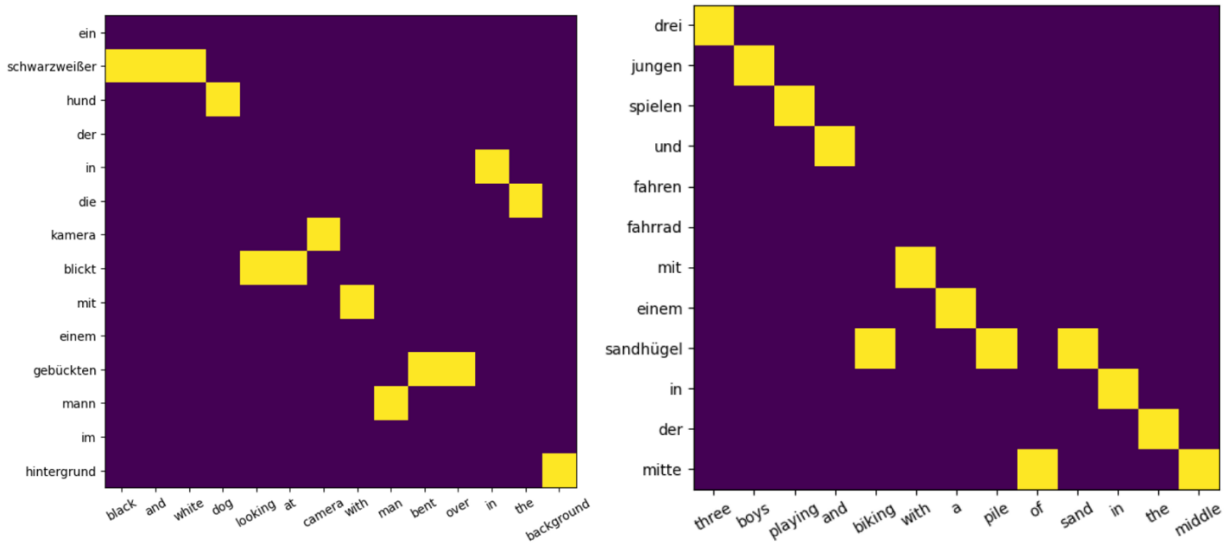
CS288
Sanghoon Lee (Department of Materials Science and Engineering)

1. HW2a



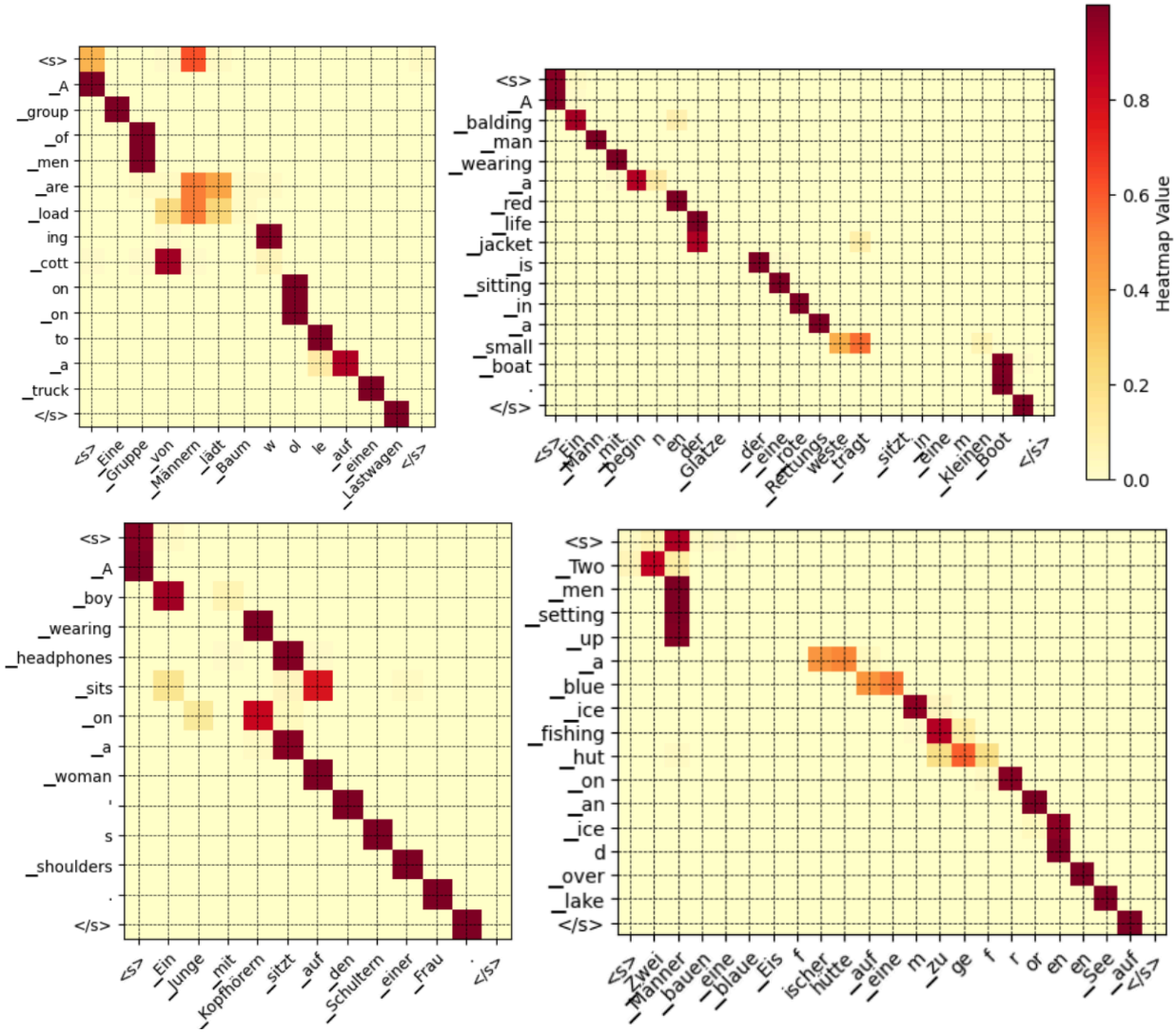
Left case is an example of a relatively well-performing case of IBM model. Other than that ‘im’ probably should be mapped to ‘im’, ‘in die’ (which does not exactly have a corresponding part in the target sentence other than ‘at’ which should be more like ‘into’) also works as an alignment. Because IBM model is dealing with an alignment pair, it would have decided that ‘in die’ is more closely alignable with ‘in the’ rather than ‘im’ which makes sense.

The case in the right is rather a failing case. This, if done right should be very sequential one-to-one map looking like an identity matrix. However, the model fails in the middle part. Especially, keywords like ‘fahrrad’ meaning ‘bike’ is skipped, while at least one of ‘fahren fahrrad’ should both light ‘biking’. Instead, ‘sandhugel’ is mapped to ‘biking’ and this is catastrophic considering that both ‘sandhugel’ and ‘biking’ are specific enough.

2. HW 2b

This attention map is not necessarily word alignment map. Attention weights are obtained by essentially the decoder output at timestep (query)’s dot product with the encoder outputs at each point (values). The heatmap of attention weights will show that which token of source sentence was similar to the vector that is a precursor of target token.

We can see clues for direct correlation. For instance, starting tokens have large correlation with another starting tokens. ‘Two’ having a large correlation with ‘Zwei’ in lower right heatmap. However, in the same example, ‘blue’ did not have large correlation with ‘blaue’.



We can also get interesting grammar-related aspects. For instance, when deciding ‘a’ in ‘a blue ice fishing hut’, the attention was pointing to ‘(f)isherhutte’ as a singular form. Although we would expect it to point to ‘eine’, it is interesting to note these points.