

Milestone 2: Quality and Error Analysis

1. Translating In-domain development and test set:

BLEU scores for de-preprocessed in-domain dev and test set:

- In-domain dev set (accurate): 21.97
- Test set: 29.69

Score for dev set is much lower, it may be so, because there are only 1000 sentences compared to 50000 in test set. Also sentences for dev set might be hand picked.

2. Analyse the Translation

2.1 Summary of the Quality of Translations

- Some incorrectly translated words (should we point that out??) (e.g. sentence ID 779)
- Abbreviation wrongly translated (HRM - PH (ID:8), PRK - PIM (ID:14), EUPCU - EUPU (ID:3))
- Missing words (ID: 73)
- Machine translation missing context (e.g. sentence ID 199)
- Some sentences were translated perfectly, but word order was heavily changed and words were changed, but meaning was kept intact. That was very interesting, low score, but good translation. (e.g. ID 567)
- Translation missing big chunks, mostly if periods/dots "." in the sentence, System most likely thinks that End of Sequence is arrived, but actually it is not there. (e.g. sentence ID 909, 123, 369, 420)
- Numbers in front of sentence like "1. 2. 3." are missing. For example for task lists or numbered lists (ID: 490, 496). However for some sentences it was translated correctly.
- S vs Z - Organize vs Organise (ID: 985)
- In some translated sentences synonyms were used, which preserved the meaning but were nonetheless imperfect. (e.g. ID: 529)
- Incorrect verb tenses (e.g. ID: 184)
- Named entities translated (e.g. ID: 333)
- In some translated sentences, words that have a very specific meaning get translated to something else, which could lead to the meaning being lost (e.g. in ID: 810 "black mass" instead of "dark matter")

2.2 Analyses of Sentences

1. ID:8

- 1.1. Source: Praeguses olukorras on direktorid ning üksuste juhid projektidega sedavõrd hõivatud, et PH arendamiseks ei leita lihtsalt aega.
- 1.2. Human: At present situation the directors and heads of units have so heavy workloads with projects that they hardly have time to develop HRM.
- 1.3. Machine: In the current situation , directors and entities are so busy with projects , so that there is no time for the development of PH .
- 1.4. Assessment:
 - 1.4.1. heads - entities (human translation vs machine translation)
 - 1.4.2. HRM - PH
 - 1.4.3. At present situation - In the current situation (similar meaning)
 - 1.4.4. Otherwise fluent sentence and no mistakes.

2. ID: 40

- 2.1. 2) Energiakasutajad tagavad nende omanduses, valduses või kasutuses olevate olemasolevate võrkude ning energiakasutusseadmete ja – varustuse tehnilise teeninduse ja hoolduse.
- 2.2. (2) Energy users shall ensure the technical service and maintenance of the existing network , and energy utilisation devices and equipment , which are in the ownership , possession or use thereof .
- 2.3. (2) Energy users shall ensure the technical service and maintenance of existing networks and equipment in possession of existing networks and equipment and equipment .
- 2.4. Assessment:
 - 2.4.1. Last half of the sentence does not make much sense.
 - 2.4.2. First half of the sentence perfect

3. ID: 73

- 3.1. Lāti riigivõlg on Euroopa Liidus üks väiksemaid ning eelduste kohaselt jääb see keskmise tähtja jooksul märgatavalt alla riigi koguvõla kriteeriumi (60% SKT-st), nagu määratletud Maastrichti lepingus.
- 3.2. The level of the general government debt in Latvia is among the lowest in the European Union and is expected to remain considerably below the gross government debt volume criterion in the medium-term (60% of GDP) as defined in the Maastricht Treaty .
- 3.3. Latvia ' s public debt is one of the smallest , and is expected to remain significantly below the State gross debt criterion in the medium term (60% of GDP) , as defined in the Maastricht Treaty .
- 3.4. Assessment:
 - 3.4.1. The meaning of sentence is preserved and fluent
 - 3.4.2. Some words left out: European Union, general government
 - 3.4.3. Debt Volume criterion - debt criterion
 - 3.4.4. Government - State

4. ID:123

- 4.1. Meie veebisaidi sellest osast leiate teavet, kuidas korraldab parlament erinevate komisjonide süsteemi abil oma tööd. Euroopa Parlamendi töö on oluline seepärast, et otsuseid uute Euroopa seaduste kohta teevad paljudes poliitikavaldkondades ühiselt parlament ja ministrite nõukogu, mis esindab liikmesriike.
- 4.2. 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 .

4.3. This part of our website will find information on how Parliament will organise its work through the various committees .

4.4. Assessment:

4.4.1. First half of sentence is translated fine, however the other half is missing.

5. ID:125

5.1. Euroopa Parlament koosneb 736 liikmest, kes on valitud laienenud Euroopa Liidu 27 liikmesriigist.

5.2. The European Parliament is made up of 736 Members elected in the 27 Member States of the enlarged European Union .

5.3. The European Parliament consists of 736 members elected from 27 Member States of the enlarged European Union .

5.4. Assessment:

5.4.1. Made up - consists

5.4.2. In the - from

6. ID:156

6.1. Lisaks sellele kutsutakse komisjoni andma iga aasta algul teistele EL institutsioonidele tagasisidet olulisematest saavutustest eelmise aasta ühiste kommunikatsiooniprioriteetide ellurakendamisel.

6.2. Finally , the Commission is invited to report back at the beginning of each year to the other EU Institutions on the main achievements of the implementation of the common communication priorities of the previous year .

6.3. In addition , the Commission invites the other EU institutions at the beginning of each year to feedback from the main achievements in implementing the common communication priorities of the previous year .

6.4. Assessment:

6.4.1. Meaning is more or less preserved

6.4.2. Phrase order different (the other EU institutes)

6.4.3. Report back - feedback from

7. ID:256

7.1. Elektroonilise arveldussüsteemi kasutamisel tuleb see number iga kord isikutuvastuseks sisestada.

7.2. You will have to use this combination of digits every time to authenticate yourself in the electronic settlement system .

7.3. In the use of the electronic settlement system , this number shall be entered in each individual case .

7.4. Assessment:

7.4.1. Important word "Authentication" missing

7.4.2. Otherwise well translated (however rephrased)

8. ID:265

8.1. Minge veebilehele www.itunes.com / Download ning laadige oma Maci või PC jaoks alla uusim iTunes'i versioon.

8.2. Go to [www . itunes . com](http://www.itunes.com) / download and download the latest version of iTunes to install on your Mac or PC .

8.3. Go on website [www . itunes . com](http://www.itunes.com) / download and download the newest iTunes version of your Maci or PC .

- 8.3.1. Maci - Mac
- 8.3.2. On your Mac - of your Mac (changes meaning)
- 8.3.3. Fluent

9. ID:320

- 9.1. Telefonileidja (find My iPhone) funktsioon aitab leida kadunud või unustatud iPhone'i ning kuvab iPhone'i ekraanile teate, et leidja saaks selle teile tagastada.
- 9.2. The Find My iPhone feature helps you locate iPhone if it's been lost or misplaced , and display a message on your iPhone to help someone return it to you .
- 9.3. The fundraiser (Find My iPhone) helps to find a missing or forgotten iPhone and display an iPhone screen on the screen to return it to you .
 - 9.3.1. Okay translation, however messed up the part about displaying a message on your iPhone
 - 9.3.2. fundraiser - telephone finder

10. ID:369

- 10.1. Niisiis, kui Richard oli seitsmeaastaseks saanud, mölliti ta Colstoni Poisslastekooli ning sai selga kuulsa sinise mantli, mis kuulutas bristollastele et ta isa on vaene ent austusväärne, kaljukindel anglikaan.
- 10.2. So when Richard turned seven he was enrolled at Colston's School for Boys and donned the famous blue coat which informed Bristolians that his father was poor but respectable , staunchly Church of England .
- 10.3. So if Richard was seven years old , he was the one with Colston Poisson School and became a famous blue coat that proclaimed the bristolans that his father was poor , but his father is honorable .
 - 10.3.1. Not very fluent sentence
 - 10.3.2. Missing the end (staunchly Church of England)
 - 10.3.3. Poisson School (poisslaste - boys) - school of boys
 - 10.3.4. Enrolled - was the one with

11. ID:420

- 11.1. Enne kui hakkame tulevikku avastama, vaatame, kuidas me jõudsimme tänapäeva. alustame sellega, et läheme tagasi aastasse 1800, kus info leidmine toimus hoopis teisiti.
- 11.2. Before we start to explore the future , let's see how we reached the present . we'll begin by going back to 1800 — when finding out information was very different .
- 11.3. Before we start exploring the future , we'll see how we made it to the day we go back to 1800 .
- 11.4. Assessment
 - 11.4.1. First half of the sentence okay and fluent
 - 11.4.2. Last phrase is missing ("when finding out information was very different")

12. ID: 490

- 12.1. 1. info saamine kaugetest kohtadest.
- 12.2. 1 . getting information from distant places .
- 12.3. Getting out of the remote places .
 - 12.3.1. Information - out
 - 12.3.2. Idea of sentence missing
 - 12.3.3. Number "1." missing

13. ID: 496

- 13.1. 7. Infoallikaks olemine.
- 13.2. 7 . being the source of information .

- 13.3. Being a source of information .
 - 13.3.1. Missing number "7."
- 14. ID: 502**
 - 14.1. Need edastatakse teile otse allikast, jättes vahele traditsioonilised meediakanalid.
 - 14.2. It is being reported directly from the source to you — bypassing the traditional media channels.
 - 14.3. They will be transmitted directly from the source, leaving traditional media.
 - 14.3.1. Leaving - bypassing
 - 14.3.2. Otherwise fluent and correct sentence
- 15. ID: 527**
 - 15.1. Aastal 2010 hakkab esile kerkima kaks uut kontsepti.
 - 15.2. In 2010, two new concepts will start to emerge.
 - 15.3. Two new concepts are emerging in 2010.
 - 15.3.1. Rephrasing
 - 15.3.2. Otherwise fluent and correct
- 16. ID: 733**
 - 16.1. Ta on sarnane nendele osakestele kui Higgsi bosonid üritavad nendega vastastikku mõju avaldada - nad koguvad massi.
 - 16.2. He is similar to these particles when Higgs bosons try to interact with them – they gain mass.
 - 16.3. He's similar to those particles, the bosses of Higgs trying to interact with them - they collect mass.
 - 16.3.1. Higgs bosons - bosses of Higgs
 - 16.3.2. Otherwise correct
- 17. ID: 881**
 - 17.1. Võimaldades rakenduskihi sujuv integreerimine infrastruktuurikihiga, saab IBM aidata teil optimeerida teie IT-keskkonda, teenida IT-investeeringutelt suuremat tulu ja tootma IT abil paremat äriväärtust ja innovatsiooni.
 - 17.2. By enabling seamless integration of the application layer with the infrastructure layer, IBM can help you optimize your IT environment, generate higher returns on IT investments, and yield greater business value and innovation from IT.
 - 17.3. Allowing the smooth integration of the application layer with the infrastructure layer, IBM can help you optimise your IT environment, earn more revenues from IT investment and generate better business value and innovation through IT.
 - 17.3.1. Fluent and correct sentence
 - 17.3.2. In some places different words used.
- 18. ID: 925**
 - 18.1. Tippige oma vastus uue kommentaari jutumulli.
 - 18.2. Type your response in the new comment balloon.
 - 18.3. Get your answer to the new comment of the chatter.
 - 18.3.1. Meaning of sentence lost
 - 18.3.2. Type your response - get your answer
 - 18.3.3. Comment balloon - comment of the chatter
- 19. ID: 985**
 - 19.1. Rakendades ennetavalt hästiorganiseeritud lähenemise teie talletusmeedia- ja andmekeskonnale, saate kasu eelistest, mis tulenevad talletamise efektiivsest infrastruktuurist.

- 19.2. By proactively establishing a well-organized approach to your storage and data environment, you can gain the advantages that come from an effective storage infrastructure.
- 19.3. Implementing a well-organised approach to your storage media and data environment will benefit from the benefits resulting from the effective infrastructure for storage.
 - 19.3.1. Organized - organised
 - 19.3.2. Gain the advantages Benefit from the benefits
 - 19.3.3. By proactively missing
 - 19.3.4. Otherwise fluent sentence

20. ID: 996

- 20.1. IBM-i talletuse ja andmete hallatud teenused - aidates teil teostada ekspertiise, mõõtmisi ja eelintegreerimist - standardised võimekused maailmatasemel teenusepakkujalt, kes saab hallata talletuskeskkonda teie nimel
- 20.2. IBM Storage and Data Managed Services - helping you leverage the expertise, scale and pre-integrated, standard capabilities of a world-class service provider that can manage the storage environment on your behalf
- 20.3. Services managed by IBM and managed services - helping you to carry out expert, measurements and pre-integration - standard capabilities from a world-class service provider who can manage the storage environment
 - 20.3.1. End is missing ("on your behalf")
 - 20.3.2. Starting is messed up.

Andre

21. ID: 1

- 21.1. Source: Käesoleva aruande eesmärk on kirjeldada samme, kuidas tuleks üles ehitada personalihalduse funktsioon – selgitada PH strateegiat ja toiminguid.
- 21.2. Human: The purpose of this report is to describe the steps how Human Resource Management function should be built up – to clarify the strategy and actions of HRM-plan .
- 21.3. Machine: The purpose of this report is to describe how the function of staff management should be established – to clarify the PH strategy and operations .
 - 21.3.1. Almost perfect translation, perfect fluency.
 - 21.3.2. One small mistake is that "PH strateegia" is translated into "PH strategy", but it is "HRM strategy" or even "S strategy", S for staff.
 - 21.3.3. The system does not know how to translate abbreviation.
 - 21.3.4. BLEU:

22. ID: 2

- 22.1. Source: Investeeringisosakonna strateegia ja lühiajaline tegevusplaan on soovitatav kord aastas läbi vaadata.
- 22.2. Human: It is recommendable to review the strategy and short term action plan of the Investment department yearly .
- 22.3. Machine: The Investment Department strategy and the short-term action plan are recommended annually .
 - 22.3.1. Machine is missing one important word - verb, like "to be overlooked"/"to be reviewed".
 - 22.3.2. Fluency is good, but the translation is not very adequate.

23. ID: 3

- 23.1. Source: Käesoleva aruande idee kujunes 22. septembri arutelus EUPCU juhi Dace Rozega.
- 23.2. Human: The aim of this report was formulated with Head of EUPCU Dace Roze in discussion on the 22nd of September .
- 23.3. Machine: The idea of this report was in the debate on the Dace Roz leader of the EUPU .
 - 23.3.1. We can see two obvious mistakes here - "Dace Rozega" -> "Dace Roz" instead of the correct "Dace Roze".
 - 23.3.2. "EUPCU" is translated into "EUPU".
 - 23.3.3. Translation lost the date and month part, that is bad.
 - 23.3.4. Bad verb in the machine version "was in the debate" vs "was formulated".
 - 23.3.5. Fluency and adequacy problems.

24. ID: 14

- 24.1. Source: Peagi on valmimas PRK (projekti rakendamise käsiraamat) ja PKJ (projekti korraldamise juhend).
- 24.2. Human: Pim (Project implementation Manual) and POG (Project Organization Guide) are under finalising .
- 24.3. Machine: Soon , the PRK (the Handbook of implementation of the project) and the PKJ (project guide) will be prepared .
 - 24.3.1. Perfect fluency. slight adequacy problems, with abbreviations and translation of terms, e.g. "PIM (Project Implementation Manual)" vs "RPK (the Handbook of implementation of the project).
 - 24.3.2. But probably quite impossible to teach this, as we probably would raise more errors if we trained system to learn those "PRK" -> "PIM" abbreviations.

25. ID: 53

- 25.1. Source: 4) Süsteemi käitajatel ei ole õigust avaldada äriteavet, mis on neile saanud teatavaks nende kohustuste täitmisel. Energiavarustusega kauplejad määratlevad äri- ja konfidentsiaalse teabe sisu, koordineerides seda reguleerijaga.
- 25.2. Human: (4) System operators are not entitled to disclose commercial information , which has become known to them in fulfilling their duties . energy supply merchants shall determine the contents of commercial and confidential information by co-ordination thereof with the regulator .
- 25.3. Machine: (4) Operators of the system shall not have the right to disclose commercial information to them in the performance of their duties .
 - 25.3.1. Clearly the second part of the input is missing, probably because our system does not know how to translate sentences with dots in the end (end of sentence). We should teach the system to handle such cases.
 - 25.3.2. Even the first part of the sentence is wrongly translated, first part is ok, but ending has bad fluency, also adequacy problems, not clear what is meant in the sentence.

26. ID: 104

- 26.1. Source: 7. septembri 2005. aasta seisuga ei olnud Läti teatanud 1,3% ELi direktiivide ülevõtmisest (või 0,5% direktiivide koguarvust).
- 26.2. Human: As of September 7 , 2005 , Latvia had not announced transposition of 1.3% of EU directives (or 0.5% of the total number of directives) .

- 26.3. Machine: As at 7 September 2005 Latvia had not notified 1.3% of the transposition of EU directives (or 0.5% of the total number of directives) .
- 26.3.1. Perfect translation, only smallest mistake with changing of with at, so minor fluency flaw, perfect adequacy.

27. ID: 106

- 27.1. Source: Siseturu direktiivide ülevõtmine ja asjakohaste õigusaktide rakendamine sõltub otseselt ELi õigusaktide ülevõtmise ühise jälgimissüsteemi rajamisest.
- 27.2. Human: Transposition of Internal Market directives and implementation of the relevant legal acts directly depends on establishing a common system for supervising the transposition of EU legal acts .
- 27.3. Machine: Transposition and implementation of internal market directives will depend directly on the establishment of a common monitoring system for transposition of EU legislation .
- 27.3.1. Fluency is perfect in my opinion.
- 27.3.2. Adequacy is also good.
- 27.3.3. Thing is that Machine shortened the translation without losing meaning. But that is my opinion.

28. ID: 109

- 28.1. Source: Eli siseturu vabaduste piirangud olid enamikus kaotatud enne Läti ühinemist ELiga, kuid piire ületavate tehingute puhul on takistused senini olemas.
- 28.2. Human: Restrictions to the EU Internal Market freedoms were largely eliminated before Latvia's accession to the EU , however , obstacles to cross-border transactions continue to exist .
- 28.3. Machine: The restrictions on the freedoms of the EU internal market were most lost before Latvia's accession to the EU , but the barriers to transactions above borders are still there .
- 28.3.1. Fluency and adequacy problems with sentence, "were most lost before" vs "largely eliminated" and "above borders" vs "cross-borders". Other mistakes are not so bad.

29. ID: 199

- 29.1. Source: Keskkonna hea tervis on eurooplaste ja nende riikide valitsuste jaoks oluline teema.
- 29.2. Human: A healthy environment is a big issue for Europeans and their governments .
- 29.3. Machine: Good health is an important issue for Europeans and governments of these countries .
- 29.3.1. Perfect fluency, but mayor adequacy problem, sentence is about environment, but Machine implies that the sentence is about humans "good health" - machine missing context.

30. ID: 200

- 30.1. Source: Seepärast on EL ülemaailmsetes keskkonna kaitsmise ja kliimamuutustega võitlemise jõupingutustes esirinnas.
- 30.2. Human: This is why the EU is spearheading world efforts to protect the environment and fight climate change .
- 30.3. Machine: The EU is therefore at the forefront of the efforts to protect and combat climate change .
- 30.3.1. Perfect translation.

31. ID: 201

- 31.1. Source: Kuna reostus ei tunne piire, on EL riigid sel alal juba aastaid ühiselt tegutsenud.
- 31.2. Human: As pollution knows no borders , EU member states have taken joint action in many areas .
- 31.3. Machine: Since pollution does not feel boundaries , the EU has been working together for years in this area .
 - 31.3.1. Fluency problems, like "pollution does not feel boundaries", "EL riigid" -> "EU" is also not good, sentence is fluent in the end.
 - 31.3.2. Adequacy is a bit bad, but still it is an okay translation, because the translation does keep the meaning.
 - 31.3.3. The second half of the sentence is quite correct

32. ID: 378

- 32.1. Source: Pealegi, sõda oli elu osa ja sõjaks läks tarvis relvi.
- 32.2. Human: Besides which , war was a part of life , and guns a part of war .
- 32.3. Machine: Besides , the war was part of life and the war needed guns .
 - 32.3.1. Minor fluency problems, articles missing, other than that perfect translation.

33. ID: 379

- 33.1. Source: Meest, kes relvi valmistada ja parandada mõistis, ei ähvardanud oht lahinguväljale kahurilihas sattuda.
- 33.2. Human: A man who could make and mend them was unlikely to become cannon fodder on a battlefield .
- 33.3. Machine: The man who made weapons and repaired the weapons didn't threaten a threat to the battlefield .
 - 33.3.1. Fluency problems with phrase "didn't threaten a threat".
 - 33.3.2. Adequacy problems in the end, as "didn't threaten a threat to the battlefield ." doesn't make much sense. Overall not a very good translation.

34. ID: 567

- 34.1. Source: Kümned tuhanded inimesed asustati ümber, paljud nendest leiavad peavarju ikka veel ajutistes põgenikekeskustes.
- 34.2. Human: Tens of thousands of people were uprooted , many of whom are still sheltering in makeshift refugee centres .
- 34.3. Machine: Tens of thousands of people were resettled , many of them still find shelter in temporary refugee centres .
 - 34.3.1. Perfect translation, words are different, but meaning is kept and fluency is great.

35. ID: 568

- 35.1. Source: Aga pikaajalisem eesmärk on taastada Georgia majandus, mis on viimastel aastatel läbi teinud suuri muutusi.
- 35.2. Human: But the long-term objective is to revive Georgia's economy , which has undergone considerable reform in recent years .
- 35.3. Machine: But the longer-term objective is to restore the Georgia economy , which has been undergoing major changes in recent years .
 - 35.3.1. Slight fluency problem "revive Georgia's economy" vs "restore the Georgia economy". Adequacy is good. Overall good translation.

36. ID: 777

- 36.1. Source: Iga anduri sees on ka magnetväli.

- 36.2. Human: There is also a magnetic field inside each detector .
- 36.3. Machine: There's a magnetic field inside every sensor .
- 36.3.1. Perfect translation.
- 37. ID: 778**
- 37.1. Source: See aitab avastada ja uurida laetud osakeste (nt elektron või prooton) trajektoore, mis on magnetvälja suhtes tundlikud.
- 37.2. Human: This helps detect and study the trajectories of the charged particles , such as the electron or the proton , which are sensitive to the magnetic field .
- 37.3. Machine: This will help detect and investigate the trajectory of particle particles (e . g . electric or proton) , which are sensitive to magnetic field .
- 37.3.1. Fluency issue with "particle particles", same thing ruins the adequacy.
- 37.3.2. Other parts of the translation are completely fine.
- 38. ID: 779**
- 38.1. Source: Teiseks seadmeks on kalorimeeter-peilimise rajaandur.
- 38.2. Human: Another device is a calorimeter-bearing track detector .
- 38.3. Machine: The second device is a calorific device .
- 38.3.1. Fluency is good, but adequacy is not.
- 38.3.2. "calorific" does not mean what is meant in source sentence.
- 39. ID: 823**
- 39.1. Source: Aga kuidas määratleda oma äri paindlikkust ja taastumisenõudmisi?
- 39.2. Human: But how do you determine the resiliency and recovery requirements of your business ?
- 39.3. Machine: What about the flexibility and recovery demands of your business ?
- 39.3.1. Perfect translation, I actually like the machine translation even better than the human one.
- 40. ID: 825**
- 40.1. Source: Kust alustada
- 40.2. Human: Where do you start ?
- 40.3. Machine: Where to start
- 40.3.1. Perfect translation.
- 41. ID: 900**
- 41.1. Source: Lihtsustab integreerimist IBM-i elutsükli metoodikate ja tööriista asjatundlikkuse kaudu
- 41.2. Human: Simplify integration through IBM lifecycle methodologies and tool expertise
- 41.3. Machine: Simplifies integration through the quality of life cycle methodologies and tools of IBM
- 41.3.1. Almost perfect translation. Interesting how the sentence is quite low BLEU score, but translation itself is good.
- 41.3.2. The translation generalizes well, so it gives meaning with different words and can change word order heavily.
- 41.3.3. The worlds are also aligned from middle to end well (Word IBM), its quite nice.
- 41.3.4. Only thing is that the verb was conjugated wrongly.
- 42. ID: 909**
- 42.1. Source: Saate sisestada kommentaare (kommentaar: märkus või kokkuvõte, mille autor või läbivaataja dokumendile lisab. Microsoft Word kuvab kommentaarid dokumendi veerisel jutumullis või läbivaatuspaanis.) jutumulli sisse

(jutumullid: küljendivaates või veebivaates kuvatakse märgistumullis märgistuselemendid (nt kommentaarid ja jälitatud muudatused) dokumendi veeristel. kasutage neid jutumulle, et lihtsalt vaadata ja vastata läbivaatajate muudatustele ja kommentaaridele.), mis ilmuvad dokumendi veeristel.

- 42.2. Human: You can insert a comment (comment : a note or annotation that an author or reviewer adds to a document . Microsoft Word displays the comment in a balloon in the margin of the document or in the reviewing Pane .) inside balloons (balloons : in print layout view or Web layout view , markup balloons show markup elements , such as comments and tracked changes , in the margins of your document . use these balloons to easily see and respond to reviewers' changes and comments .) that appear in the document margins .
- 42.3. Machine: You can enter comments (comment : note or summary of which the author of the document is annexed to the document , Microsoft Word kuvab) , in the column of the document .)
- 42.3.1. Only translated first part of the sentence, because of the dot in the source sentence.
- 42.3.2. It did try to translate more after the dot, but it got weird quite quickly, "kuvab" translated to "kuvab".
- 42.3.3. 2 closing brackets in the end.
- 42.3.4. Not the best translation, but if "kuvab" was translated correctly, first part of translation would probably make sense.
- 42.3.5. Mayor Fluency and adequacy problems.

Ian

43. ID: 5

- 43.1. Source: Investeerimisosakonna strateegia ja lühiajaline tegevusplaan on soovitatav kord aastas läbi vaadata.
- 43.2. Reference: It is recommendable to review the strategy and short term action plan of the Investment department yearly .
- 43.3. Machine: The Investment Department strategy and the short-term action plan are recommended annually .
- 43.3.1. The structure of the sentence is slightly messed up.
- 43.3.2. The translation is still fluent
- 43.3.3. The verb got lost, which leads to the meaning not being conveyed.

44. ID: 42

- 44.1. Source: 3) Süsteemi käitajatel ei ole õigust teostada tegevusi, mis ei ole otseselt seotud nende kohustustega ja mis võivad takistada või ohustada nende kohustuste täitmist.
- 44.2. Reference: 3) System operators are not entitled to perform such activities that are not directly related to their duties and may hinder or endanger the fulfilment of such duties .
- 44.3. Machine: (3) Operators shall not have the right to conduct activities which are not directly related to their obligations and which may hinder or jeopardise the performance of their duties .
- 44.3.1. The translation preserves meaning.
- 44.3.2. The translation is fluent.
- 44.3.3. The only bad part is "shall not have the right"
- 44.3.3.1. Instead of being in the present, it is kind of ambiguous whether this shall only be the case in the future.

45. ID: 117

- 45.1. Source: Praeguseni pole ükski vabatarbija elektritarnijat vahetanud.
- 45.2. Reference: None of the eligible customers has changed suppliers of electricity up to now .
- 45.3. Machine: Until now , no electricity supplier has changed the electricity supplier .
 - 45.3.1. The meaning of the sentence is not conveyed.
 - 45.3.2. The structure of the sentence is not correct.
 - 45.3.3. "eligible customers" was exchanged with "electricity supplier" which makes the sentence not understandable.

46. ID: 171

- 46.1. Source: Euroopa Parlamendi mõju EL asjaajamises suureneb ning 1979. aastal saavad kodanikud esmakordselt parlamendi liikmeid otse valida.
- 46.2. Reference: The European Parliament increases its influence in EU affairs and in 1979 all citizens can , for the first time , elect their members directly .
- 46.3. Machine: The European Parliament's influence on EU business is increasing and the citizens will be able to choose the Members for the first time in 1979 .
 - 46.3.1. The translation is fluent and the meaning is also more or less preserved.
 - 46.3.2. However, the tense of the verb is incorrect.

47. ID: 184

- 47.1. Source: Sidepidamine muutub hõlpsamaks ning üha enam inimesi asub kasutama mobiiltelefoni ja Internetti.
- 47.2. Reference: Communication is made easier as more and more people start using mobile phones and the internet .
- 47.3. Machine: Communication will become easier and more people are being used to use mobile phones and Internet .
 - 47.3.1. The overall structure of the sentence is intact, however the meaning of the sentence is not conveyed.
 - 47.3.2. The tense of the verbs is not correct.
 - 47.3.3. Instead of being the active, "people" becomes passive part of the sentence.

48. ID: 292

- 48.1. Source: 3 Sisestage salasõnad ja muu nõutav info.
- 48.2. Reference: 3 Enter passwords and other information as requested .
- 48.3. Machine: 3 enter password and other information required .
 - 48.3.1. The meaning of the sentence is conveyed correctly.
 - 48.3.2. "As requested" was substituted by "required" which can in some contexts be synonymous.

49. ID: 333

- 49.1. Aeg / ELECTROLUXi masinad vastavad üldistele ohutusstandarditele ning masinate ohutust käsitlevatele õigusaktidele.
- 49.2. Reference: The safety of AEG / ELECTROLUX appliances complies with the industry standards and with legal requirements on the safety of appliances .
- 49.3. Machine: Time / Electrolux machines meet general safety standards and safety legislation on machinery .
 - 49.3.1. The sentence is fluent and the meaning was also preserved.
 - 49.3.2. "AEG" was translated to "Time", which is wrong.

50. ID: 452

- 50.1. Source: Siiski nõustusid kõik, et see oli tulevik - uue ajastu algus.

- 50.2. Reference: However , everyone agreed that it was the future ; the dawn of a new era .
- 50.3. Machine: However , all that it was the future - the start of a new era .
- 50.3.1. The second part of the sentence was translated correctly, only with “start” substituted for “dawn”.
- 50.3.2. In the first part of the sentence the verb is missing, and instead of “everyone” the MT system chose “all”, which could be the reason for the missing verb. The meaning of the sentence is not conveyed correctly

51. ID: 465

- 51.1. Source: Sellel oli hävitav mõju info traditsioonilisele kujule.
- 51.2. Reference: This had a devastating effect on the traditional forms of information .
- 51.3. Machine: It had a disastrous impact on the traditional image of the information .
- 51.3.1. The structure of the sentence is good and it is fluent.
- 51.3.2. However the pronouns were translated wrongly.
- 51.3.3. “Image” was chosen instead of “forms”, which leads to the meaning of the sentence getting lost.

52. ID: 501

- 52.1. Source: Uudiseid ei edasta enam ajakirjanikud, kuna nüüd tuleb seda kõigilt.
- 52.2. Reference: News is no longer being reported by journalists , as now it comes from anyone and everyone .
- 52.3. Machine: The news will no longer be transmitted by journalists , because now it comes from everyone .
- 52.3.1. The sentence is fluent and meaning is preserved.
- 52.3.2. The tense of the verb is wrong.

53. ID: 525

- 53.1. Source: Kõik sisaldab mingil kujul sihtrühma.
- 53.2. Reference: Everything will incorporate some form of targeting .
- 53.3. Machine: All form the target group .
- 53.3.1. The meaning of the sentence was completely messed up.
- 53.3.2. The word “form” was translated, but as a verb, which is not the case in the reference .

54. ID: 529

- 54.1. Source: Samast teemast kirjutava viie erineva artikli lugemise asemel esitatakse teile üks, tõstes esile olulised huvipunktid.
- 54.2. Reference: Instead of reading five different articles on the same topic , you will be presented with one , highlighting the vital points of interest .
- 54.3. Machine: Instead of reading the five different articles of the same subject , you will be given one , highlighting important interest points .
- 54.3.1. The machine is fluent and conveys the meaning correctly.
- 54.3.2. Instead of “topic”, the MT chose “subject” and instead of “vital” it chose “important” which are only minor differences.

55. ID: 547

- 55.1. Source: Merepinna taseme tõus, kõrbestumine ja sagenenud ning tõsisemad loodusõnnetused võivad viia miljoneid vaesusesse ja olematuks teha aastatuhande jooksul saavutatud arengueesmärgid.
- 55.2. Reference: Rising sea levels , desertification and more frequent and severe natural disasters could push millions into poverty and undo past progress on the millennium development goals .

- 55.3. Machine: Rising levels of sea level , desertification and increased frequency and more serious natural disasters can lead to millions of poverty and unprecedented development goals .
- 55.3.1. The structure of the sentence is messed up.
- 55.3.2. Meaning is not conveyed, the translation is very bad
- 56. ID: 657**
- 56.1. Source: Siiralt, Peter Morgan
- 56.2. Reference: Sincerely , Peter Morgan
- 56.3. Machine: Sincerely , Peter Morgan
- 56.3.1. The MT is exactly the reference, but the sentence is also very short and two of the three words are named entities.
- 57. ID: 722**
- 57.1. Source: Kui mõni isik on lõpus kadunud, siis selle tõttu, et sa tead, kes sisenesisid, saab arvata, kes ei väljunud. ”
- 57.2. Reference: If a person is missing at the end , then because you know who entered you can make a guess at who did not exit . ”
- 57.3. Machine: If any person is missing at the end , because you know who entered , guess who did not . ”
- 57.3.1. The first part of the sentence was translated correctly.
- 57.3.2. In the second part of the sentence some words were not translated.
- 58. ID: 763**
- 58.1. Source : Järelikult peaksid kõik osakesed üksteist hävitama ja universum peaks olema tühi.
- 58.2. Reference : Consequently , all particles should have cancelled each other out and the universe should be empty .
- 58.3. Machine : Consequently , all particles should destroy each other and the universe should be empty .
- 58.3.1. The structure of the sentence is intact, it is fluent and the meaning of the sentence was conveyed correctly.
- 58.3.2. The tense of the verb is incorrect, as the MT chose “destroy” instead of “have cancelled out”.
- 59. ID: 810**
- 59.1. Source: Nad seetõttu postuleerivad nähtamatu musta massi olemasolu, mis on vastastikmõjus muu mateeriaga ainult selle gravitatsiooni kaudu, ja sisaldab väga stabiilseid, seega ka väga kergeid osakesi.
- 59.2. Reference: They therefore postulate the existence of an invisible dark matter that only interacts with the rest of matter by its gravitation , and consists of very stable , and therefore very light , particles .
- 59.3. Machine: They therefore possess the presence of invisible black mass , which is interchangeable with other matter only through this gravity , and contains very stable , so also very mild particles .
- 59.3.1. The fluency of the sentence is largely intact with some of the meaning being lost.
- 59.3.2. “dark matter” gets lost is the , being translated to “black mass”, since it has a unique meaning, the meaning of the sentence gets lost through this.
- 60. ID: 845**
- 60.1. Source: Ühtlusta IT-algatused ärieesmärkidega
- 60.2. Reference: Align IT initiatives with business objectives

- 60.3. Machine: Harmonise IT initiatives with business objectives
 60.3.1. Perfect translation
 60.3.2. "Align" is in this context interchangeable with "Harmonise"

61. ID: 932

- 61.1. Source: 3. loendis jaotise Käskude valimiskoht, valige käsk Kõik käsud.
 61.2. Reference: 3 . in the list under Choose commands from , select All Commands .
 61.3. Machine: In the list of 3 .
 61.3.1. The structure of the sentence is completely messed up.
 61.3.2. The machine translated sentence ends after the dot of point 3, which could be the reason for the bad .

62. ID: 967

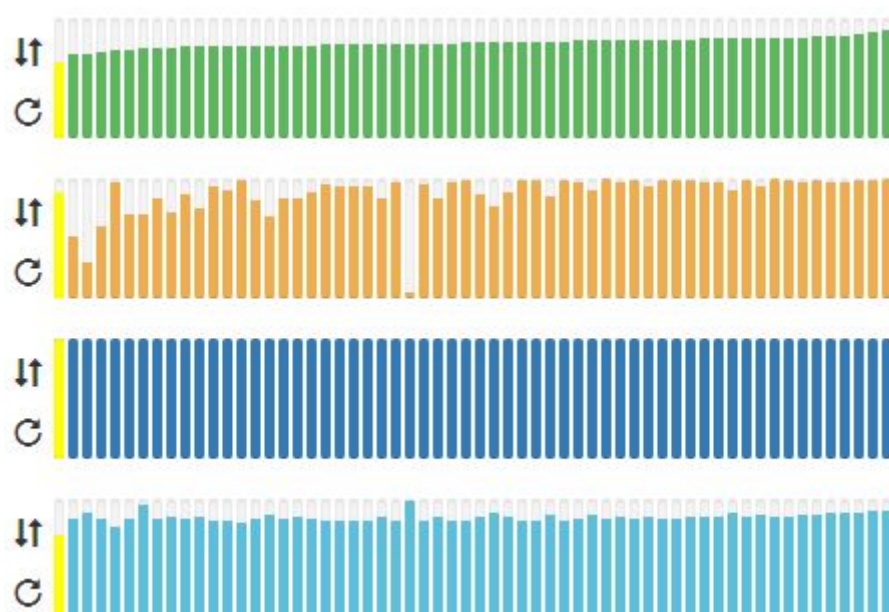
- 62.1. Source: Kommentaarides kasutatud nime lisamine või muutmine
 62.2. Reference: Add or change the name used in comments
 62.3. Machine: Inclusion or modification of the name used in the comments
 62.4. The structure of the sentence is intact.
 62.5. Two verbs were translated wrongly, with wrong tenses which are the same in English.

3. Analyse Attention

Legend:

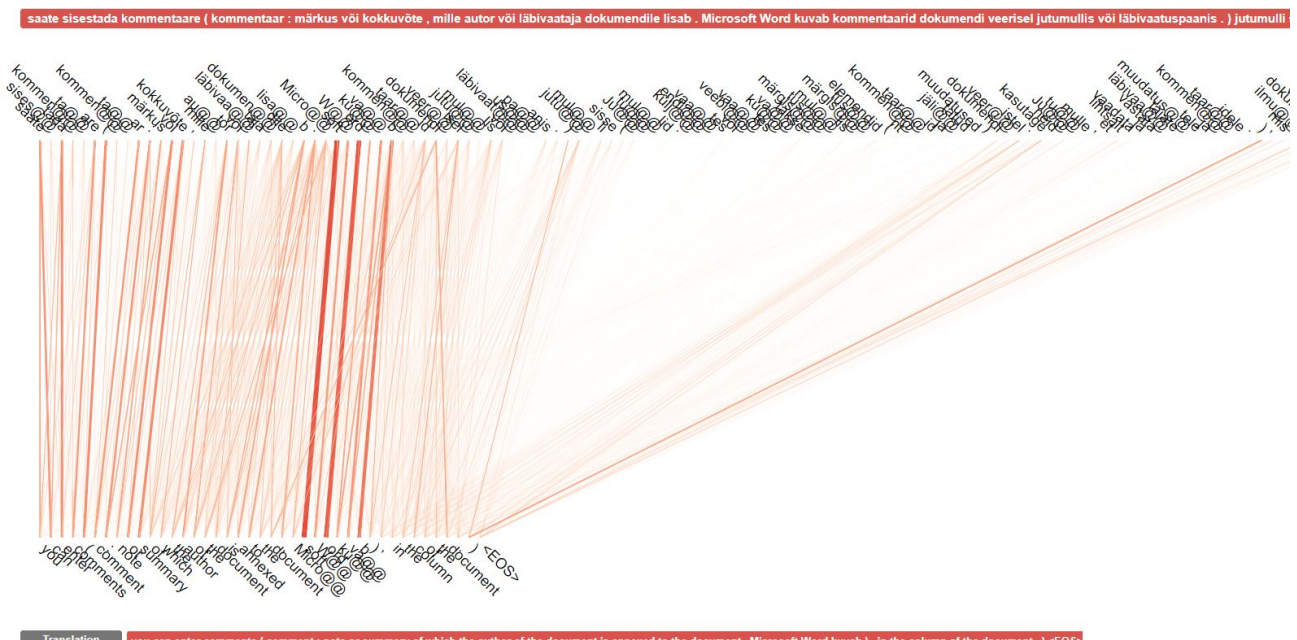
1. Green - Confidence
2. Orange - CDP (Coverage Deviation Penalty)
3. Dark Blue - AP_{out} (Absentmindedness Penalty)
4. Light Blue - AP_{in}

3.1. Sorted based on confidence metric:



The worst 3 sentences:

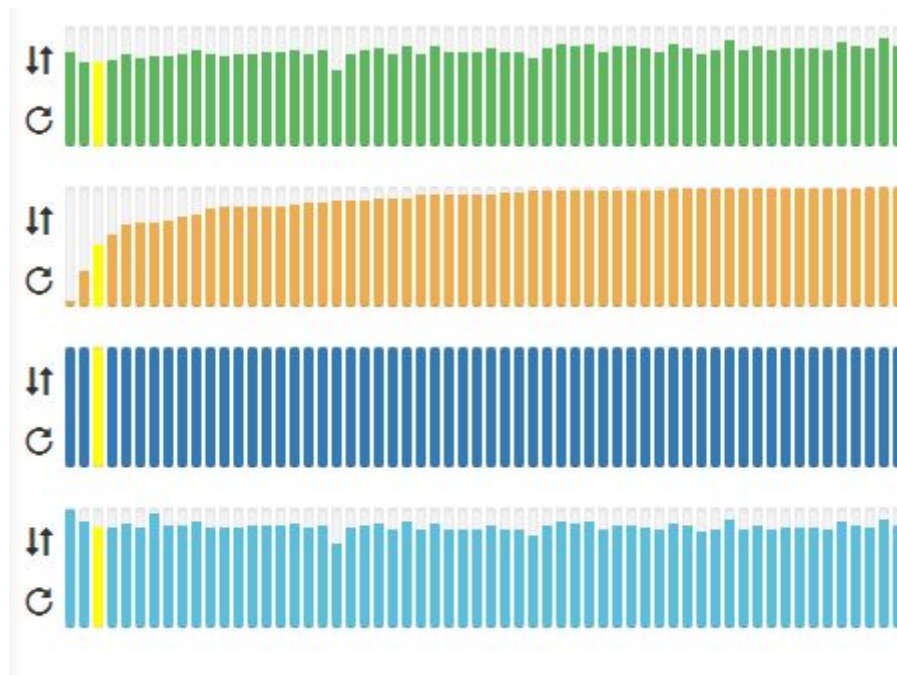
- 909 (half of sentence missing)



(this figure is really hard to read, however it is visible that half of the sentence gets only very little attention)

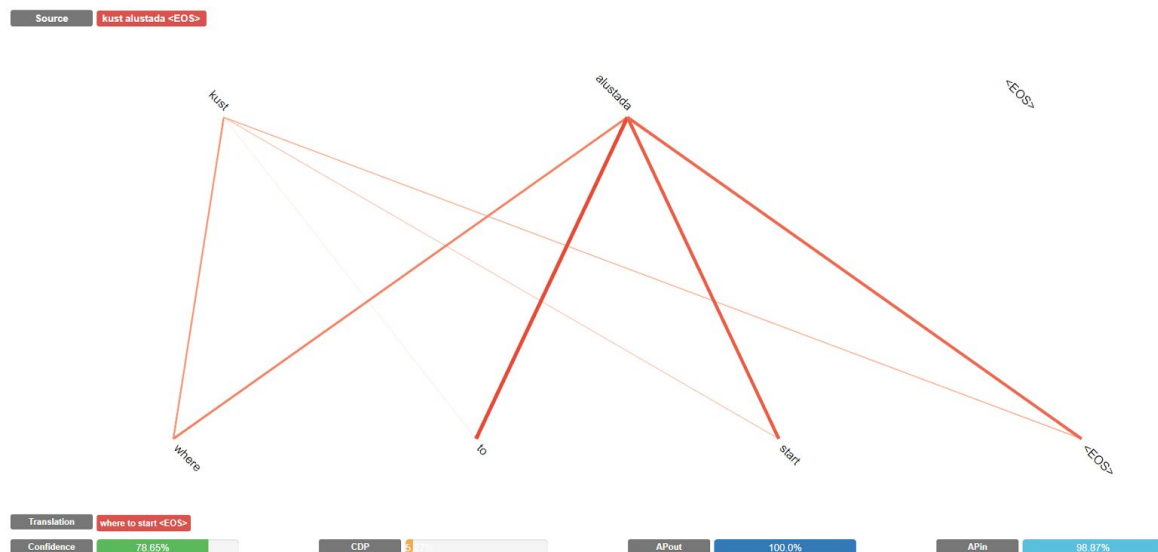
- 156 (some mistranslations and rephrasing)
- 456 (mistranslation and synonyms)

3.2. Sorted based on CDP metric



The worst 3 sentences:

- 825 (perfect translation, although different to human translation)



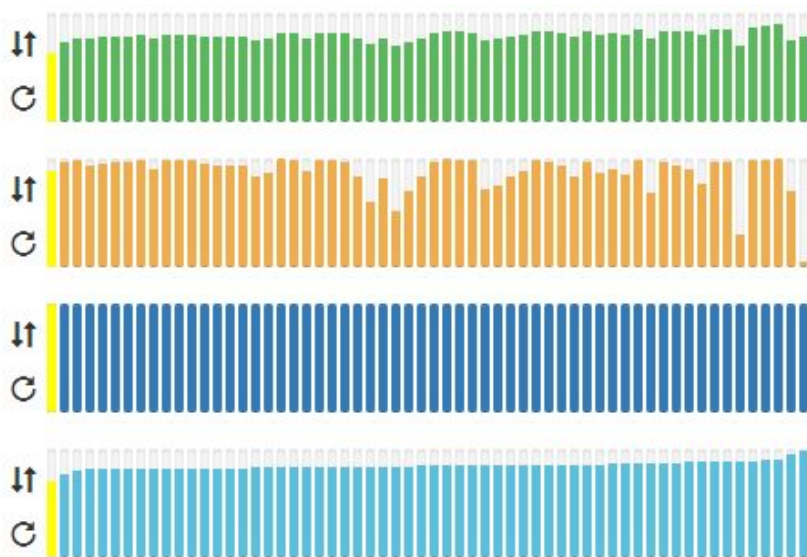
- 465
- 156

The last two of them are same as for previous sorting

3.3. Sorted based on APout metric

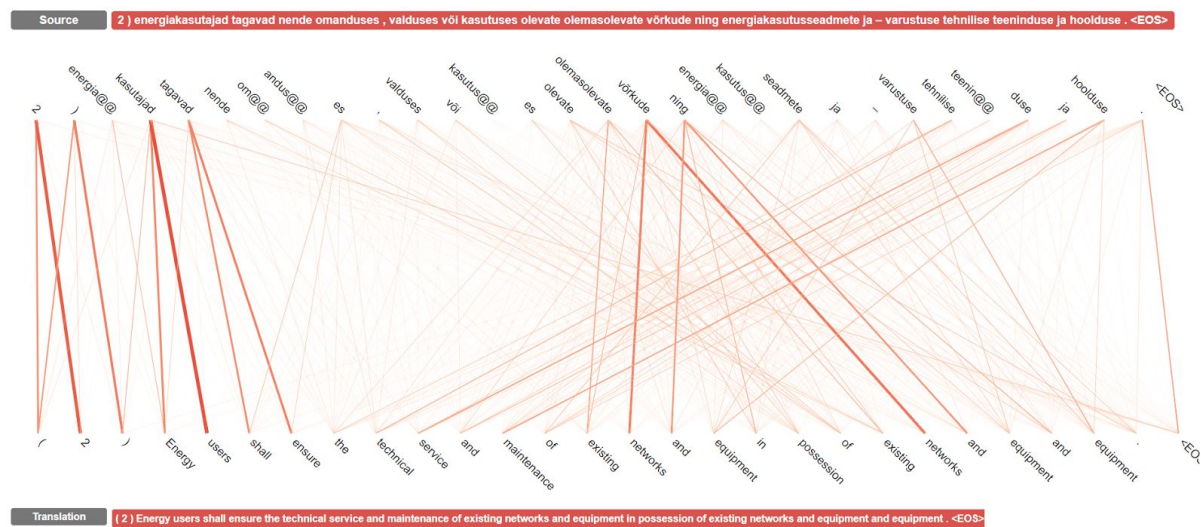
APout metric shows always 100% so makes no sense to sort based on that.

3.4. Sorted based on APin metric



The worst 3 sentences:

- 909 (half of sentence missing and mistranslations)
- 40 (last half of sentence random)



(here we can see that some words correspond to many output words)

- 420 (half of the sentence missing)

In many cases the scores were worst, when half of the sentence were missing, which is indeed major issue.