

Yorick Chern
Part 2a.

A good alignment:

de: ['eine', 'person', 'fährt', 'auf', 'einer', 'verschneiten', 'straße', 'fahrrad']
en: ['a', 'person', 'riding', 'a', 'bike', 'on', 'a', 'snowy', 'road']

eine ==> a
person ==> person
fährt ==> riding
auf ==> on
einer ==> a
verschneiten ==> snowy
straße ==> road
fahrrad ==> bike

This sentence pair was able to produce a very crisp one-to-one translation.

A bad alignment:

de: ['jungen', 'tanzen', 'mitten', 'in', 'der', 'nacht', 'auf', 'pfosten']
en: ['boys', 'dancing', 'on', 'poles', 'in', 'the', 'middle', 'of', 'the', 'night']

jungen ==> boys
tanzen ==> the
mitten ==> the
in ==> in
der ==> the
nacht ==> night
auf ==> on
pfosten ==> poles

This sentence produced a pretty bad alignment. “Mitten” should be “in the middle” and “tanzen” should be “dancing.” First of all, “mitten” in German stands for “in the middle,” and although the word “middle” seems to be more appropriate as an alignment, it is clear that “in the middle” always show up together whenever “mitten” is counted, so that’s why the model is unable to exactly identify whether it should map “mitten” to in, the, or middle. For “tanzen”, the model was able to understand that it means “dance,” but it was unable to reason that “dancing” is the same as “dance” in context, as shown by the probability analysis:

$P(\text{dance} \mid \text{tanzen}) = 0.294913456168628$
 $P(\text{dancing} \mid \text{tanzen}) = 0.02111680882816914$
 $P(\text{the} \mid \text{tanzen}) = 0.038346763508053196$

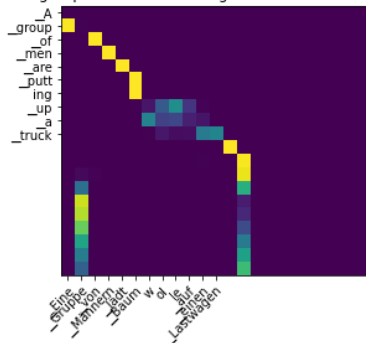
The words “the” and “dancing” probably show up at a very similar frequency, so the model was unable to tell them apart although it can tell that tanzen quite literally means “dance.”

Part 2b.

Eine Gruppe von Männern lädt Baumwolle auf einen Lastwagen

->
A group of men are putting up a truck

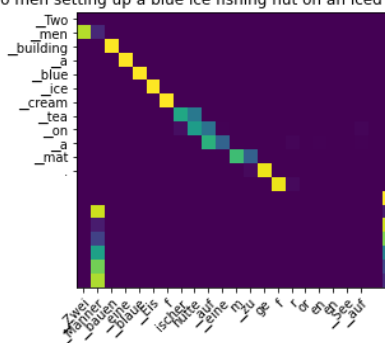
->
A group of men are loading cotton onto a truck. (trg)



Zwei Männer bauen eine blaue Eisfischerhütte auf einem zugefrorenen See auf

->
Two men building a blue ice cream tea on a mat.

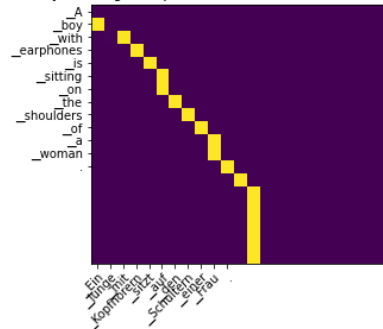
->
Two men setting up a blue ice fishing hut on an iced over lake. (trg)



Ein Junge mit Kopfhörern sitzt auf den Schultern einer Frau.

->
A boy with earphones is sitting on the shoulders of a woman.

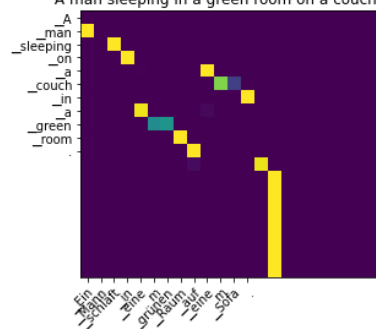
->
A boy wearing headphones sits on a woman's shoulders. (trg)



Ein Mann schläft in einem grünen Raum auf einem Sofa.

->
A man sleeping on a couch in a green room.

->
A man sleeping in a green room on a couch. (trg)



With a validation accuracy of 66%, these are pretty reasonable and expected performance of attention heat maps. Let's look at the bottom two images, where we have relatively good translations. The German word "sitzt" (actually means sit) is mapped to "is" (an incorrect one-to-one) while the German word "auf" (on) is mapped to "sitting on" together (one-to-many). This shows that the model while translating "auf" was paying attention to the pretext and the context of "sitting on" (another one-to-many). Additionally, the German word "Frau" was mapped to the words "a woman" (I deem this one-to-one because a woman is really one definition), which is consistent with the actual translation. The top right was a particularly bad translation, and we can see a peculiar attention distribution. Firstly, the model was able to correctly partially translate "Zwei männer bauen eine blaue eis" to "two men building a blue ice" before the translation turned disastrous. Since then, it appears that the words "ice cream" simply appeared much, much more often than "ice fishing," so we see the model's translation to head to the wrong direction. Interestingly, the top left translation, although pretty erroneous, we see the model focusing the word "lastwagen" on the word "truck," a great one-to-one mapping. This is probably due to the fact that there are not many words that can be confused with truck. Overall, the maps are intuitive and explain the model's "thought process" pretty clearly.