Overview of Machine Translation

Kenneth Heafield, University of Edinburgh

Edinburgh
University of Edinburgh
Natural Language Processing
NLP

Seen at Golden Acre Shopping Centre


Airport Priority Queue

Edinburgh Translation Group 1/2

Rico Sennrich
Lecturer

Barry Haddow
Postdoc

Alexandra Birch
Postdoc

Ulrich Germann
Postdoc

Antonio Valerio Miceli Barone
Postdoc

Phil Williams
Postdoc

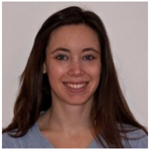
Roman Grundkiewicz
Postdoc

Proyag Pal
Research Assistant

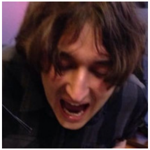
Edinburgh Translation Group 2/2

Marek Strelec
Research Assistant

Alham Aji
PhD student



Anna Currey
PhD student



Nikolay Bogoychev
Shared PhD student

Maxi Behnke
MSc+PhD student

Projects

English–isiXhosa for doctors with UCT

Potentially relevant other projects:

- Health information with UK National Health Service
- Harvesting translations from the web
- Low-resource languages
- Speech translation
- Speed

Projects

English–isiXhosa for doctors with UCT

Potentially relevant other projects:

- Health information with UK National Health Service
- Harvesting translations from the web
- Low-resource languages
- Speech translation
- Speed

Broader projects active in the group:

- Monitoring news in other languages
- Massively open online course translation
- Grammatical error correction

Open-Source Software

Marian

`https://marian-nmt.github.io/`

Neural networks in C++

Nematus

`https://github.com/EdinburghNLP/nematus`

Neural networks in Python and Theano

Moses

`http://www.statmt.org/moses/`

Phrase-based translation

Funding

EU Currently 6 projects.

UK Currently this project.

Industry Amazon, Booking.com, eBay, Facebook, Google, Intel, Microsoft, Mozilla, Samsung, WIPO

Funding

EU Currently 6 projects.
South Africa in Horizon 2020.

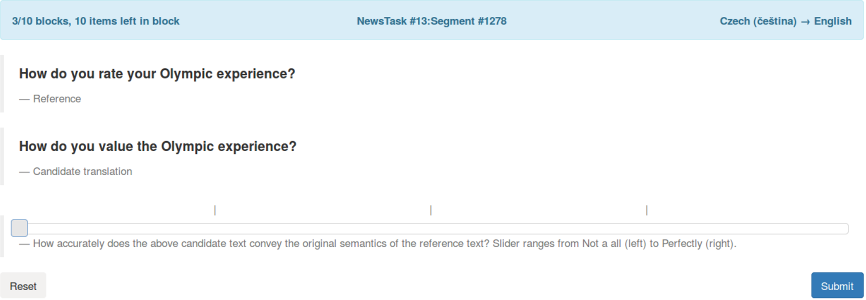
UK Currently this project.
South Africa eligible for “global challenges research fund” (10% of UK science)

Industry Amazon, Booking.com, eBay, Facebook, Google, Intel, Microsoft, Mozilla, Samsung, WIPO
Usually 1–1 arrangements.

1 Evaluating quality
2 Phrase-based models
3 Neural models

1 **Evaluating quality**
2 Phrase-based models
3 Neural models

Direct Assessment



Conference on Machine Translation:
same annotators score competing systems

Winners in 2017 Conference on MT

Constrained news task, including ties:

	From English	To English
Turkish	Edinburgh	Edinburgh
Czech	Edinburgh	Edinburgh
Chinese	Edinburgh	Edinburgh
German	Munich	Edinburgh
Russian	Edinburgh	National Research Council
Latvian	Tilde	Edinburgh
Finnish	Helsinki	UPC

(Edinburgh did not participate in Finnish)

BLEU score

Human evaluation is expensive
$\implies$ Want automatic evaluation

BLEU: how much does the output match a human translator?

- 1-word matches
- 2-word matches
- 3-word matches
- 4-word matches
- Length

Typically expressed as a percentage: 0–100%
1% BLEU increase considered publishable

1 Evaluating quality

2 **Phrase-based models**

3 Neural models

Chambre

$\downarrow$

Bedroom

présidente de la Chambre des représentants

chairwoman of the Bedroom of Representatives présidente de la $\boxed{\text{Chambre des représentants}}$

chairwoman of the $\boxed{\text{House of Representatives}}$

Phrase-based models

Extract translated phrases
Score phrases with translation probabilities
String together to form translations

Phrase-based models

Extract translated phrases
Score phrases with translation probabilities
String together to form translations

Neural is largely replacing phrase-based translation, except for low-resource

1 Evaluating quality
2 Phrase-based models
3 **Neural models**

Turning Words into Vectors

<s>	in	the	raw
$\begin{pmatrix}1\\0\\0\\0\end{pmatrix}$	$\begin{pmatrix}0\\1\\0\\0\end{pmatrix}$	$\begin{pmatrix}0\\0\\1\\0\end{pmatrix}$	$\begin{pmatrix}0\\0\\0\\1\end{pmatrix}$

Assign each word a unique row.

Recurrent Neural Network

<s>

Word $\begin{pmatrix}1\\0\\0\\0\end{pmatrix}$

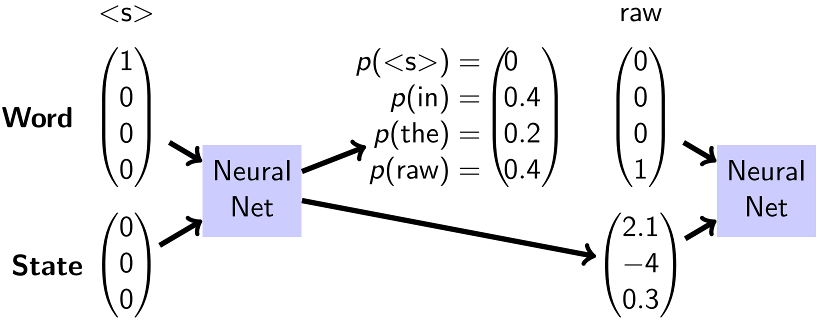
State $\begin{pmatrix}0\\0\\0\end{pmatrix}$

Neural Net

$$\begin{matrix} p(<\text{s}>) = \\ p(\text{in}) = \\ p(\text{the}) = \\ p(\text{raw}) = \end{matrix} \begin{pmatrix}0\\0.4\\0.2\\0.4\end{pmatrix}$$

$$\begin{pmatrix}2.1\\-4\\0.3\end{pmatrix}$$

Recurrent Neural Network



Recurrent Neural Networks for Translation

Read source one word at a time (throw out predictions)
Then predict target words left to right.

Recurrent Neural Networks for Translation

Read source one word at a time (throw out predictions)
Then predict target words left to right.

Problem: sentences don't fit in a 1024-dimensional vector
$\rightarrow$ Look up source word ("attention") while producing target word

Transition to Neural Models

What types of models won pairs in the Conference on Machine Translation?

Year	Neural	Phrase	Rule
2015	1	9	1
2016	6	6	1
2017	14	0	0

Neural is better at agreement

Source Byl to bratr, který bral věci takové, jaké jsou.
Reference He was the brother that went with the flow.
Phrase It was a brother who took things as they are.
Neural He was a brother who took things the way they are.

Decisions based on whole sentence, not just local context.

Word selection

Source	Seit Jahrzehnten fördert Langer den Nachwuchs.
Reference	Langer has been encouraging up-and-coming talent for years.
Phrase	For decades, Langer promotes the offspring.
Neural	For decades, Langer has been promoting the young.

Larger context $\implies$ generally better at fluency.

Rare words are hard

Source	Jennifer Aniston: Ich werde immer in Schubladen gesteckt
Reference	Jennifer Aniston: I'm always pigeonholed
Phrase	Jennifer Aniston: I am always plugged in drawers
Neural	Jennifer Aniston: I'll always be put in drawers

$\Longrightarrow$ Translations covering terminology are important.

Style

Source Erkek kardeşim, her duruma uyum sağlardı.
Reference He was the brother that went with the flow.
Phrase My brother, sağlardı fit in every situation.
Neural My brother was harmonised in every situation.

Neural fits style more closely:

Great if you add in-domain data
Awkward if you don't. Subtitles swear a lot.