An LDA Topic Analysis: How the WHO and Trump framed the COVID-19 Pandemic

Sofia Lugli

Abstract

This project presents an analysis of the discourse around Covid-19, based on a corpus of speeches given by the two relevant public figures: the WHO Director-General Tedros Adhanom Ghebreyesus and former President of the United States of America Donald Trump, during the first two months of the pandemic (March-April 2020). Using topic modelling, it was first analysed the topics around which the discourse about Covid-19 can be classified. Then, it was investigated to what extent the discourse was framed figuratively, in terms of a WAR and JOURNEY metaphors, which have been pervasive in healthcare discourse when talking about diseases and viruses. Finally, this project explored the connection between these frames and topics, showing a relative tendency to employ the WAR frame to discuss safety measures and treatment, while the JOURNEY frame to discuss topics related to the population and patients. However, it is important to note that the boundaries between these frames and topics are fuzzy and they often overlap and intersect with one another.

1 Theoretical Framework

Conceptual metaphors are pervasive in everyday language, playing a significant role in communication, cognition and decision making. According to the Conceptual Metaphor Theory (CMT) (Lakoff & Johnson, 1980), metaphors are defined as systematic sets of mappings across conceptual domains, whereby a target domain, which is more abstract and difficult to understand, is partly structured in terms of a different source domain, which usually defines more concrete and common concepts. However, the set of correspondences between the two domains is not neutral since the source domain inevitably foregrounds some aspects and downgrades others of the target domain (Kövecses 2017). When particular uses of metaphor become the dominant way of talking about a particular topic, they may be extremely difficult to perceive and challenge, since they come to represent the 'commonsense'; however, always using the same conventional metaphor limits the way we think of a concept and the way we act in the world (Semino 2008). The framing power of metaphor is especially important in areas such as healthcare, where the choice of different frames of illness can impact patients' well-being (Semino et al., 2016). Within studies of health communication, the two frequently used metaphors to describe diseases are WAR and JOURNEY metaphors (Gibbs 2017). The usefulness of war metaphors derives from the transformation of war making into an occasion for mass ideological mobilization (Sontag 1989), whose intents can be reached after defeating an "enemy". Given the life-threatening nature of war, war metaphors connote an unmistakable seriousness of purpose in a simple and comprehensible manner, which allows to draw immediate attention to the issue in question. However, this metaphor has been often considered as inappropriate and rejected by patients and health workers, since it may lead to a sense of helplessness and guilt vis-à-vis the disease. On the other hand, JOURNEY metaphors are usually more accepted by patients, because they refer to progress, goals and moving forward, which tends to corresponds to positive change, development and success. Within this metaphor, it is possible to reach the objective by changing path and direction, hence, defeat is not necessarily implied. Nonetheless, the destination is not always easy to reach, there may be obstacles along the way and one may feel discouraged towards an uncertain outcome.

Recent corpus studies have estimated that the proportion of words used metaphorically ranges from 5% to 20% (Steen et al., 2010). Given the prevalence and importance of metaphoric language, metaphor study has become an important research topic also in the field of natural language processing. However, the research of modelling metaphors is not an easy task. According to Tong et al. (2021), the research on metaphors could be divided into three sub-fields: metaphor identification, metaphor interpretation and metaphor generation. This project focus on metaphor detection. Within this field, previous work employed extensive manually-annotated linguistic resources (Gedigian et al., 2006; Turney et al., 2011; Tsvetkov et al., 2013) and corpus-based approaches (Shutova et al., 2013). More recently, Shutova et al. (2016) proposed a novel approach integrating meaning representations learned from linguistic and visual data. The present study relies on the LDA topic modelling technique. Heintz et al. (2013), Huang (2013) and Strzalkowski et al. (2013) focused on modeling topical structure of text to identify metaphors. Their main hypothesis was that metaphorical language (coming from a different domain) would represent atypical vocabulary within the topical structure of the text.

The present project follows the procedure of the work by Wicke and Bolognesi (2020), which, through LDA, analysed the discourse around Covid-19 on a corpus of tweets during March and April 2020. They focused on the metaphoric framing effect and the persuasive power of metaphors (Lakoff, 1991; Lakoff and Wehling, 2012). With the aim to reduce relying in manual metaphor identification methods, such as MIP and MIPVU, they proposed an approach that could be applied to large corpora. Their results show the way in which a wide selection of American-English speakers conceptualized the pandemic on Twitter, revealing a pervasiveness of the WAR frame in shaping public discourse.

2 Research Questions and Study Design

Referring to the speeches delivered by the WHO Director-General and the USA President in the first two months of the Covid-19 pandemic, this project addressed the following research questions:

- What type of topics are discussed in relation to Covid-19 pandemic?
- To what extent are the WAR and JOURNEY figurative frames used to talk about Covid-19? Which lexical units are used within these metaphorical frames and which are not?
- Are specific frames used to discuss specific topics around Covid-19?

To address these questions, the first step involved an investigation of the topics present in the corpus employing the topic modelling technique of Latent Dirichlet Allocation (LDA). Consequently, it was explored the occurrences and usage of war- and journey-related terms. Finally, it was conducted an analysis to examine whether specific frames are employed to address particular topics around Covid-19 (among the topics identified in the first part).

2.1 Corpus

The corpus of the present study consists of 83 public speeches delivered during March and April 2020, the first two months of the Covid-19 pandemic. The corpus is divided as follows: (i) 38 speeches given by the WHO Director-General Tedros Adhanom Ghebreyesus (from March 1st, 2020 to April 29th, 2020); and (ii) 45 public speeches given by the former President of the United States of America Donald Trump (from March 2nd, 2020 to April 30th, 2020). The transcriptions of each speech are freely available respectively on the WHO website[1] and on the American speech-to-text company Rev website. [2] The speeches consist of either opening remarks or complete speeches. Transcripts of interviews, where questions and conversations with others could potentially influence the speaker's speech, were excluded. Therefore, only speeches without any interaction among multiple speakers were selected. The prepocessing phase of the corpus encompassed the following steps: converting the speeches in a list of tokens; removing stopwords and tokens with less than 3 characters; excluding Covid-19 words ("covid-19", "covid", "ncov", "coronavirus", "corona", "ncov19", "virus") from the topic modelling because they do not add information about the topics themselves; lemmatizing the corpus for the LDA; turning them into a bag-of-words for the LDA processing.

3 Identifying Topics through LDA

In order to identify the probable topics from the corpus, the Latent Dirichlet Allocation algorithm (LDA) (Blei et al., 2003) was employed, in particular the Gensim LDA-Multicore algorithm. LDA is a

[1] https://www.who.int/news-room/speeches
[2] https://www.rev.com/blog/transcript-category/donald-trump-transcripts

three-level hierarchical Bayesian model. The basic idea is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words (Blei et al., 2003). As an unsupervised learner, LDA needs to be given the number of topics (k) into which it will divide the data. Looking at the literature, there is not a standardized manner to decide the number of topics. Some suggest (Syed et al., 2017; Hasan et al., 2021) to evaluate the different models according to the coherence score, which measures the degree of semantic similarity between high scoring words in each topic. Working with the gensim library, I implemented the C_v which ranges from 0 (complete incoherence) to 1 (complete coherence). I tested the corpus changing the k parameter, ranging from 2 to 10 topics, after measuring the coherence score, I chose the model with the highest coherence score. Given that the number of documents of the corpus is small for the method, the coherence score does not go above 0.5. For each model, the number of passes was 10 and the words which occurred in less than 5 documents or more than 50% of the documents were filtered out, in order to exclude the most rare and common words. Moreover, due to the randomness involved in the training and inference process of the LDA model, training a new model with the same parameters will always result in slightly different topic distributions, to avoid that and ensure replicability, a random state of 100 was set.

3.1 WHO Corpus: LDA topic modelling

Within the LDA models trained on the WHO corpus, the evaluation of the coherence score showed that the model with k=5 had the highest coherence score of 0,34, as shown in Figure 1.

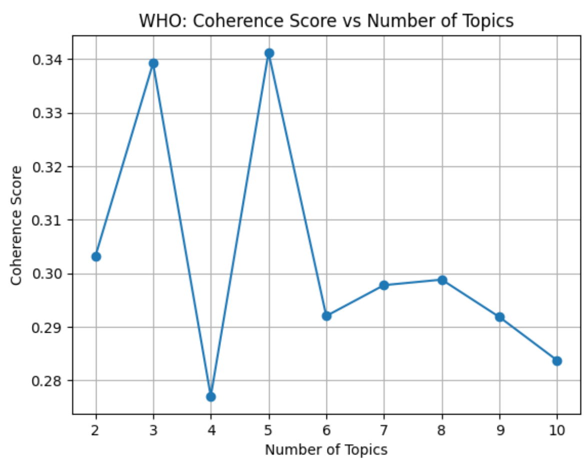

FIGURE 1: Correlation between Coherence score and number of topics of LDA models trained on the WHO corpus.

The weights (importance) and words for each of the five topics allocated by the LDA model were the following:
Topic 1: 0.017 epidemic + 0.010 treatment + 0.009 approach + 0.009 testing + 0.008 prevent + 0.007 action + 0.007 plan + 0.007 facility + 0.007 infection + 0.007 cluster
Topic 2: 0.013 state + 0.012 emergency + 0.012 plan + 0.012 information + 0.011 january + 0.011 member + 0.010 preparedness + 0.010 united + 0.009 nation + 0.009 fund
Topic 3: 0.015 mask + 0.014 medical + 0.008 place + 0.008 population + 0.008 treat + 0.008 globally + 0.008 said + 0.008 region + 0.007 economic + 0.007 thousand
Topic 4: 0.013 service + 0.010 fund + 0.009 economic + 0.009 year + 0.008 foundation + 0.008 child + 0.008 information + 0.008 plan + 0.008 family + 0.007 crisis
Topic 5: 0.016 food + 0.012 evidence + 0.011 trial + 0.009 place + 0.009 important + 0.008 effective + 0.008 restriction + 0.008 tool + 0.008 drug + 0.008 mask

The LDA algorithm does not provide labels for the identified topics. They are also visualized in Fig.2, in which greater size of the word indicates greater occurrence in the corpus.

FIGURE 2: Word clouds for the 5 WHO topics

Among the 38 documents of the WHO corpus, the distribution of the dominant topic for each document was the following: topic 1 was the dominant topic for 34.21% of the corpus (13 documents); topic 2 for the 15.79% (6 documents); topic 3 for the 15.79% (6); topic 4 for the 18.42% (7); topic 5 for the 15.79% (6).

The ones mentioned above represent the dominant topic for each document, however, it is important to note that each document may contain multiple topics. Fig. 3 illustrates the topic distribution across documents. For instance, topic distribution for document 1 is (1, 0.99335665), where the first digit refers to the topic number, and the percentage is the distribution of such in that document, meaning that topic 1 is 99%; whereas the topic distribution for document 3 is (1, 0.5226145), (2, 0.3136306), (3, 0.15767215), as such multiple topics have been identified in a single document.

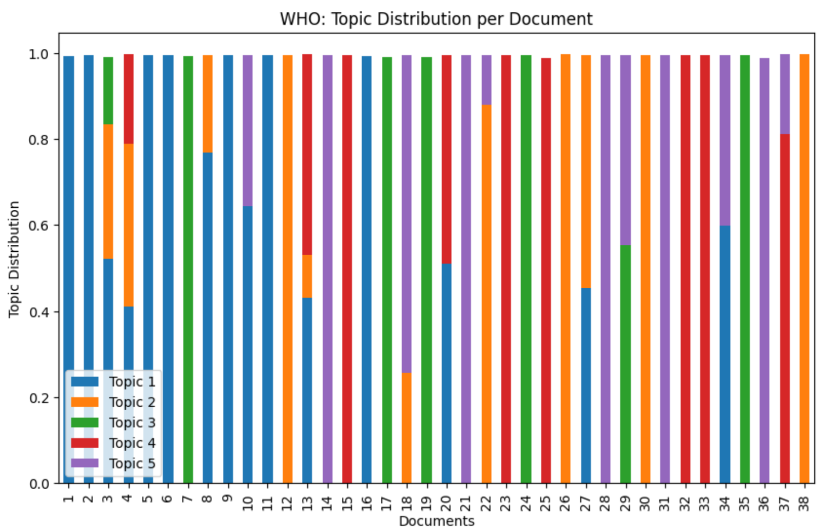

FIGURE 3: WHO corpus: topics allocation to documents

3.2 TRUMP Corpus: LDA topic modelling

The same procedure conducted on the WHO corpus, has been carried out also on the TRUMP corpus. In this case, the LDA model with the highest highest coherence score was the model with two topics k=2, whose coherence score was equal to 0,26 as shown in Fig. 4.

The weights (importance) and words for the two topics allocated by the LDA model where the following:
Topic 1: 0.009 test + 0.006 year + 0.006 capacity + 0.005 drug + 0.005 economy + 0.005 talking + 0.004 fda + 0.004 approved + 0.004 together + 0.004 bed
Topic 2: 0.009 mask" + 0.005 hit + 0.005 patient + 0.005 test + 0.005 bed + 0.004 military + 0.004 happy + 0.004 jersey + 0.004 fema + 0.004 large

Fig. 5 provides a word cloud visualization of the topics. Among the 45 documents of the TRUMP corpus, the distribution of the dominant topic for each document was the following: topic 1 was the dominant topic for 46.67% of the corpus (21 documents); topic 2 for the 53.33% (24 documents);

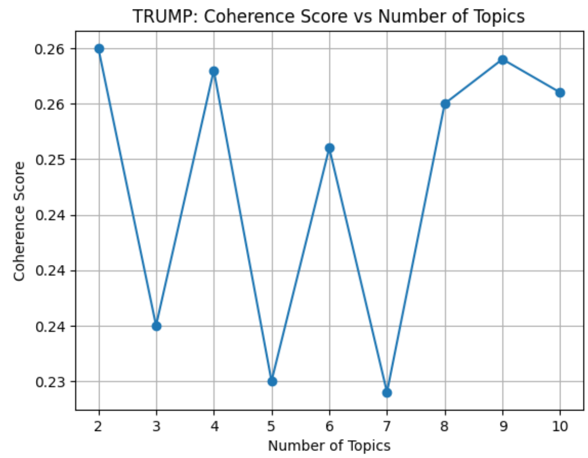

FIGURE 4: Correlation between Coherence score and number of topics of LDA models trained on the TRUMP corpus.

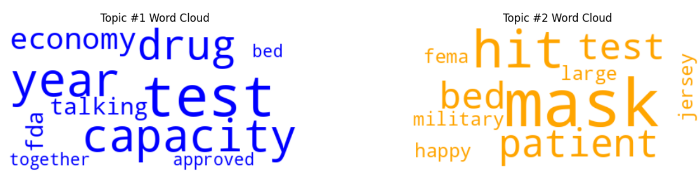

FIGURE 5: Word clouds for the 2 TRUMP topics

Also in this corpus, some document contained more than one topic. Fig.6 shows for instance that the topic distribution for document 1 is (1, 0.990969); whereas the topic distribution of document 3 is (1, 0.9348291), (2, 0.06517091).

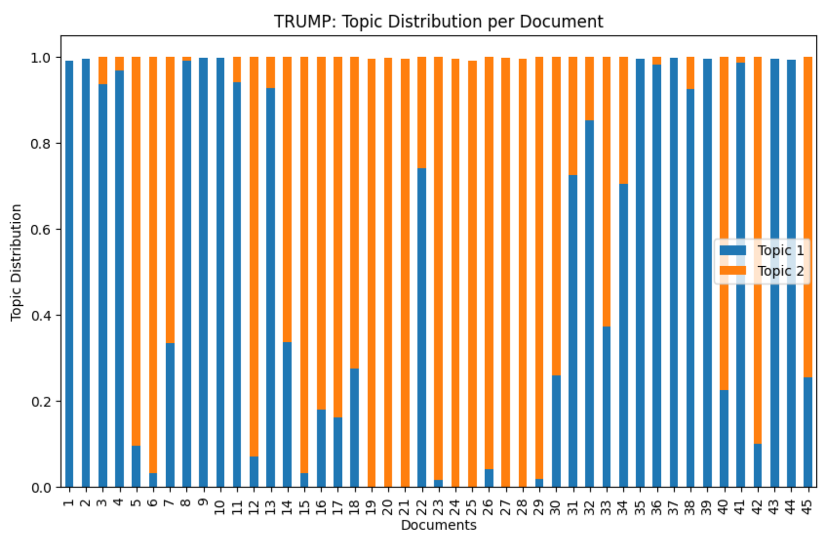

FIGURE 6: TRUMP corpus: topics allocation to documents

Finally, in the attached files, it is provided an interactive online tool to explore the results of the LDA models for the 5 LDA topics of the WHO corpus and for the 2 LDA topics of the TRUMP corpus. These rendering have been produced using pyLDAvis.

4 WAR and JOURNEY Frames

In this section, it is investigated to what extent WHO Director-General and President Trump use the WAR and JOURNEY frames to talk about Covid-19 in the speeches given during the first two months of the pandemic (March - April 2020). The assumption made in this project is that war- and journey-related terms would be used metaphorically within speeches addressing Covid-19. It is important to acknowledge that some of these terms may have been used literally, representing a limitation to this approach.

To explore the lexical units associated with the WAR frame, I considered the set of words collected by Wicke and Bolognesi (2020). To collect the war terms, they used two web-services *relatedwords.org* and *ConceptNet*; moreover they exploited the MetaNet repository of conceptual metaphors and frames. The result was 91 lexical units related to the WAR frame:

WAR (91): allied, allies, armed, armies, army, attack, attacks, battle, battlefield, battleground, battles, belligerent, bloodshed, bomb, captured, casualties, combat, combatant, combative, conflict, conquer, conquering, conquest, crusade, defeat, defend, defenses, destruction, disarmament, enemies, enemy, escalation, fight, fighter, fighting, foe, fortify, fought, grenade, guerrilla, gunfight, holocaust, homeland, hostilities, hostility, insurgency, invaded, invader, invaders, invasion, liberation, military, peace, peacetime, raider, rebellion, resist, resistance, riot, siege, soldier, soldiers, struggle, tank, threat, treaty, trench, trenches, troops, uprising, victory, violence, war, warfare, warrior, wars, wartime, warzone, weapon, alliance, ally, arsenal, blitzkrieg, bombard, front, line, minefield, troop, vanquish, vanquishment.

To collect the lexical units related to the target word of JOURNEY, the same resources were used as proposed by Wicke and Bolognesi (2020), excluding the MetaNet repository, in which the JOURNEY frame page has not been made available to the public. The journey related terms are the following:

JOURNEY (74): adventure, arriving, arrival, arrive, approaching, byway, climb, coming, commute, cruise, crossroad, detour, destination, direction, digression, distance, embark, excursion, expedition, forward, heading, itinerary, jaunt, journey, journeying, junket, map, meandering, motion, move, moving, movement, odyssey, passage, path, paths, pilgrimage, process, pursuing, pursuit, quest, reach, reaching, return, returning, ride, roadmap, route, sail, shift, step, steps, stretch, through, tour, touring, tourism, tourist, tours, transit, travel, travelling, travels, trek, trip, trudge, turn, voyage, walk, walking, way.

In the following subsections it is presented the analysis of occurrences in the corpus of the lexical units related to the WAR and JOURNEY frames across the two corpora.

4.1 WHO Corpus: WAR & JOURNEY frames

In the WHO corpus, only 19 words of the 91 war related terms were identified. The 19 war terms found are hereby reported with relative percentage to all war terms and number of occurrences:

WAR (19): fight (28.46%, 37), threat (18.46%, 24), enemy (7.69%, 10), violence (7.69%, 10), defeat (6.92%, 9), front (4.62%, 6), fighting (4.62%, 6), line (3.85%, 5), attack (3.85%, 5), escalation (3.08%, 4), struggle (3.08%, 4), conflict (2.31%, 3), battle (0.77%, 1), captured (0.77%, 1), war (0.77%, 1), combat (0.77%, 1), attacks (0.77%, 1), fought (0.77%, 1), alliance (0.77%, 1).

Of the 38 speeches, only 5 did not contain any war related terms from the list.

Among the WHO corpus, only 24 JOURNEY related terms were identified, among the list of 74 terms. The terms found are hereby reported with relative percentage to all journey terms and number of occurrences:

JOURNEY (24):way (24.67%, 38), reach (9.74%, 15), coming (9.09%, 14), forward (7.79%, 12), movement (7.79%, 12), move (5.19%, 8), travel (4.55%, 7), turn (3.89%, 6), process (3.89%, 6), reaching (3.24%, 5), moving (3.24%, 5), step (2.59%, 4), distance (2.59%, 4), arrive (1.94%, 3), roadmap (1.94%, 3), shift (1.29%, 2), walk (1.29%, 2), stage (1.29%, 2), path (0.65%, 1), climb (0.65%, 1), approaching (0.65%, 1), ride (0.65%, 1), sail (0.65%, 1), return (0.65%, 1).

Only one document did not present any journey related terms, among the 38 analysed.

Fig.7 illustrates the comparison between the war- and journey-related terms found in the WHO corpus. The sum percentage of war-related terms across WHO speeches is approximately 24.11%; the sum of all the journey-related percentages is 26.76%.

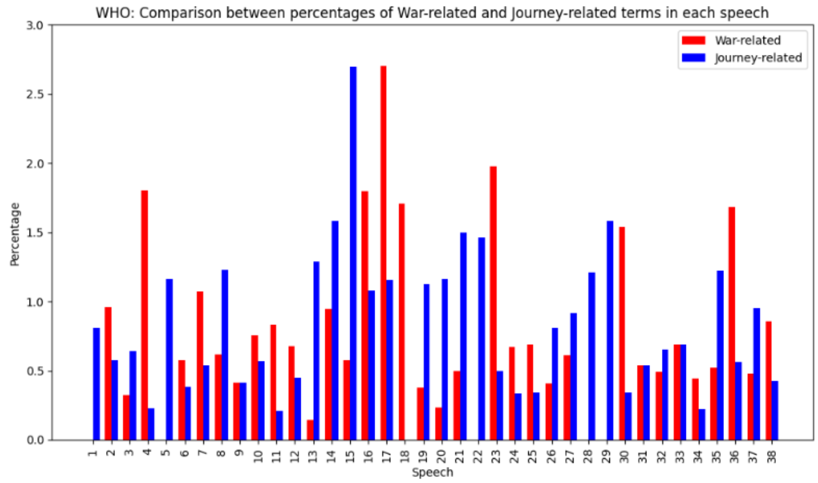

FIGURE 7: Comparison between the war- and journey-related terms in the WHO corpus.

To answer to the third research question of whether specific frames are used to discuss specific topics around Covid-19, with a cross examination of the percentage of war and journey terms found in each speech and the dominant topic for each speech, I concluded that the speeches, which had topic 1 as the dominant topic, had the highest percentage of both war and journey related terms. The sum of the percentages of war related terms for the speeches whose dominant topic is topic 1 is 8.54%; for topic 2 is 4.09%; for topic 3 is 5.34%; for topic 4 is 5.05%; and for topic 5 is 5.39%. Whereas, the sum of the percentages of journey related terms for the speeches whose dominant topic is topic 1 is 8.79%; for topic 2 is 4.41%; for topic 3 is 6.96%; for topic 4 is 7.53%; for topic 5 is 5.39%. Fig. 8 shows the topic distribution of war and journey-related speeches.

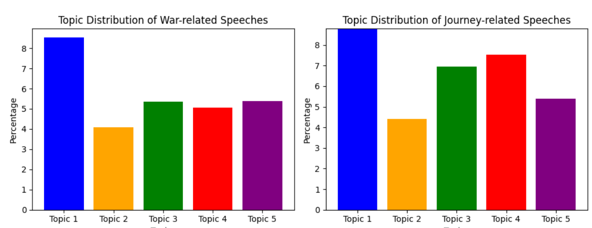

FIGURE 8: The speeches with more war and journey related terms percentage have topic 1 as their dominant topic.

4.2 TRUMP Corpus: WAR & JOURNEY frames

In the TRUMP corpus, only the following 29 lexical units of the war frame were present. Of the 45 speeches, only 4 did not contain war terms from the list.

WAR (29): war (17.69%, 66), military (11.53%, 43), enemy (9.38%,35), army(9.12%, 34), battle (6.97%, 26), fight (6.70%, 25), victory (5.63%,21), defeat(5.36%,20), line(4.29%, 16), front (4.02%,15), fighting(2.95%, 11), vanquish (2.14%, 8), threat (1.88%, 7), soldiers (1.6%, 6), weapon (1.34%, 5), combat(1.34%, 5), homeland (1.34%,5), attack (1.34%, 5), allies (1.07%, 4), struggles (0.80%, 3), troops (0.80%,3), destruction (0.54%, 2), armed (0.54%, 2), captured (0.27%, 1), arsenal (0.27%, 1), defend (0.27%, 1), conquer (0.27%, 1), fought (0.27%, 1), conflict (0.27%, 1).

Concerning the JOURNEY frame, 33 journey-related terms were identified; each speech contained at least one journey related terms from the list.

JOURNEY(33): way (24.59%, 105), coming (14.75%, 63), travel (7.96%, 34), move (7.96%, 34), process (7.03%, 30), moving (6.56%, 28), forward (4.68%, 20), steps (4.22%, 18), step (2.81%, 12), turn (2.34%, 10), return (2.11%, 9), direction (1.64%, 7), distance (1.64%, 7), arriving (1.64%, 7), reach (1.17%, 5), arrive (1.17%, 5), heading (0.94%, 4), walk (0.94%, 4), walking (0.70%, 3), stage (0.70%, 3), cruise (0.70%, 3), route (0.70%, 3), shift (0.47%, 2), quest (0.47%, 2), destination (0.23%, 1), pursuing (0.23%, 1), movement (0.23%, 1), trip (0.23%, 1), path (0.23%, 1), embark (0.23%, 1), tour (0.23%, 1), returning (0.23%, 1), passage (0.23%, 1).

In Fig. 9, it is shown the comparison between war- and journey-related terms. The sum percentage of war-related terms across TRUMP speeches is approximately 35.98%; whereas, the sum percentages of journey-related terms is 47.41%.

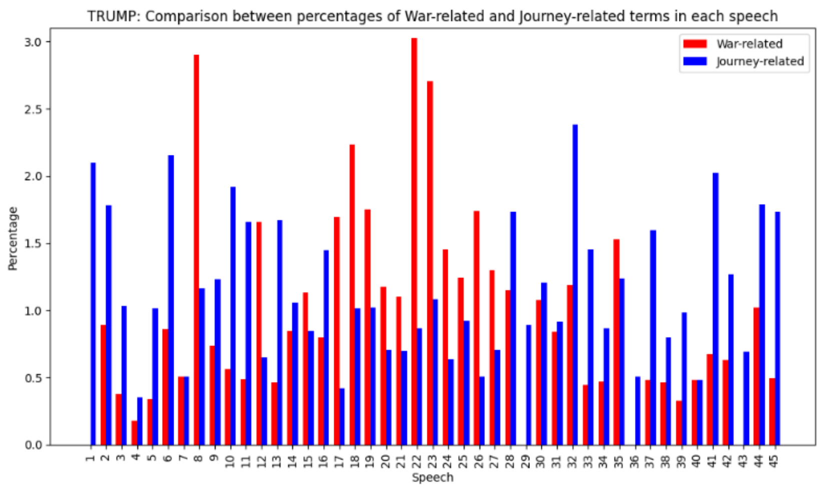

FIGURE 9: Comparison between the war- and journey-related terms in the TRUMP corpus.

With a cross examination of the percentage of war term present in each speech and the dominant topic for each speech, it resulted that more speeches containing war terms have as dominant topic topic number 2 (24 speeches) with a war-related terms percentage of 26.19%; whereas 21 speeches containing war terms have topic 1 as their dominant topic, with 14.19% as the sum percentage. Regarding the journey-related terms, the opposite is true: the sum of the percentages of journey-related speeches which have as dominant topic 1 is 23.70%; whereas the sum percentages of the documents of dominant topic 2 is 23.52%. These results are shown in Fig.10.

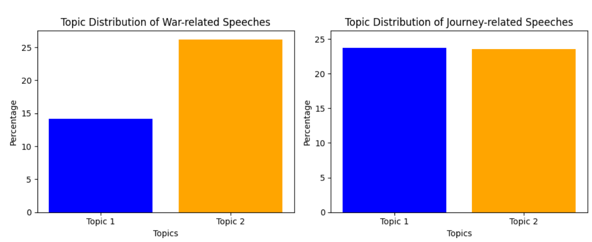

FIGURE 10: The speeches with more war-related terms have topic 2 as their dominant one. The speeches with more journey-related terms have topic 1 as their dominant.

5 Discussion

In this project, it has been investigated through LDA the type of topics which had been discussed in relation to Covid-19 by the WHO General Director and the USA President. Furthermore, it has been explored the usage of WAR and JOURNEY frames and whether they had been used to discuss specific topics, among the topics identified through LDA.

Within the WHO corpus, the top five most frequently used words of the WAR frame are "fight", "threat", "enemy", "violence" and "defeat". All these words carry a highly negative valence, indicating that the situation is conceptualized as a threat that can be overcome by defeating the enemy, that is the virus. Additionally, within the TRUMP corpus, the top five most frequently used war-related words are "war", "military", "enemy", "army" and "battle". These words appears to have a stronger association with warfare, compared to those in the WHO corpus, implying a more tangible conflict involving armies and militants fighting battles against the enemy. When comparing the 2 corpora, it is interesting to note the differences in war-related terms. Trump's corpus encompasses a broader variety of war-related words (29 compared to 19 in the WHO corpus), with "war" being the most frequently used one, while it only occurred once in the WHO corpus. This may indicate that the WHO is less explicit in employing this frame compared to Trump. Additionally, the WHO corpus does not include terms such as "military", "weapon", "soldiers", "troops" and "homeland", which are present in the TRUMP corpus. This discrepancy could be attributed to the fact that Trump, as a country's president, has a more patriotic and aggressive perspective compared to the WHO, which, as being a world organization, is responsible for international public health. In contrast, there is more coherence within the two corpora regarding the JOURNEY frame. In the WHO corpus, the top five journey-related words are "way", "reach", "coming", "forward" and "movement", whereas in the TRUMP corpus, they are "way", "coming", "travel", "move" and "process". Despite the fact that these words are relatively less informative compared to the war-related terms, they convey the concepts of progress and forward movement, which could be interpreted in terms of policy changes aimed at improving the situation, as well as the arrival of the virus and its spread across countries.

Moreover, another important point to note is that the JOURNEY frame was more prevalent across the two corpora compared to the WAR frame. This may be attributed to two factors. Firstly, the journey-related terms encompass verbs such as "going" and "coming", which are highly employed in the English language and it is very likely that they have been used in the corpus in their literal sense as well. Secondly, the JOURNEY metaphor has become so conventionalized that it becomes more challenging to recognize as such. Another point worth nothing is that speeches by Trump exhibit a higher degree of metaphorical language, with a greater prevalence of war and journey frames compared to the WHO corpus. This difference in metaphorical usage could be attributed to a stylistic choice and a deliberate persuasive effect employed by Trump.

In relation to the topic modelling of the WHO speeches, topic 1 had the highest percentage of both war- and journey-related terms. Topic 1 addresses aspects related to the reactions to the pandemic, such as the safety measures proposed, the dominant words are indeed "treatment", "testing", "prevent", "action", "plan", "approach". This may suggest that war- and journey-related terms are employed to express aspects of the Covid-19 pandemic related to the safety measures and actions that are needed to take to oppose to the pandemic. Topics 2 and 5 show comparable percentages of war and journey frames, containing words such as "emergency", "plan", "preparedness" and "trial", "restriction", "mask". On the other hand, topics 3 and 4 exhibit a higher degree of journey-related terms. It may be interesting to point out that topic 3 includes words such as "mask", "treat" and "medical", which could still be associated with safety measures, as in topic 1, which, instead, registered a higher percentage of journey-related terms. In contrast, topic 4 contains words such as "child" and "family", which refer to the familiar sphere evoking the need of togetherness and social compassion, therefore the use of war-related terms might be perceived as too intense. Nonetheless, topic 4 still encompasses words related to the economic damages caused by the pandemic, including terms like "economic", "fund", "plan" and "crisis". It is possible that the discourse surrounding the actions taken to restore economic losses may not be framed in terms of war.

Regarding the topic modelling of Trump's speeches, topic 2 stands out with a higher percentage of war-related terms, compared to topic 1. Topic 2 encompasses words directly associated with warlike actions, such as "military", "hit" and reference to emergency situations with the inclusion of "FEMA" (Federal Emergency Management Agency). However, topic 2 also contains words such as "patient", "bed" and "test", which relate to delicate aspect of healthcare and well-being of patients. Utilizing the war frame in such context may be particularly inadequate because it could evoke feelings of guilt and hopelessness in the process of treating the disease and recovering from it. It is surprising to find the word "happy" as one of the most dominant ones in topic 2, which appears to be an outlier in such a dramatic situation. However, despite topic 2 having a higher percentage of war frame, it also exhibits a significant presence of journey-related terms. This combination of war and journey frames could explain the inclusion of words that are not directly associated with war in this topic. On the other hand, topic 1 registers the highest percentage of journey-related terms, including words such as "bed" and "together", "drug", "test", "approved" and "FDA" (Food and Drug Administration). These terms pertain to the health aspects of the virus, encompassing the well-being and treatment of patients, who are constrained to bed. Additionally, they convey a sense of unity and social compassion, emphasizing the importance of collective efforts and individuals taking responsibility in combating the virus. Moreover, these terms may convey a sense of optimism and hope for patients as they might refer to the approval of new drugs and tests by the FDA.

6 Conclusion

Taken together the results suggest the pervasiveness of the WAR and JOURNEY frames in the discourse on Covid-19, as previous literature pointed out the use of these frames in discourses on diseases and viruses. It has been shown a relative tendency of the use of war-related terms to talk about specific aspects such as treatment and diagnostics; whereas the use of journey-related terms to talk about the patients and social compassion during the pandemic. However, both WAR and JOURNEY metaphors are strictly connected with one another and using one does not eclipse the other. In fact, different frames are apt to elaborate the discourse on Covid-19 around different aspects or on the same aspects but from different points of view. Metaphor is a double-edged sword, having both positive and negative implications in shaping public perception about the pandemic, whether people are aware of it or not. By using multiple frames, it is more likely to enable the effective description and discussion of different aspects related to the Covid-19 reality.

Finally, it is important to point out the limitations of this project. Firstly, the size of the corpus used might be small for the LDA topic analysis, which would likely perform better on a larger corpus. Additionally, this project is limited to the speeches delivered during the first two months of the pandemic (March - April 2020). Different results may arise when considering different time periods and the evolution of the pandemic. Furthermore, as already mentioned above, the assumption made in this project is that war- and journey-related terms would be used metaphorically within speeches addressing Covid-19; however, it is possible that some terms were used in their literal sense, in particular the journey-related terms.

References

- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.

- Gedigian, M.; Bryant, J.; Narayanan,S.; Ciric, B. 2006. Catching metaphors. In In Proceedings of the 3rd Workshop on Scalable Natural Language Understanding, pages 41–48, New York

- Gibbs, R. W., Jr. (2017). Metaphor wars: Conceptual metaphors in human life. Cambridge University Press.

- Hasan, M., Rahman, A., Karim, M.R., Khan, M.S.I., Islam, M.J. (2021). Normalized Approach to Find Optimal Number of Topics in Latent Dirichlet Allocation (LDA). In: Kaiser, M.S., Bandyopadhyay, A., Mahmud, M., Ray, K. (eds) Proceedings of International Conference on Trends in Computational and Cognitive Engineering. Advances in Intelligent Systems and Computing, vol 1309. Springer, Singapore. https://doi.org/10.1007/978-981-33-4673-4-27

- Heintz, I.; Gabbard, R.; Srivastava, M.; Barner, D.; Black, D.; Friedman, M.; Weischedel, R. 2013. Automatic extraction of linguistic metaphors with lda topic modeling. In Proceedings of the First Workshop on Metaphor in NLP, pages 58–66, Atlanta, Georgia

- Lakoff, G. & Johnson, M. (1980). Conceptual metaphor in everyday language. *The Journal of Philosophy* (pp. 453–486).

- Semino, E. (2008). Metaphor in Discourse. Cambridge University Press

- Semino, E., Demjén, Z., & Demmen, J. (2016). An integrated approach to metaphor and framing in cognition, discourse and practice, with an application to metaphors for cancer. Applied Linguistics, 39, 625-645. DOI: 10.1093/APPLIN%2FAMW028

- Shutova, E.; Sun, L. 2013. Unsupervised metaphor identification using hierarchical graph factorization clustering. In Proceedings of NAACL 2013, Atlanta, GA, USA

- Shutova,E.; Kiela, D.; Maillard, J. 2016. Black Holes and White Rabbits: Metaphor Identification with Visual Features. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 160–170, San Diego, California. Association for Computational Linguistics.

- Sontag, S. (1979). Illness as metaphor. Allen Lane.

- Sontag S. (1989). AIDS and its Metaphors. Farrar, Straus and Giroux.

- Steen GJ, Dorst AG, Herrmann JB, Kaal AA, Krennmayr T, Pasma T. A method for linguistic metaphor identification. From MIP to MIPVU. 2010. Amsterdam: John Benjamins

- Strzalkowski, T.; Broadwell, G.; Taylor, S.; Feldman, L.; Shaikh, S.; Liu, T.; Yamrom, B.; Cho,K.; Boz, U.; Cases, I.; Elliot, K. 2013. Robust extraction of metaphor from novel data. In Proceedings of the First Workshop on Metaphor in NLP, pages 67–76, Atlanta, Georgia

- Syed, S. and Spruit, M. "Full-Text or Abstract? Examining Topic Coherence Scores Using Latent Dirichlet Allocation," 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Tokyo, Japan, 2017, pp. 165-174, doi: 10.1109/DSAA.2017.61

- Tong, X., Shutova, E., & Lewis, M. (2021, June). Recent advances in neural metaphor processing: A linguistic, cognitive and social perspective. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 4673-4686).

- Tsvetkov, Y.; Mukomel, E.; Gershman, A. 2013. Cross-lingual metaphor detection using common semantic features. In Proceedings of the First Workshop on Metaphor in NLP, pages 45–51, Atlanta, Georgia

- Turney,P.; Neuman Y.;Assaf, D., Cohen, Y. 2011. Literal and metaphorical sense identification through concrete and abstract context. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP '11, pages 680–690, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Wicke P, Bolognesi MM (2020) Framing COVID-19: How we conceptualize and discuss the pandemic on Twitter. PLoS ONE 15(9): e0240010. https://doi.org/10.1371/journal.pone.0240010