Beispiel-Ablauf für Task 1

Input Satzgruppe mit 5 Transkriptionen

* Bi mer esches so, dass mi s Thema sehr interessiert und au immer interessiert het Aso i machs nüd wegem Job nocher, i weiss o no gär nüd wo assi gärn här wett nochäne Ehm. Aso mi interessierts würklech und es isch cool.
* Es isch so das mich sThema interessiert hät aso ich mags nöd weg em Job noch, \*\*\* aso mi intressiert würklich und es isch cool\*\*\*
* dem isch scho so, das mis s Thema sehr interessiert und au imer interessiert hett. aso, ich machs nüd wegem Dschob noch i wäss no gär nid won ig will here nochane. äm- also mi interessierts würkli und es isch kuul.
* thurgauer thurgauer thurgau
* de meh isch scho so dass mi s Thema intessiert und scho immer ineressiert het aso i machs nü wegem Job ich wäss no gä nö woni herewel nochine aso mi intresiets wöckli und es isch kuul

1. **Satzgruppe in Liste abspeichern**  
   Group = ['Bi mer esches so, dass mi s Thema sehr interessiert und au immer interessiert het Aso i machs nüd wegem Job nocher, i weiss o no gär nüd wo assi gärn här wett nochäne Ehm. Aso mi interessierts würklech und es isch cool.', 'Es isch so das mich sThema interessiert hät aso ich mags nöd weg em Job noch, \*\*\* aso mi intressiert würklich und es isch cool\*\*\*', 'dem isch scho so, das mis s Thema sehr interessiert und au imer interessiert hett. aso, ich machs nüd wegem Dschob noch i wäss no gär nid won ig will here nochane. äm- also mi interessierts würkli und es isch kuul.', 'thurgauer thurgauer thurgau', 'de meh isch scho so dass mi s Thema intessiert und scho immer ineressiert het aso i machs nü wegem Job ich wäss no gä nö woni herewel nochine aso mi intresiets wöckli und es isch kuul']  
     
   scores\_of\_whole\_group = bleu\_ratings(group)  
   # bleu\_ratings(group) bewertet alle Sätze in der Gruppe und gibt eine Liste mit den Scores zurück, hier sieht man zum Beispiel, dass Satz mit Index 3 eine sehr tiefe Score hat.

[0.7955705875934534, 0.5296049840973046, 0.813912716804547, 0.009606403793569373, 0.7107904003877338]

1. **Transkriptionen bewerten und schlechte ausfiltern**goodTranscriptions = bleu\_score.getGoodTranscriptions(group)  
   # getGoodTranscriptions ruft bleu\_ratings(group) auf und übergibt dies an reject\_outliers(group, scores, minscore=0.5) wo dann alle Sätze welche unter minscore sind, aus der Gruppe genommen werden. Zurückgegeben wird ein Tupel mit den Sätzen und den Scores dazu.  
     
   group = goodTranscriptions[0]  
   ['Bi mer esches so, dass mi s Thema sehr interessiert und au immer interessiert het Aso i machs nüd wegem Job nocher, i weiss o no gär nüd wo assi gärn här wett nochäne Ehm. Aso mi interessierts würklech und es isch cool.', 'Es isch so das mich sThema interessiert hät aso ich mags nöd weg em Job noch, \*\*\* aso mi intressiert würklich und es isch cool\*\*\*', 'dem isch scho so, das mis s Thema sehr interessiert und au imer interessiert hett. aso, ich machs nüd wegem Dschob noch i wäss no gär nid won ig will here nochane. äm- also mi interessierts würkli und es isch kuul.', 'de meh isch scho so dass mi s Thema intessiert und scho immer ineressiert het aso i machs nü wegem Job ich wäss no gä nö woni herewel nochine aso mi intresiets wöckli und es isch kuul']  
     
   scores = goodTranscriptions[1]  
   [0.7955705875934534, 0.5296049840973046, 0.813912716804547, 0.7107904003877338]
2. **Bester Satz aus Gruppe finden**best\_index = bleu\_score.max\_index(scores)  
   2

group[best\_index]

dem isch scho so, das mis s Thema sehr interessiert und au imer interessiert hett. aso, ich machs nüd wegem Dschob noch i wäss no gär nid won ig will here nochane. äm- also mi interessierts würkli und es isch kuul.

1. **Schlechtester Satz aus Gruppe finden**  
   worst\_index = bleu\_score.max\_index(scores)  
   1

group[worst \_index]  
Es isch so das mich sThema interessiert hät aso ich mags nöd weg em Job noch, \*\*\* aso mi intressiert würklich und es isch cool\*\*\*

1. **Schlechter Satz verbessern**improved\_sentence = align.improve(group, worst \_index, align.HUNALIGN)  
   #die improve function nimmt einen Aligner als Argument, zur Zeit haben wir HUNALIGN und BLEUALIGN implementiert.  
   #In einem ersten Schritt wird der Schlechte Satz mit den anderen aligniert.  
   aligned\_words = align\_one\_sentence\_to\_the\_others(…)  
   [['isch', 'esches', 'scho'], ['so', 'so,'], ['das', 'dass'], ['mich', 'mi s', 'mis'], ['sThema', 'Thema', 's Thema'], ['interessiert', 'ineressiert'], ['hät', 'het', 'hett.'], ['aso', 'Aso', 'aso,', 'also'], ['mags', 'machs'], ['nöd', 'nüd'], ['weg em', 'wegem', 'weg'], ['Job'], ['noch,', 'nocher,', 'noch i'], ['mi'], ['intressiert', 'interessierts', 'intresiets'], ['würklich', 'würklech', 'würkli', 'wöckli'], ['und'], ['es'], ['cool\*\*\*', ‘kuul’, 'cool.'], ['Es'], ['ich', ’ig’], ['weg'], ['em'], ['\*\*\*', 'här', 'äm-', 'woni']]  
     
   #Wörter welche wir ersetzen wollen  
   bad\_words = ["\*\*", "\*\*\*", "\*\*\*\*", "??", "???"]  
     
   #Mit einer gewichteten Levensteindistanz suchen wir weitere schlechte Alignments und die darin schlechtesten Wörter  
   levenshtein.score\_alignment(group), group)  
   0.75 ['\*\*\*', 'här', 'äm-', 'woni']

0.9016058016058016 ['intressiert', 'interessierts', 'intresiets']

0.8607142857142858 ['cool\*\*\*', 'cool.', 'kuul.', 'kuul']  
  
bad\_words.append(levenshtein.worst\_word(group))  
['\*\*', '\*\*\*', '\*\*\*\*', '??', '???', '', '', '\*\*\*', 'intresiets', 'cool\*\*\*']  
  
#Für den zu verbessernden Satz gehen wir jedes Wort durch und ersetzen die bad\_words durch das Beste aus dieser Gruppe  
words[INDEX\_OF\_BAD\_WORD] = levenshtein.best\_word(group)  
#Ausgabe von improved\_sentence  
Es isch so das mich sThema interessiert hät aso ich mags nöd weg em Job noch, äm- aso mi intressiert würklich und es isch kuul.

Nochmals zum Vergleich  
bad sentence:

Es isch so das mich sThema interessiert hät aso ich mags nöd weg em Job noch, \*\*\* aso mi intressiert würklich und es isch cool\*\*\*

Improved sentence:

Es isch so das mich sThema interessiert hät aso ich mags nöd weg em Job noch, äm- aso mi intressiert würklich und es isch kuul.