

Supplementary material for French

1 Attention

Figure 1 presents the encoder attention graphs from the main paper and the paper’s appendix for French, visualising all data (a, b, c), the subset of sentences where there is an identical match between a PIE’s typical surface form and the sentence (d, e, f) and the subset of sentences containing PIEs that are both in the *fig-par* and the *lit-wfw* part of the dataset (g, h, i).

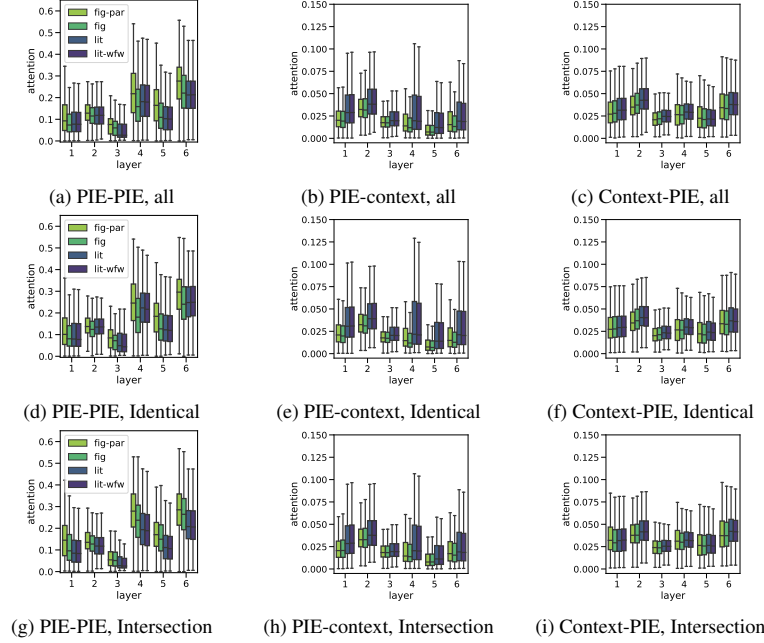


Figure 1: Encoder self-attention distributions, illustrating attention within the PIE and the interaction between the PIE and its context, for the identical and intersection data subsets.

Figure 2 shows the cross-attention graphs for French, visualising all data (a, b, c), the subset of identical matches (d, e, f) and the subset of sentences containing PIEs that are in *fig-par* and *lit-wfw* (g, h, i).

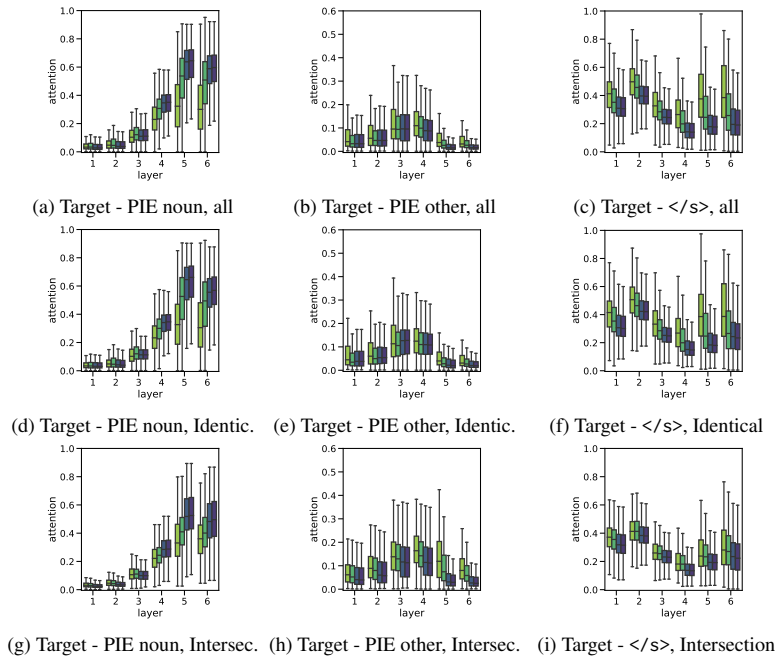


Figure 2: Cross-attention distributions from the translation of one PIE noun on the French target side to the indicated tokens on the English source side, for all data, and the identical and intersection data subsets.

2 Hidden states

Figure 3 displays the SVCCA similarities of PIE and non-PIE nouns when comparing representations from consecutive layers to each other. For non-PIE nouns, there are hardly any differences between subsets, but for the PIE nouns, the figurative cases (and the *fig-par* subcategory, in particular) stand out from the literal ones.

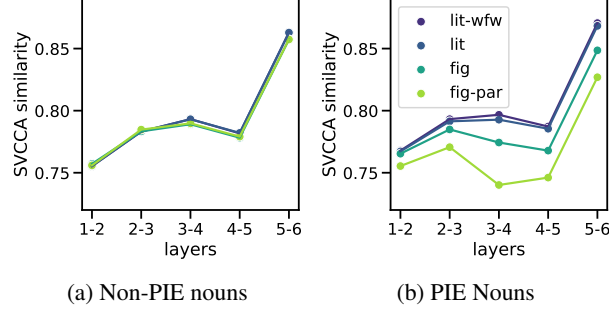


Figure 3: SVCCA similarity for layer l and layer $l + 1$, for PIE and non-PIE nouns within the English-French model. The languages comparison displays the difference in similarity between *lit-wfw* and *fig-par*.

Figures 4a and 4b indicate the influence of masking PIE nouns on other PIE tokens, and nouns in the context, respectively. Figures 4c and 4d show the impact of masking a context token on PIE tokens, and other context tokens, respectively.

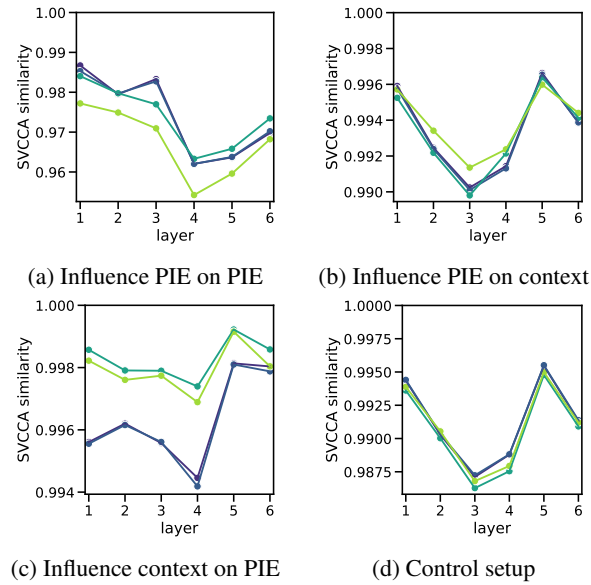


Figure 4: Impact of masking a PIE noun in the attention on (a) other PIE tokens, (b) other context tokens. Impact of masking a non-PIE noun on (c) PIE tokens and (d) other non-PIE token