

Classifying Speech Acts in Political Communication: A Transformer-based Approach with Weak Supervision and Active Learning

Klaus Schmidt*, Andreas Niekler*, Cathleen Kantner[†], Manuel Burghardt*

*Leipzig University, Email: [name].[surname]@uni-leipzig.de

[†]University of Stuttgart, [name].[surname]@sowi.uni-stuttgart.de

Abstract—We present a study on the automatic classification of speech acts in the domain of political communication, based on J. R. Searle’s classification of illocutionary acts. Our research involves creating a dataset using the US State of the Union corpus and the UN General Debate corpus (UNGD) as data sources. To overcome limited labelled data, we employ a combination of weak supervision and active learning techniques for dataset creation and model training. Through various experiments, we investigate the influence of external and internal factors on speech act classification. In addition, we discuss the potential for further analysis of speech act usage, using the trained model on the UNGD corpus. The findings demonstrate the effectiveness of Transformer-based models for automatic speech act classification, highlight the benefits of weak supervision and active learning for dataset creation and model training, and underscore the potential for large-scale statistical analysis of speech act usage in the domain of political communication.

I. INTRODUCTION

WHEN we use language, we often want to go beyond the mere conveying of information, but rather want to accomplish a communicative goal and perform a social function, such as making a promise or issuing a threat. In his seminal work “How to do things with words”, J. L. Austin has called these kinds of utterances illocutionary acts, which is not an act of just saying something, but an act of doing something by saying something [1]. According to Austin, there are two types of utterances: constatives, which make statements about the world, and performatives, which are acts performed by the speaker with the intention of fulfilling a social function. This concept was later developed further by Searle and the general term *speech act* has been developed since for such performative utterances [2]. Unlike interrogative and imperative sentences that are marked with either a “?” or “!” at the end of a sentence, performatives are not orthographically highlighted as such. Indeed, a central insight of Austin is the fact that there is apparently no trivial indicator differentiating constatives from performatives. As humans, we thus have to rely on linguistic and contextual information to properly recognize performatives and to act upon them accordingly.

While speech acts have an important socio-linguistic function in everyday communication, they are also playing a crucial role in political communication, as they are used by politicians to perform specific actions and influence their audience [3, 4].

Examples include promises for the purpose of getting elected, or requesting legislative action. By analysing the use of speech acts in political discourse, researchers can gain a better understanding of the intentions and motivations of political actors and how they seek to influence public opinion.

In this study, we present a robust computational framework for detecting and correctly identifying speech acts in political communication based on large transformer-based language models for automated and scalable analysis of speech acts in large datasets. The contributions of this paper are as follows:

- Our approach is based on Transformer-based language model classifiers [5], and we employ a combination of weak supervision and active learning techniques for dataset creation and model training.
- Through various experiments, we investigate the influence of external and internal factors on speech act classification.
- We show a concrete use case that demonstrates the use and potential of automatic speech act annotation in large text corpora containing political communication.

The use of speech acts in political communication has up to now received only little attention despite their pervasiveness and utility in this mode of communication (see [6, 7]).

Underlying this concept is the idea that the interconnected network of deontic moral forces within society are established and sustained through specific types of speech acts. J.R. Searle postulates that declarative speech acts, in particular, play a pivotal role in creating institutions and institutional facts. Nevertheless, it has not yet resulted in significant changes, possibly due to the substantial challenges involved in operationalizing his ideas for empirical research.”

The current literature on the use of speech acts in political communication has mainly focused on individual sentences or small corpora, lacking a larger, macroscopic perspective. However, with the advent of digital resources that are freely accessible and machine-readable, there is an opportunity to analyse larger datasets of political speeches, for instance parliamentary debates [8] or speeches of singular politicians [9]. This availability of material opens up new perspectives for rather empirical research questions along the lines of political sciences, international relations, computational social science and corpus linguistics.

An issue surrounding this topic is the vast amounts of data that need to be processed in order to gain valuable insights into macroscopic patterns of their uses among political actors. Especially since certain types of utterances are quite infrequent and sporadic, the criterion applies that manual sampling would certainly not be able to produce representative distributions for quantitative evaluations in any reasonable amount of time. With the advances of computational approaches for accessing different language phenomena from Natural Language Processing (NLP) and Artificial Intelligence (AI) made in recent years, it is now possible to find this language phenomenon in an automated way and with good quality in large amounts of data.

Future applications will be the quantitative analysis of political communication in international institutions [10]. This will provide a computational method for studying how to do international politics with words [11]. With the methodology shown here, the following research questions, for example, become conceivable: How can speech act theory be applied to understand declarations of war, inaugurations, pardons, government statements, etc.? How do international obligations and the self-binding of sovereign states to international norms and rules emerge under conditions of anarchy?

II. RELATED WORK

Automatic speech act classification has been a subject of research for some time, focusing on dialogue acts [12, 13, 14]. Earlier studies on Korean employed various methods, including Hidden Markov Models [15], maximum entropy models [16], and supervised machine learning algorithms [17]. Unlike Austin and Searle’s speech acts, dialogue acts specifically target synchronous language used in direct communication and many of the classification schemes used in research do not align with Austin’s or Searle’s speech act classification.

In recent years, researchers have increasingly utilized Deep Neural Networks (DNNs) in Natural Language Processing (NLP) due to advancements in computing capabilities. This has led to the adoption of more sophisticated approaches in addressing the issue at hand. Notably, recurrent neural networks (RNNs) [18] and convolutional neural networks (CNNs) [19] have been explored as viable options for tackling the challenge of speech act detection. Existing research on speech and dialogue act detection primarily focuses on synchronous language in online communication. However, little attention has been given to speech act detection in political communication, except for notable exceptions such as [20]. The emergence of the Transformer architecture [5] and pre-trained Transformer-based language models like BERT [21] has significantly changed the way NLP is practised, and automated processing of even difficult language problems is becoming increasingly possible. To the best of our knowledge, no previous work exists on classifying speech acts according to Searle’s classification of illocutionary acts.

III. EXPERIMENT DESIGN

In our experiments, we developed a Transformer-based model specifically designed to classify sentences according to Searle’s classification of illocutionary acts, referred to as “speech acts” in this study. We constructed the dataset using weak supervision and further refined it through active learning techniques. Additionally, we explored possible avenues for future research, which encompass improving the model’s performance and assessing its applicability and relevance in computational political science studies.

A. Corpora

To construct a suitable dataset for model training and initial analysis, we have selected two corpora of similar yet distinct sub-domains within political communication: the US State of the Union (SOTU) addresses and the speeches delivered at the UN General Debate. The SOTU corpus comprises the State of the Union presidential addresses given annually from 1790 to 2017.¹ In our experiments, the SOTU corpus’ temporal range was limited to between 1990 and 2017. The UN General Debate corpus (UNGD) comprises speeches delivered by national representatives, including presidents and foreign ministers, from each UN member country during the annual UN General Debate [8]. Spanning the years 1970 to 2020, this corpus provides a rich collection of political speeches from a diverse range of international stakeholders. Aside from the transcriptions of speeches in English, the UNGD corpus also includes a variety of metadata for each speech such as the year and session of the UN General Debate that the speech was held in, the name of the speaker, their position, the original language of delivery as well as the represented country. We have adopted a standardized format for referencing individual speeches in the UNGD corpus, using the year of the speech and a three-letter country code based on ISO 3166-1 alpha-3 to denote the speaker’s country he or she represents, such as “1979_IRQ” for the 1979 speech of the representative of Iraq.

We formulated the task of identifying speech acts as a multi-class classification problem. To achieve this, we have employed the theoretical framework of illocutionary acts introduced by Searle [2] and identified five classes of speech acts and one open class for sentences where none of the labels can be assigned:

- *ASSERTIVE*, the speaker commits to the truth of a stated expression.
- *EXPRESSIVE*, the speaker expresses their personal thoughts and feelings
- *COMMISSIVE*, the speaker commits themselves to a future action.
- *DIRECTIVE*, the speaker issues orders or instructions to the recipients.
- *DECLARATIVE*, the speaker, using granted institutional powers, alters or defines (social) realities.
- *NONES*, an open class for sentences that contain none of the above speech act types.

¹<https://www.kaggle.com/datasets/ratman/state-of-the-union-corpus-1989-2017>

B. Heuristic Labelling using Weak Supervision

When creating a dataset to train our model, we need to define the assignment of labels by existing linguistic properties of each class. To assign speech act annotations to linguistic expressions containing the linguistic properties just mentioned, we use the weak supervision library *skweak* [22]. This library enables the generation of so-called weak labels through annotator functions. Although these labelling functions rely on simple heuristics and may not achieve particularly high precision, they effectively enable us to efficiently search for speech acts employed in political communication during the dataset creation process.

A crucial aspect of Austin’s seminal work on speech acts pertains to the notion that certain verbs are closely linked to a specific class of speech act, which he coined *performative verbs*. We relate closely to this basic concept in our definition of the relevant linguistic features for speech acts, in order to define heuristics for the labelling functions. We chose the following verb relationships to supervise the labelling process:

- **Assertive:** *think, know, believe, convince, presume, assume, admit*
- **Expressives:** *apologize, condole, lament, deplore, forgive, welcome, thank, forgive, boast*
- **Commissives:** *promise, vow, guarantee, offer, will, refuse, volunteer*
- **Directives:** *ask, order, command, request, beg, plead, pray, invite, permit, advise, must, should*
- **Declaratives:** *announce, declare, dismiss, nominate, pronounce, pass, adopt, support, oppose, advocate, condemn*

To capture general speech act features, we developed several labelling functions that assign a general “speech act” label [SA] which is added to the matching sentences. This approach allows us to assign weak labels to encompass general features that are not confined to a specific speech act. Sentences containing performative verbs are assigned the speech act label considered characteristic for them, as shown in the list above. We derived the labelling rules from sample sentences in Austin’s work “*How to do things with words*” [1]. Note, that multiple labelling functions can be applied to a single sentence if the criteria are met. In a final step, these weak labels, which represent a supervision signal, are subsequently aggregated to form final labels. This aggregation model takes into account the varying degrees of confidence associated with each weak label. By leveraging the sequential dependencies of the weak labels between sentences, the aggregation ensures a more accurate labelling outcome. The labelling functions can be briefly described as follows:

- **Subject is 1st person:** Assigns the label [SA] if the subject of a sentence is either “I” or “we”.
- **Main verb is present tense:** Assigns the label [SA] if the main verb of a sentence is in present tense.
- **Object is 2nd person:** Assigns the label [SA] if the object of the sentence is the pronoun “you”.
- **Sentence contains imperative:** Assigns the label [DIRECTIVE] if the sentence is an imperative without an

overt subject.

- **Sentence contains interrogative:** Assigns the label [DIRECTIVE] if the main verb precedes the subject and the sentence ends with a question mark “?”.
- **Sentence contains performative verb:** Assigns the label associated with the performative verb [ASSERTIVE, EXPRESSIVE, COMMISSIVE, DIRECTIVE, DECLARATIVE] if the lemma of a performative verb is present in the sentence.

Processing the SOTU as well as the UNGD corpus using this weak supervision approach yields a list of sentences that likely contain speech acts.

C. Active Learning

Active learning is an approach that utilizes query strategies to select the most informative samples from an unlabelled pool of data, guided by a classifier trained on a set of existing labelled data. The goal is to intelligently choose data points for labelling, in order to improve the performance and efficiency of the learning process. Over the years, many query strategies have been proposed by various researchers, most of them using prediction based query strategies [23]. Active learning follows a process that involves two pools: a labelled pool and an unlabelled pool. The goal is to transfer data from the unlabelled pool to the labelled pool. In prediction-based query strategies, a primary model is trained on the labelled pool, which is then used to identify the most informative samples to query from the unlabelled pool. These queried samples are then presented to a domain expert (also known as oracle in active learning jargon), who manually confirms or rejects the labels of the primary model. The labelled samples are subsequently added to the labelled pool. This process continues iteratively until a predefined stopping criterion is met. With this approach, active learning allows the progressive enhancement of the model’s performance by actively selecting the most informative samples for annotation.

We used the active learning library Small-Text [24] to facilitate experiments with active learning. This library provides a user-friendly and consistent interface which makes it very simple to set up the experiments. After some initial testing, we selected the prediction-based query strategy *Prediction Entropy* as our choice and set a fixed iteration size of 10 iterations as the stopping criterion. In each iteration, we queried 20 samples.

To effectively utilize the library, it is essential to choose a transformer-based language model as the underlying framework for the classification task. Among the various models we tested, we ultimately opted for the light-weight DistilBERT [25] with the default configuration provided through HuggingFace.² This decision was primarily driven by DistilBERT’s significantly reduced computational overhead, with minimal performance drawbacks compared to the resource-intensive BERT model. Since the iterative approach involves many training processes, the model accelerates them significantly,

²<https://huggingface.co/distilbert-base-uncased>

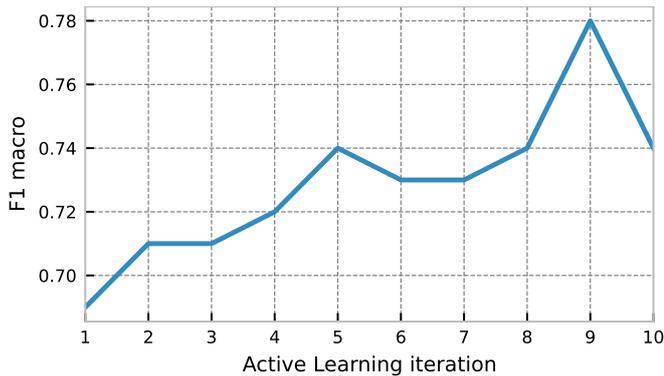


Figure 1. Learning curve of f1 macro performance on test set over all classes

making them more manageable in an interactive setting with domain experts.

The results from the weak supervision process were subsequently filtered by a domain expert, resulting in a valid semi-manual training dataset for initializing the active learning process. Despite the linguistic features serving as a basic codebook for annotating queried samples, a considerable number of samples proved to be non-trivial. Given the fuzzy nature of speech act use and their interpretation, disagreements are unavoidable. For training, we stratified the dataset by class and performed a split, allocating 60% of the data for training and 40% for evaluation. This choice of a 60:40 split was made to ensure that both sets capture a representative sample of each class. By evaluating each iteration’s performance using F1-scores for individual classes as well as the macro average, we could effectively track the model’s progress over iterations and visualize it through a learning curve.

IV. EXPERIMENTS AND RESULTS

A. Active Learning and Classifier Performance

After applying the weak supervision approach and correcting its output, our initial dataset consisted of 175 samples in the labelled pool and 118 samples in the evaluation set. Once the initial model was trained using the samples found in the labelled pool, we further refined it through the active learning process. In each iteration, we queried, annotated, and then added 20 newly labelled samples to the labelled pool. After each addition, the model was retrained using the augmented labelled pool. This iterative process continued until the stopping criterion of 10 iterations was met.

Looking at Fig. 1, after performing ten active learning iterations, we observed a rise in F1-macro performance from 0.69 of the initial model to 0.78 of the best model in the ninth iteration before the performance cratered somewhat again in the tenth iteration. In Table I we show detailed information about the classification performance of the model trained in the ninth iteration.

Examining the individual F1-scores, we noticed large discrepancies between the highest score of the commissive class and lowest score of the expressive class. In terms of precision,

Class	Precision	Recall	F1	Support
Assertive	0.71	0.89	0.79	19
Expressive	0.82	0.60	0.69	15
Commissives	0.83	1.0	0.91	24
Directives	0.78	0.96	0.86	26
Declaratives	1.0	0.67	0.80	18
Nones	0.80	0.50	0.62	16
macro average	0.82	0.77	0.78	118

Table I

TABLE PRESENTING THE PERFORMANCE OF THE CLASSIFIER FROM THE 9TH ACTIVE LEARNING ITERATION OVER ALL SPEECH ACT CLASSES

assertives do have a medium performance, indicating that the training examples provide very ambiguous features to the classifier. In terms of recall, expressives and declaratives show comparatively medium performance. In the subsequent experiments presented in sections IV-B, IV-D as well as the showcase demonstrated in section IV-E, we utilized the model trained during the 9th iteration, which incorporated an additional 180 training samples obtained through the active learning process.

B. Adding context sentences in the training data

The current model employed in our research utilizes single sentences as the units of classification. While this approach may simplify the classification task, it is not necessarily evident that speech acts occur independently of each other. Therefore, it is important to acknowledge the potential limitations of this local single sentence approach.

To better understand the relationships between single sentences, we conducted an investigation by visualizing the occurrence of all speech acts from five randomly selected speeches. Figure 2 reveals a strong clustering behaviour in the use of speech acts, even from this small sample.

We conducted an additional analysis by calculating the pointwise mutual information (PMI) between each speech act and its neighbouring co-occurring speech acts in a sample of 100 randomly selected speeches. The PMI measures the statistical association between two events, in this case, speech acts and their neighbouring counterparts. Positive PMI scores indicate a higher than random likelihood of co-occurrence. On the other hand, negative scores indicate a lower than expected likelihood of co-occurrence. Our analysis revealed a significantly higher co-occurrence between speech acts of the same class, as would be expected by random chance. Furthermore, we observed moderately positive associations between directives and commissives, as well as between expressives and commissives. These findings suggest that considering the context of neighbouring sentences might be crucial for accurate classification of speech acts.

To test this hypothesis, we investigated the effectiveness of extending the classification unit by incorporating a window of neighbouring context sentences around a sentence from the trainings set. This approach enables the classifier to capture the contextual information surrounding a sentence. However, as depicted in Figure 3, our findings indicate that this approach

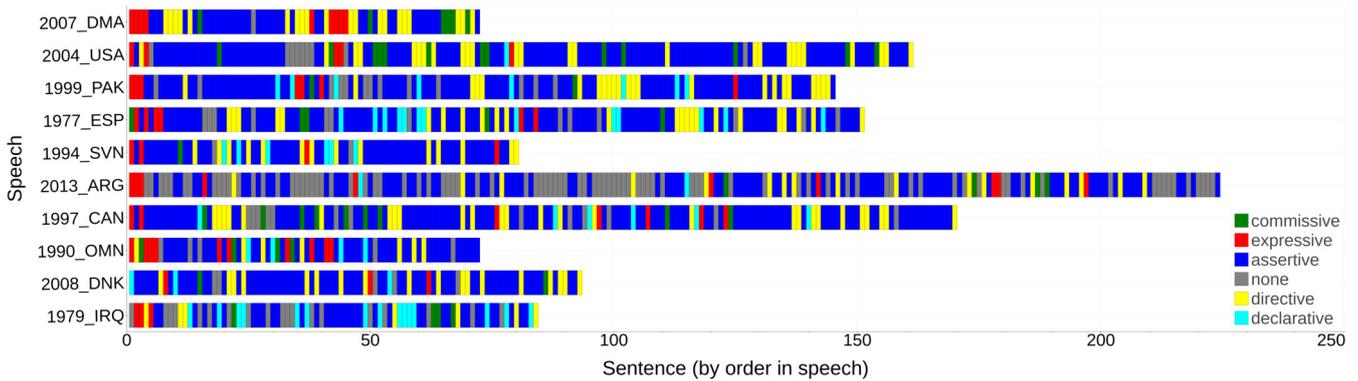


Figure 2. Speech act usage pattern over the course of a speech. The bars represent a speech in full length and the colours mark a respective speech act. It is recognizable that the same speech acts often occur consecutively in the sequence.

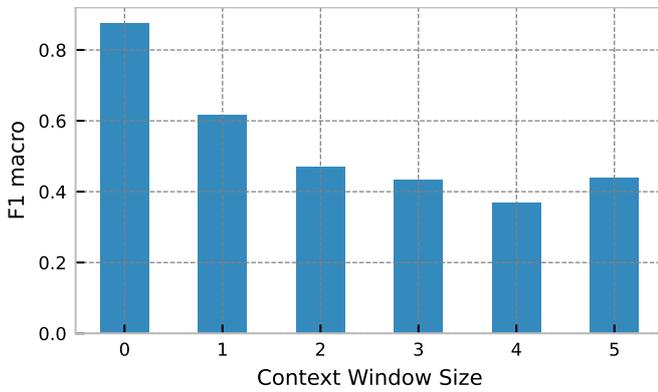


Figure 3. Model performance on the entire dataset (train+test) for various context sizes. The Number of context sentences is equivalent to $2 \times$ Context Window Size.

did not yield a substantial improvement over using a single sentence window.

C. Feature Importance of Metadata

In the preceding section, we demonstrated the significance of linguistic features in the classification of speech acts, as evidenced by the rather promising results obtained from the transformer model. In addition to the raw speeches, the UNGD Corpus includes various metadata, such as the name and position of the speaker, the country represented, and the original language of the speech. To explore the potential significance of metadata in relation to the occurrence of speech acts within the speeches, we trained a separate model in the form of a Random Forest (RF) regressor. This model is aimed to predict the distribution of each speech act class based on the available metadata alone. The objective of this approach is to examine the potential influence of extra-linguistic features, such as metadata, on the distribution of speech acts by examining the *importance* of each feature. The hyper-parameters for the RF model were left at their default values provided by scikit-learn, which consisted of using 100 trees and the Gini impurity criterion.

We compared the performance of the trained model to a baseline dummy classifier and observed a substantial improvement in goodness of fit for commissives (reducing the mean squared error from 10.75 to 2.4) and directives (reducing the mean squared error from 42.8 to 9.1). These results indicate a significant correlation between the metadata and the distribution of commissives and directives.

To identify the most influential features, we conducted an analysis of feature importances in the model. In RFs, feature importance quantifies the average decrease in impurity across all trees for each feature [26]. Figure 4 showcases the results of this analysis for commissives and directives, providing interesting insights. The feature importance analysis indicates that certain variables have a notable impact on the distribution of speech acts. For commissives, the role of the speaker is a significant variable, with the category of *president* standing out prominently. Additionally, the country represented by the speaker also plays a crucial role, with *USA* and *Japan* being particularly influential factors.

Similarly, for both commissives and directives, the original language of the speech exhibits a strong effect on the distribution of speech acts. Notably, speeches delivered in English demonstrate a substantial influence on the observed distributions.

D. Linguistic Features for Qualitative Assessment

A long-lasting shortcoming of using deep neural networks architectures such as Transformers is the black box nature of these models. Different approaches under the banner of *Explainable AI* such as CAPTUM [27] and shapley values [28] have been investigated in order to elucidate which features are important in a model’s prediction. In our investigations, we used CAPTUM which is implemented with the *transformers_interpret* library.³ This approach assigns *attribution scores* to each of the (sub-)token features which quantify the importance of each feature for the classification of a particular class. These scores are scaled between 1, indicating maximum attribution for classifying this as a particular class and -1,

³<https://github.com/cdpierce/transformers-interpret>

Speech Act	Example
<i>Assertive</i>	[CLS] today the international trade and monetary crisis , which is still un ##res ##olved , threatens to undermine that strategy and casts a pal ##l of uncertainty over the prospect of attending the goals of the development decade itself . [SEP]
<i>Commissive</i>	[CLS] we will stay the course on reform , which is the only road to peace and prosperity for our country . [SEP]
<i>Commissive</i>	[CLS] we will soon be un ##ve ##iling a new law and anti - corruption squad . [SEP]
<i>Declarative</i>	[CLS] eighth ##ly , we support the efforts of the organization of african unity and the countries of the horn of africa to restore peace and stability in the region . [SEP]
<i>Directive</i>	[CLS] let us put reason before blood ##shed . [SEP] 0.05pt
<i>Expressive</i>	[CLS] i would also like to take this opportunity to express our gratitude to the previous president of the assembly , ambassador ismail , who showed exceptional commitment to the reform initiatives throughout the fifty - first session . [SEP]
<i>Directive (low confidence)</i>	[CLS] we remain fully confident both in the united nations and in the international community . [SEP]

Figure 5. Example sentences visualized with the library CAPTUM [27] where the positive contributing features are shown in green and the negative contributing features are shown in red. The brightness indicates the importance of the features. The tokenization is taken from the transformer model.

E. Showcase: speech acts as time series data

This section will emphasize the practical application of the model through a demonstrative analysis in the UNGD corpus. We will use sentence-based annotations of speech acts to quantitatively measure the shift in communicative patterns within the UN General Debate. Specifically, we want to demonstrate a method for exploring potential influences of significant political events on speech act usage patterns. It is important to note that this demonstration serves as an illustrative example rather than a rigorous scientific investigation of speech act usage. By highlighting the viability of our approach, we aim to emphasize its potential as a valuable research tool in the fields of political and social sciences. For this purpose, we decided to focus on the political events that unfolded in Ukraine from 2013 to 2014. Later being known as the Maidan Uprising, a series of protests and civil unrest unravelled on November 21, 2013, primarily centred around Kyiv’s Maidan Nezalezhnosti (Independence Square) [29]. These protests ultimately led to the Revolution of Dignity in February 2014, albeit with a heavy toll of 108 protesters and 13 police officers losing their lives in the clashes [30]. Consequently, the Ukrainian parliament, with a significant

majority of 328 out of 450 votes (approximately 73% of the votes), voted to remove President Viktor Yanukovich from power. Yanukovich disputed the legality of the vote and sought assistance from Russia. Russia criticized the events, labelling them a ”coup”. Subsequently, pro-Russian protests erupted in southern and eastern Ukraine. Russia occupied and eventually gained control of Crimea, while armed pro-Russian groups seized government buildings and declared Donetsk and Luhansk as independent states. This resulted in the onset of the Donbas war, prompting international efforts to diplomatically address and find resolution to the unfolding events within the international community.

To investigate the change in the strategic political communication following those events, we analysed the speech act usage of *directives*, which primarily attempt to motivate the addressees to take an action, of Ukrainian representatives in the UN General Debate. By choosing this scenario, we aim to clearly demonstrate the ways in which events can affect the language used in international political discourse by identifying and quantifying speech acts as an indicator variable.

To identify a statistically significant change in usage pat-

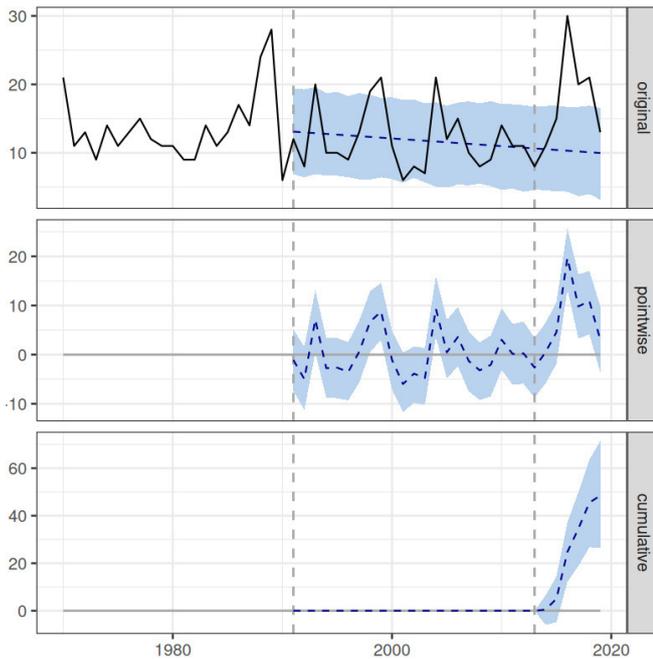


Figure 6. CausalImpact plot comparing directive speech act usage in absolute numbers between the years 1991-2013 and 2014-2019

terms of *directives* between the period starting in 2014 and the most recent available speech in 2019, we utilized the R library *CausalImpact* [31]. Specifically, we compared the *directive* usage frequency of Ukrainian representatives after the Maidan events (2014-2019) with those during the period before the events, which spans from the dissolution of the Soviet Union in 1991 up until 2013. The software creates a Bayesian structural time-series model which is used to predict how the response would have evolved after an event, called intervention, if the event had never occurred. The result is the cumulative effect of the intervention on the time series.

In the model, we see a positive effect on the use of *directive* speech acts in the period between 2014-2019. The first panel in Figure 6 showcases the observed data and a calculated prediction derived from the data of the pre-intervention period. The pre-intervention period is enclosed by two dotted vertical lines and serves as the baseline for comparison with the post-intervention period. The second panel shows the pointwise causal effect of the model, which is the calculated difference between the observed data and the prediction. The third panel presents the cumulative effect of the model, which is the summation of the pointwise causal effect over time. The positive effect on the use of *directive* speech acts observed during the period after the Maidan protests in Ukraine in 2014 is statistically significant and unlikely to be due to random fluctuations. It is important to note that during this period, not only the frequency of *directive* speech acts increased significantly, but the speeches also showed a general increase in length. This suggests that the involvement and activation of the international community by Ukrainian representatives

has increased and become a larger part of Ukraine's political communication.

V. DISCUSSION

In this paper, we investigate for the first time how the linguistic phenomenon of speech acts described by Searle can be automatically identified in text corpora using a transformer-based approach and be made accessible for quantitative and qualitative analyses. In doing so, the study focuses on the identification of speech acts within very large datasets of political communication, which makes the results interesting for political science research. Furthermore, the study presents very transparently which linguistic properties are included for the definition of speech acts. Our experiments demonstrate the effectiveness of using a combination of weak supervision and active learning for dataset creation in the classification of speech acts. Weak supervision enables the identification of samples that satisfy defined linguistic features, while active learning can provide new examples from an unlabelled dataset. The results show a 0.09 increase in F1 score within 9 iterations of the model, indicating the usefulness of these techniques in expanding small datasets and improving model performance. The problems we encountered in identifying *expressives* and *declaratives* in terms of recall indicate that there may be variance between the examples in the training data and the test data, or the linguistic characteristics of these speech acts may be ambiguous, making it difficult for the classifier to properly delineate between these and other classes. However, evaluating the entire process still poses challenges, especially with the limitations of our evaluation set, which has a rather small size. The properties of this set might introduce a high degree of variance and possible biases in the evaluation. To address these challenges, future research will focus on developing a robust evaluation framework and larger evaluation sets for the methods employed.

Aside from linguistic features, our observations indicate that extra-linguistic factors play a role in influencing the distribution of speech acts. Specifically, variables such as the speaker's role, the country represented, and the original language of the speech have shown significant influence on the occurrence and distribution of speech acts. These findings emphasize the importance of considering contextual factors when studying speech acts in political discourse. Understanding the impact of these variables can contribute to the development of more accurate speech act classification frameworks and provide valuable insights into the factors shaping the use of speech acts.

Additionally, our findings suggest tentative evidence of relationships between speech acts, as utterances of the same class tend to occur in proximity to each other. This observation not only contributes to our understanding of how speech acts are used in actual communication, but also provides valuable insights for the classification task. While our initial attempts at utilizing neighbouring utterances did not yield satisfactory results, further exploration in this area holds promise for enhancing the accuracy of speech act classification. One

possible approach is to employ techniques such as linear chain Conditional Random Fields (CRFs) modelling [32] or similar sequential models. Such modelling can capture the sequential dependencies among speech acts and exploit the contextual information provided by neighbouring utterances. By incorporating such techniques, we can potentially improve the classification performance by considering the sequential nature of speech acts.

The application of our model concerning the use of *directives* by the representatives of Ukraine demonstrate the potential of utilizing the trained model for quantitative research. We presented a statistical analysis technique to explore the impact of political events on speech act usage patterns. By examining specific periods and comparing them to relevant baselines, we can gain insights into how political events shape language use in international discourse. The approach presented here can be extended to investigate other geopolitical events and their influence on speech act patterns, opening the potential for providing a deeper understanding of the dynamics of political communication, which is crucial for researchers and practitioners in fields such as political science, international relations, and diplomacy. In addition, our analysis of the relevant features with CAPTUM demonstrated that the qualitative evaluation of speech acts has the potential to enhance our understanding of the strategic use of speech acts. Automatic speech act detection can thus contribute to various fields by facilitating empirical studies on persuasive strategies employed by political actors. It not only enables the monitoring of shifts in rhetoric and discourse, but also provides insights into the motivations and intentions behind political speeches. By employing and further developing this technology, researchers can delve deeper into understanding the intricate dynamics of political communication and gain a comprehensive understanding of the strategies and objectives employed by politicians.

REFERENCES

- [1] J. L. Austin, *How to do things with words*. Cambridge, Mass., Harvard University Press, 2003., 1962.
- [2] J. R. Searle, "A classification of illocutionary acts," *Language in society*, vol. 5, no. 1, pp. 1–23, 1976. doi: 10.1017/s0047404500006837
- [3] P. L. Berger and T. Luckmann, *Die gesellschaftliche Konstruktion der Wirklichkeit: Eine Theorie der Wissenssoziologie*. Frankfurt am Main: Fischer, 1966.
- [4] J. Habermas, *Theorie des kommunikativen Handelns: Handlungsrationalität und gesellschaftliche Rationalisierung*. Frankfurt: Suhrkamp, 1995, vol. 1.
- [5] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. doi: 10.5555/3295222.3295349. [Online]. Available: <https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>
- [6] S. S. M. Hashim and S. Safwat, "Speech acts in political speeches," *Journal of Modern Education Review*, vol. 5, no. 7, pp. 699–706, 2015. doi: 10.15341/jmer(2155-7993)/07.05.2015/008
- [7] M. Ulum, D. Sutopo, and W. Warsono, "A comparison between trump's and clinton's commissive speech act in america's presidential campaign speech," *English Education Journal*, vol. 8, no. 2, pp. 221–228, 2018.
- [8] A. Baturo, N. Dasandi, and S. J. Mikhaylov, "Understanding state preferences with text as data: Introducing the un general debate corpus," *Research & Politics*, 2017. doi: 10.1177/2053168017712821
- [9] G. Peters and J. T. Woolley, "The state of the union, background and reference table," The American Presidency Project, Santa Barbara, CA, 1999–2021. [Online]. Available: <https://www.presidency.ucsb.edu/node/324107/>
- [10] J. Duffield, "What are international institutions?" *International Studies Review*, vol. 9, pp. 1–22, 2007. doi: 10.1111/j.1468-2486.2007.00643.x
- [11] C. Daase, S. Engert, M.-A. Horelt, J. Renner, and R. Strassner, *Apology and Reconciliation in International Relations: The Importance of Being Sorry*. London: Routledge, 2015.
- [12] C. Moldovan, V. Rus, and A. C. Graesser, "Automated speech act classification for online chat." *MAICS*, vol. 710, pp. 23–29, 2011.
- [13] B. Bayat, C. Krauss, A. Merceron, and S. Arbanowski, "Supervised speech act classification of messages in german online discussions," in *The Twenty-Ninth International Flairs Conference*, 2016.
- [14] Y. Liu, K. Han, Z. Tan, and Y. Lei, "Using context information for dialog act classification in DNN framework," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Copenhagen, Denmark: Association for Computational Linguistics, Sep. 2017. doi: 10.18653/v1/D17-1231 pp. 2170–2178. [Online]. Available: <https://aclanthology.org/D17-1231>
- [15] S. Lee and J. Seo, "Korean speech act analysis system using hidden markov model with decision trees," *International Journal of Computer Processing of Oriental Languages*, vol. 15, no. 03, pp. 231–243, 2002. doi: 10.1142/s0219427902000625
- [16] W. S. Choi, H. Kim, and J. Seo, "An integrated dialogue analysis model for determining speech acts and discourse structures," *IEICE TRANSACTIONS on Information and Systems*, vol. 88, no. 1, pp. 150–157, 2005. doi: 10.1093/ietisy/e88-d.1.150
- [17] N. Song, K. Bae, and Y. Ko, "Effective korean speech-act classification using the classification priority application and a post-correction rules," *Journal of KIISE*, vol. 43, no. 1, pp. 80–86, 2016. doi: 10.5626/jok.2016.43.1.80
- [18] R. Li, C. Lin, M. Collinson, X. Li, and G. Chen,

- “A dual-attention hierarchical recurrent neural network for dialogue act classification,” in *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*. Hong Kong, China: Association for Computational Linguistics, Nov. 2019. doi: 10.18653/v1/K19-1036 pp. 383–392. [Online]. Available: <https://aclanthology.org/K19-1036>
- [19] D. Yoo, Y. Ko, and J. Seo, “Speech-act classification using a convolutional neural network based on pos tag and dependency-relation bigram embedding,” *IEICE Transactions on Information and Systems*, vol. 100, no. 12, pp. 3081–3084, 2017. doi: 10.1587/transinf.2017edl8083
- [20] S. Subramanian, T. Cohn, and T. Baldwin, “Target based speech act classification in political campaign text,” in *Proceedings of the Eighth Joint Conference on Lexical and Computational Semantics (*SEM 2019)*. Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019. doi: 10.18653/v1/S19-1030 pp. 273–282. [Online]. Available: <https://aclanthology.org/S19-1030>
- [21] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019. doi: 10.18653/v1/N19-1423 pp. 4171–4186. [Online]. Available: <https://aclanthology.org/N19-1423>
- [22] P. Lison, J. Barnes, and A. Hubin, “skweak: Weak supervision made easy for NLP,” in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations*. Online: Association for Computational Linguistics, Aug. 2021. doi: 10.18653/v1/2021.acl-demo.40 pp. 337–346. [Online]. Available: <https://aclanthology.org/2021.acl-demo.40>
- [23] C. Schröder and A. Niekler, “A survey of active learning for text classification using deep neural networks,” 2020. [Online]. Available: <https://arxiv.org/abs/2008.07267>
- [24] C. Schröder, L. Müller, A. Niekler, and M. Potthast, “Small-text: Active learning for text classification in python,” 2021. [Online]. Available: <https://arxiv.org/abs/2107.10314>
- [25] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, “Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter,” 2019. [Online]. Available: <https://arxiv.org/abs/1910.01108>
- [26] G. James, T. Hastie, R. Tibshirani, and D. Witten, *An Introduction to Statistical Learning: With Applications in R*. Springer, 2013. ISBN 978-1-4614-7137-0
- [27] N. Kokhlikyan, V. Miglani, M. Martin, E. Wang, B. Alsallakh, J. Reynolds, A. Melnikov, N. Kliushkina, C. Araya, S. Yan, and O. Reblitz-Richardson, “Captum: A unified and generic model interpretability library for pytorch,” 2020.
- [28] S. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” 2017.
- [29] “Euromaidan,” page Version ID: 1153492594. [Online]. Available: <https://en.wikipedia.org/w/index.php?title=Euromaidan&oldid=1153492594>
- [30] “Revolution of dignity,” page Version ID: 1154376283. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Revolution_of_Dignity&oldid=1154376283
- [31] K. H. Brodersen, F. Gallusser, J. Koehler, N. Remy, and S. L. Scott, “Inferring causal impact using Bayesian structural time-series models,” *Annals of Applied Statistics*, vol. 9, pp. 247–274, 2015. doi: 10.1214/14-aos788
- [32] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, “Conditional random fields: Probabilistic models for segmenting and labeling sequence data,” in *Proceedings of the Eighteenth International Conference on Machine Learning*, ser. ICML ’01. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001. doi: 10.5555/645530.655813. ISBN 1558607781 p. 282–289.