Attitudes in Diplomatic Speeches: Introducing the CoDipA UNSC 1.0

Mariia Anisimova, Šárka Zikánová
Charles University
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics
Malostranské náměstí 25
118 00 Praha, Czech Republic
{anisimova, zikanova}@ufal.mff.cuni.cz

Abstract

This paper presents CoDipA UNSC 1.0, a Corpus of Diplomatic Attitudes of the United Nations Security Council annotated with the attitude-part of Appraisal theory. The speeches were manually selected according to topic-related and temporal criteria. The texts were then annotated according to the predefined annotation scenario. The distinguishing features of the diplomatic texts require a modified approach to attitude evaluation, which was implemented and presented in the current work. The corpus analysis has proven diplomatic speeches to be consistently evaluative, offered an overview of the most prominent means of expressing subjectivity in the corpus, and provided the results of the inter-annotator agreement evaluation.

Keywords: Appraisal theory, diplomatic discourse, corpus linguistics, CoDipA UNSC 1.0

1. Introduction

This paper is aimed at describing the CoDipA UNSC 1.0, a corpus of the thematically and temporally selected diplomatic speeches of the United Nations Security Council (Schoenfeld et al., 2019), annotated with the adaptation of the attitude-part of Appraisal theory (Martin and White, 2005). It describes the annotation scenario together with the annotated data, and offers an overview of the corpus statistics, evaluation of the double annotations, and the future of the project.

The need for such a corpus derives from the specific features of multilateral diplomatic communication, which influence the development of a distinctive type of subjectivity expression, that is rarely addressed.[1]

Diplomatic speeches form a distinctive group of texts that are different from other types of discourse in many aspects. These texts are highly *formalized* and *structured*, typically preserving the main outline components in a set order independent of the topic of the meeting or the length of a document.

The syntactic complexity of these texts is mainly dependent on the communicative goal of the speaker, who may either choose shorter and simpler formulations if they wish to be concise and clear or opt for complex syntactic structures and complicated style if their goal is to avoid being specific (Stanko, 2001).

Other prominent characteristics of these texts are the understated tone (Stanko, 2001) and indirectness, which result in implicit formulations, complex syntax, and passivization. These pragmatic features prove to be very important to how diplomats express opinions, which are most frequently not of their own but of the political body they represent (Swain, 2017). It is also because of them, that the diplomatic attitudes require their own approach in the process of annotation.

The format of multilateral communication set in the Security Council does not allow for a direct dialogue between the speakers, causing the argumentation to be rather one-sided and monologic (Swain, 2017).

In our previous publications (Anisimova and Zikánová, 2022; Anisimova, 2021) we have discussed the notion of attitude in diplomatic discourse and described our view on the most suitable annotation schemes for its evaluation, explained the annotation process and environment, as well as the criteria for selecting the data for our corpus of diplomatic speeches. We have then provided the outcomes of the first annotation experiment, which was then utilized for redefining the annotation scenario based on problematic and unclear annotation cases. The described work has led to the creation of the language resource, presented in this paper.

The structure of the paper includes the two main sections, namely:

- Approach, which offers an overview of Appraisal theory, our selected approach to it, the description of the annotation process, and the basic principles of the annotation scenario;
- and Corpus analysis, which provides infor-

[1] The corpus and the guidelines are ready for publication after the anonymity period.

ISA-20 Proceedings @LREC-COLING-2024, pages 17–26

mation about the corpus statistics, and inter-annotator reliability.

1.1. Related work

Due to its extensive informativity, Appraisal theory (as described in subsection 2.1) has long been applied to various types of discourses. The detailed description of various aspects of emotionality and opinion makes it useful for both qualitative and quantitative analysis. Appraisal theory is applied in various areas of linguistic research, for instance for analyzing argumentative essays (Lam and Crosthwaite, 2018), literary studies (Busetto and Delmonte, 2019), translation studies (Tajvidi and Arjani, 2017), political (Zhang and Pei, 2018) and diplomatic (Lian, 2018) text analysis, as well as movie, book, and consumer product reviews (Kolhatkar et al., 2020). The extensiveness of the list of possible areas of application corresponds with the versatility of the approach.

Particular practical aspects of annotating appraisal-bearing expressions were described by Read et al. (2007) and Fuoli (2018). In their work, Read et al. (2007) have offered a view on methodology for annotating appraisal, and an overview of the use of this methodology to annotate the corpus of book reviews. An inter-annotator agreement study and the considerations of instances of systematic disagreement are particularly useful for developing an appraisal-related annotation framework.

Another work related to the practical aspects of annotating appraisal was developed by Fuoli (2018). This study offers a step-wise method for the manual annotation of appraisal and covers some of the problematic aspects of this type of annotation, such as challenges in identifying appraisal, challenges in classifying appraisal, and questions of reliability, replicability, and transparency of the annotation process. As for practical applications, one of the bigger available resources is the Simon Fraser University Review Corpus (Kolhatkar et al., 2020) that offers 150 movies, books, and hotel reviews annotated with subjectivity types.

2. Approach

Our approach is based on the attitude part of Appraisal theory (Martin and White, 2005). During the first stage of the corpus creation, we have carried out the first trial annotations and designed the annotation scenario in accordance to the text type and annotation task. After that, the scenario was edited according to the annotators' comments to unify the possible inconsistencies in the approaches to the annotation process. This section provides a description of Appraisal theory, annotation scenario, and annotation process, as well as the data selection process.

2.1. Appraisal theory

Appraisal theory is an approach to analyzing expressions of subjectivity in a written text (Martin and White, 2005). The theory is located within a framework of Systemic Functional Linguistics (Halliday, 2004), and aims at providing a piece of extensive information about the various types of meanings conveyed by a subjective expression. The three main subsystems of Appraisal theory are:

- **attitude**, referring to feelings as they are construed in texts by distinguishing between emotion, ethics, and aesthetics; the values by which speakers pass judgements and associate emotional/affectual responses with participants and processes;
- **engagement**, providing resources for positioning the speaker's/author's voice with respect to the various propositions and proposals conveyed by a text;
- and **graduation**, describing the resources that allow for graduating the interpersonal impact of an expression (White, 2020).

The framework could be summarized as a comparably extensive tree of choices providing information on various aspects of subjectivity (Taboada, 2017).

2.1.1. Attitude

For the annotation of our corpus, we have selected the attitude part of Appraisal theory. The subsystem of attitude according to Appraisal theory (Martin and White, 2005) provides a framework for the analysis of evaluative expressions by categorizing them into three main attitude types, being an *affect* (an emotional reaction), a *judgement* (a reaction of ethical evaluation), or an *appreciation* (an evaluation of aesthetics), as well as attitude polarity, attitude force, and explicitness. Each category is then subdivided into its own tree of choices making the system a complex and informative structure.

The authors offer a variety of subcategories within each of the types of attitude, which allows for detailed expression of subjectivity. In our approach, we decided to focus on the three main subcategories, namely **affect, judgement, and appreciation** and their types, as well as categories of **sentiment polarity**, and **explicitness**. Our approach to the attitude framework is presented in Table 1.

In our experience, the range of parts of speech that the attitudes could be expressed with include

mainly adjectives (*proper*), verbs (*violate*), and adverbs (*interestingly*), but also other parts of speech, while the annotated sequence may range from one token to a whole sentence – especially in case an attitude was expressed in an implicit way. However, as per Martin and White (2005) the borders of an attitude may be spread across a discourse unit, irrespective of grammatical boundaries.

Resource	Type
Affect *expression of one's feelings*	happiness security satisfaction inclination
Judgement *attitude towards behaviour*	normality capacity tenacity veracity propriety
Appreciation *evaluation of semiotic and natural phenomena*	impact quality balance complexity valuation
Polarity	positive negative
Explicitness	inscribed invoked

Table 1: Overview of the selected aspects of the attitude system based on Martin and White (2005)

2.2. Approach to data selection

The corpus of annotated speeches consists of 100 texts that were manually selected from the UN Security Council Debates dataset (Schoenfeld et al., 2019). The language of the data is English, and the speeches were either originally presented in English, or included in a form of the official UN translations. The information about the original language of the speech, as well as the speaker's affiliation and sex, the topic of the session, and its year are stored in the metadata of each text.

The text selection was based on certain criteria, to ensure the data represent diplomatic discourse of the given time period in a balanced way.

The first criterion for the data selection was the *topic* of the meeting. We have decided to focus on international military conflicts at the turn of the century, and among those that are present in the dataset the following topics were selected given their representation within the period of time, covered in the dataset:

- the Palestinian topic, comprising the Israeli–Palestinian conflict;
- the Yugoslavian topic, comprising the meetings dedicated to the Yugoslav wars;
- the Ukrainian topic, comprising the meetings dedicated to discussing the Russo-Ukrainian war;
- the Georgian topic, comprising the War in Abkhazia of 1992-1993, as well as the Russo-Georgian War of 2008;
- and the Iraqi topic, comprising the discussions of the 9/11 terrorist attack (2001) and the subsequent Iraq War, as well as the Gulf War.

Each topic was devoted an equal proportion of space within a corpus, which means that we have selected 20 speeches from the meetings devoted to discussing each of the topics.

The second criterion is connected to the selection of particular meetings that would be representative of the topic. After we have grouped the available speeches and according to the topic, we have selected the meetings that would be included in the corpus. At this stage, our aim was to ensure that the corpus is representative of various stages of each of the included conflicts. Each topic is therefore represented by four sessions of the Security Council, spanning within the given conflict and dataset time frame.

The third criterion in speech selection was the speaker's presumed position towards the topic under discussion. We have differentiated between three types of speakers, namely

- the representatives of the countries that are directly participating in the conflict;
- their allies (if possible among permanent members of the Security Council);
- and a representative of a state, whose international political interests appear to be further from the discussed events (typically among non-permanent UNSC members).

The combination of the three criteria allowed for the creation of a more balanced corpus containing various appraisals of the selected topics, and focused on international armed conflicts of the selected time period.

2.3. Annotation process

1. The first trial annotation was completed by two non-native English speakers with background in linguistics, one of whom is among the authors of the presented paper (annotator A and annotator C). The annotators were instructed to follow the description of attitude subtypes and polarity from Martin and White (2005) and Oteíza (2017). The annotations were conducted following an xml-like

scheme that is described in Table 1, except for the categories of explicitness, which were not yet added to the framework. The achieved dataset consisted of ten speeches with double annotations (around 10000 tokens).

2. This step was followed by calculating the inter-annotator agreement to assess the reliability of the first version of the annotation scenario. The assessment was conducted according to the three levels of depth of the attitude scheme:

- The *complete* agreement, if the annotators agree on the presence of the attitude, attitude-type, subtype, and sentiment polarity on the exact segment of the text;

  the F1 for this category is 0.265.

- the *core* category refers to the agreement on levels on annotators agree on the presence of an attitude, and attitude-type;

  The F1 for the core agreement is 0.691.

- and the results for the *general* category refer to agreement on the presence of an attitude;

  the F1 for the general agreement is 0.713.

Results of this experiment supported the hypothesis that even though subjectivity identification task is complicated there would be quite high agreement between the annotators, whereas the more fine-grained categories may need further development to be understood uniformly.

3. After analyzing the agreement and comparing the annotations, we have proceeded with the creation of the second version of a formal annotation scenario.

4. 80% of data (60490 tokens) was then annotated again by one annotator (annotator A) in the selected environment (see Section 2.3.2) according to the updated annotation guidelines which led to their further improvement. In addition to the above-mentioned improvements, it was decided to add the dichotomous category of explicitness to further enrich the corpus.

5. After the annotation scenario was updated, we have proceeded with the annotation of the whole corpus.

Similar to the very first experiment, the annotations were completed by two annotators with background in linguistics. Both of the annotators are non-native English speakers with a high command of this language.

The main annotator (annotator A) has had the task of annotating the whole dataset (105592 tokens), whereas the annotator B has annotated a smaller subset of texts (ca. 10000 tokens) with the aim of the inter-annotator agreement estimation.

2.3.1. Annotation scenario

The annotation scenario was developed for intrasentential annotation of the attitudes in the diplomatic speeches of the United Nations Security Council. The document provides an extensive step-by-step description, which guides an annotator through the following annotation steps:

- Attitude identification: It first provides two approaches to attitude identification, namely:

  1. identifying attitudes by first identifying all of the available subjectivity meanings, which relies on SentiWordNet (Baccianella et al., 2010) for the identification of explicit attitudes;
  2. and a context-dependent approach, which requires annotators to first read the whole contextual unit (a sentence) and decide on the presence/absence and the borders of an attitude based on their subjective perception of a text.

     This approach allows capturing various implicit expressions of subjectivity as for instance *"That evaluation has been transformed into a brutal reality"*, in example of Judgement, that would be perceived as Affect if not analyzed together with the surrounding context.

- Identification of attitude explicitness

The annotators are asked to distinguish between the explicit and implicit attitudes, as in the following examples of text fragments, annotated with the category of affect: *"We are concerned"* as opposed to *"I would like to use this opportunity to express our serious concern"*.

- Identification of attitude sentiment polarity

At this stage of annotation, the annotators are asked to decide if the attitude conveys positive or negative sentiment, as in the following opposition of positive and negative appreciation excerpted from the corpus: *"the best"* as opposed to *"the most challenging"*.

- Select the appropriate length of the annotated fragment

Depending on the context, it may be necessary to annotate units, which are larger than one token to capture the appraisal-bearing meanings (Read et al., 2007). We advise deciding on the appropriate fragment length based on the attitude explicitness. In this approach, the annotated fragment would either include only the tokens that express the meaning of an attitude in a direct explicit way (the *inscribed* tag), or allow for the inclusion of all

tokens that are required to fully capture the meaning if an attitude is expressed implicitly (the *invoked* tag).

Let us take a look at this distinction by the following examples of the fragments annotated as judgement. The inscribed judgement may take as little as one token to fully capture the meaning (as in *"tireless"*) whereas an invoked judgement requires more context (as in *"The conflicts that have raged over the past few days must be completely stopped"*).

- Select between the three main categories and their subtypes

Annotators were to choose between a variety of categories (first presented in the Table 1). One of the challenges of this type of annotation is the fact that the diplomatic attitudes often differ from the textbook examples (Martin and White, 2005), therefore the annotators were provided with detailed descriptions of attitude, judgement, and appreciation together with their subcategories, as well as the observed doubtful annotation cases.

2.3.2. Annotation tool

During the corpus design stage, we have considered various available annotation tools, which would be compatible with our annotation scenario.

Our initial requirements were:

- support of span annotation, preferably allowing to annotated fragments to overlap as well as span over unannotated tokens;
- support of tree-like annotation schemes;
- convenient import and export of documents;
- support of MacOS or Linux;
- convenient format of the exported documents and annotations, as this would matter at the stage of annotation analysis.

After considering various annotation tools that were available at the time and conducting test annotations, it was decided to proceed with the doccano annotation tool (Nakayama et al., 2018). Doccano is a web-based open-source annotation tool that supports sequence labelling and allows the creation of one's own annotation scenarios. It also provides basic statistics and supports auto-labelling. Another useful feature of this tool is collaborative annotation.

For our project, we have selected the sequence labelling annotation type, together with an additional feature that would allow overlapping entities.

3. Corpus analysis

3.1. Corpus statistics

The corpus consists of one hundred manually annotated speeches, namely of 105592 tokens and 7296 types. The total number of sentences in this corpus is 3296. On average, one text in the corpus consists of 33 sentences, while the average length of a sentence is 32 tokens.

The metadata includes:

- the speaker's **name**;
- their **gender** (title-based distinction);

The ratio of female to male speakers in the corpus is 7 to 93. This study does not focus on gender-specific aspects of the diplomatic discourse, however, our dataset shows, that diplomacy is still a mainly male-dominated area, therefore the number of speeches from women diplomats is much lower.

- the **country** or institution represented;

The full list of all the affiliations alongside the number of their texts is available in Table 2. Most of the speeches belong to permanent UNSC members, such as the United Kingdom, the United States of America, the Russian Federation, France, and China. The dataset includes speeches affiliated with the main countries, connected to the selected topics: Palestine, Georgia, Ukraine, Iraq, and Palestine. It was, however, impossible to include them on the same scale, as none of them are permanent UNSC members.

Country/Organization	Number of texts
Argentina, Brazil, Czech Republic, Secretary-General or Deputy Secretary-General, IAEA Director, Ireland, Italy, Jamaica, Japan, Lebanon, Nigeria, Pakistan, Republic of Korea, Romania, Turkey	1
Bosnia and Herzegovina, Iraq, Yugoslavia, Croatia, Jordan, Syrian Arab Republic	2
Israel, Palestine	3
Georgia	5
Ukraine	6
France, China	9
United Kingdom	11
United States of America	12
Russian Federation	15

Table 2: The distribution of countries and organizations within the corpus

- the **topic** of the meeting (as per the UNSC meeting records);
- the **conflict** the meeting was devoted to;
- the **year** of the meeting;

The number of selected speeches in relation to the chronological measures of the UNSC dataset (Schoenfeld et al., 2019) is represented in Figure 1. The corpus aims at covering the time span of the conflicts, although the surge in the number of texts is always connected to real-life events and their discussions.

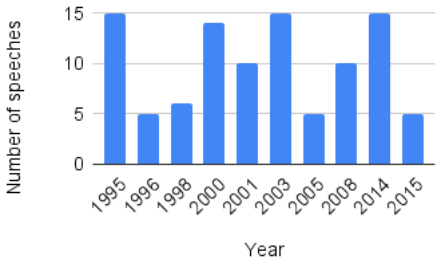

Figure 1: Distribution of the texts over the years

- the meeting identifier;
- the speech identifier.

3.2. Annotation statistics and their interpretation

The following subsection provides information on annotation statistics of the gold data provided by the annotator A.

The total number of attitudes in the CoDipA UNSC 1.0 texts is 1938, with an average of 1.7 attitudes per sentence. The three main categories are represented in the following way:

- Affect: 422 instances
- Judgement: 980 instances
- Appreciation: 536 instances

The average length of an annotated fragment varies from 4.3 tokens for affect to 1.8 tokens for appreciation, and 17.8 tokens for judgement, while the overall average length of a fragment for the corpus is 10.5 tokens. The difference in length corresponds to the preference of either inscribed or invoked modes of expressing subjectivity, with inscribed fragments spanning on average over 2.77 tokens, whereas the invoked fragments - over 16.1.

Let us now take a closer look at the distribution within each of the main categories throughout the corpus.

The distribution of the subcategories of affect is shown in the Figure 2. The two most prevalent categories representing emotional response are *inclination* and *satisfaction*.

Prevalence of the inclination signifies the importance of expressing the diplomats' preferences within the discussed context (if they *incline* and support the events under discussion or other people present during a Council Session), while the category of satisfaction is commonly utilized in the first paragraph of a speech to express the diplomats' emotions towards the other participants of the meeting (with 76.8% of all instances conveying positive sentiment).

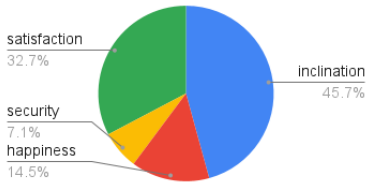

Figure 2: Distribution of the subcategories of affect

The most prevalent subcategory within the judgement subsystem is propriety (Figure 3), which constitutes more than a half of all annotated occurrences. This subcategory is utilized to mark the ethics of the other's (or self) behaviour and belongs to judgement type of social sanction. A curious distribution of the sentiment polarity within the text spans that were identified as propriety (62.7% of tags are positive, and 37.3% are negative) lead us to conclude that the Council members are more interested in advising others on the appropriate course of actions rather than criticize their intentions or behaviour, or praise their and their allies' decisions and actions.

36% of judgements of propriety are formulated by using modalities of ability, permission, obligation, and advice, as in *"the war must be stopped", "our Council should be seized of the matter", "we must demonstrate that we are capable"*.

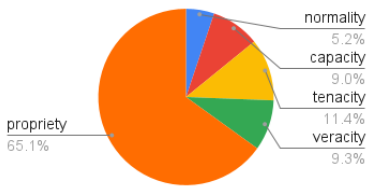

Figure 3: Distribution of the subcategories of judgement

The subcategories of appreciation are represented in a comparatively more diverse way (Figure 4). Here, the three most frequent subcategories are valuation (33.6%), complexity (24.8%), and quality

(20.3%). The entities are being assigned a subjective evaluation based on how valuable they are (value), how well they are put together or are hard to follow (complexity), as well as based on personal preference (quality). In the diplomatic discourse of the UNSC, a part of these expressions is constituted by a set of diplomatic cliches, which repeatedly occurred throughout the corpus (for instance, *"grave consequences", "clear violation", "comprehensive and just solution"*, etc.) and is constituted by a sentiment-bearing adjective. [2]

The sentiment polarity of the appreciation category is rather positive (68.7% of positive entities and 31.3% of negative entities).

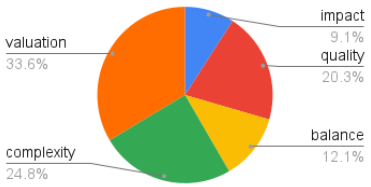

Figure 4: Distribution of the subcategories of appreciation

On a corpus level, the positive evaluations prevail over the negative ones with 64% of all evaluations being positive and 36% being negative. In our opinion, such a sentiment distribution does not completely go in line with the selected topics and could be explained by positive sentiment prevailing in the speeches of the diplomats, who do not represent the countries directly involved in the selected international military conflicts, as well as sufficient amount of subjectivity being directed towards praising themselves and their allies (as in *"the Secretary-General and his Special Envoy have made tremendous efforts"*.

3.3. Inter-annotator agreement

As our corpus was annotated by two annotators, we have selected the Cohen's Kappa as a measure of inter-annotator agreement. The results of the inter-annotator agreement evaluation are presented in the Table 3.

The agreement refers to a sentence-level comparison representing the presence or absence of a designated label in each sentence of a text. The results of the experiments suggest that the inter-annotator agreement is:

- **Fair** for Attitude identification and the subcategory of Affect;
- **Moderate** for the subcategories of Judgement and Appreciation.[3]

	Cohen's kappa
Attitude	0.41
Affect	0.44
Judgement	0.31
Appreciation	0.32

Table 3: Inter-annotator agreement

The agreement level reflects the difficulty of the task as well as the unavoidable subjectivity of attitude evaluation. After a careful manual evaluation of the double annotations,[4] we have concluded that most cases of disagreement stem from an absence of annotation, which underlines the problematic nature of the attitude identification process as mentioned by Fuoli (2018). However, when annotators do agree on the presence of an attitude in a sentence (with possible small variation in a number of tokens selected) they tend to agree on both on attitude polarity (94.8%), and on the attitude type (79.1%).

3.3.1. Exploring confusion matrices and the main cases of inter-annotator disagreement

Let us take a look at the confusion matrices to summarize the major cases of inter-annotator agreement and disagreement.

As the Figure 5 shows, the annotators mostly agree on the sentences, where they both detect presence of attitudes (93 instances), whereas the biggest disagreement comes from Annotator A not detecting attitudes in sentences, where the Annotator B does (53 instances).

Now, let us illustrate the agreement for three main subcategories of the attitude system with confusion matrices for affect 6, judgement 7, and appreciation 8.

Within the matrices, it is visible that the annotators generally tend to agree on the absence of a tag in a sentence.

The best agreement is observed for the category of affect, while the agreement for the categories of judgement and appreciation is much lower. The reason for the relatively low agreement in general, stems from the subjectivity of the assigned task: it may often appear doubtful which level of semantic meaning should be chosen for the annotation.

Another reason for the lower agreement of the judgement and appreciation is connected to the

[2] A diplomatic cliche is an expression that is meant to support the topic of a speech in a standardized way. Such expression could constitute a greeting, an expression of one's condolences, etc.

[3] The evaluation of the agreement was derived from the classification described by Koch and Cruz (2004).

[4] The detailed results of the manual analysis of the double annotations will be published separately.

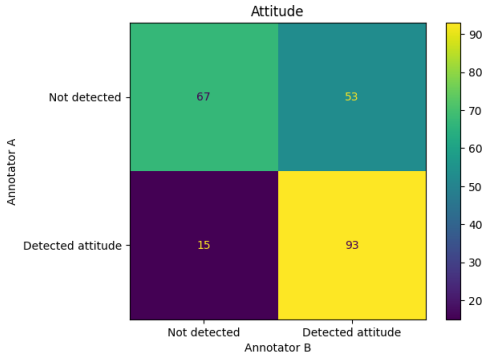

Figure 5: Confusion matrix for the category of attitude

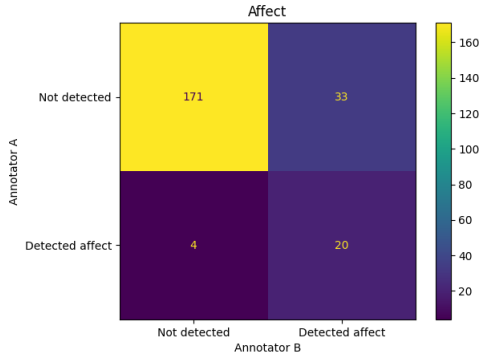

Figure 6: Confusion matrix for the subcategory of affect

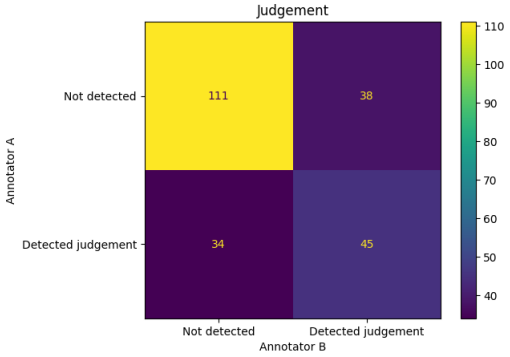

Figure 7: Confusion matrix for the subcategory of judgement

fact that these meanings are often less direct in the diplomatic communication. Affect is still being represented in a way that is quite close to canonical representation of this category (Martin and White, 2005), whereas judgement and appreciation are very often represented in an implicit form, with hidden indirect meanings.

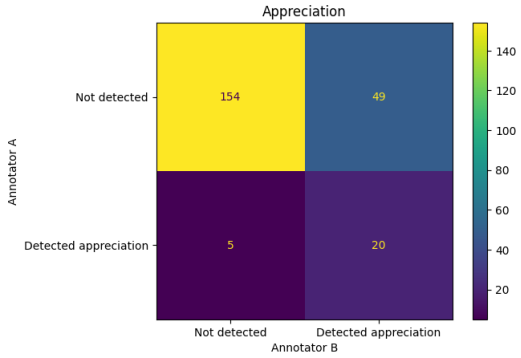

Figure 8: Confusion matrix for the subcategory of appreciation

4. Conclusion

This work has introduced the CoDipA UNSC 1.0, a new language resource stemming from Schoenfeld et al. (2019) that provides the data and the framework for analyzing attitudes in diplomatic texts based on Appraisal theory (Martin and White, 2005).

Our corpus offers the annotated dataset that not only proves that the usage of attitudes is consistent throughout the texts, and suggests that the diplomatic texts are highly subjective and evaluative, but also allows for finer-grained attitude analysis based on topical, temporal, and functional criteria.

The most quantitatively significant means of expressing an attitude in the diplomatic speeches of UNSC is judgement, as this category occurs almost two times more often than the other two. The sentiment polarity of the annotations suggests that even though the selected meetings were devoted to discussing the ongoing armed conflicts, diplomats tend on average to keep the positive appearance. This may be explained by various reasons, such as the parties of a conflict being non-prevalent in the corpus, quite significant amount of self-praise, or have other, non-linguistic, explanations.

In our future work, we will focus on further analysis of the obtained language resource from the point of view of possible typical combinations of the attitudinal categories in a text, as well as train a classifier to distinguish between the types of attitudes automatically.

5. Acknowledgements

The research described in this paper has been funded by the Project of the Czech Science Foundation “LuSyD” (No. GX20-16819X), and partially supported by SVV project number 260 698.

6. Ethical considerations and limitations

The constraints of this work lie in its limited scope tied to specific selection criteria, the potential subjectivity inherent in manual annotation, and the likelihood of biases stemming from the selective criteria affecting representational comprehensiveness and introducing variability.

7. Bibliographical References

Mariia Anisimova. 2021. An introductory overview of evaluating facts and attitudes in diplomatic discourse. In *2nd Workshop on Automata, Formal and Natural Languages – WAFNL 2021 Open Session Proceedings*, pages 1–4. Faculty of Mathematics and Physics, Charles University, Prague.

Mariia Anisimova and Šárka Zikánová. 2022. Attitude in diplomatic speeches: a pilot study. In *3rd Workshop on Automata, Formal and Natural Languages – WAFNL 2022 Open Session Proceedings*, pages 151–158.

Nicolò Busetto and Rodolfo Delmonte. 2019. Annotating shakespeare's sonnets with appraisal theory to detect irony. In *Italian Conference on Computational Linguistics*.

Matteo Fuoli. 2018. A step-wise method for annotating appraisal. *Functions of Language*, 25(2):229 – 258.

Michael A.K. Halliday. 2004. *An Introduction to Functional Grammar*. Arnold, London, UK.

Sabine Koch and Robyn Cruz. 2004. *Issues of validity and reliability in the use of movement observations and scales.*

Suet Ling Lam and Peter Crosthwaite. 2018. Appraisal resources in l1 and l2 argumentative essays: A contrastive learner corpus-informed study of evaluative stance. *Journal of Corpora and Discourse Studies*.

Yun Lian. 2018. Analysis of xi's diplomatic speeches from the perspective of appraisal theory. *Journal of Language Teaching and Research*, 9:759–764.

James R. Martin and Peter R. R. White. 2005. *The Language of Evaluation: Appraisal in English*. Springer.

Hiroki Nakayama, Takahiro Kubo, Junya Kamura, Yasufumi Taniguchi, and Xu Liang. 2018. doccano: Text annotation tool for human. Software available from https://github.com/doccano/doccano.

Teresa Oteíza. 2017. The appraisal framework and discourse analysis. In *The Routledge handbook of systemic functional linguistics*, pages 481–496. Routledge.

Jonathon Read, David Hope, and John Carroll. 2007. Annotating expressions of appraisal in english. In *Proceedings of the Linguistic Annotation Workshop*, LAW '07, page 93–100, USA. Association for Computational Linguistics.

Nick Stanko. 2001. Use of language in diplomacy. In H. Slavik J. Kurbalija, editor, *Language and Diplomacy*, pages 39–48. DiploProjects, Mediterranean Academy of Diplomatic Studies, University of Malta, Msida MSD 06, Malta.

Elizabeth Swain. 2017. Interpersonal style(s) in diplomatic argumentation online a study of argument schemes and evaluation in press releases of unsc permanent members. In G. Garzone C. Ilie, editor, *Argumentation across communities of practice : multi-disciplinary perspectives*, pages 127–148. John Benjamins Publishing Company, Amsterdam, Netherlands.

Maite Taboada. 2017. System network for appraisal.

Gholam-Reza Tajvidi and S. Hossein Arjani. 2017. Appraisal theory in translation studies: An introduction and review of studies of evaluation in translation. *Journal of Research in Applied Linguistics*, 8:3–30.

Peter R. R. White. 2020. The appraisal website.

Siyou Zhang and Zhongwen Pei. 2018. Analysis of political language based on appraisal theory: The mutual construction of language and powermtaking xi jinping and donald trumprs speeches at world economic forum as examples.

8. Language Resource References

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining.

Varada Kolhatkar, Hanhan Wu, Luca Cavasso, Emilie Francis, Kavan Shukla, and Maite Taboada. 2020. The SFU Opinion and Comments Corpus: A Corpus for the Analysis of Online News Comments. *Corpus Pragmatics*, 4:155–190.

Mirco Schoenfeld, Steffen Eckhard, Ronny Patz, Hilde van Meegdenburg, and Antonio Pires. 2019. The UN Security Council Debates.