BPE

Answer the following questions:

1. How many “complete tokens” (according to the previous definition of token) end up in the BPE vocabulary for each of the languages?
2. What are some learned vocabulary items in each language that are larger than a single character, smaller than a complete token, and “make sense” linguistically? briefly discuss why you think they make sense.
3. What are some learned vocabulary items in each language that are larger than a single character, smaller than a complete token, and do not “make sense” linguistically? briefly discuss why you think they do not make sense.
4. Are there learned vocabulary items in either of the languages that “make sense” linguistically when used to split some words, but do not “make sense” linguistically when used to split some other words? show such examples if they exist and discuss briefly.