



University of
Zurich^{UZH}

Institute of Computational Linguistics

Machine Translation

13 "Applications"

Mathias Müller

Last time

Grösse des Suchraums: Beispiel

$V = \{ \text{EOS, can, everything, except, I, resist, temptation} \}$

$|V| = 7$

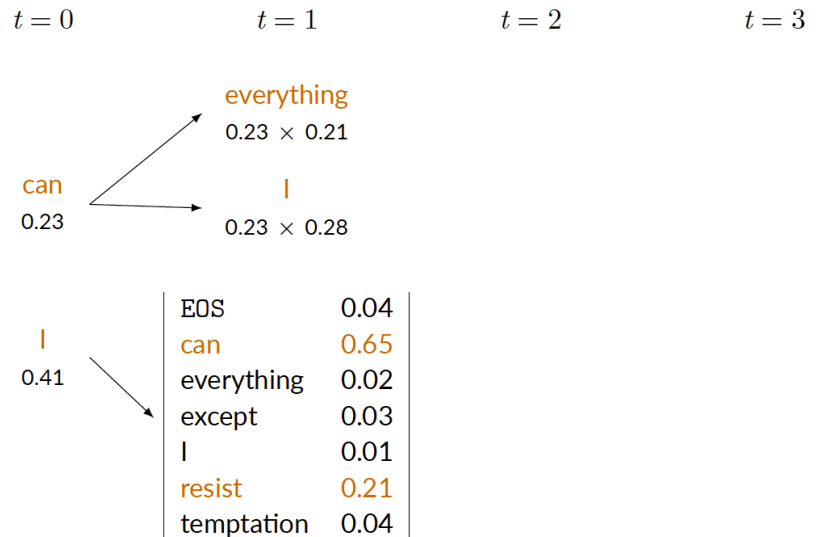
$\text{max_len} = 15$

Grösse des Suchraums: $7^{15} = 4,747,561,509,943$ Sätze

→ Welches ist der beste Satz?

Random Sampling: Beispiel

- $V = \{ \text{EOS, can, everything, except, I, resist, temptation} \}$
- Eingabe: *ich kann allem widerstehen, ausser der Versuchung*
- Mögliche Ausgaben (bei mehreren Durchläufen):
 - *everything except temptation I can resist*
 - *I can resist everything except temptation*
 - *except temptation I can resist everything*
 - *can I resist everything except temptation*
 - *I resist can everything except temptation*



Exam questions

- On May 28, we will have an exam Q&A
- Until May 28, **please** post on OLAT:

Exam question that would be fair in your opinion

- We will discuss exactly those questions that day.

Topics of this lesson

Outcomes:

- LVB outcome
- Exercise 5 competition outcome!

Current NMT:

- WMT19 manual evaluation
- Recent advances that improve NMT

*the
most*



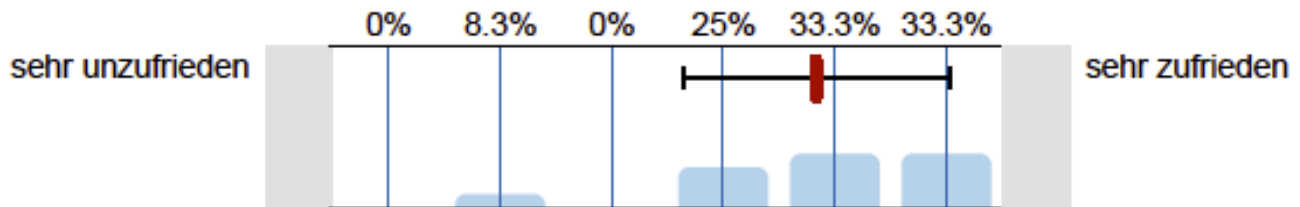
**University of
Zurich** ^{UZH}

Institute of Computational Linguistics

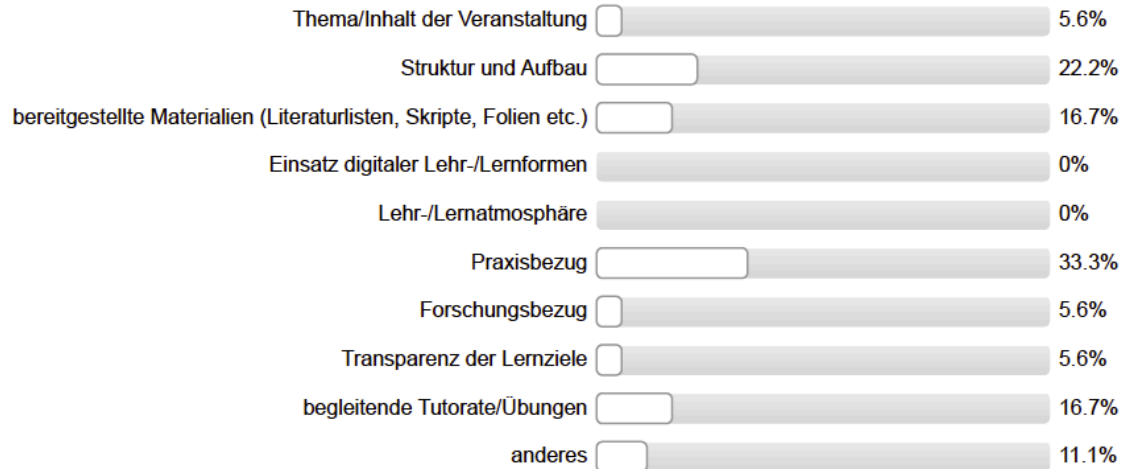
LVB Outcome

LVB

Wie zufrieden sind Sie mit dieser Veranstaltung insgesamt?

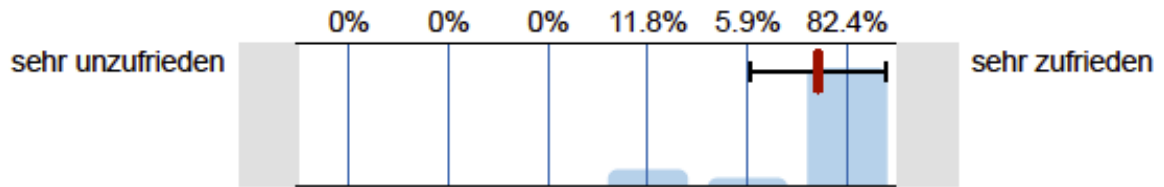


In welchem Bereich hinsichtlich dieser Veranstaltung wünschen Sie sich **Verbesserungen**? (Mehrfachantworten möglich)



LVB

Wie zufrieden sind Sie mit der/dem Dozierenden insgesamt?



In welchem Bereich hinsichtlich der/des Dozierenden wünschen Sie sich **Verbesserungen**? (Mehrfachantworten möglich)



Mehrfach gewünschte Verbesserungen:

- Ausführlichere Slides
- Mehr Einführung in Tensorflow
- Mehr Praxisbezug

- a) Etwas ausführlichere Kommentare und Beschreibungen auf den Folien wären für die Prüfungsvorbereitung hilfreich. Ansonsten sind die Materialien aber sehr gut!
b) Ausführlichere Instruktionen bei den Übungen 4 und 5 (romanesco und daikon) wären hilfreich gewesen: z.B. Einführung in Tensorflow, Code besprechen
- Besonders bei den letzten Übungen hätte ich mir eine aufwendigere Einführung in Tensorflow gewünscht. Das generelle Setup eines Modell & dessen Trainings finde ich nicht intuitiv verständlich und Dokumentation hat auch nicht immer geholfen. Das fehlende Verständnis meinerseits hat dann vor allem zu Frustration geführt, und weniger dass ich Daikon/Romanesco als lehrreiche Anwendungen explorieren konnte. Ich könnte jetzt kein neuronales System mit Tensorflow bauen und darum fehlt für mich auch ein bisschen der Praxisbezug.



**University of
Zurich** ^{UZH}

Institute of Computational Linguistics

Exercise 5 Competition

Winners (BLEU)

BL: 11

1. Sarah Kiener (16.5)

Winners (BLEU)

1. Sarah Kiener (16.5)
2. Eva Bühlmann (15.8)

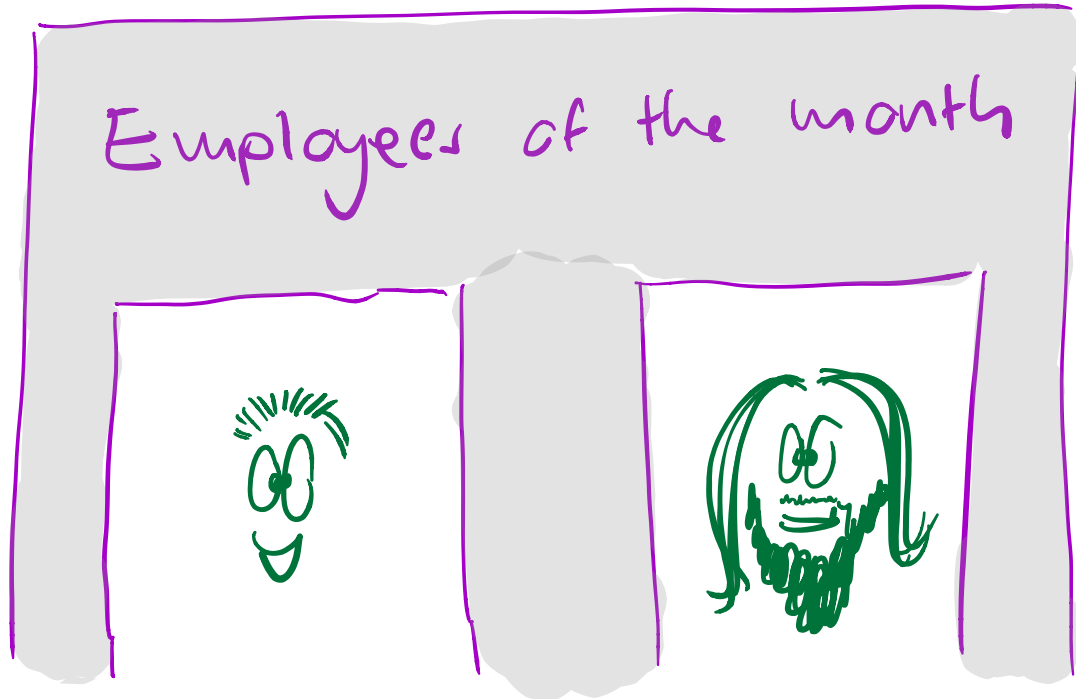
Winners (BLEU)

1. Sarah Kiener (16.5)
2. Eva Bühlmann (15.8)
3. Benjamin Suter (15.5)

*Congrats to
everyone!*

Tied for first place! (Student Satisfaction)

1. Dominik Martinez (100.00)
Nicolas Spring (100.00)





University of
Zurich ^{UZH}

Institute of Computational Linguistics



Matt Post

WMT19 Human Evaluation

WMT19

- WMT is the most important conference on machine translation
- Most prestigious shared task: news translation

NEWS TRANSLATION TASK

This shared task will examine translation between the following language pairs:

- English-Chinese and Chinese-English
- English-Czech
- English-Finnish and Finnish-English
- English-German and German-English
- English-Gujarati and Gujarati-English
- English-Kazakh and Kazakh-English
- English-Lithuanian and Lithuanian-English
- English-Russian and Russian-English
- French-German and German-French

WMT19 human evaluation campaigns

This year, there are 3 human evaluation campaigns:

- Reference-based, segment-level evaluation on AMT, by crowd workers
- Source-based, document-level evaluation on Appraise, by research teams
- **Reference-based, segment-level evaluation on Turkle, by research teams**

Reference-based vs. source-based

Obama is great.

Obama are great.

green sentence
same meaning as
blue sentence ?

Obama ist echt gut.

Obama is realism.

green sentence
same meaning as
blue sentence ?

Segment-level vs. document-level

Obama is great.

Obama are great.

green sentence
same meaning as
blue sentence ?

Obama is great. His campaign
really took off after the
July debate.

Obama are great. His
campaigns really taken
after the July fight.

green document
same meaning as
blue document ?

Segment-level, reference-based campaign

Currently running at

<https://wmt19.waypost.net>



**University of
Zurich** ^{UZH}

Institute of Computational Linguistics

Current NMT

Recent advances that make NMT work well

Broad categories:

- 1) **Architecture changes:** deeper models, **non-recurrent models**
- 2) **Regularization techniques:** **dropout**, residual connections, label smoothing, layer normalization, weight tying
- 3) **Data augmentation:** **back-translation**

Non-recurrent encoder-decoder models

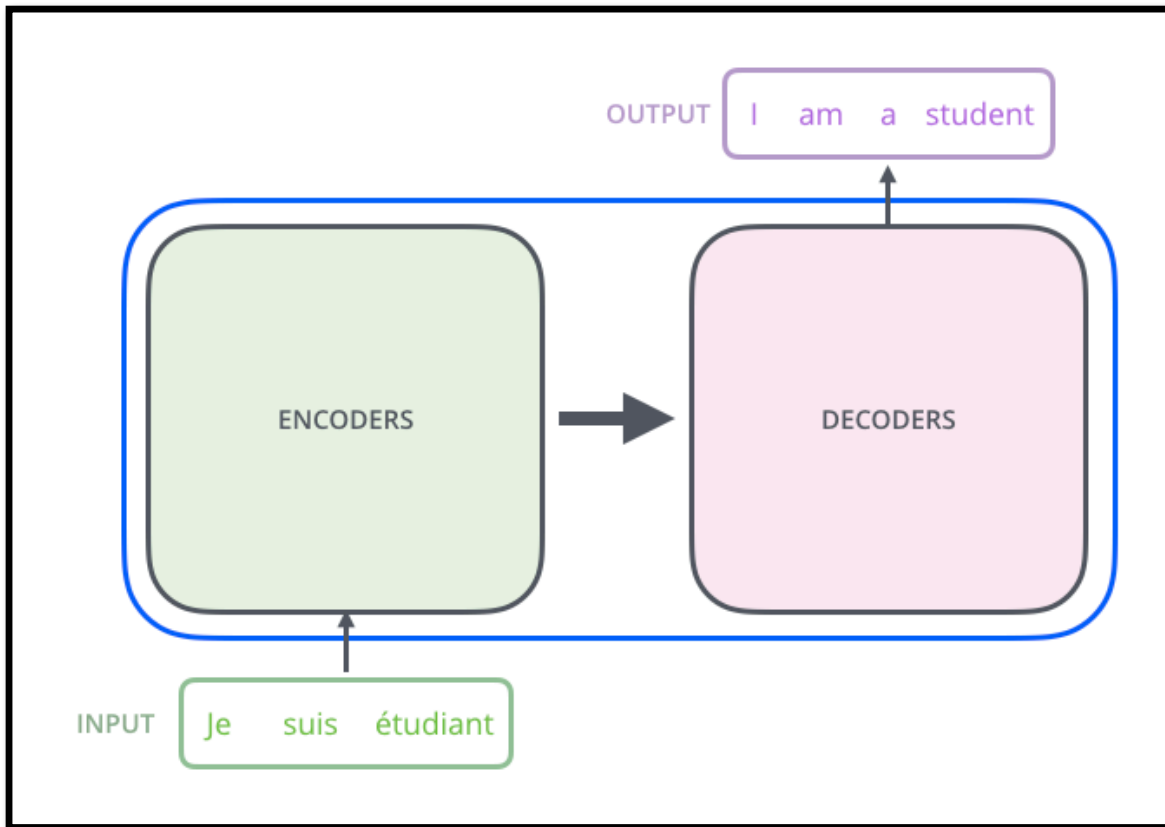
There are non-recurrent NMT models:

- **convolutional** models: use convolution layers, like image recognition

FB 2017

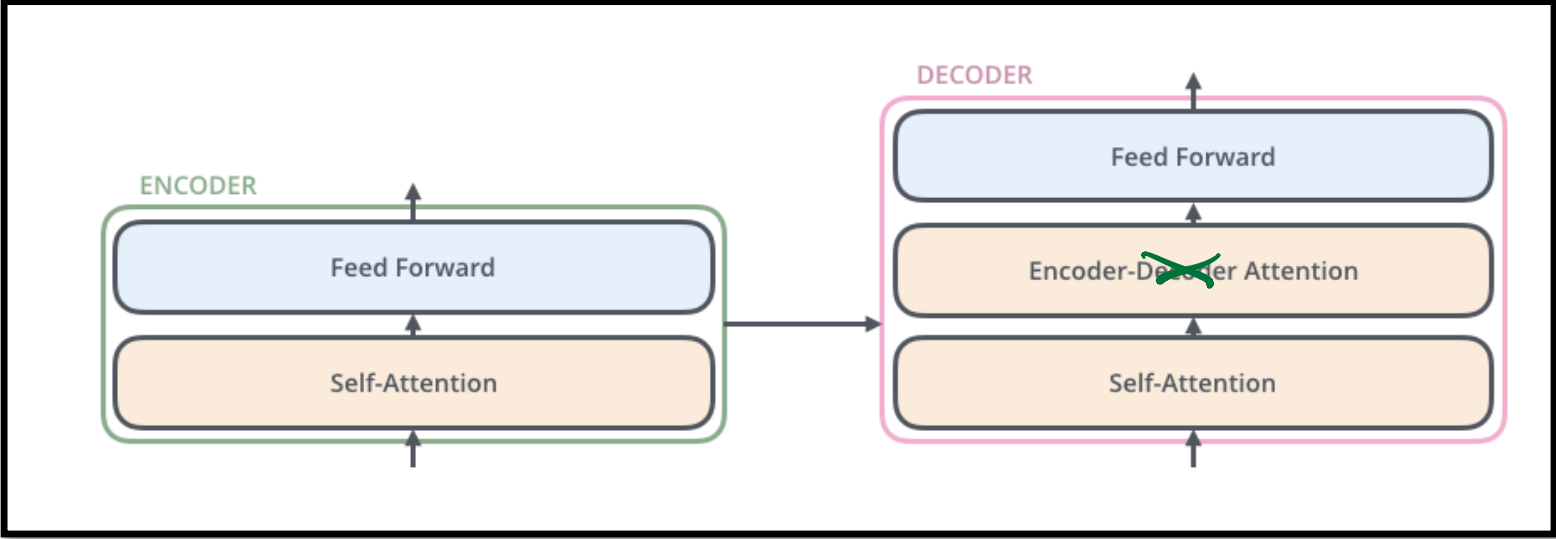
- **self-attentional** or (“Transformer”) models: use feed-forward networks only

Transformer models



<http://jalammar.github.io/illustrated-transformer/>

Transformer models



<http://jalammar.github.io/illustrated-transformer/>

RNN Encoder

$[:] \rightarrow [:] \rightarrow [:]$

RNN Encoder

$[:] [:] [:]$

Je suis étudiant

Transformer Encoder

$[:] [:] [:]$

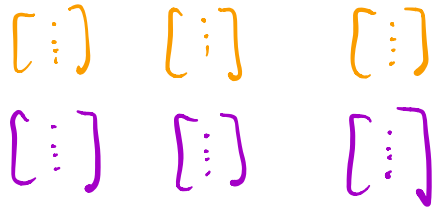
Transformer Encoder

$[:] [:] [:]$

Je suis étudiant

$$\begin{bmatrix} \vdots \end{bmatrix} = W \begin{bmatrix} \vdots \end{bmatrix}$$

Transformer encoder



$\begin{bmatrix} \vdots \end{bmatrix}$ $\begin{bmatrix} \vdots \end{bmatrix}$ $\begin{bmatrix} \vdots \end{bmatrix}$
Je suis étudiant

for "suis" :

$\begin{bmatrix} \cdot \\ \cdot \\ \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot \\ \cdot \\ \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot \\ \cdot \\ \cdot \end{bmatrix}$
Je	suis	étudiant
0.08	0.9	0.002

after self-attention

$$\begin{bmatrix} \cdot \\ \cdot \\ \cdot \end{bmatrix}_{\text{suis}} = 0.08 \times \begin{bmatrix} \cdot \\ \cdot \\ \cdot \end{bmatrix}_{\text{Je}} + 0.9 \times \begin{bmatrix} \cdot \\ \cdot \\ \cdot \end{bmatrix}_{\text{suis}} + 0.002 \times \begin{bmatrix} \cdot \\ \cdot \\ \cdot \end{bmatrix}_{\text{étudiant}}$$

Regularization techniques

Definition:

Place additional constraints on model parameters or training procedure.

Example:

Dropout, very universal regularization technique

Dropout

$$x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad W = \begin{bmatrix} 3 & 1 & -1 \\ 2 & -2 & 0 \\ 2 & 3 & 4 \\ 1 & -1 & 1 \end{bmatrix} \quad y = \text{RELU}(Wx)$$

25% dropout

$$y = \begin{bmatrix} 2 \\ 0 \\ 20 \\ 2 \end{bmatrix} \xrightarrow{\text{dropout}} \begin{bmatrix} 0 \\ 0 \\ 20 \\ 2 \end{bmatrix} \sim 75\%$$

Dropout

Data augmentation

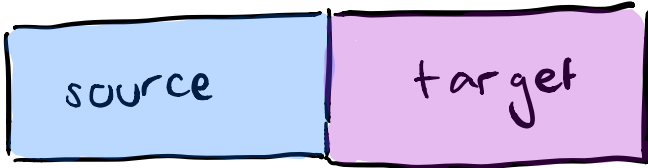
Definition:

Create more training data in a clever way

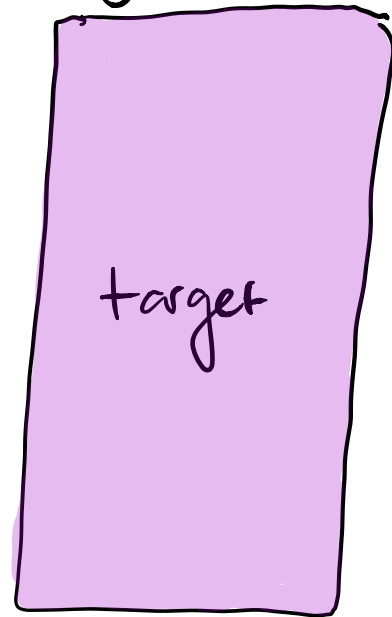
Example for NMT: **Back-translation**

Back-translation

parallel training data

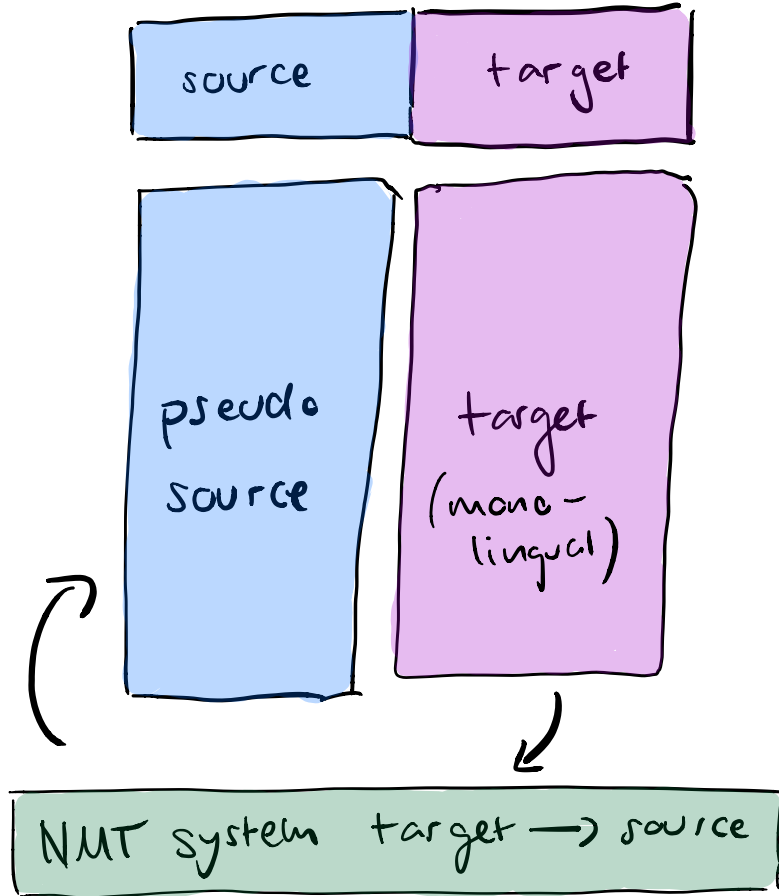


monolingual data?



Back-translation

knowledge
distillation



Summary

- WMT19 News translation task evaluation: human evaluation for all submitted systems
- What makes current NMT work well are new
 - Architectures, e.g. **Transformers**
 - Regularization, e.g. **Dropout**
 - Augmentation, e.g. **Back-translation**

Recommended software

My personal recommendation: work with

~~daikon~~ Sockeye ✓

- Well-written, open-source toolkit
- Good documentation
- Fast, and high engineering standards

Further reading / links

- Look at instructions for WMT 19 human evaluation:
<https://github.com/bricksdont/WMT19RefDA>
- Sockeye: <https://github.com/aws-labs/sockeye/>
- Deeper models paper:
<https://arxiv.org/abs/1707.07631>
- Dropout paper:
<http://jmlr.org/papers/volume15/srivastava14a.old/srivastava14a.pdf>
- Backtranslation paper:
<https://arxiv.org/abs/1511.06709>
- Convolutional NMT paper by Facebook:
<https://arxiv.org/abs/1705.03122>
- Self-attentional, Transformer model paper:
<https://arxiv.org/abs/1706.03762>
- The illustrated Transformer:
<http://jalamar.github.io/illustrated-transformer/>
- The annotated Transformer:
<http://nlp.seas.harvard.edu/2018/04/03/attention.html>

— read together!

Next time

	Cloud Plattform		
30.04.	Encoder-Decoder-Modell	NMT Kapitel 5	Übung 5
07.05.	Attention-Mechanismus, bidirektionales Encoding, Byte Pair Encoding	NMT Kapitel 5-6	
14.05.	Decoding-Strategien	NMT Kapitel 5.4	Übung 6
21.05.	Maschinelle Übersetzung in der Praxis (Anwendungen)		
28.05.	Zusammenfassung, Q&A Prüfung		
Eventuell: Gastvortrag Prof. Artem Sokolov			
Cancelled! Prof. Sokolov had to decline the invitation.			
Prüfung (schriftlich)			
18.06., AND-2-48, 16.15 bis 18:00 Uhr			