**Language Model/ Dataset Paper**

**Venue & Timeline:**

ACL 2021,

|  |  |
| --- | --- |
| Abstract deadline (*long & short papers*): | January 25, 2021 |
| Submission deadline (*long & short papers*): | February 1, 2021 |

Dec 1-11 – Find novelty

Dec 12 – 30 – Experiments

Jan – Write up experiments

Feb – Paper writing

**Paper to reference for style:** GoEmotions: A Dataset of Fine-Grained Emotions

**Data:** Current set of annotations where speaker tells the truth against (Reasoning/ Game Move/ Rapport/ Share Information)

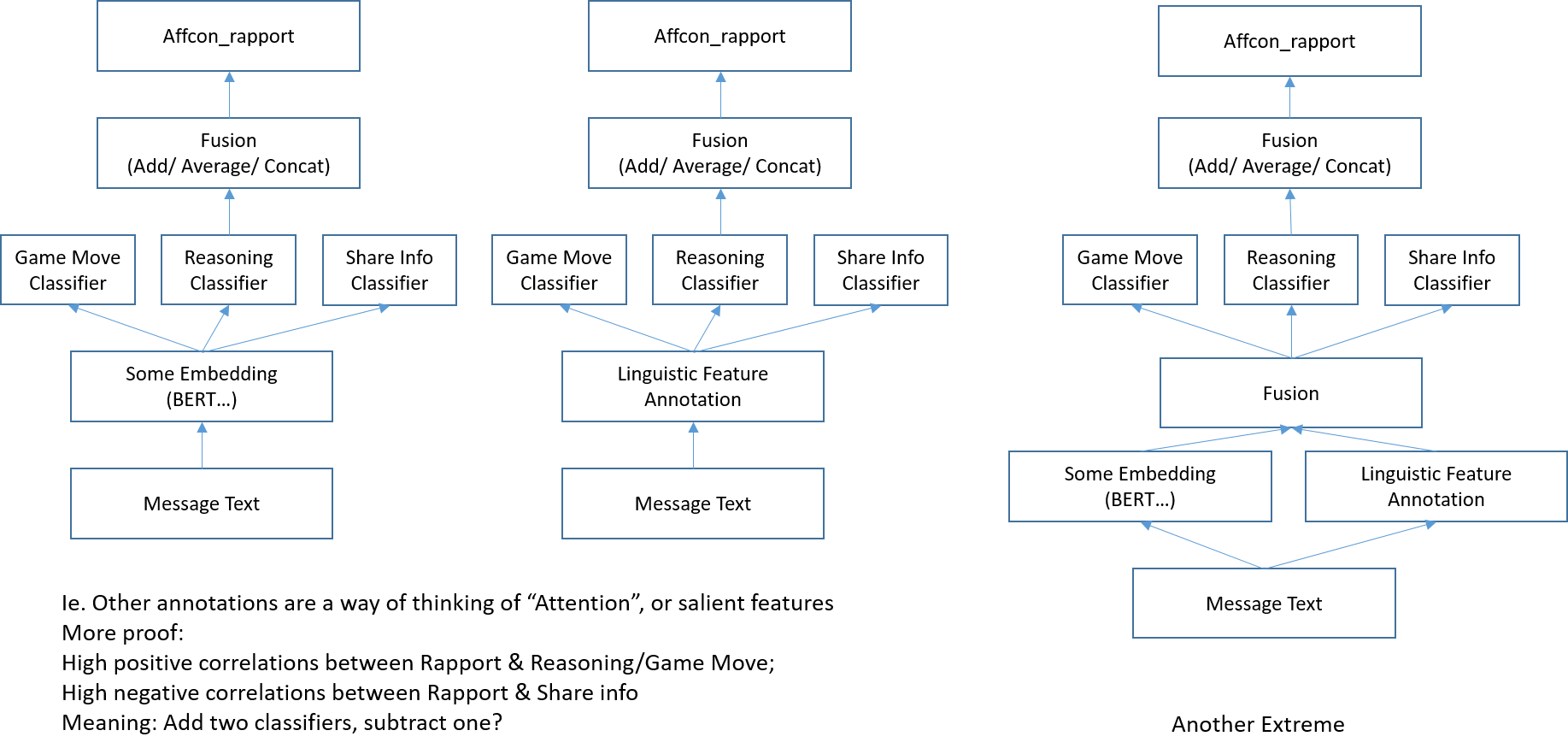
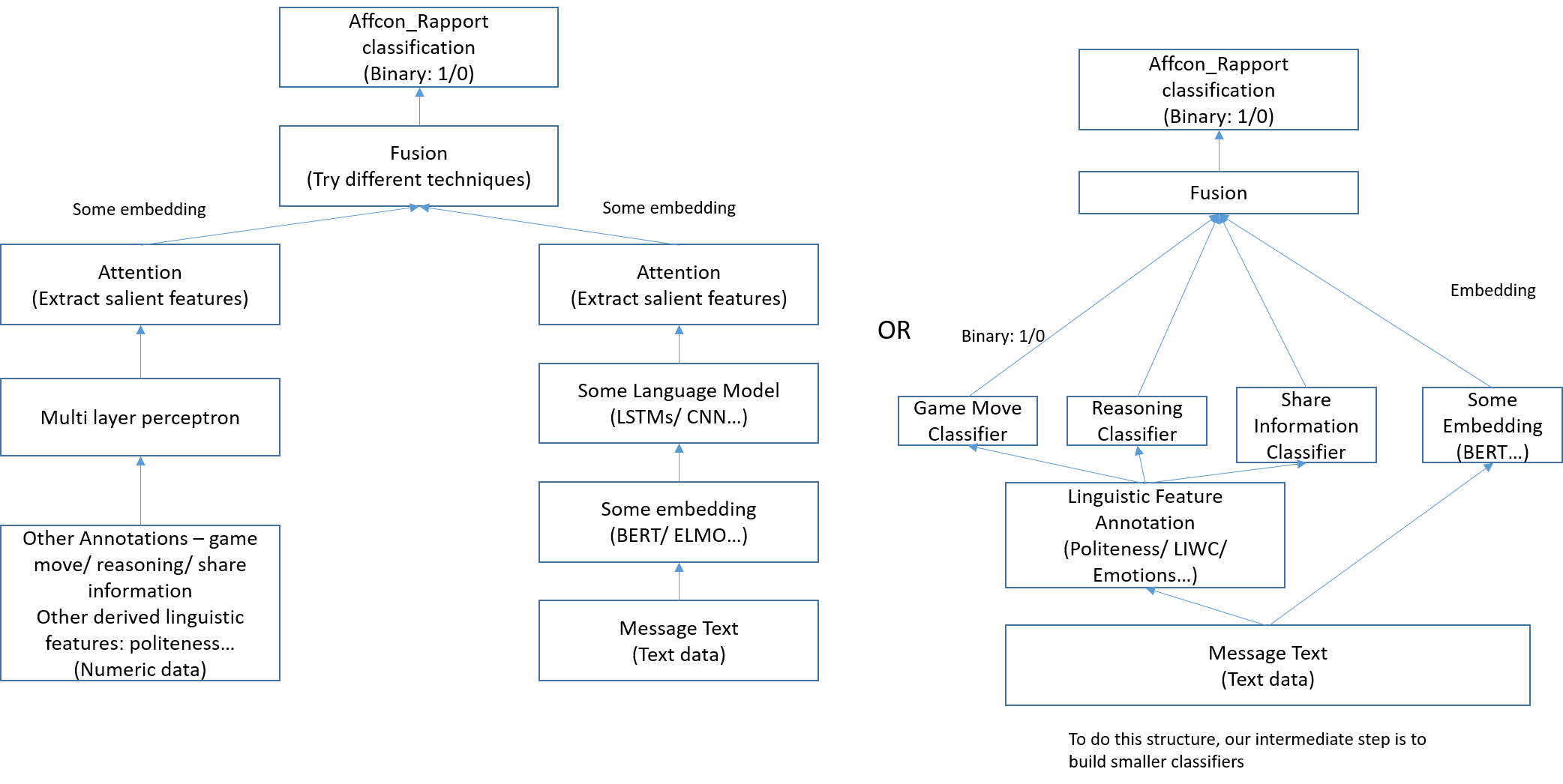
**Possible Direction**: What is the best way to enrich the data for the model? **[Novelty could be in enrichment construction or fusion construction]**

Enrich stuff in hierarchical models – identify different points in the NN pipeline to incorporate language features & examine which works out best

+ Figures with different training set size/ epochs

Enrich datasets with generative adversarial networks

**Proposed Structure (Lynnette)**



Note: can remove linguistic feature annotation to just an embedding

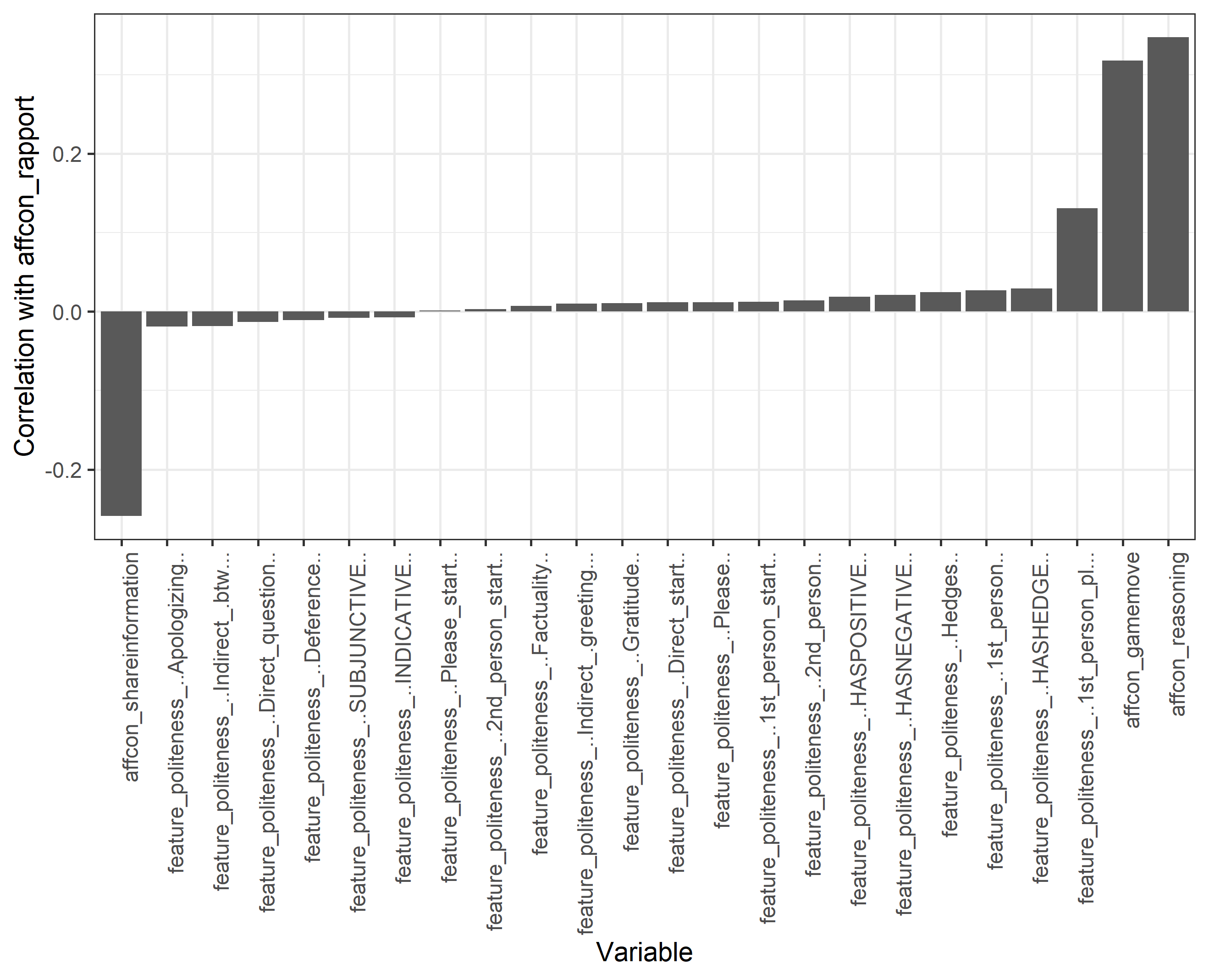
**4-grams**

|  |  |
| --- | --- |
| **Affcon\_rapport = True** | **Affcon\_Rapport = False** |
| Move east to counter  Mostly relying on the  Move into Serbia keeping  Most likely to be  Move combinations if  Move either direction  Move cautiously  Move by  Move either  *Could we say a lot on game movement discussion, and gives choices (“either”), and adjectives (“cautiously”)* | Mind terribly if moved  Minute they are asking  Military lined up at  Might be more worth it  Mind moving into  Misclicked when placing  Might move  Mind terribly  Miss it  Misery yknow  *Could we say a lot of minding, trying to be nice but actually shooting a dagger HA* |

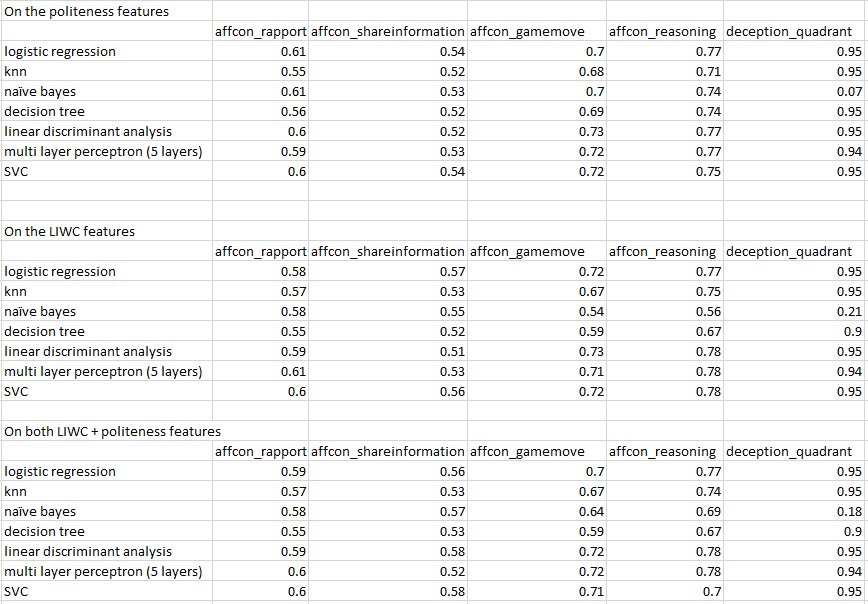
**Pronouns**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **First person %** | **Second Person %** | **Third Person %** |
| Average |  |  |  |
| Affcon\_rapport=True | 3.93847 | 3.325816 | 1.024625 |
| Affcon\_rapport=False | 4.15769 | 3.41641 | 1.0598 |
|  |  |  |  |
| Per text |  |  |  |
| Affcon\_rapport=True | 5.7562  Std 5.9032 | 3.7712  Std 5.2289 | 1.0721  Std 3.057 |
| Affcon\_rapport=False | 4.8014  Std 5.7312 | 4.0585  Std 5.6784 | 1.0823  Std 3.1328 |

**Correlation Plots**



**Lexical Features**



**Against affcon Rapport (only Text) + Early stopping**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Number of epochs | Batch Size | Precision | Recall | F1 Score |
| ROBERTA + Roberta Transformer | 3 | 64 | 0.586 | 0.586 | 0.586 |
| BERT + 3 layers BiDirectional LSTM | 32 | 64 | 0.5063 | 0.91485 | 0.6517 |
| BERT + 3xCNN | 32 | 64 | 0.62 | 0.62 | 0.62 |
| BERT + RNN | 32 | 64 |  |  |  |
| BERT + CNN-LSTM | 32 | 16 | 0.7422 | 0.5393 | 0.6183 |
| BERT + LSTM | 32 | 64 | 0.4999 | 0.9999 | 0.6665 |
| BERT + BERT Transformer | 32 | 64 | 0.4956 | 0.4956 | 0.4956 |
| ConvLSTM | 32 | 64 | OUT OF MEMORY GG |  |  |
| GPT2 + GPT transformer | 32 | 64 | 0.3957 | 0.3957 | 0.3957 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Number of epochs | Batch Size | Precision | Recall | F1 Score |
| Numerical MLP | 32 | 64 | 1.00 | 0.586 | 0.4061 |
| Combined  Text LSTM +  Numerical MLP | 32 | 16 |  |  | 0.50 |

**References To Consider:**

<https://arxiv.org/pdf/2004.13609.pdf>

**Our Paper Structure: +cool title name**

Introduction: Intro to the Diplomacy game, rapport (+sample texts)

Related work: Past Diplomacy papers by Cornell lab, some stuff on Rapport/affcon

Dataset description:

* Constructing dataset for annotation, i.e. splitting them into sentences, deduplication those that are similar (eg moves)
* Explain Rapport/ Game Move/ Reasoning/ Share Information – paraphrase from how we put in MTurk + examples
* When we construct this dataset, we seek to maximize the following objectives… OR what is our objectives of the dataset
* Annotation using MTurk, rater interface, summary statistics

Data Analysis

* Summary statistics + table/ distribution graphs
* Interrater correlation
* Correlation among annotated categories
* Performing LIWC + correlation among categories (correlation plots)
* Rapport vs talkativeness

Modelling

* N-grams of each annotated category (reference Table3 of GoEmotions)
* Pronouns
* Language model architecture + parameters + experiments + results
* **Novelty experiment**: fusion of data

Conclusion

* Conclusion, future work
* Data disclaimer/ limitations