$ decoder feature

Machine Translation Evaluation

How to use the LM?

- rescore n-best list $<$ rescore lattice $<$ decoder feature

Hierarchical-phrase based decoder (cdec)

- *Baseline:* Kneser-Ney LM feature
- *Test:* Kneser-Ney LM feature + CLBL feature

MACHINE TRANSLATION EVALUATION

How to use the LM?

- rescore n-best list $<$ rescore lattice $<$ decoder feature

Hierarchical-phrase based decoder (cdec)

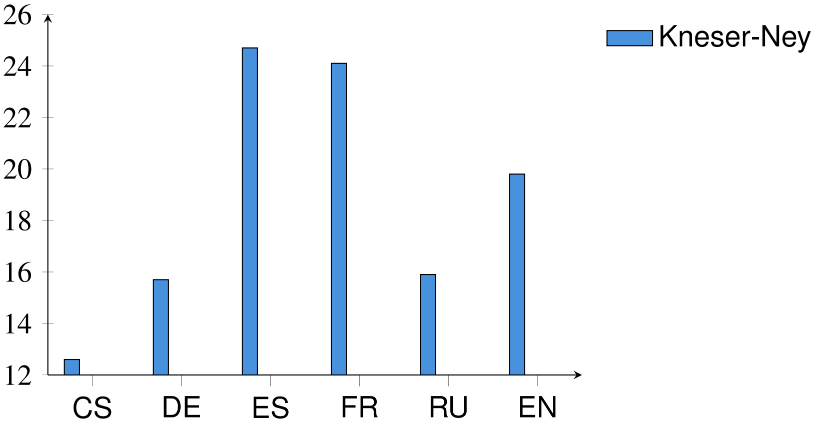
- *Baseline:* Kneser-Ney LM feature
- *Test:* Kneser-Ney LM feature + CLBL feature

CLBL speed-up from:

- class decomposition
- cache normalisers on-the-fly

TRANSLATION QUALITY (BLEU)

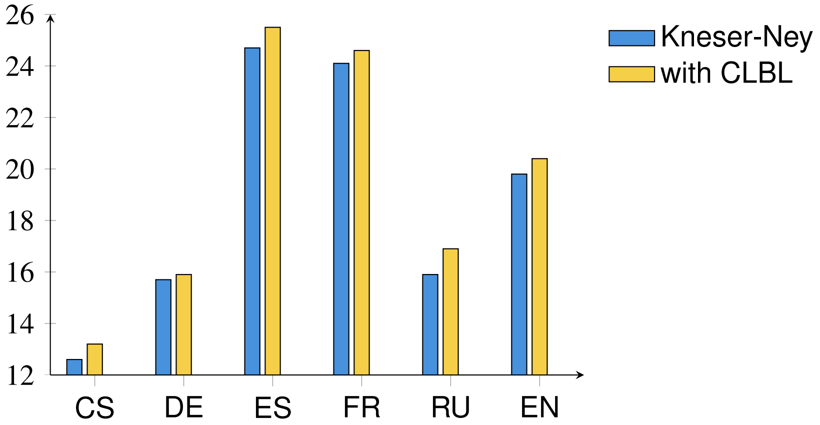
FOR TRANSLATING INTO GIVEN LANGUAGE



higher better

TRANSLATION QUALITY (BLEU)

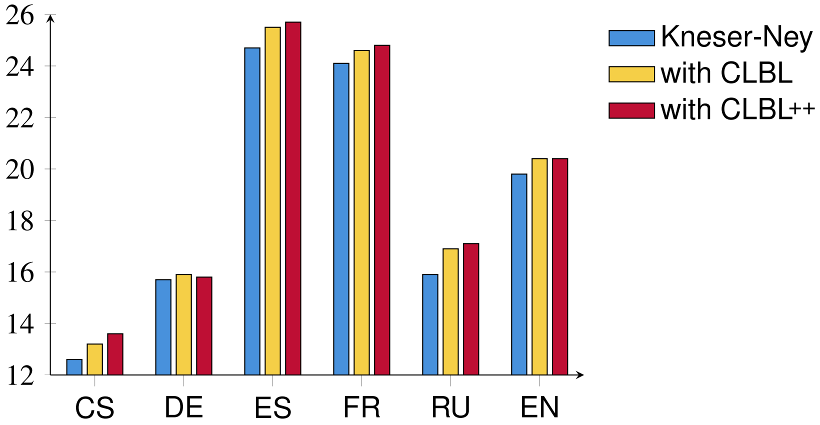
FOR TRANSLATING INTO GIVEN LANGUAGE



higher better

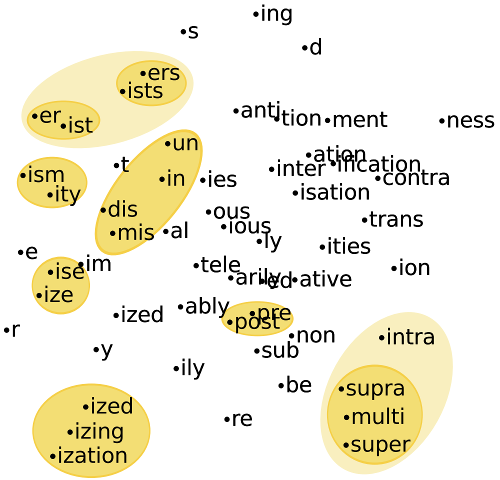
Translation Quality (Bleu)

For translating into given language



higher better

QUALITATIVE EVALUATION: ENGLISH AFFIX VECTORS



SUMMARY

Simple,
scaleable,
unsupervised
method for integrating morphology into vector-based LM

- improvements in three evaluation settings
- translation with normalised NLM works

SUMMARY

Simple,
scaleable,
unsupervised
method for integrating morphology into vector-based LM

- improvements in three evaluation settings
- translation with normalised NLM works

Software released shortly

`www.clg.ox.ac.uk/resources`

{Jan.Botha,Phil.Blunsom}@cs.ox.ac.uk

UNIVERSITY OF OXFORD