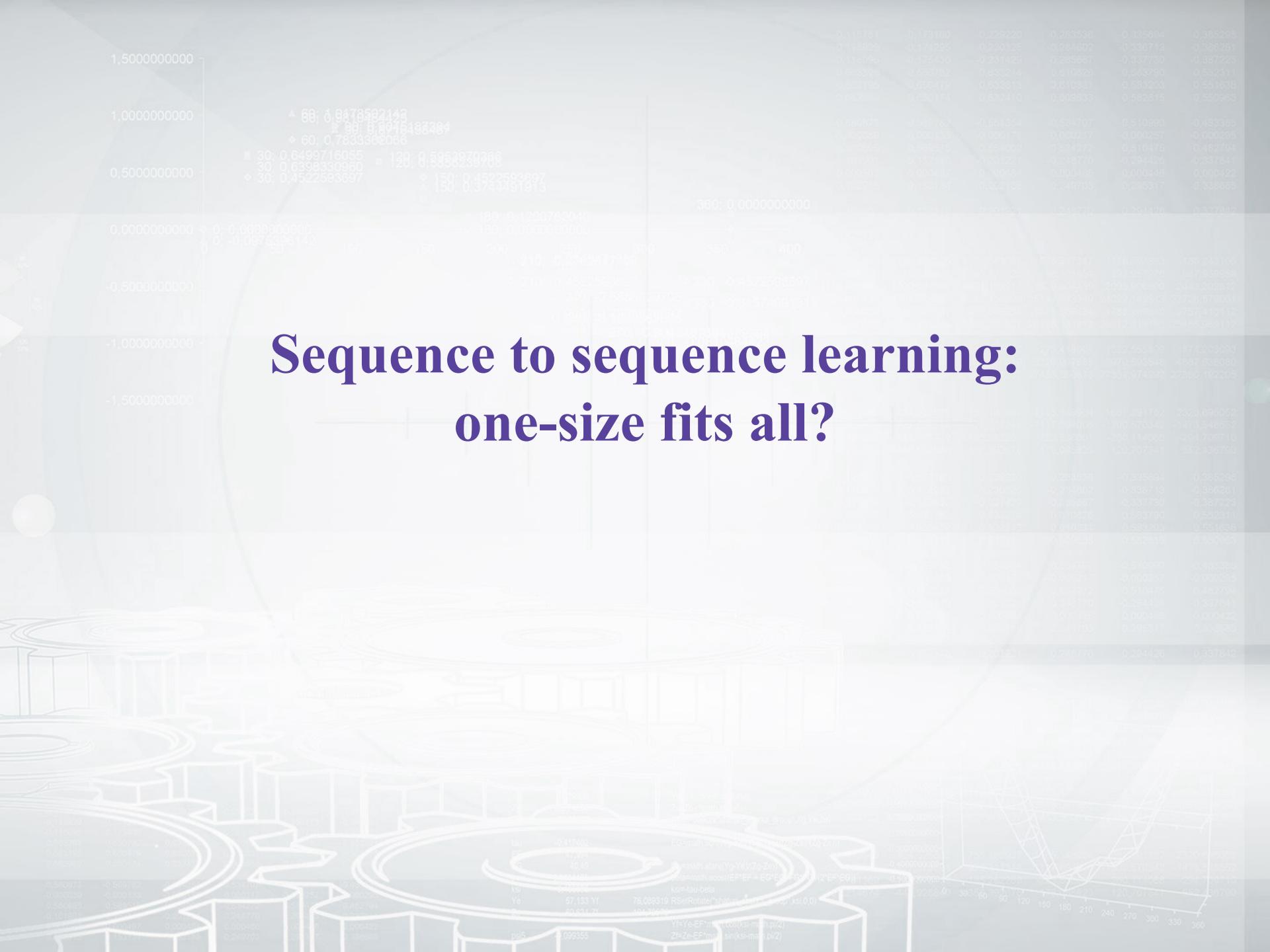


Sequence to sequence learning: one-size fits all?



Sequence to sequence

- Machine Translation
- Summarization
- Text simplification
- Language to code
- Chit-chat bot
- Question answering
- Listen, attend and spell: speech recognition
- Show, attend and tell: image caption generation
- ...

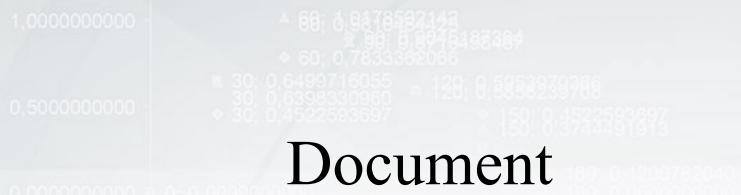


Sequence to sequence

- Machine Translation
- Summarization
- Text simplification
- Language to code
- Chit-chat bot
- Question answering
- Listen, attend and spell: speech recognition
- Show, attend and tell: image caption generation
- ...

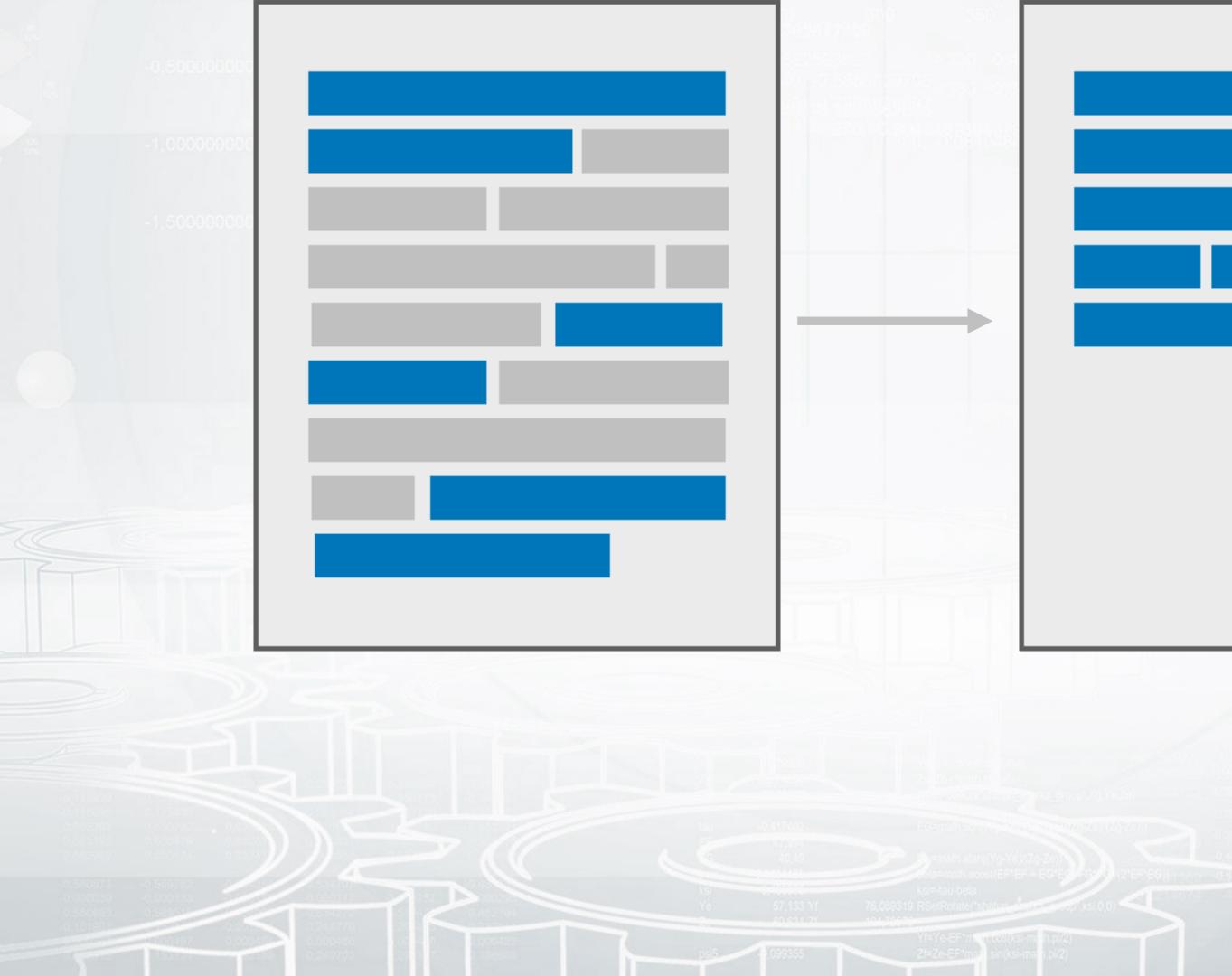


Summarization



360; 0.0000000000

Summary



Summarization

Original Text

Alice and Bob took the train to visit the zoo. They saw a baby giraffe, a lion, and a flock of colorful tropical birds.

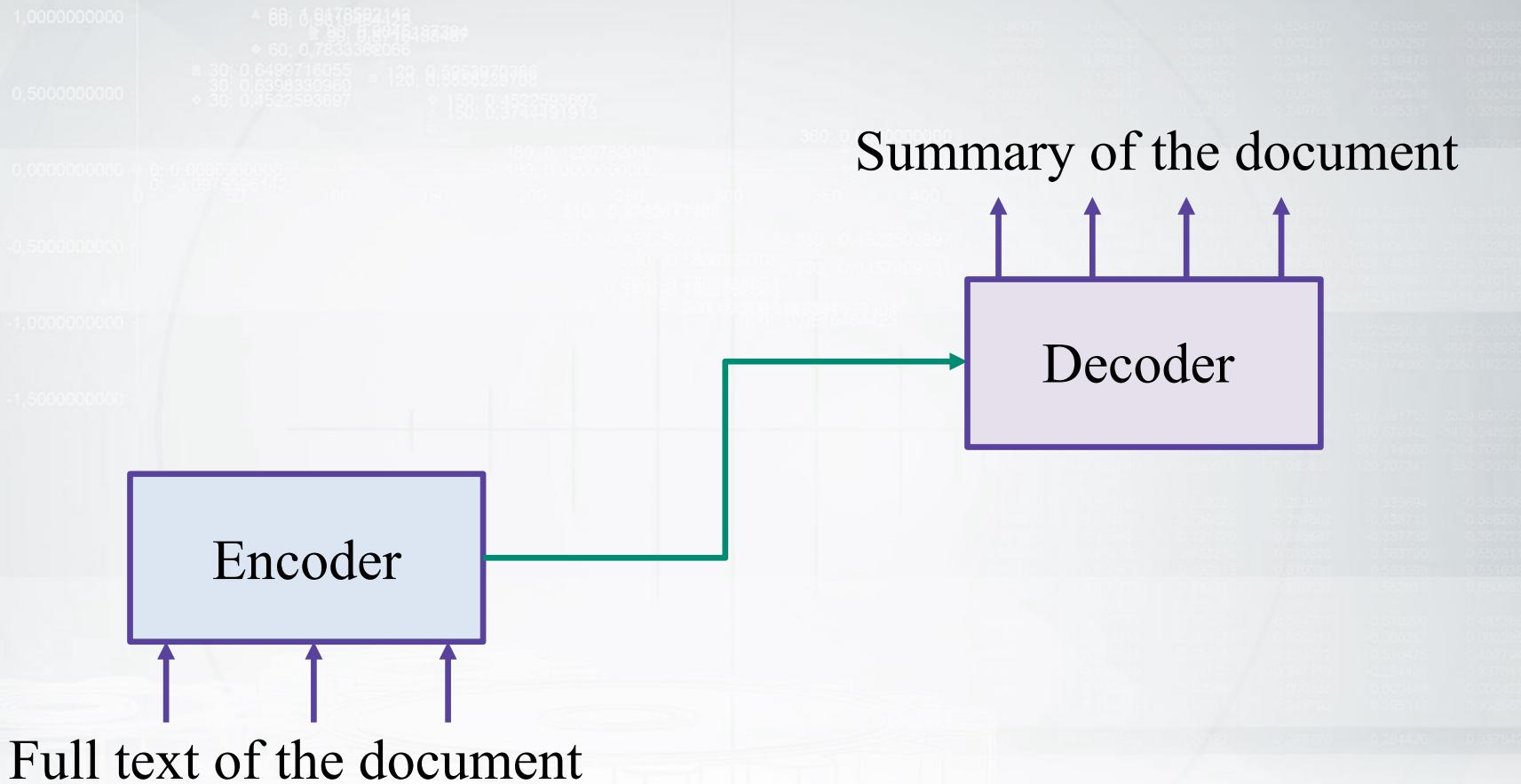
Extractive Summary

Alice and Bob visit the zoo. saw a flock of birds.

Abstractive summary

Alice and Bob visited the zoo and saw animals and birds.

Sequence to sequence!



From Google research blog

Dataset: Annotated English Gigaword – 10 mln. documents
catalog.ldc.upenn.edu/LDC2012T21

Model: sequence to sequence with attention + beam search

Code: open-source TF implementation
github.com/tensorflow/models/tree/master/research/textsum

Results?

From Google research blog

Input: Article 1st sentence	Model-written headline
<i>metro-goldwyn-mayer reported a third-quarter net loss of dlrs 16 million due mainly to the effect of accounting rules adopted this year</i>	<i>mgm reports 16 million net loss on higher revenue</i>
<i>starting from july 1, the island province of hainan in southern china will implement strict market access control on all incoming livestock and animal products to prevent the possible spread of epidemic diseases</i>	<i>hainan to curb spread of diseases</i>

<https://research.googleblog.com/2016/08/text-summarization-with-tensorflow.html>

Encoder-decoder framework

- Machine Translation
- Summarization
- **Text simplification**
- Language to code
- Chit-chat bot
- Question answering
- Listen, attend and spell: speech recognition
- Show, attend and tell: image caption generation
- ...

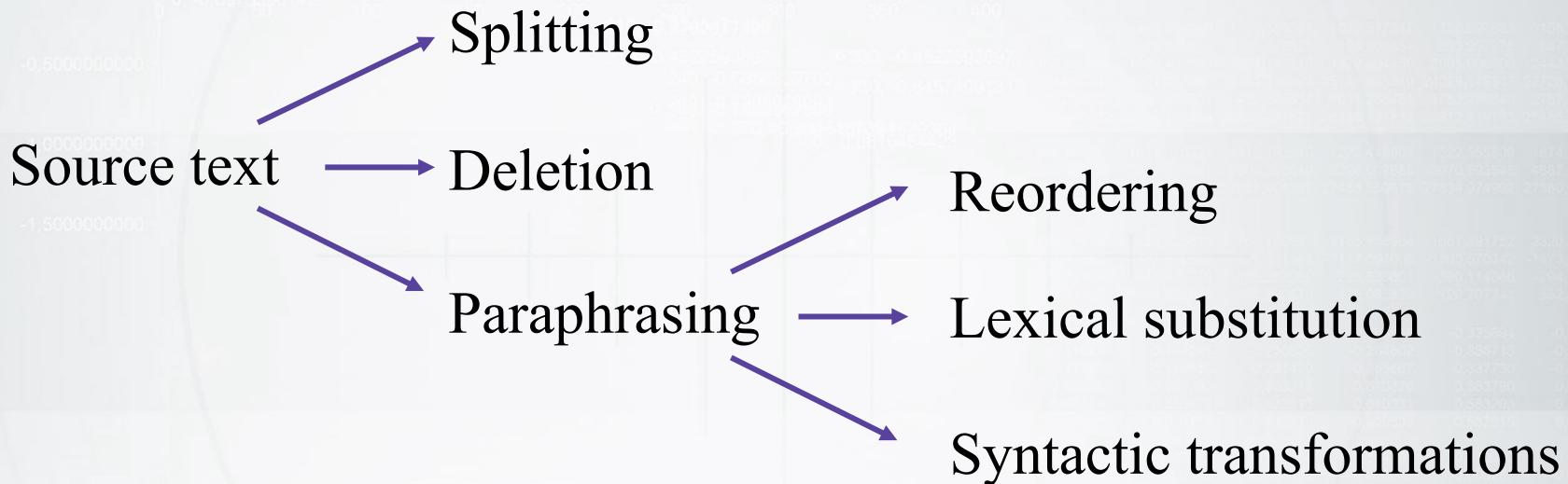


Simplification

Text simplification – reducing the lexical and syntactical complexity of text.

- | | |
|----|--|
| a. | <p>Normal: As Isolde arrives at his side, Tristan dies with her name on his lips.</p> <p>Simple: As Isolde arrives at his side, Tristan dies while speaking her name.</p> |
| b. | <p>Normal: Alfonso Perez Munoz, usually referred to as Alfonso, is a former Spanish footballer, in the striker position.</p> <p>Simple: Alfonso Perez is a former Spanish football player.</p> |
| c. | <p>Normal: Endemic types or species are especially likely to develop on islands because of their geographical isolation.</p> <p>Simple: Endemic types are most likely to develop on islands because they are isolated.</p> |

Operations to simplify text



Xu et. al. Optimizing Statistical Machine Translation for Text Simplification, 2016.

Tong Wang et al. Text Simplification Using Neural Machine Translation, AAAI-16

Rule-based approach for paraphrasing

Lexical	[RB]	solely	→	only
	[NN]	objective	→	goal
	[JJ]	undue	→	unnecessary
Phrasal	[VP]	accomplished	→	carried out
	[VP/PP]	make a significant contribution	→	contribute greatly
	[VP/S]	is generally acknowledged that	→	is widely accepted that
Syntactic	[NP/VP]	the manner in which NN	→	the way NN
	[NP]	NNP 's population	→	the people of NNP
	[NP]	NNP 's JJ legislation	→	the JJ law of NNP

- Synchronous context-free grammar (SCFG) rules
 - Uppercase indicates non-terminal symbols
 - Paraphrase Database <http://www.cis.upenn.edu/~ccb/ppdb/>

Simplification

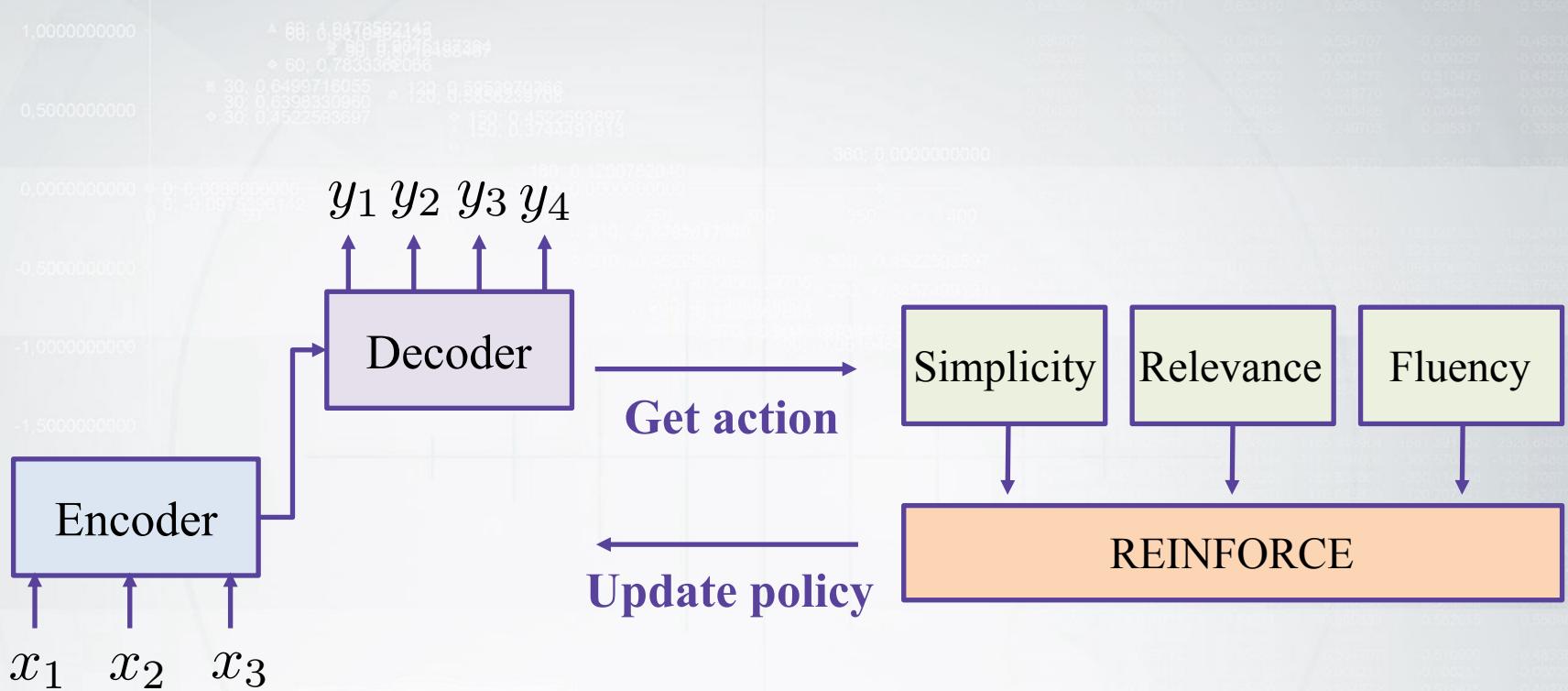
Encoder-decoder framework – yes, but the network might learn just to **copy** the content... How do we force it to **simplify**?

Reinforcement learning can be used to do **weak supervision**.

- **Action:** output next word y_j
- **Policy:** $p(y_j | \mathbf{x}, y_1, \dots, y_{j-1})$
- **Reward:** Adequacy + Fluency + Simplicity

Rewards come only when the whole sequence is generated.

Simplification



Zhang, Lapata. Sentence Simplification with Deep Reinforcement Learning, 2017.

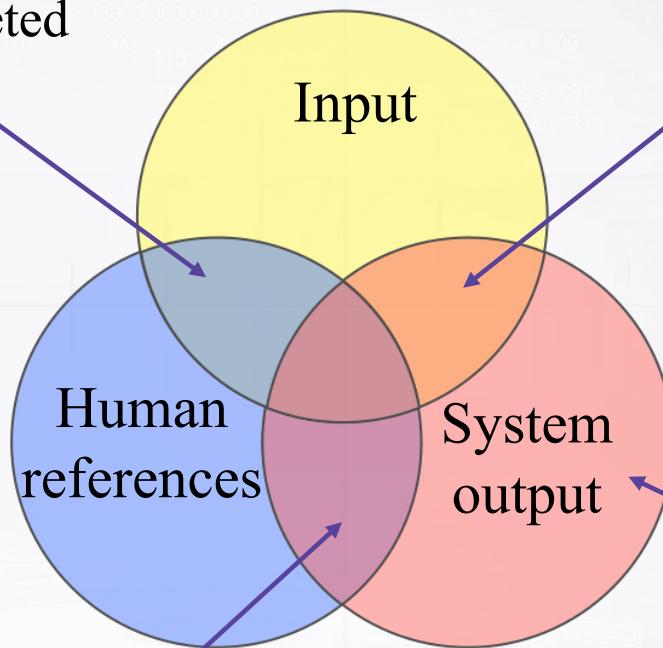
How to measure simplicity?

Input that is retained in the references, but was deleted by the system

Input that is unchanged by the system and which is not in the references

Input that was correctly deleted by the system, and replaced by content from the references

Potentially incorrect system output



How to measure simplicity?

SARI (system against references and input) –

arithmetic average of n-gram precision and recall of

- addition
- copying
- deletion

For example, precision for **addition**:

$$\text{precision} = \frac{\sum_{g \in O} [g \in (O \cap \bar{I} \cap R)]}{\sum_{g \in O} [g \in (O \cap \bar{I})]}$$

SARI: example

INPUT: About 95 species are currently accepted.

REF-1: About 95 species are currently known.

*REF-2: About 95 species are **now** accepted.*

REF-3: 95 species are now accepted.

OUTPUT-1: About 95 you now get in. → **0.2683**

*OUTPUT-2: About 95 species are **now** agreed.* → **0.7594**

OUTPUT-3: About 95 species are currently agreed. → **0.5890**

Compare with BLEU

INPUT: About 95 species are currently accepted.

REF-1: About 95 species are currently known.

*REF-2: About 95 species are **now** accepted.*

REF-3: 95 species are now accepted.

OUTPUT-1: About 95 you now get in. → **0.1562**

*OUTPUT-2: About 95 species are **now** agreed.* → **0.6435**

OUTPUT-3: About 95 species are currently agreed. → **0.6435**

BLEU does not distinguish between outputs 2 and 3.