**Project Tekst Analysis assigment 4**

Robert van Timmeren, Kylian de Rooij and Remi Thüss

**Inter-annotator agreement results**

Below are the results of the inter-annotator agreement by calculating the precision, recall and f1-score. These scores also include “None” when at least one annotator annotated a word as something else as “None” and at least one annotator annotated the word as “None”. This same method was also used for the confusion matrix.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Precision** | **Recall** | **F1-score** | **Support** |
| **CIT** | 0.810 | 0.895 | 0.850 | 19 |
| **COU** | 0.925 | 0.790 | 0.852 | 62 |
| **ENT** | 0.111 | 0.059 | 0.077 | 17 |
| **NAT** | 0.500 | 1.000 | 0.667 | 1 |
| **ORG** | 0.790 | 0.711 | 0.749 | 90 |
| **PER** | 0.644 | 0.927 | 0.760 | 41 |
| **None** | 0.292 | 0.326 | 0.308 | 43 |
|  |  |  |  |  |
| **Accuracy** |  |  | 0.674 | 273 |
| **Macro avg** | 0.582 | 0.672 | **0.609** | 273 |
| **Weighted avg** | 0.678 | 0.674 | 0.669 | 273 |

*Table 1. Inter-annotator agreement results*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **CIT** | **COU** | **ENT** | **NAT** | **ORG** | **PER** | **None** |
| **CIT** | <17> | 2 |  |  |  |  |  |
| **COU** | 3 | <49> |  |  | 2 |  | 8 |
| **ENT** |  |  | <1> |  | 11 |  | 5 |
| **NAT** |  |  |  | <1> |  |  |  |
| **ORG** |  |  | 8 |  | <64> |  | 18 |
| **PER** |  |  |  |  |  | <38> | 3 |
| **None** | 1 | 2 |  | 1 | 4 | 21 | <14> |

*Table 2. Confusion matrix*

**Discussion of difficult cases**

The majority of differences between the annotators were caused by similar differences in annotations. These differences can be divided into 5 groups. These differences and the solution are shown in the table below.

|  |  |
| --- | --- |
| **Difference** | **Solution** |
| Provinces were sometimes annotated as COU (country) and sometimes as CIT (city). | The gold standard always contains all provinces as COU (according to the guidelines). |
| Continents where not consistently annotated as COU. | The gold standard always contains continents as COU. |
| Sometimes words that did not have a Wikipedia page were annotated as just the class (without the Wikipedia link). | The gold standard always contains a link if a class is found. If there is no link for a particular class, the class is removed. |
| Places of interest that do not belong to any of the classes were sometimes given a class (for example The Vietnam Veterans Memorial). | The gold standard only contains words that directly relate to a class. Cases like The Vietnam Veterans Memorial were not annotated for that reason. |
| Sometimes the title and the person that has that title were annotated given a different Wikipedia link. For example French President Jaques Chiraq was annotated as:  French President -> French President Wikipedia  Jaques Chiraq -> Jaques Chiraq Wikipedia | The gold standard contains the most specific class, meaning “French President Jaques Chiraq” is all linked to the Wikipedia of Jaques Chiraq. |