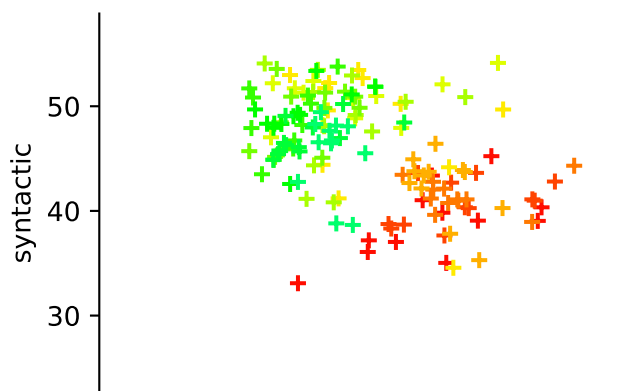
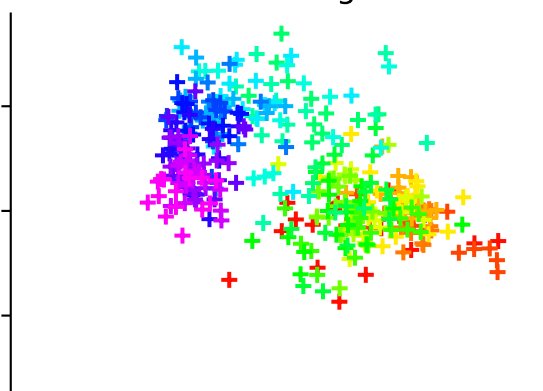


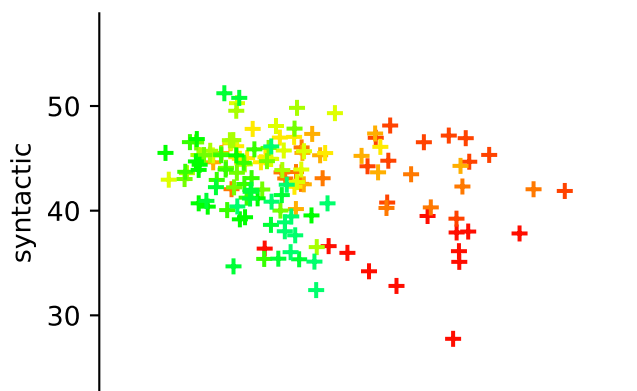
BERT-Base



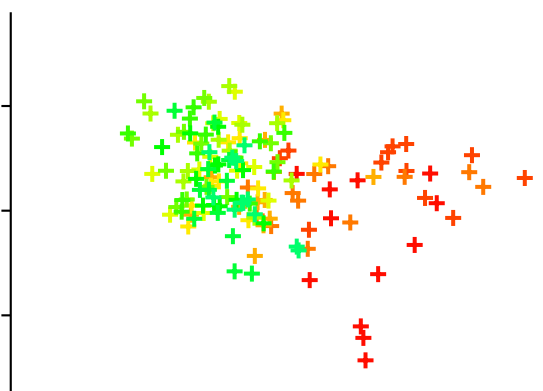
BERT-Large



GPT



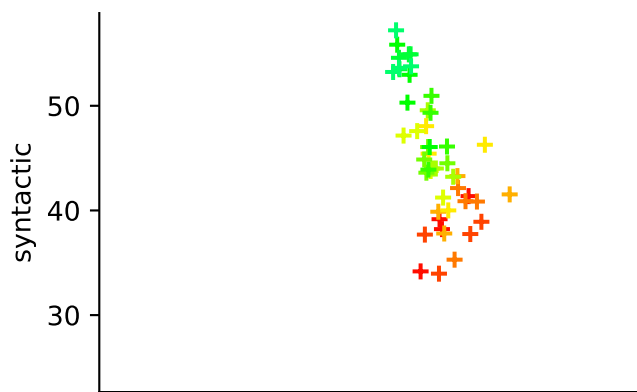
GPT2



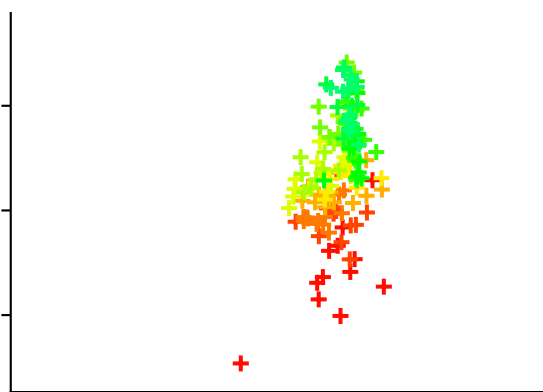
layer

- +
- +
- +
- +
- +
- +
- +
- +
- +
- +
- +
- +
- +
- +
- +
- +
- +
- +
- +
- +
- +
- +
- +
- +
- +

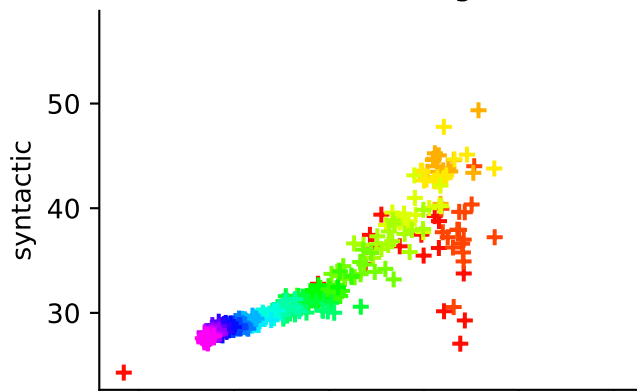
ELECTRA-small



ELECTRA-Base



ELECTRA-Large



surface