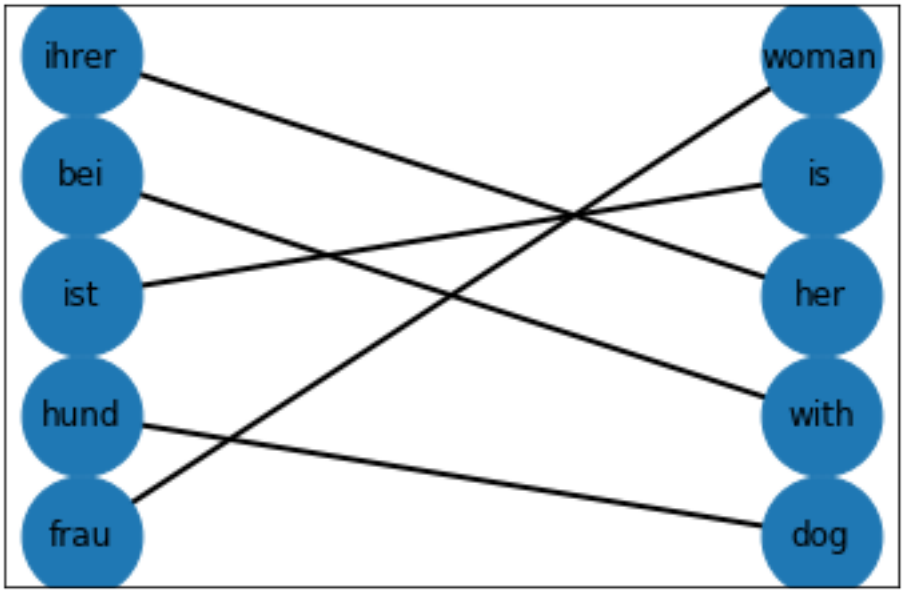
HW2: Neural Machine Translation

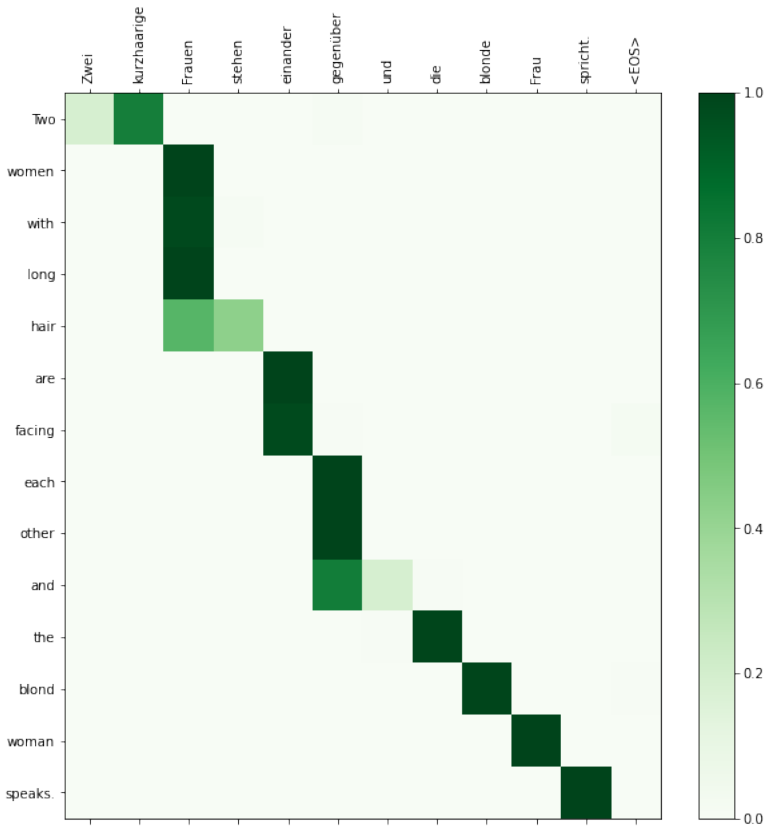
Shreshta Bhat

February 25, 2022

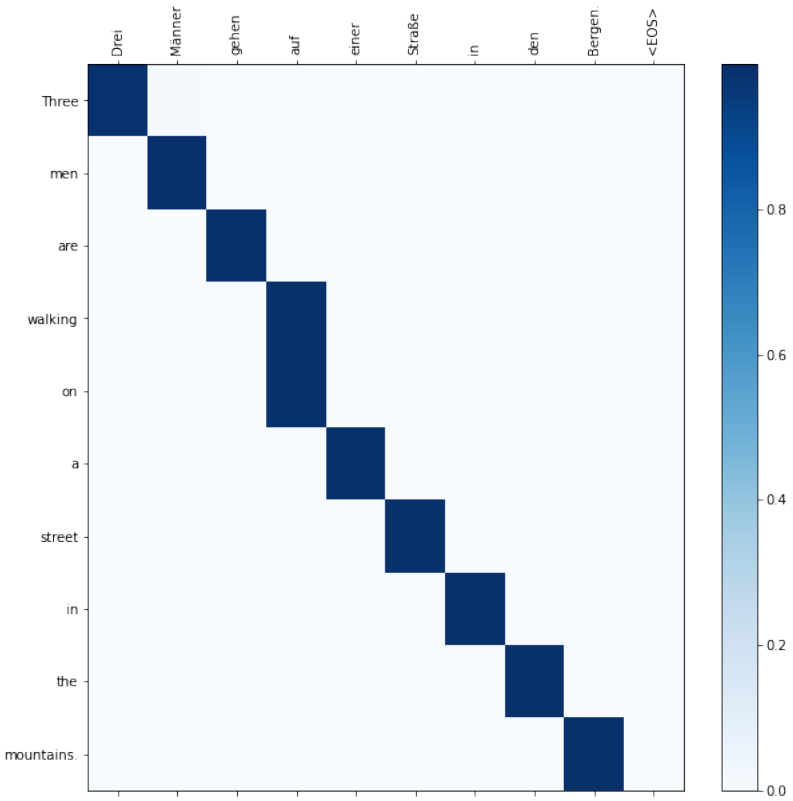
1. The alignment model does well on simple translations with one-to-one mapping as shown in the visualization below. The major shortcoming of model 1 is the lack of many-to-many alignment and is that it considers $P(a_j = 1)$ as uniform, when there should be positional information encoded into the model.

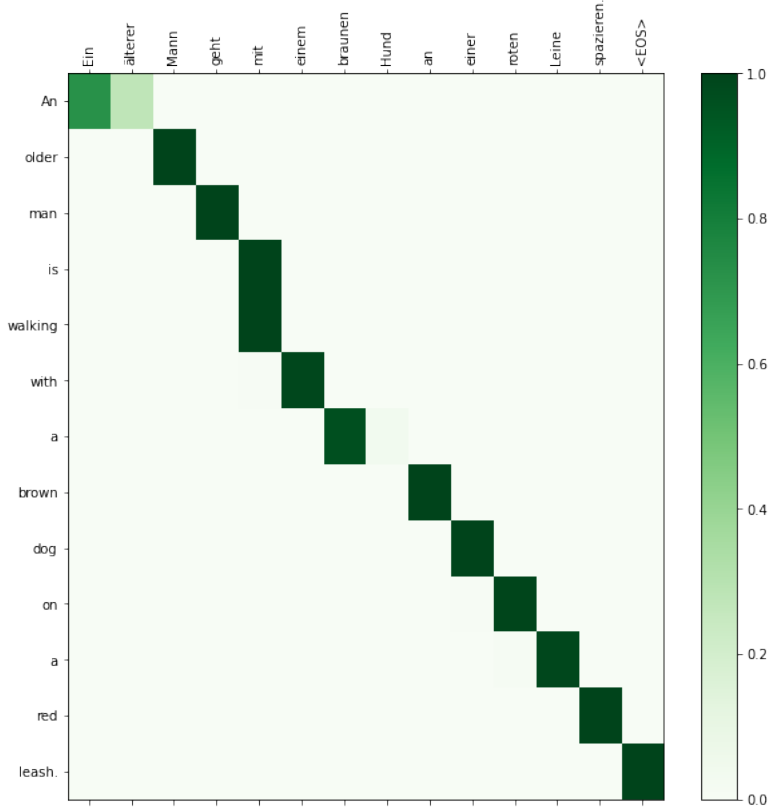


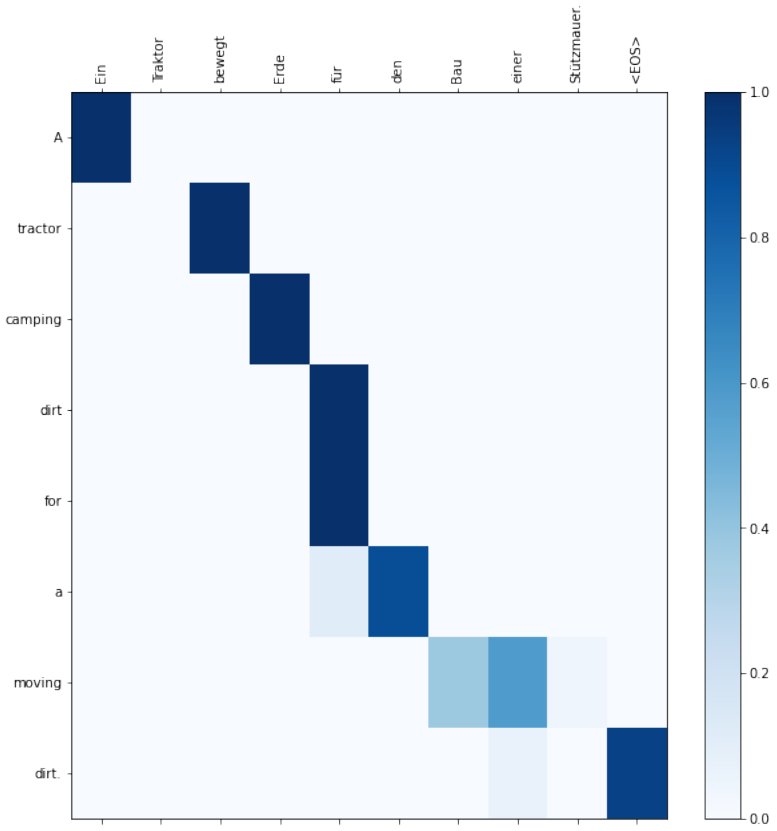
2. Following are the visualization plot for a random sequence of 4 sentences.



Lining up with the realizations of part 1's shortcomings, there seems to be a very high evidence of positional alignment as most of the weights are concentrated along the diagonal. There are also many to many alignments (as we have a softmax distribution), some of them are very obviously stronger than the others, most noticeable in the middle of the sentences. Many target positions, especially in a sequence of 2-3, seem to strongly align to the same source word, but the inverse is not as common, at least in this model. There is also a weak context in the second position for a few examples, although I am not sure why as it isn't common across a majority.


Drei
Männer
gehen
auf
einer
Straße
in
den
Bergen.
<EOS>
Three
men
are
walking
on
a
street
in
the
mountains.
0.0
0.2
0.4
0.6
0.8


Ein
älterer
Mann
geht
mit
einem
braunen
Hund
an
einer
roten
Leine
spazieren.
<EOS>
An
older
man
is
walking
with
a
brown
dog
on
a
red
leash.
0.0
0.2
0.4
0.6
0.8
1.0


Ein
Traktor
bewegt
Erde
für
den
Bau
einer
Stützmauer.
<EOS>
A
tractor
camping
dirt
for
a
moving
dirt.
1.0
0.8
0.6
0.4
0.2
0.0