Example 1

BiLSTM translation

According to the translation, the couch is “black and black”. In the heatmap for this example, we can see that the model mistakenly associates “red” with “en”, instead of its translation, “rouge”. This might be because in French, many expressions that refer to colours start with “en”.

Transformer translation

The translation uses “qui est vêtus de rouge et noir”. In French, “est vêtus de” can only refer to how persons are dressed, with the “s” in “vêtus” being a marker of the plural. “est vêtus de rouge et noir” therefore means “are wearing red and black clothes”. This suggests that the model might think “red and black” is related to the boys instead of the couch.

The heatmap shows that attention is mainly given to “red and black” when generating “est vêtus de” (is wearing), which explains why the model is unable to associate the colours with the couch.

In both cases, the use of coordination at the end of the sentence does not appear to be the main cause of errors.

Example 2

In this example, both models successfully translated the full sentence, including the verb "sont assis" which correctly use the plural form.

BiLSTM assigns low attention values when the source token is "are", and high attention to both "sont" and "assis" for "sitting". This is interesting, as the word "sitting" alone does not provide information about whether the subject is singular or plural.

Similarly, the Transformer model assigns more attention to nouns than to articles, and more attention to the past participle of a verb than to the auxiliary.

It appears that models are better at handling coordinations when they are located at the beginning of a sentence.

Example 3

BiLSTM translation

In French, when both boys and girls are involved in an action, the masculine form prevails. The BiLSTM translation returns "Deux garçons et des filles sont assises sur un canapé rouge", with the verb "assises" in the feminine plural form. This shows that the model associated "are sitting" with "girls" instead of "two boys and girls".

This means that coordination at the beginning of the sentence is not always properly understood by the BiLSTM model. The attention heatmap shows that the model rarely

assigns attention to previous words in a sentence, which may explain this lack of grammatical coherence.

Transformer translation

The Transformer model correctly translated the sentence. In addition to other important words such as "assis" ("sitting") and "couch", some heads of the model appear to be specialised in processing the word "et" ("and"), which suggests that the impact of the initial coordination on the rest of the sentence is effectively taken into account by the model.

Example 4

BiLSTM translation

"A group of men" is correctly translated into "Un groupe d'hommes". However, when generating the ending for the verb "charge", the model assigns attention to the English ending "-ing" and the next word, and disregards the previous "is", which indicates that "a group of men" is the subject of a conjugated verb rather than a gerund.

This phenomenon results in the incorrect gerund form "chargeant" and disrupts with the rest of the translation, with "chargeant à la main d'un camion" roughly translating to "loading at the hand of a truck".

Transformer translation

The same mistake can be found in the Transformer translation. In all the heads of the model, little attention is paid to the word "is" in "is loading", even though it is crucial for understanding the structure of the sentence.

The model also translates "cotton" as "barbe" ("beard"). "Cotton candy" in French is "barbe à papa", so we can infer that "cotton candy" was likely part of the training data and misled the model in this context.

Example 5

BiLSTM translation

The model decided to use the verb "passer" (to go past) instead of "marcher" (to walk), which is acceptable. However, the heatmap shows that the model used the "-ed" ending from the "walked" to predict the following word "devant", but not when translating "walked", which caused it to omit the ending for past tense and produce "passent" instead of "passaient".

Transformer translation

The Transformer model also used the present tense instead of the past tense and computed "marchent" instead of "marchaient". Some heads assigned a small amount of

attention to the "-ed" ending when translating the verb, but it was apparently insufficient.