

Self-Attention Architectures for Answer-Agnostic Neural Question Generation

Thomas Scialom^{*‡}, Benjamin Piwowski^{◇‡} and Jacopo Staiano^{*}



reciTAL.

◇ CNRS

‡ LIP6 - Sorbonne Universités - UPMC Paris 06

* reciTAL, Paris, France

{thomas,jacopo}@recital.ai
benjamin.piwowski@lip6.fr



Task: Question Generation

The ability to ask questions is linked to reading comprehension level.
Generate correct and relevant questions given an input text.

Research Questions

We observed poor performances of a Vanilla Transformer architecture:
⇒ **which mechanisms could be added to a Transformer?**
⇒ **how do they affect the performances on the task?**

Answer-Agnostic Setting

Same *end-to-end* setting used in Du et al. [1]. We train a Transformer (Vaswani et al. [2]) model on the QG task (using SQuAD dataset) without constraining generation to a pre-selected answer span.

Sample #1

Source: Under Rockne, the Irish would post a record of 105 wins, 12 losses, and five ties.

Human: What was the amount of wins Knute Rockne attained at Notre Dame while head coach?

Ours: how many losses did the irish have ?

Sample #2

Source: Chopin was of slight build, and even in early childhood was prone to illnesses.

Human: What was Frédéric prone to during early childhood as a result of his slight build?

Ours: what type of disease did chopin have ?

Sample #3

Source: The Montana Act led to the arrest of over 200 individuals and the conviction of 78, mostly of German or Austrian descent.

Human: How many people were arrested from the Montana Act?

Ours: how many individuals were killed in the montana act ?

Human Evaluation

100 context-question pairs from the test set, randomly sampled. N=3.

	Correct	Fluent	Sound	Answerable	Relevant
<i>Transformer_base</i>	4.49	4.02	3.33	1.7	2.51
<i>+PH+Copy+ELMO</i>	4.5	4.12	3.78	2.87**	3.59*
Du et al. [1]	4.53	4.15	3.64	2.45	3.27

Table 1: Two-tailed t-test for our best method compared to [1]; **: p<.005; *: p<.05.

Implementation Details

N = 2 # number of blocks
d_model = 256 # hidden state dimension
d_ff = 512 # position-wise feed-forward net dimension
h = 2 # number of attention heads

Potential applications

Chat-bots, AI-supported learning, Information Retrieval,
Summarization, Data Augmentation

Pipeline

Context-free embeddings (base):

Word-level tokenization and (GloVe) representations (Pennington et al. [3])

Copy:

Pointer-softmax (Gulcehre et al. [4]) to select words to be copied from the source sentence. The generation probability $p_{gen} \in [0, 1]$ at time-step t is calculated as:

$$p_{gen} = \sigma(W \cdot (h^* \oplus s_t \oplus x_t))$$

Contextualized Embeddings (ELMO):

The context-free vectors are concatenated, at the encoding stage, with ELMO contextual representations (Peters et al. [5])

Placeholding (PH):

Enforce the copy of named entities from the source to the target language, often used in MT systems (Crego et al. [6])

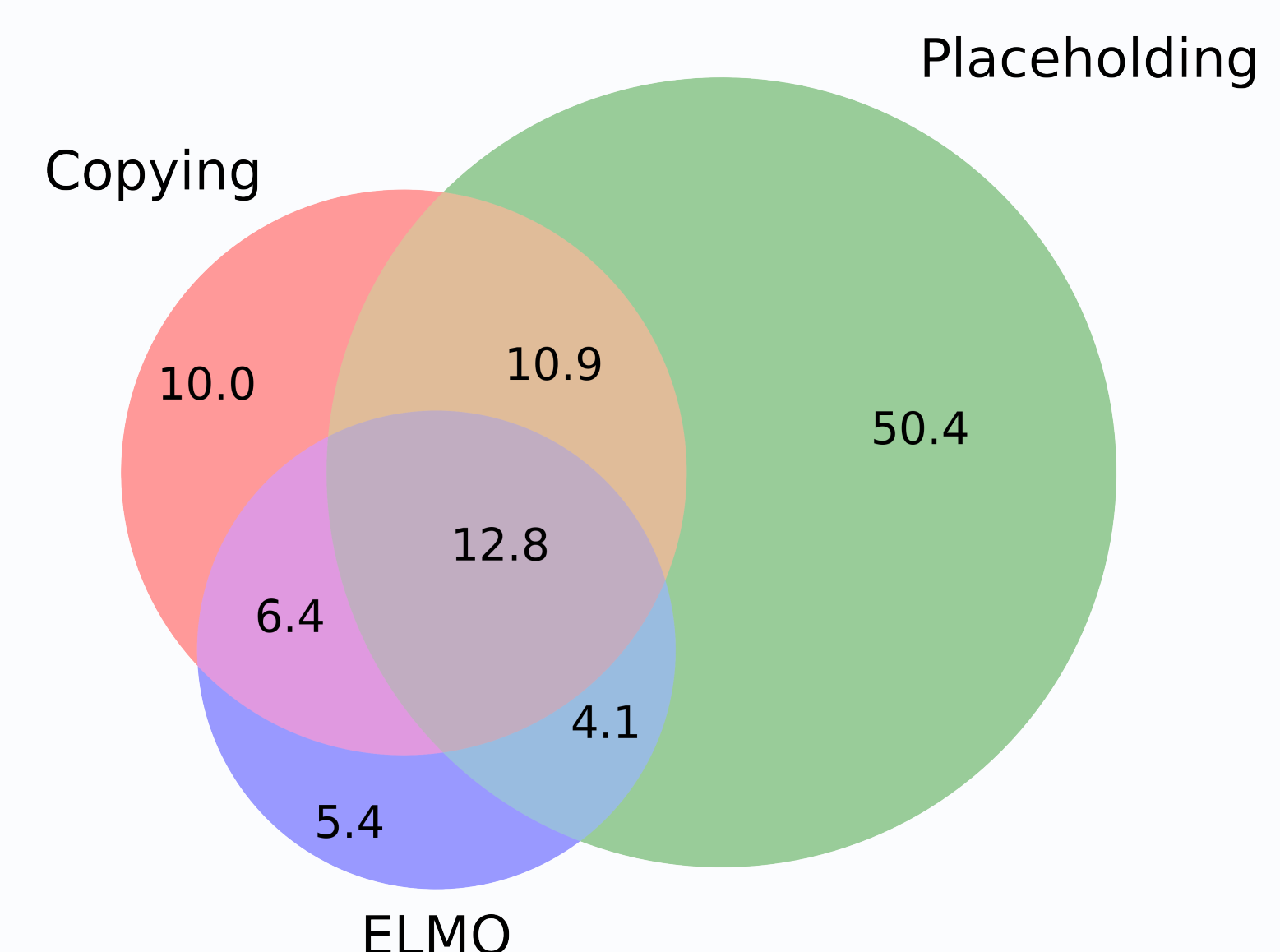
Ablations

	BLEU1	BLEU2	BLEU3	BLEU4	ROUGE-L
<i>Vanilla Transformer</i>	36.13	17.77	10.04	6.04	33.17
<i>Transformer_base</i>	38.74	20.54	12.26	7.66	35.69
<i>+Copy</i>	39.81	22.47	14.25	9.32	37.28
<i>+ELMO</i>	40.44	23.87	15.74	10.62	38.32
<i>+Copy+ELMO</i>	41.72	25.07	16.77	11.58	39.22
<i>+PH</i>	41.54	25.52	17.56	12.49	39.26
<i>+PH+ELMO</i>	42.2	26.2	18.14	12.92	40.23
<i>+PH+Copy</i>	42.72	26.52	18.28	13.0	39.63
<i>+PH+Copy+ELMO</i>	43.33	26.27	18.32	13.23	40.22
Du et al. [1]	43.09	25.96	17.50	12.28	39.75

Table 2: Results obtained under different ablations.

Effects on Copying Behavior

Percentage of OOV tokens copied by the different mechanisms and combinations thereof, over all OOV tokens copied.



References

- [1] Learning to Ask: Neural Question Generation for Reading Comprehension. Du et al. In *ACL 2017*.
- [2] Attention is all you need. Vaswani et al. In *NIPS 2017*.
- [3] Glove: Global Vectors for Word Representation. Pennington et al. In *EMNLP 2014*.
- [4] Pointing the Unknown Words. Gulcehre et al. In *ACL 2016*.
- [5] Deep Contextualized Word Representations. Peters et al. In *NAACL 2018*.
- [6] Systran's Pure Neural Machine Translation Systems. Crego et al. 2016 *arXiv:1610.05540*.