Original

This kind of experience is
part of Disney’s efforts
to ”extend the lifetime
of its series
and build new relationships
with audiences via digital platforms
that are becoming ever more important
,” he added.

Translation

Ce type d’expérience fait
partie des initiatives du Disney
pour ”prolonger la durée de vie
de ses nouvelles
et de développer des liens
avec les lecteurs numériques
qui deviennent plus complexes

How can I evaluate a translation?

Original

This kind of experience is
part of Disney’s efforts
to ”extend the lifetime
of its series
and build new relationships
with audiences via digital platforms
that are becoming ever more important
,” he added.

Retranslated - old

With Google Translate / DeepL

This type of experience is
part of Disney's initiatives
to “extend the life
of its stories”
and develop relationships/connections
with digital readers,
which are becoming more complex/sophisticated.”

Neural machine translation by jointly learning to align and translate

Published: ICLR 2015

Authors: Dzmitry Bahdanau, KyungHyun, Cho YoshuaBengio

Goals:

- Why the old Encoder-Decoder method struggled with long sentences
- How the Attention mechanism improves the system

Original

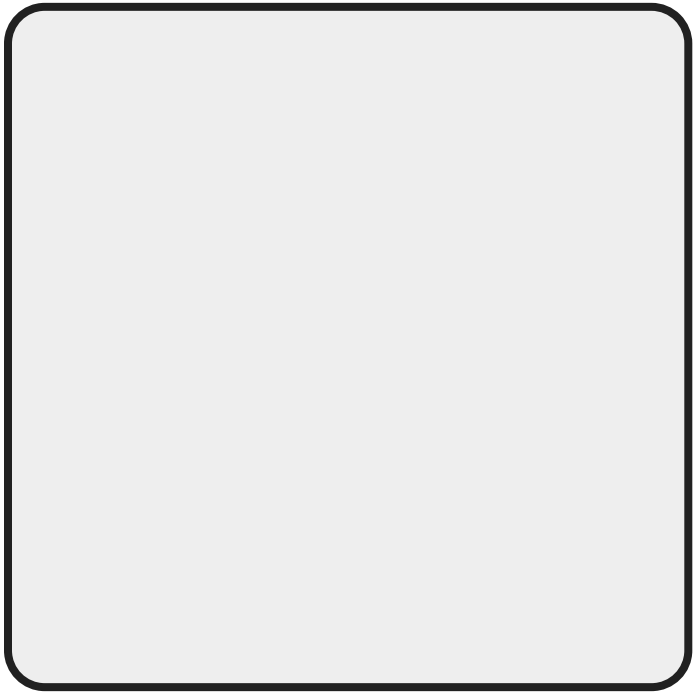
This kind of experience is
part of Disney’s efforts
to ”extend the lifetime
of its series
and build new relationships
with audiences via digital platforms
that are becoming ever more important
,” he added.

Retranslated - PBSMT

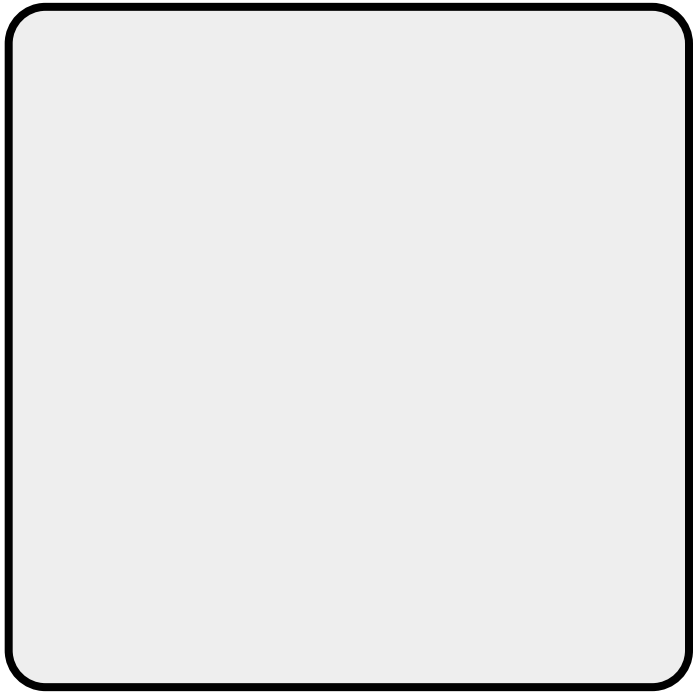
With Google Translate / DeepL

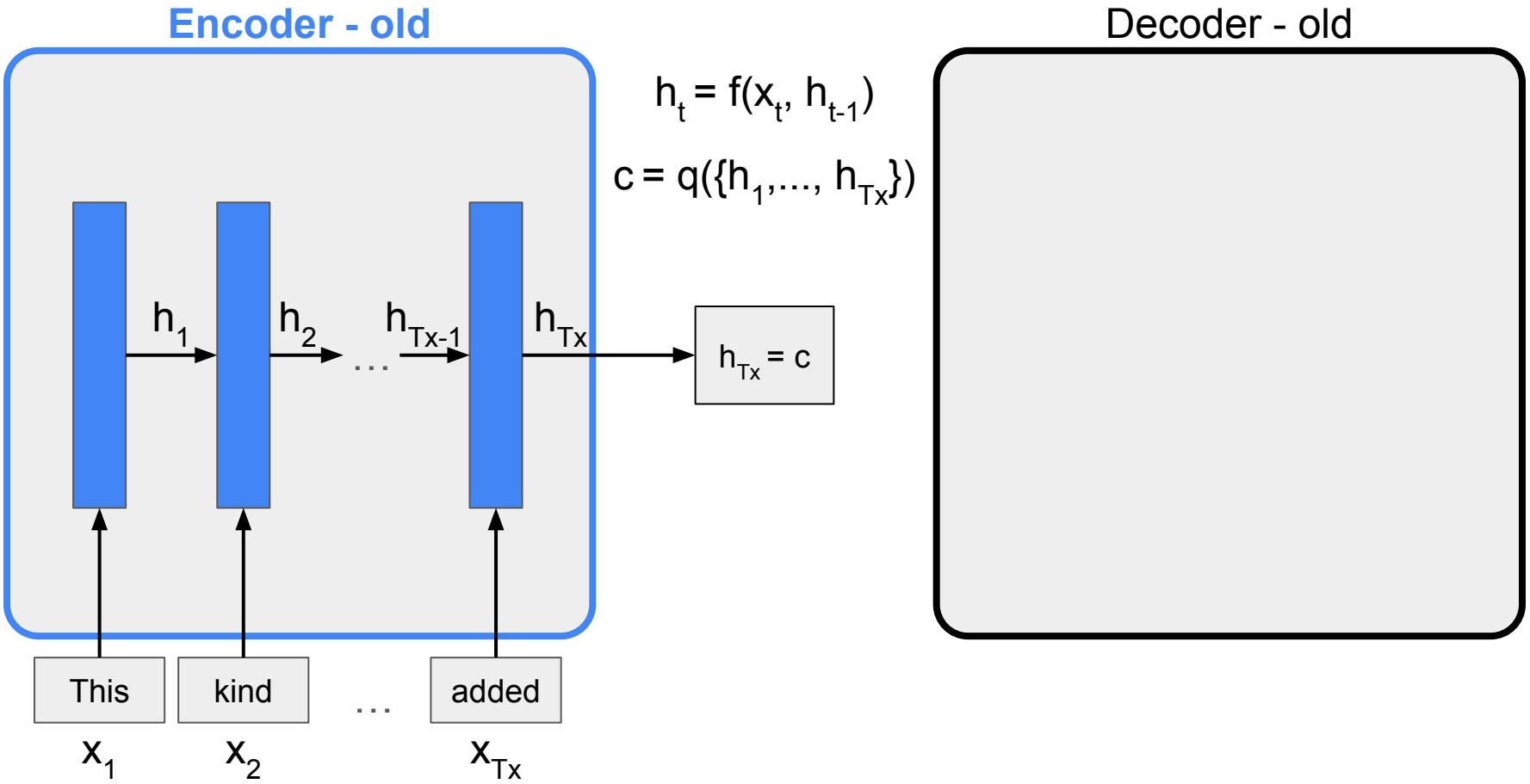
This type of experience is
part of Disney's efforts
to "extend the life
of its series
and create new relationships
with audiences through/via
increasingly important digital platforms
," he added.

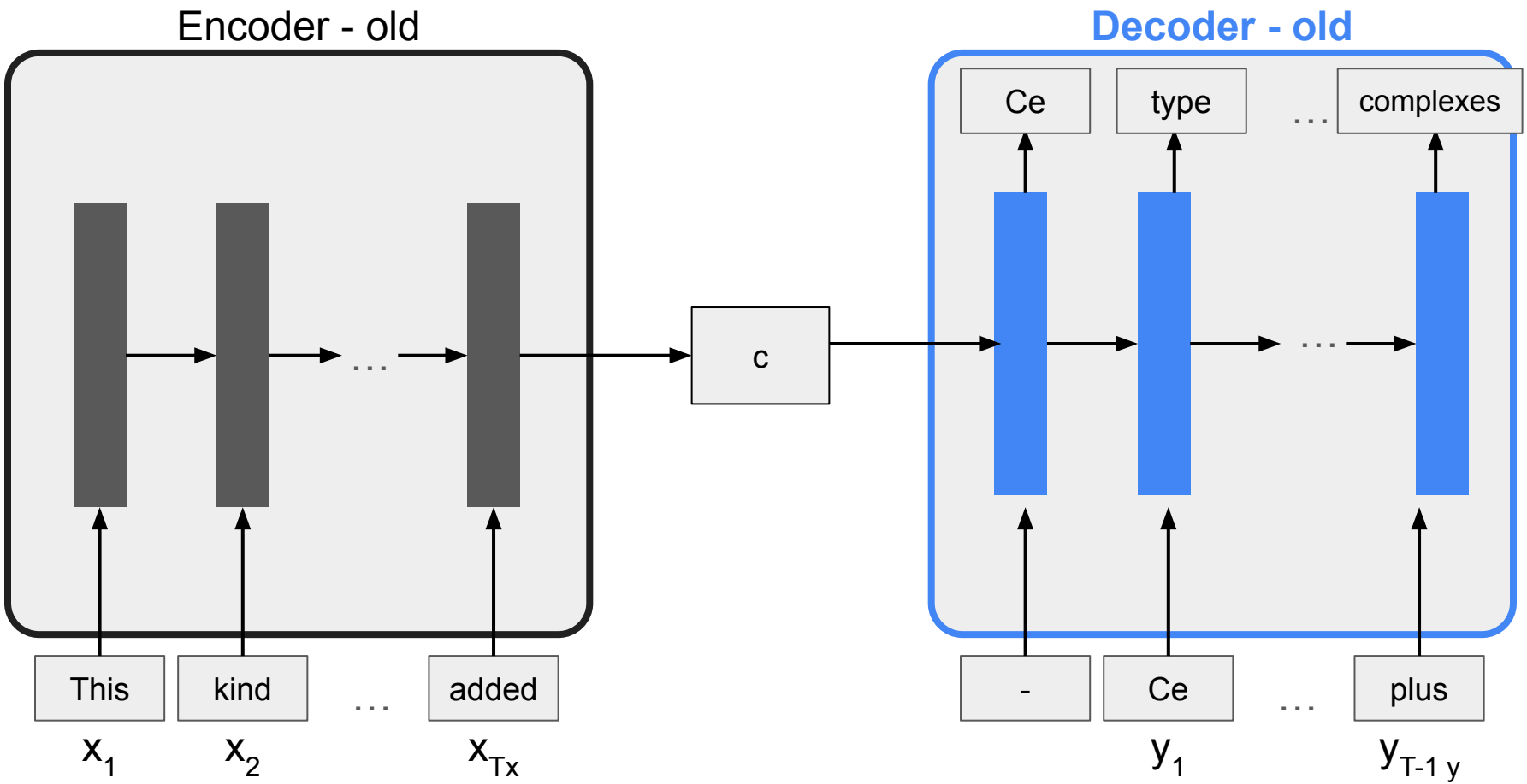
Encoder - old



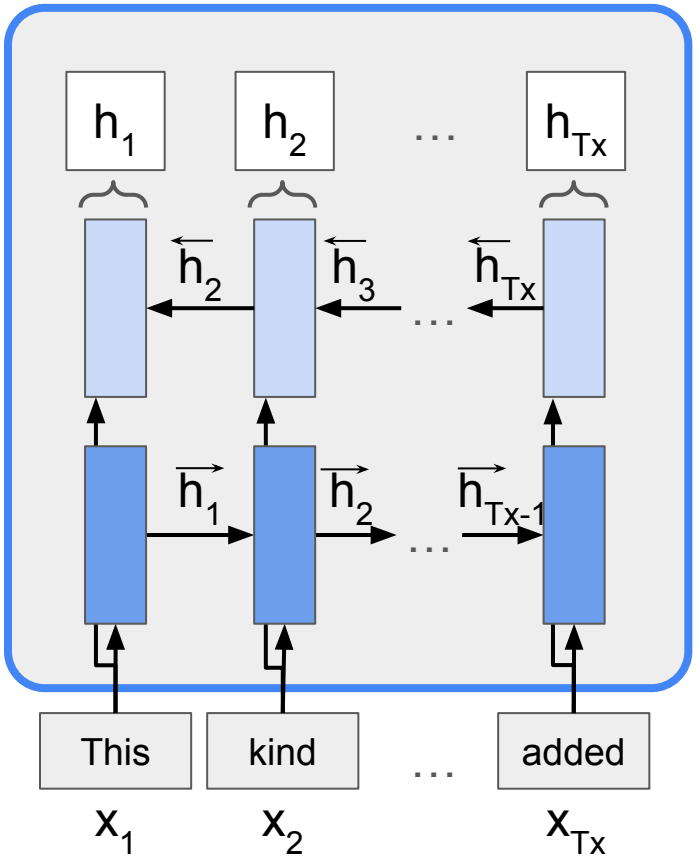
Decoder - old




Encoder - old
Decoder - old
$h_t = f(x_t, h_{t-1})$
$c = q(\{h_1,..., h_{Tx}\})$
h_1
h_2
h_{Tx-1}
h_{Tx}
$h_{Tx} = c$
This
kind
added
x_1
x_2
x_{Tx}


Encoder - old
Decoder - old
Ce
type
...
complexes
c
...
This
kind
...
added
-
Ce
...
plus
x_1
x_2
x_{Tx}
y_1
y_{T-1} y

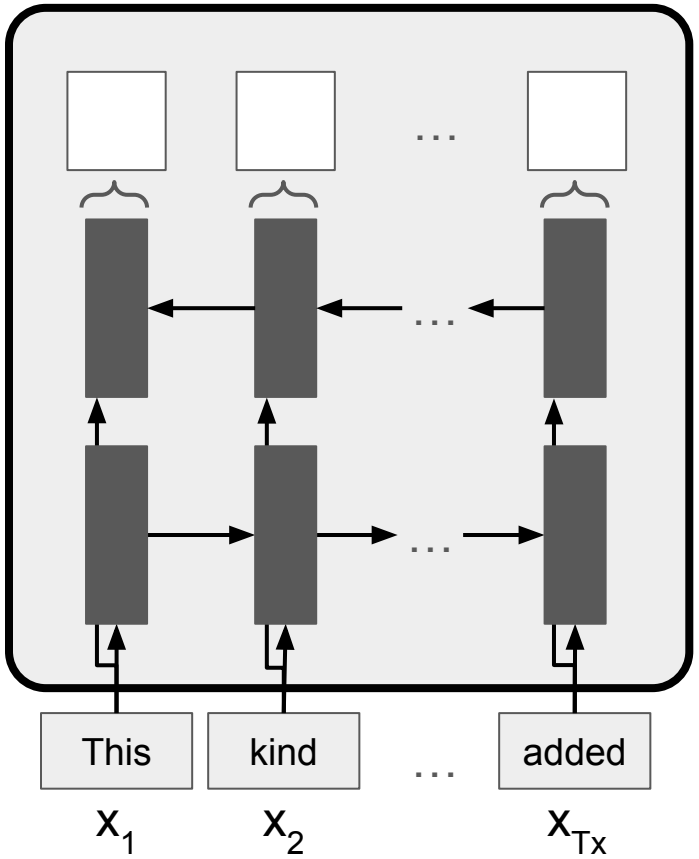
Encoder - new



$h_t = [\overrightarrow{h_t}^T; \overleftarrow{h_t}^T]^T$

Decoder - new

Encoder - new

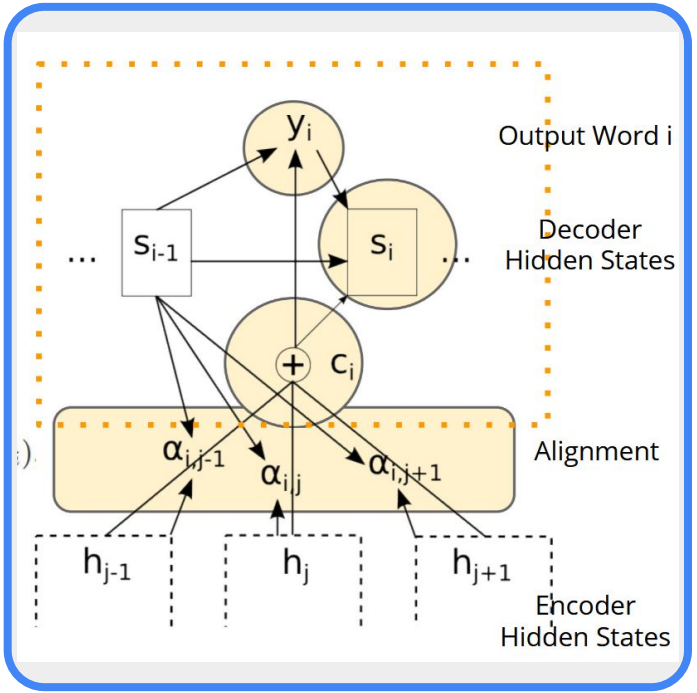


$$c_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j$$

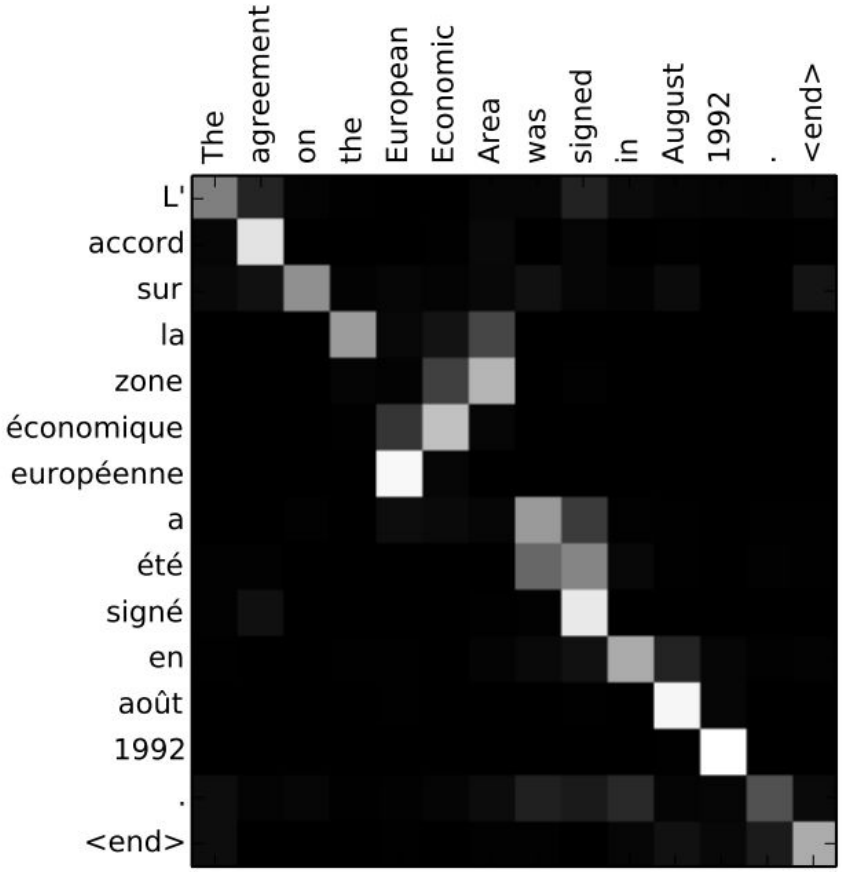
$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})},$$

$$e_{ij} = a(s_{i-1}, h_j)$$

Decoder - new



https://courses.grainger.illinois.edu/cs546/sp2018/Slides/Mar15_Bahdanau.pdf



Original

English → French

Translation

(a)

Original

This kind of experience is
part of Disney’s efforts
to ”extend the lifetime
of its series
and build new relationships
with audiences via digital platforms
that are becoming ever more important
,” he added.

Retranslated - New

With Google Translate / DeepL

This type of experience is
part of Disney's efforts
to "extend the life
of its series
and create new relationships
with audiences through/via
increasingly important digital platforms
," he added.

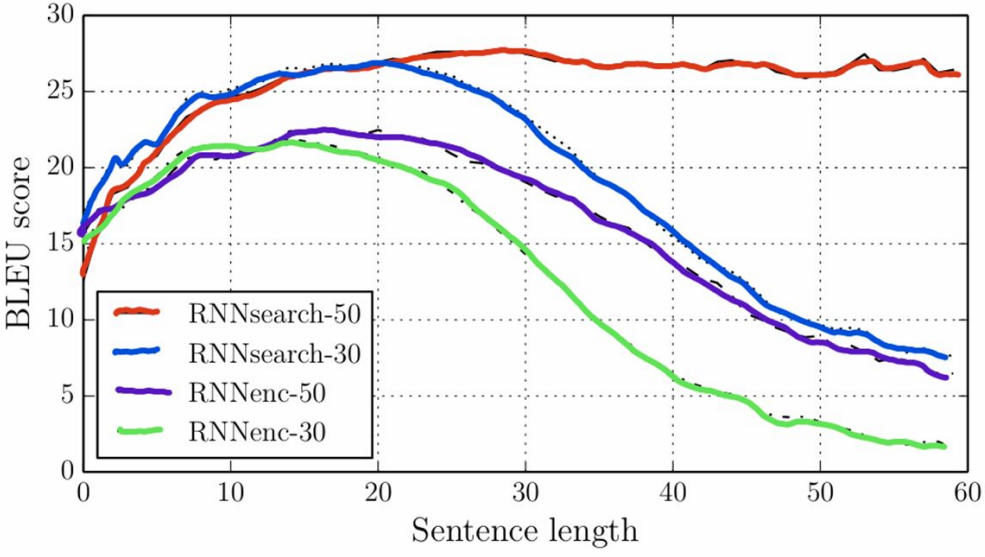
Experiment Settings

- For the dataset they used the WMT 14 the standard dataset for translation evaluations
- They also used a shortlist of 30 000 most frequent words in each language to train their models
- Any word not included in the shortlist was mapped to a special token (UNK)

Experiment Settings

- They used 2 different models for the task
  - RNNsearch (the new model with the attention mechanismus)
  - RNNencdec (classical model without attention mechanismus)
- They trained both models twice
  - One time with sentences of length up to 30 words
  - One time with sentences of length up to 50 words
- The trained every model approximately 5 days

Results



Model	All	No UNK°
RNNencdec-30	13.93	24.19
RNNsearch-30	21.50	31.44
RNNencdec-50	17.82	26.71
RNNsearch-50	26.75	34.16
RNNsearch-50*	28.45	36.15
Moses	33.30	35.63

Results

- Much better BLEU-Score than RNNencdec especially for longer sentences
- The soft alignment allows to recognize phrases and translate them correctly
- The new model is as good in the BLEU-Score as the conventional phrase based models (Moses) when only sentences with known words are considered
- The model works on it‘s own and generates a translation from a source sentence directly

Still remaining problems

- Handling of unknown or rare words
- Slow and unstable training
- Modell is working sentence wise without the context of a complete document
- Attention isn‘t perfect explainable
- Shows where the model is looking at but not why

Conclusion

- The RNNsearch outperforms the RNNencdec significantly
- This paper was the starting signal for the use of end-to-end neural systems
- Basics for all modern sequence to sequence models (transformer, GPT, etc.)

https://de.pinterest.com/pin/867435578202750883/