

# Named Entity Recognition in Peace Agreements

Myanmar as a case study

Niamh Henry  
PeaceRep  
University of Edinburgh  
niamh.henry@ed.ac.uk

## 1 INTRODUCTION

The PA-X database [3] hosts over 1,900 peace agreements from over 150 peace processes from 1990-present. Within this database, there is data provided on the content of each peace agreement outlining what the content includes such as the date it was signed, the stage of the peace process it is at and the themes covered within the agreement. Another useful measure from these agreements is explicitly stating what actors are included within the agreement, as it can provide valuable insights into how these actors, and their actions, influence what is included in the contents of the agreements. Currently, the actors coded explicitly within the data provided in the PA-X database are those who signed the agreement, as it is written in the agreement, under the variable *parties* and *third-parties*. However, this is restricted to only those who signed the agreement, rather than any actors or organisations that are mentioned throughout the text, that may be involved or hold influence over the contents.

To provide further data and detail on the entities (actors) referred throughout the text, automatic recognition of entities would provide to be useful to extend this database to further investigate those involved within peace agreements. Further information on the people and organisations that are actors, or mentioned, within these peace agreements can provide methods to link this valuable data with external data sources, in particular, data on violent and conflict events (such as ACLED [8] or UCDP [9]). Particularly, an end-goal of this activity would be to map all actors within the PA-X database in a network graph, to link and clearly visualise the impact of specific actors (people, and organisations/group) that are present within peace agreements, and the influence they have on the contents of the agreements, and their resulting impacts. This would be a valuable asset to studies on peace processes, as it would allow us to see what actors are mentioned, or take part in the peace agreements, and what context they are surrounded by.

To address this need for entities to be recognised within the texts of peace agreements, a subset of PA-X corpus, namely, all peace agreements from Myanmar, is utilised to test automatic named entity recognition methods. There is a large volume of texts available in PA-X, therefore a subset of this provides to be a good indication if these methods can be extended to the entire database. This report outlines the methods and analyses the results of entity recognition from the agreements of Myanmar, and discusses how these methods can be extended to the entire database in the future.

## 2 DATA AND METHODS

### 2.1 Data

From the PA-X database, all the agreements found under the query of Country/Entity as Myanmar were extracted using the search tool

on the PA-X website [2]. This included 36 agreements spanning from 2011-2020. Myanmar provides to be a useful sample of the agreements available from PA-X to test automatic named entity recognition as it is a good sized sample of an ongoing and active peace process, with a high number of actors involved. Additionally, the texts have been translated to English (like many of the agreements from PA-X), which will provide a useful indication of how the models work on translated documents.

The full text of each agreement was extracted from the PA-X corpus export tool and the dataset outlining the finer details of these agreements was also extracted in csv format. The full text of each agreement was copied into Microsoft Word document files, one per agreement, and the documents were named as the agreement ID as stated in the database. This provided the data in the required format for the methods outlined below.

### 2.2 Methods

To execute automatic named entity recognition (NER) on these texts, Python was utilised to read all texts and run the required models. Each document including the full texts were read into Jupyter notebook and segmented into sentences. These were then serialised to disk as JSON and stored as csvs (one per document/peace agreement) which allowed each sentence to be read into the Jupyter notebook to be inputted into the NER model.

To automatically detect named entities within the files containing these agreements, the Natural Language Processing (NLP) model SpaCy [6] was utilised. Other models, such as Stanford's Stanza NER model [7], were considered, however, it provides IOB (Inside-outside-beginning) tagging for each entity, which requires high levels of post-processing of results to aggregate all entities. Therefore, SpaCy was selected as it labels entities as a whole, and has shown to be reliable and robust for NLP, particularly in large texts. SpaCy provides a pre-trained model to perform NER using a pipeline that tokenises the text to ensure multi-term phrases are read correctly, which limits the pre-processing required to run the model. Nevertheless, the full-text of the agreements required some cleaning to optimize the model. This included stripping the texts of unnecessary whitespace (such as paragraph breaks), and removing the stop-words within SpaCy's predefined dictionary.<sup>1</sup> This helps to reduce the noise in the data, and allows the model to focus on meaningful text.

SpaCy's predefined NER categories selected to be extracted from the texts included organisation, GPE (geopolitical entities), location and person to represent all possible actors (organisations, people and other entities) and to provide location data to further improve

<sup>1</sup>The full list of stop-words can be found here: [https://github.com/explosion/spaCy/blob/master/spacy/lang/en/stop\\_words.py](https://github.com/explosion/spaCy/blob/master/spacy/lang/en/stop_words.py)

Named Entity Label	Total Recognised	Total Unique Entities
Organisations	992	326
Geopolitical Entity	197	78
Location	9	9
Person	179	117

**Table 1: The number of entities recognised automatically by the SpaCy model on all peace agreements, and the unique values of each entity per category.**

the geo-coding of the agreements in PA-X in the future. The model was executed on each agreement iteratively, and each entity recognised was labelled with the agreement ID from PA-X, the sentence number in the text where the entity was found, the tokens of the entity and the label (either ORG, GPE, LOCATION or PERSON)<sup>2</sup>. These raw results were then exported as a csv file for further analysis, for which the results are outlined in Section 3 below.

### 3 RESULTS

#### 3.1 Raw Data Results

From the 36 peace agreements inputted to the model, 1,377 tokens were identified as named entities of either organisations, GPE, locations or person names. This model provides to be extremely useful, as it recognised a large number of entities within the large volumes of text at a much quicker rate than a human manually coding (execution time of model >2 minutes). A table outlining the number of entities recognised per category (and the unique number of entities) can be seen in Table 1.<sup>3</sup> However, after manually checking the results there were some discrepancies found with the results, therefore the results of this model were not 100 percent accurate. For instance, some people’s names were recognised as organisations, only parts of peoples names or titles were picked up correctly, and some place names were recognised as names of people. To correct some of these discrepancies, manual dictionaries were created for organisations, persons and their affiliated organisations. Although this will not provide 100 percent accuracy in the results, it provides a good indication of all the actors present in the agreements, specifically for organisations, and people mentioned. These manual labelling dictionaries are further outlined in Section 3.2 below.

#### 3.2 Labelling of results

As the end-goal of using this data is to eventually identify the key actors within agreements to map a network of actors, it was decided to focus on only the organisations and people’s names identified within the texts. Thus, several dictionaries were created to recognise what tokens that were identified actually represent the same identities for organisations, people, and their affiliated organisations. Firstly, an organisation dictionary was developed to match the different alias’ of organisations to represent the one organisation. This aids in recognising abbreviations of organisations, and different branches of them. For instance, the group Chin National Front was

often referred to in the texts as CNF, or it is often used as an umbrella term for the Chin National Army (sometimes referred to as CNA). The same applies to Karen National Union (KNU) and its military branch the Karen National Liberation Army (KNLA). Through this method 30 organisations were coded with different alias’.<sup>4</sup> For visualisation purposes, these organisation names were then mapped to their abbreviated terms using another dictionary.<sup>5</sup> This resulted in a full dataset of corrected labelled organisations that are mentioned within all the agreements, which provides 486 instances of organisations throughout the 36 peace agreements. This dataset can be found here: [https://github.com/niamhhenry/NER\\_Myanmar/blob/main/Datasets/Organisations/org\\_in\\_agreements.csv](https://github.com/niamhhenry/NER_Myanmar/blob/main/Datasets/Organisations/org_in_agreements.csv).

Further dictionaries were developed to identify the person entities identified by the model, to the proper full names and titles of the people the recognised entity is referring to. This was deemed necessary as in some cases the model picked up parts of names, or additional/unnecessary text that appeared next to the names. For instance, ‘Dr. Tin Myo Win’ was recognised by the model as ‘Tin Myo Win Representative’. The process of developing this dictionary allowed for persons who were identified as organisations to be correctly coded as people, and to manually populate the dataset with persons not identified by the model.<sup>6</sup> Additionally, another dictionary was created to map the people identified to their affiliated organisation/group (who they were representing when partaking in these agreements). This resulted in a dataset recording 104 instances of people mentioned throughout the texts, with those who’s affiliated organisations were known (from PA-X or Google searches) also recorded. This dataset of persons recognised within all the texts can be found here [https://github.com/niamhhenry/NER\\_Myanmar/blob/main/Datasets/People/all\\_people\\_in\\_agreements\\_withManual.csv](https://github.com/niamhhenry/NER_Myanmar/blob/main/Datasets/People/all_people_in_agreements_withManual.csv).

#### 3.3 Analysis of organisation occurrences

**3.3.1 Occurrences over all agreements.** Using the corrected and labelled datasets of organisations identified within the peace agreements, we can see some interesting trends by aggregating the occurrences of each organisation found in the texts overall. Table 2 shows the overall mentions of all identified organisations throughout all 36 texts of peace agreements. Figure 1 visualises this data in the form of a bar chart, to illustrate the range of organisations and how many times they are mentioned. From this, we can see that the Chin National Front (CNF) is the organisation mentioned the most (64 times) throughout all the texts. Thus, it is evident the CNF have a high involvement and influence in peace agreements in Myanmar. The four organisations that follow (different versions of the Joint Ceasefire Monitoring Committees) are expected as these committees were a direct result from these agreements. Therefore, the next highest mentions organisation/group that are actors within the agreements are the Shan State Progressive Party (SSPP) with 32 mentions overall, followed by the Karen National Union (KNU) with 28 mentions. From this data we can assume that these groups have

<sup>2</sup>The Jupyter notebook where the model was run can be accessed here: [https://github.com/niamhhenry/NER\\_Myanmar/blob/main/NER\\_spacy\\_Myanmar.ipynb](https://github.com/niamhhenry/NER_Myanmar/blob/main/NER_spacy_Myanmar.ipynb)

<sup>3</sup>The csv of the raw results from this model can be found on GitHub here: [https://github.com/niamhhenry/NER\\_Myanmar/blob/main/Datasets/all\\_entities\\_myanmar.csv](https://github.com/niamhhenry/NER_Myanmar/blob/main/Datasets/all_entities_myanmar.csv)

<sup>4</sup>This dictionary of alias’ can be found on GitHub here:

<sup>5</sup>This abbreviation dictionary can be found here: [https://github.com/niamhhenry/NER\\_Myanmar/blob/main/Dictionaries/org\\_abr\\_dict.csv](https://github.com/niamhhenry/NER_Myanmar/blob/main/Dictionaries/org_abr_dict.csv)

<sup>6</sup>17 persons were identified in this manual check that were not picked up by the spaCy model. Some were due to an error in reading the last page of agreement text number 1841, however, the reasons for others not identified have not yet been realised.

high involvement in peace agreements, and perhaps, in the conflict surrounding them. However, to gain more insights into these actors it is useful to investigate the mentions of these organisations per agreement.

**3.3.2 Occurrences per agreement.** Disaggregating the occurrences of the identified organisations per agreement text<sup>7</sup> allows us to investigate what organisations are involved in specific agreements, and the range of agreements that organisations are involved in. For instance, we can see in Figure 2 that although the CNF have the highest number of mentions in agreements, the Government of the Republic of the Union of Myanmar is mentioned throughout a much higher range of agreements. This is expected, as usually peace agreements would be with the government of the state in question. This graph is useful to illustrate how some agreements are specific to organisations, for instance the Arakan Army (AA on the graph) are only mentioned in one agreement, whereas the SSPP/SSA (Shan State Army) and the Tatmadaw (the armed forces of Myanmar) are both mentioned in 7 different agreements. Evidently, these two army's have a much higher involvement in the Myanmar peace processes (and perhaps, Myanmar conflicts) than the Arakan Army.

An interesting perspective of the involvement of these organisations, is to view the occurrences of them by the different stages of the Myanmar peace process, as coded in PA-X database. Figure 3 illustrates the frequency of mentions of occurrence of each group in peace agreements, by the stage of the peace process the agreement took part at. We can immediately see from this plot that the majority of organisations were involved during the ceasefire stages, which is expected. Interestingly, we can see the few organisations that were involved in the pre-negotiation/process phase, with most being peace-committees, with the exception of the KNU/KNLA and the SSPP/SSA. Additionally, from this chart we can see that the CNF have the most mentions in the 'SubPar' stage, which includes agreements that concerns parties that are engaged in discussion and agreeing to substantive issues to resolve the conflict, but only deal with some of the issues in ways that appear to contemplate future agreements to complete [3].

**3.3.3 Occurrences over time.** Organising the mentions of organisations per agreement allows us to view the occurrences of the organisations in agreements over time, as seen in Figure 4. From this plot we can easily identify some trends in the mentions of organisations and groups in Myanmar's peace agreements. Around the year 2012, we can see that a cluster of data points (representing organisations), which represents the many individual peace agreements made with specific groups. Within PA-X, 16 peace agreements were signed in Myanmar in 2012, and most of these explicitly state in the agreement name the organisation/group the agreement directly applies to, for example, the Rakhine State Liberation Party (RSLP) and State Government Peace-Making Group 5-point State-level Agreement, the Chin National Front (CNF) and Government 9-point State-level Peace Agreement and the Shan State Army-South (SSA-S), Government 11-Point Peace Agreement to name a few. This clearly indicates a increasing shift in peace agreements around this time (as seen in Figure 4) as the current Myanmar peace process was

beginning to materialise, after the election of President U Thein Sein in 2011 [5] (no peace agreements are recorded in PA-X for Myanmar prior to 2011).

Another trend that is evident on this plot in Figure 4 is the signing of the National Ceasefire Agreement (NCA) between the Myanmar government and Ethnic Armed Organisations (EAO) in 2015. This can be clearly seen as there are a range of datapoints representing occurrences of organisations at the same time (in the same agreement). The highest organisation occurring in this agreement is EAO which was an umbrella term for the 10 signatories of the NCA, including a range of the organisation mentioned in the texts - namely, the CNF, KNU, DKBA-5, ABSDF, Lahu Democratic Union, New Mon State Party, RCSS, Arakan Liberation Party and the Pa-O National Liberation Army. From this plot, this agreement seems to be somewhat successful, as there is not another peace agreement recorded in PA-X until August 2020 - the Union Accord Part III (Pyidaungsu Accord).

## 3.4 Analysis of person occurrences

**3.4.1 Persons and their affiliated groups.** The resulting dataset of people recognised within all the agreements from Myanmar consists of 104 instances of mentions of persons names within the text, of 78 different people. The total mentions for each person within the texts can be found in this dataset found here [https://github.com/niamhhenry/NER\\_Myanmar/blob/main/Datasets/People/persons\\_mentions\\_overall.csv](https://github.com/niamhhenry/NER_Myanmar/blob/main/Datasets/People/persons_mentions_overall.csv), and is illustrated in Figure 5 in a bar chart. From this plot we can see few people are mentioned more than once. Hence, we can assume that those mentioned more than once are key actors within the Myanmar peace process. For instance, the person with the most mentions throughout all the texts is U Aung Min, and from the legend it is known that he is a government representative. From a search of the name, it was found that he was the Minister of the Presidents office, and the Chairperson of Myanmar Peace Centre<sup>8</sup>, thus, it is expected that he was heavily involved within the peace agreements in Myanmar. From Figure 5 we can see that government representatives are those who are named most throughout all the texts. The next highest mentions are in the group with no known affiliation labels, nevertheless we can assume from some of their titles such as Col. Aung Thu. that they are involved with the Tatmadaw (which is linked to the government) or some other military group.

Viewing this data per affiliated organisation, rather than per person in Figure 6, we can see that the Government has the widest range of persons representing them with a total of 27 different representatives (two are coded as 'gov'). Additionally, they are mentioned significantly more than any other group. From the known values, this is then followed by the KNU/KNLA who have 6 known different representatives (5 for KNU/KNLA and one just KNU), and then the CNF with 4 known different representatives. This echoes the findings from the previous section, that the KNU/KNLA and the CNF were heavily involved in the peace agreements. Surprisingly, the SSPP only have one person coded as affiliated with the party, despite being mentioned numerous times as a whole. However, some of the unknown persons may be representatives of them, or they may have had representatives who were not named.

<sup>7</sup>Full dataset of this can be found here: [https://github.com/niamhhenry/NER\\_Myanmar/blob/main/Datasets/Organisations/orgs\\_identified\\_pax.csv](https://github.com/niamhhenry/NER_Myanmar/blob/main/Datasets/Organisations/orgs_identified_pax.csv)

<sup>8</sup>According to his Wikipedia page: [https://en.wikipedia.org/wiki/Aung\\_Min](https://en.wikipedia.org/wiki/Aung_Min)

Organisation	Total Mentions
Chin National Front (CNF)	64
State Level Joint Ceasefire Monitoring Committee (JMC-S)	57
Joint Monitoring Committee (JMC)	39
Union Level Joint Ceasefire Monitoring Committee (JMC-U)	39
Union Peace Dialogue Joint Committee (UPDJC)	33
Shan State Progress Party (SSPP)/Shan State Army (SSA)	32
Local Level Joint Ceasefire Monitoring Committee (JMC-L)	30
Karen National Union (KNU)/Karen National Liberation Army (KNLA)	28
Ethnic Armed Organizations (EAO)	28
Government of the Republic of the Union of Myanmar	22
Joint Ceasefire Monitoring Committee (JCMC)	22
Tatmadaw	20
Pa-O National Liberation Organisation (PNLO)	12
Hluttaw	11
Karenni National Progressive Party (KNPP)	6
United Nations Office on Drugs and Crime (UNODC)	6
National Socialist Council of Nagaland (NSCN)	6
Rakhine State Liberation Party (RSLP)	6
Union Peacemaking Working Committee (UPWC)	5
New Mon State Party (NMSP)	5
Union Election Commission (UEC)	4
Kachin Independence Organisation (KIO)	3
Arakan Army	2
United Nationalities Alliance (UNA)	1
National League Democracy (NLD)	1
United Nationalities Federal Council (UNFC)	1
United Nations	1
Food and Drug Administration (FDA)	1
Lahu Democratic Union	1

**Table 2: The number of mentions of particular entities recognised in all peace agreements, aggregated using manual dictionaries in Jupyter notebook.**

**3.4.2 Mentions of persons per agreement.** Disaggregating the data to investigate occurrences of people per each peace agreement shows interesting insights. Viewing the mentions of people over time in Figure 7, we can see the same theme around 2012 as we did viewing the occurrences of organisations over time in the previous section (in Figure 4), as many people from different organisations are mentioned in the peace agreements. This reinforces the trend of the number of agreements with different groups in 2012 as the Myanmar peace process was beginning to take form. Unlike Figure 4 of the organisations occurrences, there is not a peak of mentions of people in 2015 when the NCA was signed. This emphasises the use of the term EAO to represent all groups involved in the ceasefire agreement, rather than specific people holding discussions in place of their affiliated organisation.

Investigating the occurrence of persons names per stages of the peace process in Figure 8 overwhelmingly shows that people were involved to represent their affiliated organisations at the ceasefire stages of the peace process. Interestingly, only government representatives were named in the pre-negotiation phases, and only 6 different organisations had people recognised within the texts at the SubPar (substantive partial) stages. Comparing with the mentions

of organisations per stage in Figure 3, we can see that there are some similarities in the mentions of people and organisations, as the CNF have approximately half of their mentions in the ceasefire and SubPar stages in both plots. However, for government representatives, there is a much higher number of specific people mentioned in the ceasefire stages, rather than the organisational entity as a whole, where Figure 3 shows that the mentions are split almost half and half between the stages. Comparing the two plots shows that the overall organisations have more mentions than the individual people who represent them, which is expected as individuals are there to represent a group of people. Nevertheless, there are a lack of mentions of people's names in the pre-negotiation stages, compared to organisations. If one were to solely look at Figure 8, it could be (wrongly) assumed that the government did not negotiate with any other groups in the pre-negotiation stages of the peace process.

### 3.5 Comparison of results with other data

To further contextualise this actor data extracted from peace agreements in PA-X, it is useful to compare actors activity in violent and conflictual events within the same time period (2011-2021). The

external datasources Armed Conflict and Location Dataset (ACLED) [8] which records violent events, and the Uppsala Conflict Data Program (UCDP) [9] which records conflict events (using the definition of a conflict of more than 25 deaths per year) are utilised for this comparison. The data extracted from these sources are limited as neither are available for direct comparison using the timeframe 2011-2021, as the ACLED basic data export account is limited to extracting events from only the most recent 3 years<sup>9</sup>, and the UCDP does not have any conflict events coded for Myanmar after 2019. Thus, they can be used in harmony to fill in the gaps for which they are both limited to. Despite the time constraints of both these datasets, we can still see some interesting trends within the data relative to the actors involved in PA-X peace agreements.

**3.5.1 Aggregation of events by actors.** Aggregating all the events by the actors in both datasets, reinforces to us the involvement of specific actors in the conflicts in Myanmar. For instance, in Figure 9 a bar chart displays the total number of events actors were involved in within the UCDP dataset from 2011-2019. The Government of Myanmar (Burma) were clearly involved in more conflict events than any other actor, with the next highest actor involved being civilians. The government is always coded as *side\_a* in the UCDP GED due to all state-based violence coded as this according to the UCDP codebook [4]. Other actors that were recognised in PA-X agreements that appear in this dataset of conflict events include the KNU, RCSS, KNPP, DKBA-5 and ABSDF. Interestingly, despite being widely mentioned throughout the peace agreements, the CNF do not appear as actors in the UCDP dataset. Moving to more recent violent events in the ACLED dataset, for which events are aggregated per actor in the bar chart in Figure 10, we can see that there is a much larger range of actors present in this dataset. For the purposes of visualisations, actors who were involved in 5 events or less in this dataset were removed for this plot. Yet, the number of events for many actors are still difficult to see in this static image due to the large number of violent events involving protesters, unidentified armed groups and rioters<sup>10</sup>. Nevertheless, we can still see that the Government of Myanmar has been involved in 78 violent events over the last 3 years (government are coded as two separate actors in this dataset for the post-2016, and post-2021 governments). Additionally, we can see that the KNU/KNLA have been involved in 239, and the KNPP/KA involved in 63 violent events over the last three years, indicating a violation of the NCA signed in 2015. Notably, no other actors identified in the PA-X agreements are present as actors in violent events in ACLED for these years.

**3.5.2 Events by actors per year.** A further comparison of timing of these events, with the timing of the signing of peace agreements, is useful to indicate if such events have an impact on the creation of a peace agreement. The events recorded within the UCDP dataset can be seen per actor for the years 2011-2019 in Figure 11. From this plot three overall trends are evident in the timing of events;

- **2012:** The number of events drop for every actor in the dataset. This coincides with the large number of peace

agreements signed in 2012 that are recorded in PA-X, indicating that they directly resulted in a decrease of conflict.

- **2015:** There is a spike in conflict events for almost all actors in 2015, thus providing reason for the ceasefire agreement signed that year.
- **2018:** From 2018-19 there is a spike in the number of events for a number of actors (namely, the government, civilians, the PSLF (Palaung State Liberation Front) and a sharp increase in activity from the ULA (United League of Arakan)). This increase in events may give reason for the revision of the ceasefire agreement - the Union Accord Part III (Pyidaungsu Accord) in 2020 (as seen in Figure 4, there were no peace agreements in the 5 year period prior.)

Using the ACLED dataset to fill in the blanks of events that actors were involved in from 2018 onward, we can see in Figure 12 that there was also a spike in the number of events from 2018-19 for many actors. Particularly, the Military Forces of Myanmar were involved in 914 events in 2019 - an percentage change of 1,329% from 64 events in 2018. The most significant shift in the number of events in this plot is from 2020-2021, as the number of events involving protesters (red line in plot) significantly increases from 179 events in 2020 to 5,729 events (so far) in 2021<sup>11</sup> - a percentage change of 3,100%. Additionally, violent events involving civilians vastly increases in this time from 542 in 2020 to 4,006 in 2021. This growth in protest movement can be seen in the number of protests that were held in opposition to the coup d'état on 1 February 2021, when the members of the National League for Democracy (NLD) were deposed and detained by the Tatmadaw, and the handling of the third wave of the COVID-19 pandemic. The treatment of protesters and bystanders in these protests, where it is estimated by the Assistance Association for Political Prisoners that 1,325 were killed and 7,616 arrested, charged or sentenced due to the events post-coup [1], may also give reason for the severe increase of violent events that included civilians seen in Figure 12. Notably, the only other actor to increase as much in this time frame are the Unidentified Armed Group (purple line on plot), for which further information on this is unknown.

## 4 CONCLUSION

From the results of this case study of using automatic NER on the full-texts of peace agreements surrounding Myanmar that were extracted from PA-X database, it is evident that the SpaCy NER model provides to be extremely useful to quickly, and relatively accurately, identify key actors (namely, people and organisations) within the agreements. The manual dictionaries developed on Python using PA-X and other search tools enrich this data by mapping alias' and abbreviations of organisations/groups that are mentioned within the text to the one organisation, and provides datasets of people involved in the agreements, and their affiliated organisations. This provides extremely useful data<sup>12</sup> for future use into the investigation of actors in peace agreements in Myanmar. This data can be mapped in a network of actors to illustrate, and provide grounds for further research into actors involvement, and influences on peace agreements and peace processes/transitions. Comparing the

<sup>9</sup>This is using my personal ACLED key - I will apply for another academic account to gain access to more data for the future.

<sup>10</sup>Please view the interactive versions of these plots here: [https://github.com/niamhhenry/NER\\_Myanmar/blob/main/visualisations.html](https://github.com/niamhhenry/NER_Myanmar/blob/main/visualisations.html)

<sup>11</sup>ACLED data is up until 2nd December 2021

<sup>12</sup>Publicly available on GitHub here:

actors found within the peace agreements, to actors named within conflict/violent event databases shows that some actors that are heavily present in peace agreements are also heavily involved in these events. Additionally, the timings of peaks in these events also correlate with the timings of peace agreements which provides context and perhaps reason as to why the agreements were developed. Further research into the specific actors involved in these peace agreements, and further inquiry into the events they are involved in will provide to be extremely valuable to studies on peace and conflict resolution. Possible steps to further investigate the results from this case study, and to extend these methods to other peace processes are outlined in Section 5 below.

## 5 FUTURE WORK

- Mapping within a network the actors involved in all peace agreements in Myanmar, and the number of times they are mentioned.
- Linking the content of the peace agreements (using data in PA-X) to investigate actors influences on the contents of peace agreements (for a range of themes coded in PA-X).
- Further investigating the link between actors in PA-X and actors in violent/conflict event databases, particularly with regard to timings of events, ceasefire violations, and the number of fatalities as a result of the events, and how these impact the peace agreements/processes.
- Extending the NER methods to other peace processes in PA-X. This would require extracting the full texts of agreements to the format required to be read by the model (in word documents).
- Improving the results of the NER model to limit the manual cross-checking required, and size of the manual dictionaries required to understand the results. This may be through training the spaCy model on a subset of PA-X database that is manually labelled for a gold-standard (however, this may be an issue as actors are country-specific).

Named Entity Recognition in Peace Agreements

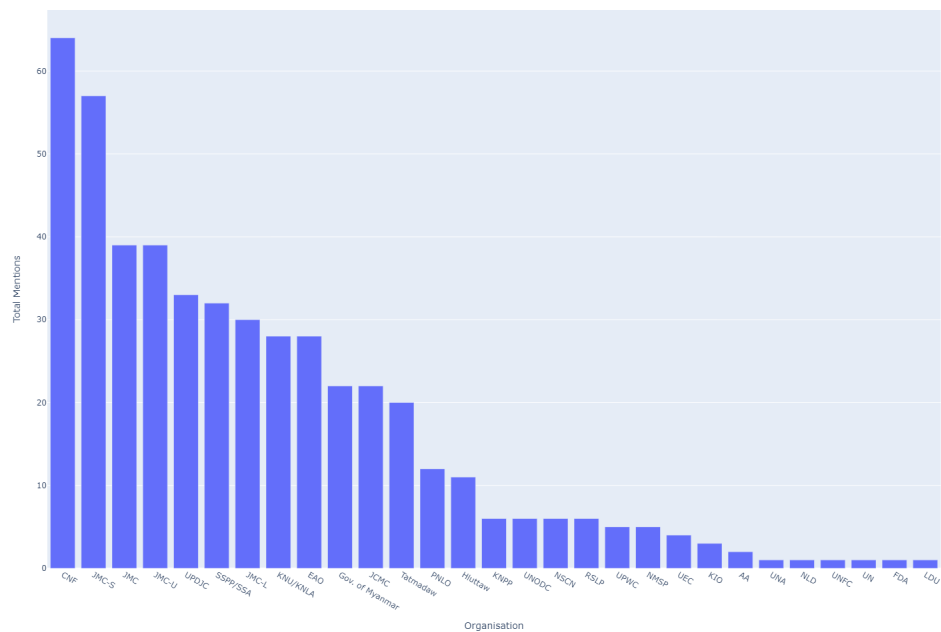


Figure 1: Bar chart of frequency of mentions of each recognised organisation in all 36 peace agreements from Myanmar

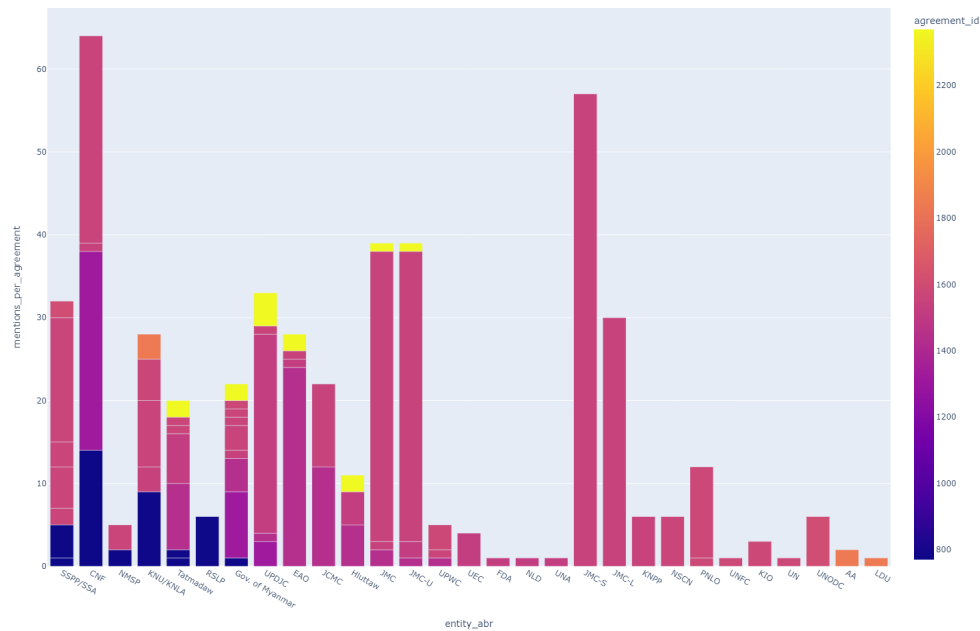
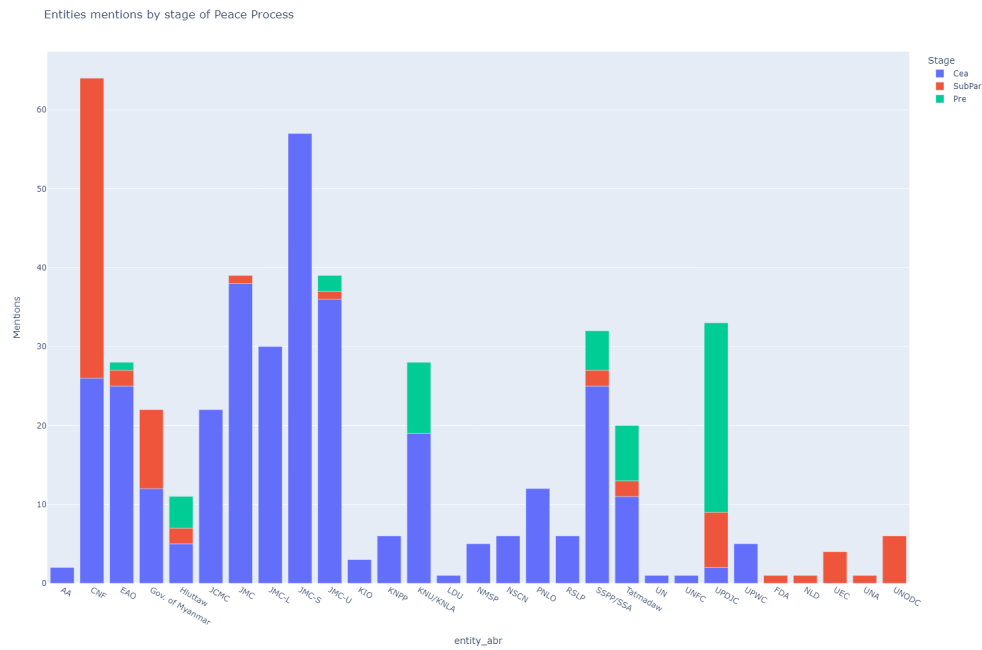


Figure 2: Bar chart of frequency of mentions of each recognised organisation, coloured per peace agreement from Myanmar.



**Figure 3: Bar chart of occurrences of organisations/groups in Myanmar peace agreements viewed by the stage of the peace process.**



**Figure 4: Scatter plot of organisations in peace agreements over time (2011-2021)**



Named Entity Recognition in Peace Agreements

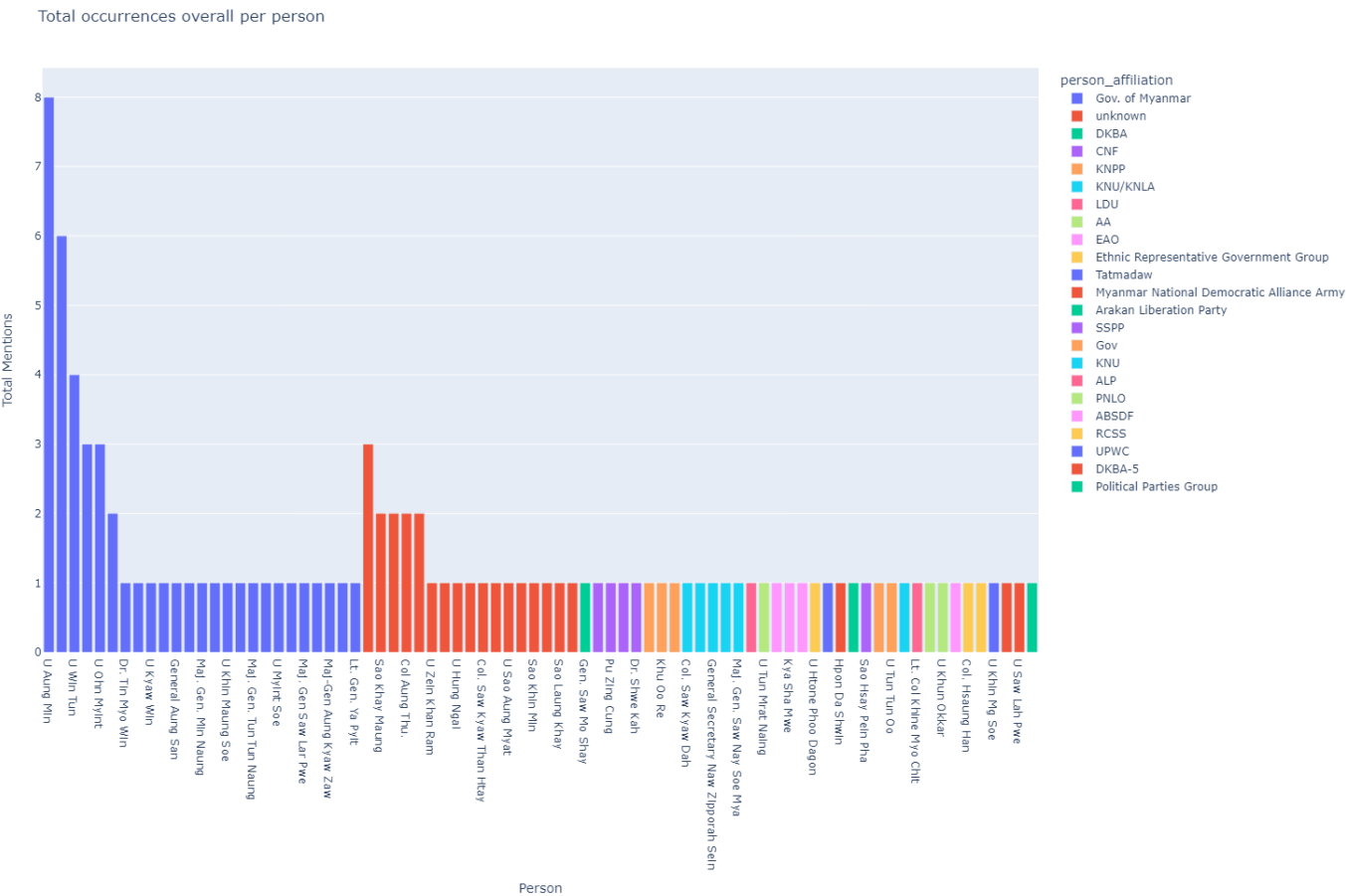
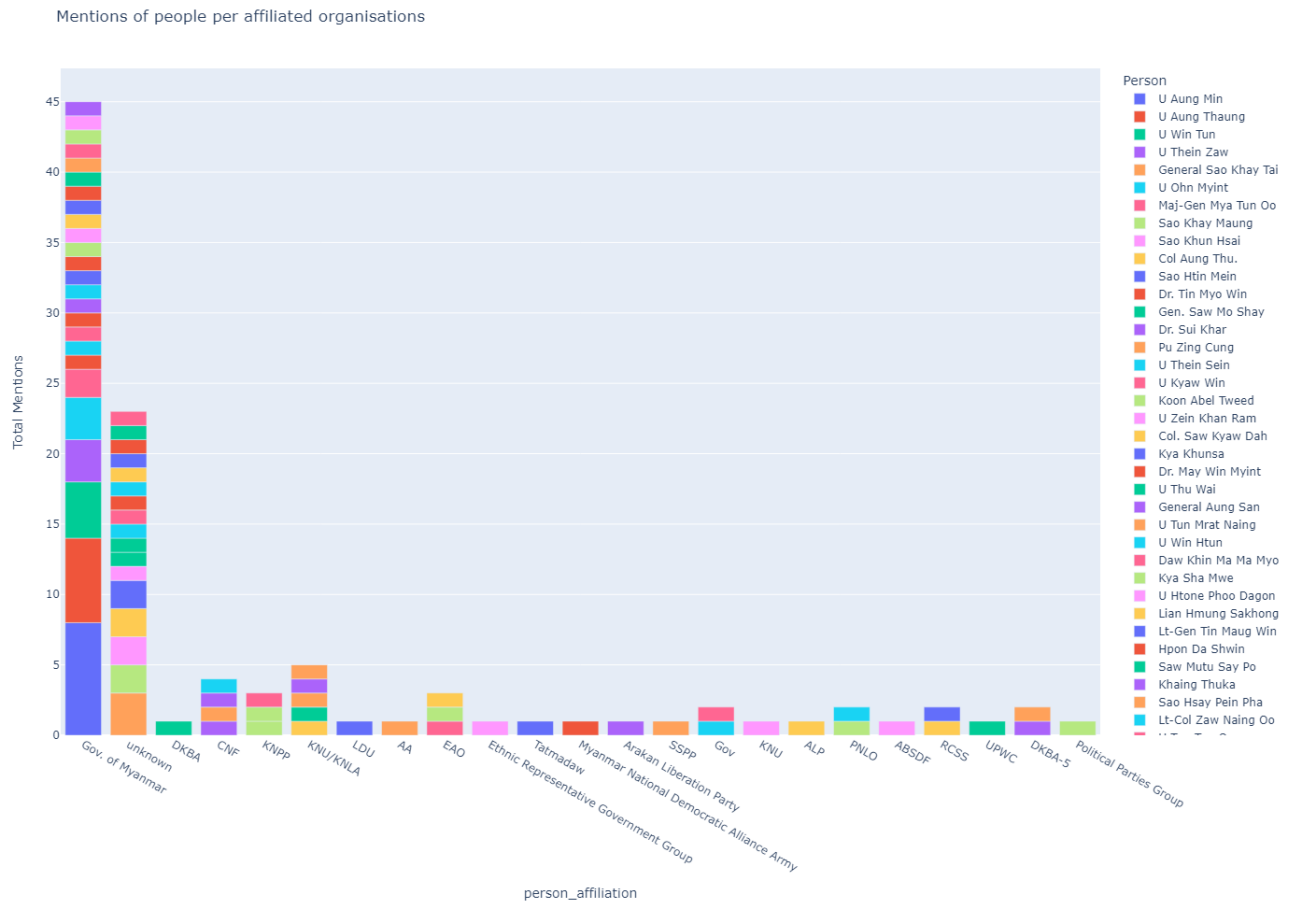


Figure 5: Bar chart of the total occurrences of each person recognised within the texts, coloured by their affiliated organisation.



**Figure 6: Bar chart of the total occurrences of affiliated groups of the persons recognised within the texts, coloured by the persons names.**

Named Entity Recognition in Peace Agreements

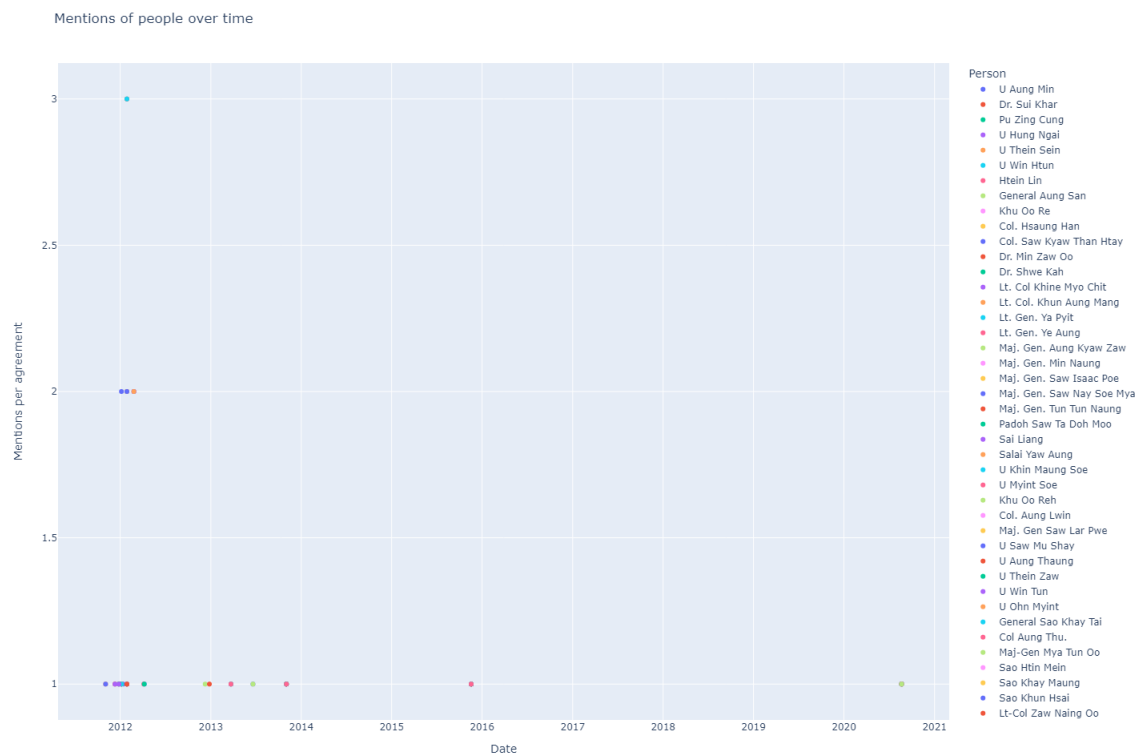


Figure 7: Scatter plot of persons recognised in peace agreements over time (2011-2021)

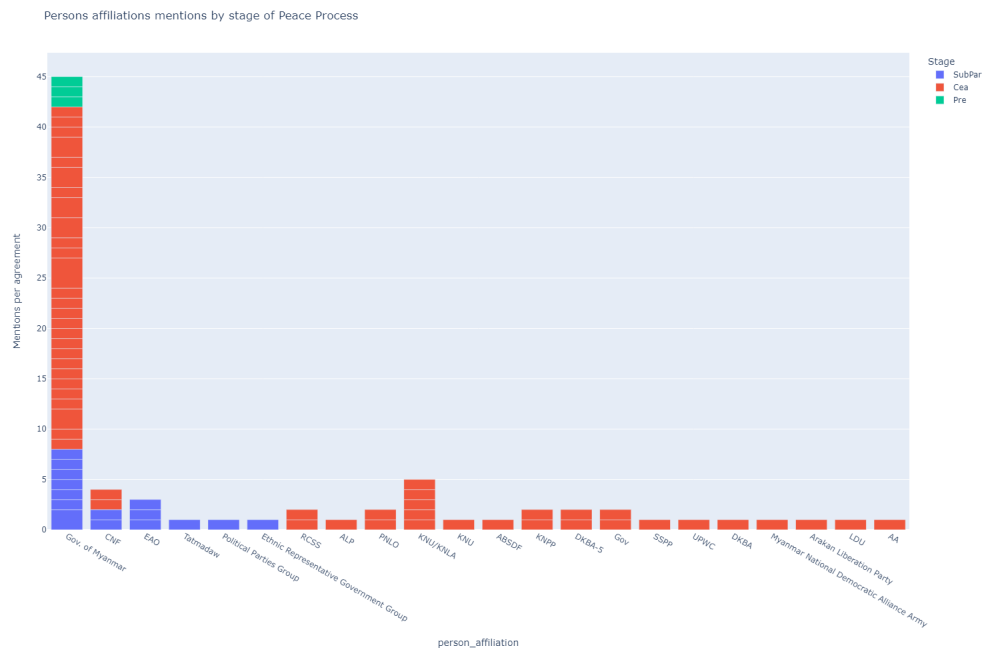
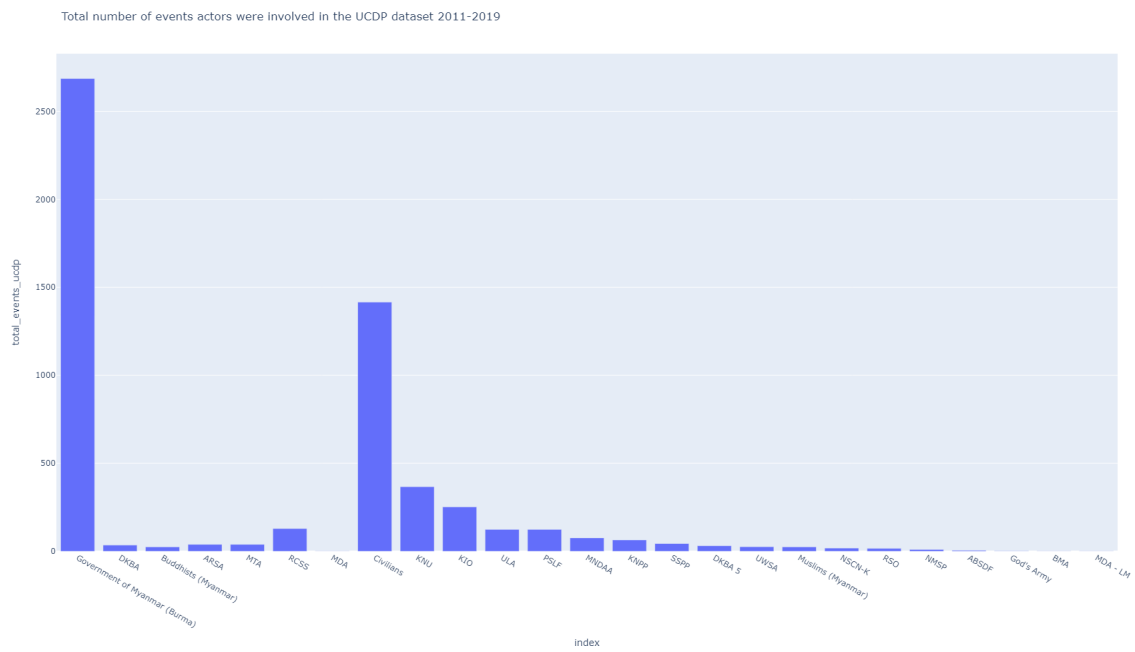


Figure 8: Bar chart of the total occurrences of affiliated groups of the persons recognised within the texts, coloured by the stage of the peace process the agreement took place.



**Figure 9: Bar chart of the total number of conflict events named actors were involved in from 2011-2019, as recorded by UCDP [9].**

Named Entity Recognition in Peace Agreements

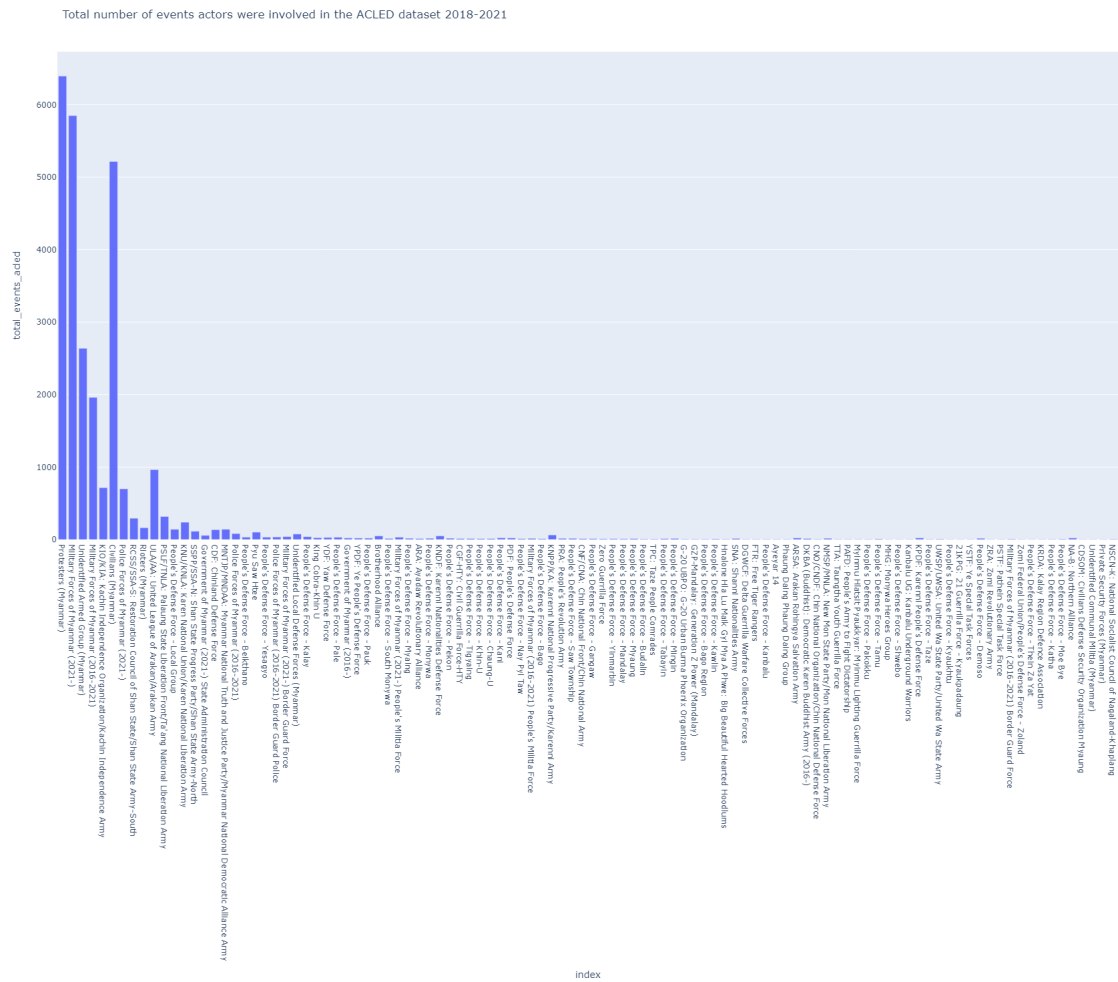
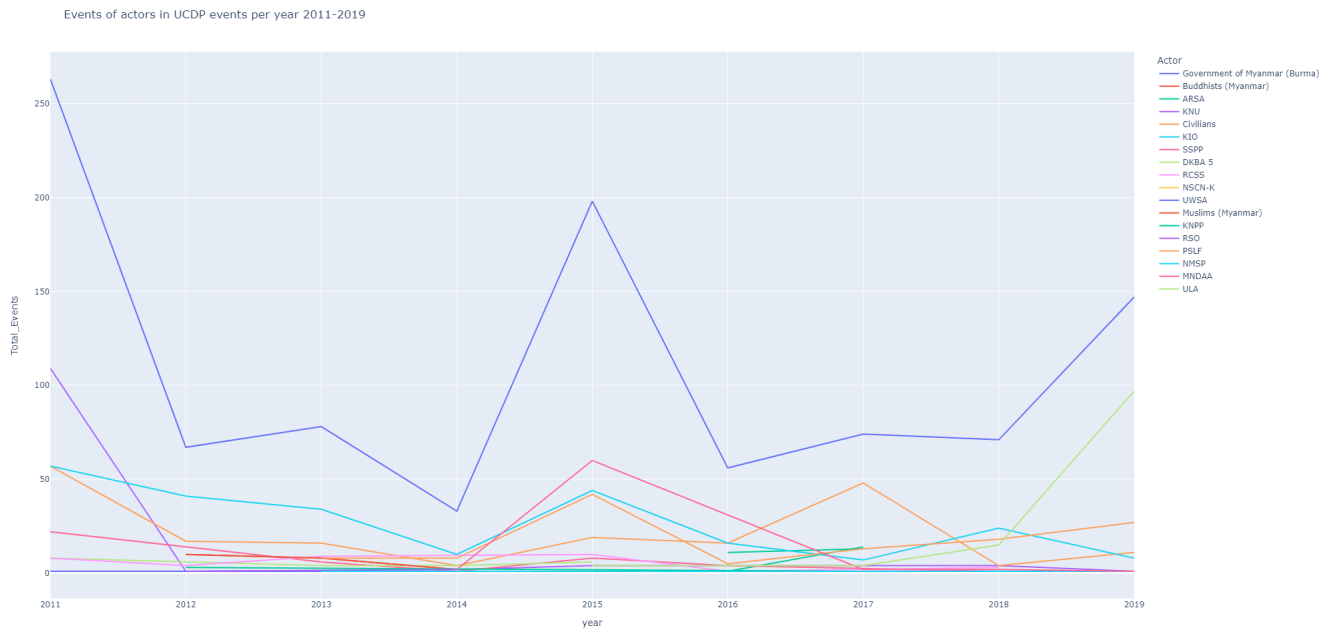
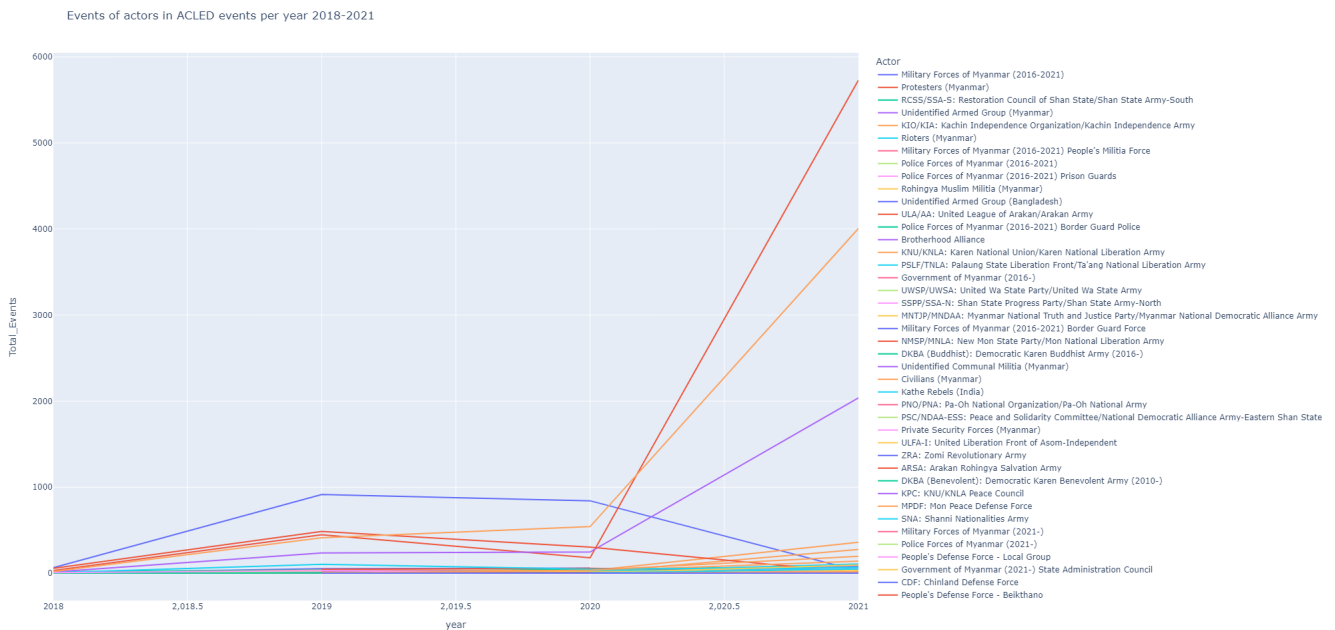


Figure 10: Bar chart of the total number of violent events named actors were involved in from 2018-2021, as recorded by ACLED [8].



**Figure 11:** Line graph of the total number of conflict events named actors were involved in per year from 2011-2019, as recorded by UCDP [9].



**Figure 12:** Line graph of the total number of conflict events named actors were involved in per year from 2018-2021, as recorded by ACLED [8].

## REFERENCES

- [1] [n.d.]. Assistance association for political prisoners. <https://aappb.org/>
- [2] [n.d.]. Peace Agreements Database. <https://www.peaceagreements.org/search>
- [3] Christine Bell and Sanja Badanjak. 2019. Introducing PA-X: A new peace agreement database and dataset. *Journal of Peace Research* 56, 3 (2019), 452–466.
- [4] Mihai Croicu and Ralph Sundberg. 2015. UCDP georeferenced event dataset codebook version 4.0. *Journal of Peace Research* 50, 4 (2015), 523–532.
- [5] Rachel Gasser, Julia Palmiano Federer, and Danielle Furgler. 2016. Perspectives on the Myanmar Peace Process 2011-2015. (2016).
- [6] Matthew Honnibal and Ines Montani. 2017. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. (2017). To appear.
- [7] Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D Manning. 2020. Stanza: A Python natural language processing toolkit for many human languages. *arXiv preprint arXiv:2003.07082* (2020).
- [8] Clionadh Raleigh, Andrew Linke, Håvard Hegre, and Joakim Karlsen. 2010. Introducing ACLED: an armed conflict location and event dataset: special data feature. *Journal of peace research* 47, 5 (2010), 651–660.
- [9] Ralph Sundberg and Erik Melander. 2013. Introducing the UCDP georeferenced event dataset. *Journal of Peace Research* 50, 4 (2013), 523–532.