Proposed modifications (team D)

Problem 1: bad splitting

Translation: In this part of our website , you can find information on how the Parliament organises its work , through a system of specialised committees . The work of the European Parliament is important because in many policy areas , decisions on new European laws are made jointly by Parliament and the Council of Ministers , which represents Member States .

Hypothesis: This part of our website will find information on how Parliament will organise its work through a system of different committees .

Possible solution: changes in preprocessing

- not a problem of the model - initial segmentation is wrong
- try to split during preprocessing

Problem 2: wrong tense

Translation	Hypothesis
organises its work	will organise its work
influence	will have an impact
citizens can elect	citizens will be able to choose
helps ensure	will help to ensure

Possible solution: incorporate linguistic information

- words are annotated with morphological features, e.g. tense
- morphological tags are added to each word, so that the system sees both words and their tags
- tags are kept for each subword unit

Problem 3: parts of compound words are missing

Source	Translation	Hypothesis
tööprotsesse	working processes	processes
pädevusvaldkondade	competence areas	competences
tervishoiusüsteemi	health- care system	health system

Possible solution: split compound words

- examine words which are longer than some threshold (>10?)
- find frequent stems (>2?)
- split words into frequent components - will probably force the system to translate them separately

Problem 4: lack of special knowledge & terms

Translation	Hypothesis
For each particle , there is an antiparticle that	For each part , an antibody is
lead-ion beams	the iion speeds
Quarks are said to be confined .	They say they're in prison .

Possible solution: add more in-domain texts

our data comes from multiple domains -> we just need a bigger corpus

there is no bigger corpus -> ¯_(ツ)_/¯