Problematic cases of attitude annotation in diplomatic speeches

Mariia Anisimova[1,*], Šárka Zikánová[1]

[1]Charles University
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics, 6 Malostranské náměstí 25, 118 00 Praha, Czech Republic

Abstract

This paper provides an overview of challenges that occurred during the creation of the annotation scenario for the attitude annotation in diplomatic speeches of the UNSC. The scenario followed the attitude part of the Appraisal theory. The various challenges in annotating the speeches such as the extent of arguments, identification of attitude in verbal forms, and complex structures, were classified, and, in part, resolved. This paper would be useful for anyone considering this type of attitude analysis when working with diplomatic and political texts.

Keywords

Appraisal theory, attitude analysis, manual annotation, diplomatic discourse, corpus linguistics

1. Introduction

The research on annotating attitudes [?] and its various categories has long been an ongoing process [?], [?]. As described before, annotating attitudes is a complex procedure not only due to elaborated annotation schemes but also due to the lack of definitive criteria for the identification and categorization of attitudes and the other appraisal labels.

Another problematic side of this annotation type is that the researchers often omit the step of creating a formalized annotation scenario, which is why motivation and consistency of their choices are hard to follow [?]. This, in turn, causes issues with testing the consistency of the annotations, and test-retest reliability [?].

This study aims to present a discussion on the issues observed while creating the discourse-specific annotation scenario suited for annotating types of attitudes and their subcategories as defined by [?]. We classify the issues and particularities observed during the creation of the scenario as well as the annotation process and compare them to the issues observed by [?]. We then offer our perspective on the choice of a possible solution and discuss further experiments.

Diplomatic speeches form a very particular and peculiar group of texts that are very different from other types of discourse. The prominent characteristics of these texts are the understated tone [?] and indirectness. 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. It is also because of them, that the diplomatic attitudes form a separate group of attitude-bearing expressions and require a comprehensive approach in the process of annotation.

In our previous publications, 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 batch of annotation [?], which was then utilized for redefining the annotation scenario based on problematic and unclear cases of annotation. In the most recent addition to our project [?], we have also presented the Corpus of Diplomatic Attitudes (CoDipA 1.0), which includes a description of the selected speeches from the United Nations Security Council annotated with the attitude part of the Appraisal theory and expands on corpus and inter-annotator statistics, such as our calculations and commentary on the Cohen's kappa which varies between 0.44 and 0.32 depending on the selected category.

The objectives of this research are to define and exemplify the challenging parts of the annotation scenario and offer solutions for the observed issues to make the annotation process reliable and stable, resulting in a satisfactory inter-annotator agreement. The clarity and reproducibility of the annotation process would also help further automatize the processing of attitudes.

The structure of the paper is outlined in three main

ITAT (Information technologies – Applications and Theory), Workshop on Automata, Formal and Natural Languages 2024 (WAFNL 2024)
*Corresponding author.
anisimova@ufal.mff.cuni.cz (M. Anisimova);
sarka.zikanova@mff.cuni.cz (Š. Zikánová)
https://ufal.mff.cuni.cz/mariia-anisimova (M. Anisimova);
https://ufal.mff.cuni.cz/sarka-zikanova (Š. Zikánová)
0000-0002-2478-2815 (M. Anisimova); 0000-0002-7805-9649 (Š. Zikánová)

sections, namely the brief description of the annotation scenario, the detailed description of the issues observed, a brief outline of the future work, and the main conclusions. Sections 5 and 6 correspond to the Limitations and Ethical considerations of the presented research.

2. Description of the annotation scenario

The annotation scenario was developed to annotate the attitudes [**?**] in the diplomatic speeches of the United Nations Security Council [**?**]. The scenario first offers a brief introduction to the Appraisal theory by first describing its three main subsystems: attitude, engagement, and graduation. The three subsystems refer to the description of our feelings and emotional reactions, sourcing of attitudes and the play of voices around opinions in discourse, and grading the phenomena [**?**]. It then focuses on the detailed description of the subsystem of attitude, which was selected to define emotion and subjectivity in diplomatic speeches. Finally, a practical how-to guide for setting up the annotation environment in the doccano annotation tool [**?**] is offered along with a set of annotation labels and references.

The subsystem of attitude according to the Appraisal theory provides a framework for the analysis of evaluative expressions by categorizing them as being an affect (an emotional reaction), a judgement (an expression of ethical evaluation), or an appreciation (an evaluation of aesthetics) and defining their polarity and explicitness. Each category is then subdivided into a separate tree of choices making the system a complex and informative structure.

The dataset that is to be annotated, is also described in detail in the scenario. It consists of a corpus of 100 diplomatic speeches, selected according to a set of criteria that allows for proportional representation of diversity according to the topic of a speech, the year of the meeting, and the country the speaker represents. The language of the data is English, and the speeches were either originally presented in English, or were 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 corpus consists of 105592 tokens and 7296 types, and the average length of a sentence is 32 tokens.

Our initial goal was for the guidelines to be of a quality that would allow for the full disambiguation of the task, which means the annotators would select the same text span and assign completely identical labels to it in terms of attitude type and subtype, sentiment polarity, and explicitness. We therefore apply the strictest approach to annotations assessment at the time as it is more convenient to start with, while gradually adapting the evaluation criteria to the task.

2.1. On identifying attitudes and the scope of annotated fragment

The pilot annotation task completed to support the scenario creation has first shown a set of challenges, which vary in nature, and difficulty and have proven to be either general to this type of text analysis [**?**], or particular to the chosen text type.

The first task that the annotator should be able to do is to find the attitudinal fragments. This task is not straightforward and lacks direct instructions and explicit description in the original Appraisal theory documentation [**?**]. This has been posing a general problem for researchers working with various discourse types [**?**], in diplomatic discourse this issue is also of critical importance.

2.2. Attitude identification in diplomatic texts

From our empirical experience, there can be several approaches to the process of identification of an attitude.

It is, first of all, possible to judge the presence or absence of an attitude based on the token's polarity, as if judging each token individually by deciding if it may signify the presence of an attitudinal expression. Such an approach, however, does not allow for capturing all of the available attitudinal meanings. Let's take a look at Example 1. and analyze it by first highlighting all of the positive and negative connotations available to estimate if the available sentiment in separate tokens correlates with the identification of expressions of attitude.

Example 1: *Even items subject to control would go to Iraq once there is [confidence: POS] that they would not be used to rebuild Iraq's [[weapons: NEG] of [mass [destruction: NEG]: NEG]] or [[improve: POS] its military capabilities].*

Such complex lexical structures may contain multiple layers of semantic connotations, which may lead to the annotator's confusion due to the superposition of the labels. The annotators may be drawn to find the attitudinal meaning of every available token conveying attitudinal polarity which often leads to viewing an annotated fragment in isolation of the broader text meaning.

Another approach to attitude identification is based on the subjective evaluation of a textual fragment under consideration based on the contextual boundaries of the syntactic structure that bounds it, as well as the context of the whole text (which in the case of our study is a speech).

Let us take another look at the same sentence, where the possible locations of attitude are defined by its subjective interpretation and context of the whole structure:

Even items subject to control would go to Iraq [once there is confidence that they would not be used to rebuild Iraq's weapons of mass destruction or improve its military capabilities: JUDGEMENT-INVOKED].

Here, the speaker provides an implicit evaluation of the future possible behaviour of Iraq, expressing their hope and inclination towards Iraq not rebuilding its weapons of mass destruction. This scenario is presented as a subjective evaluation of the possible future behaviour, which means that the attitude in question should be perceived as an implicit evaluation of the desired appropriate action of Iraq.

Even though this way of finding attitudes leads us to a somewhat more vague definition of the borders of the annotated expression, it also provides us with more contextual meaning as opposed to the cumulative polarity of separate tokens.

Let us exemplify this approach again with the Example 2:

Example 2: *[A number of these have been under discussion for some time in this Council: AFFECT-INVOKED], and [if we were able to agree on this package, it would be an important step forward in that regard as well: JUDGEMENT-INVOKED].*

In this example, the speaker implies their emotional state of unsatisfaction by the amount of time that an issue is already under discussion while providing their implicit evaluation of the appropriate course of future actions.

Our preferred approach is to judge the span of the annotated fragment by the context and the explicitness of the evaluation, which would lead to a clearer logic and accountability of the annotation process.

2.3. On identifying the borders of inscribed and invoked attitudes

Attitude examples observed in the corpus were proven to make use of both explicit and implicit (inscribed and invoked) ways of expressing an attitude. The diplomatic discourse is known for its indirectness [**?**] and a particularly subtle expression of subjectivity. However, the inscribed and invoked attitudes are not the same regarding the additional context needed for discerning the message's meaning.

The inscribed attitudes only require the explicitly subjective tokens to be annotated, as shown in Example 3:

Example 3:*France [fully supports: AFFECT-INSCRIBED] the search for a political solution.*

This decision would lead to a sufficient increase in the expected average length of the annotated fragment for the invoked annotated attitudes in comparison to the inscribed attitudes.

Our main solution for annotating invoked attitudes is to capture as much meaning as possible by annotating as few tokens as needed.

Let's analyze an excerpt from one of the diplomatic speeches provided in Example 4:

Example 4: *Experts in our countries must immediately begin to analyse it and then draw the appropriate conclusions.*

Here, there is a definite presence of positive judgement coded by the directive 'must'. *experts must begin to analyse* would be our first choice to annotate. The problem is, such a fragment, if non-explicit, does not convey any attitude on its own, therefore inclusion of the whole context necessary for understanding of the chosen label is advised: beginning of analysis is only good if it leads to drawing appropriate conclusions.

2.3.1. On annotating articles

Another required decision is related to the annotation of determiners. As most of the annotated entities constitute either an evaluative adjective or a collocation of an evaluative adjective and an evaluative noun, it is necessary to decide whether or not articles defining the evaluative phrase should be included in the annotation or not.

Our solution to this question is related to the overall viewpoint on the attitude framework and depends on whether the particular article is important for evaluating the expression under consideration.

If an attitude is expressed explicitly (**inscribed** tag) as in 'a *capable* management', 'we are *happy* about the resolution', 'the *valuable* lesson', etc.) and the perceived meaning of the adjective+noun collocation is enough to understand it in the full, annotation of any additional tokens should be omitted. Usually, it is enough to only mark one token or collocation that is used for expressing an attitude. There is no need to include any additional tokens in the annotated fragment, for example, if an attitude is found within a phrase [they have thought of a *wonderful* solution], we would advise only marking the evaluative adjective and omit to annotate the article that precedes it and the noun that follows. We also advise against article annotation when a noun that is in collocation with the attitude-bearing adjective has no attitudinal meaning itself (*the unfortunate circumstances*).

However, if an attitude is a part of the superlative form (*the best*) or is expressed implicitly and the broader context is necessary for understanding the expression, any needed number of tokens within one syntactic clause could be considered (as in *the veto today will not prevent that*).

2.4. Overlapping attitudes

Another situation that may cause confusion and subsequent mistakes in annotation includes the cases, where the two annotated fragments overlap. An overlap happens if one attitude occurs within the scope of the an-

notated fragment of another attitude. This may happen in many ways, however, if there are only two attitudinal elements, the overlap may look one of the following ways:

1) The first option is for the second label to overlap with the first one as in *(a(b))*;

2) The second option is for the first element to overlap with the second one as in *((a)b)*;

3) The third option is the inclusion of an element (b) in the middle of the first element (a) following a scheme (a(b)a) as in Example 5:

Example 5: *There is no doubt that we all* ***want [Affect-inclination]*** *to resolve this problem [Judgement-tenacity].*

Here, the whole sentence is an invoked judgement with an inclusion of an inscribed affect (in bold) corresponding to the expression of the speaker's desire.

To technically deliver this solution it is necessary to allow for overlapping annotation in doccano [**?**] when creating a project.

2.5. Interrupted attitudes

The issue of interrupted annotation refers to a situation, where a span of attitudinal text includes a sequence of non-attitudinal tokens. Annotating unnecessary tokens may lead to a decrease in the quality and reliability of the future corpus. As per our observations, such an annotation scenario occurs solely with the explicitly formulated attitudes, therefore deciding on this type of label is advised as a first step.

Here is an example of an interrupted attitude (Example 6), an excerpt from another diplomatic speech. An invoked judgment of the previous lack of actions of the Security Council is interrupted by a referral to one of the resolutions, which does not add any additional attitudinal meaning to the annotated fragment and should therefore be excluded from it.

Example 6: *Unfortunately, in the past hundred days, the very [limited suspension of the sanctions: JUDGEMENT-INVOKED-PART1] established by the Security Council resolution 943 (1994) [has also not been entirely fulfilled: JUDGEMENT-INVOKED-PART2].*

Our solution to resolving it is purely technical and includes the creation of an additional label [none] that is to be assigned to the tokens excluded from the annotated fragment.

3. Related ongoing and future work

The acquired annotation scenario has helped us further define our approach to the category of attitude as it is understood by the Appraisal theory [**?**] and adapt it to the needs of the diplomatic discourse analysis, which has led to a parallel annotation experiment as well as to the creation of a background for the newly-published corpus of evaluation in diplomatic speeches of the UNSC (CoDipA 1.0).

We are specifically interested in investigating the attitude development processes throughout the selected time frame and in comparing the findings on an inter-conflict scale. Another expected development is the practical application of the acquired data on a fine-tuned large language model (such as Bert or GPT 4) to establish whether there is a potential for expanding the analysis to a bigger scale.

The results of both quantitative and qualitative evaluation of the annotation outcomes are expected to be published as well.

4. Conclusions

In this research paper, we have outlined some of the challenging parts of annotating attitudes according to the Appraisal theory scheme [**?**] in the diplomatic speeches and provided our viewpoint on their resolution, as well as technical comments on how to implement this type of annotation in practice.

The first problem that an annotator would encounter if they would seek attitude in diplomatic speeches, would be attitude identification and deciding on the textual borders of an attitudinal fragment. We have outlined these processes and provided examples from our corpus.

In the next sections, we have offered an overview of the technically challenging parts of the attitude annotation process, such as overlapping and interrupted types of attitudes, as well as their solution.

Limitations

This paper presents an up-to-date analysis of the ongoing research based on the annotation project, including the development of a discourse-specific annotation scenario. The findings from the presented annotation scenario should not yet be considered as final, they may be updated in the final version of the scenario before its publication.

Ethics Statement

We honour the ethical code set out in the ACL Code of Ethics and there are no special ethical issues involved during the creation of this research paper.

5. Acknowledgements

The research reported in this paper was supported by 207-01/207PROV and the Czech Science Foundation (project no. 24-11132S, Disagreement in Corpus Annotation and Variation in Human Understanding of Text), and partially supported by SVV project number 260 698; a part of the used data comes from the project no. LM2018101 by the Czech Ministry of Education, Youth and Sports (Digital Research Infrastructure for Language Technologies, Arts and Humanities).

References

[1] F. Mráz, D. Pardubská, M. Plátek, J. Šíma, Pumping deterministic monotone restarting automata and DCFL, in: M. Holeňa, T. Horváth, A. Kelemenová, F. Mráz, D. Pardubská, M. Plátek, P. Sosík (Eds.), Proceedings of the 20th Conference Information Technologies – Applications and Theory (ITAT 2020), volume 2718 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 51–58. URL: http://ceur-ws.org/Vol-2718/paper13.pdf.

[2] M. Plátek, F. Mráz, D. Pardubská, D. Průša, J. Šíma, On separations of LR(0)-grammars by two types of pumping patterns, in: B. Brejová, L. Ciencialová, M. Holeňa, F. Mráz, D. Pardubská, M. Plátek, T. Vinař (Eds.), Proceedings of the 21st Conference Information Technologies – Applications and Theory (ITAT 2021), volume 2962 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021, pp. 140–146. URL: http://ceur-ws.org/Vol-2962/paper05.pdf.

[3] M. Plátek, F. Mráz, D. Pardubská, D. Průša, On pumping RP-automata controlled by complete LRG(¢, $)-grammars, in: L. Ciencialová, M. Holeňa, R. Jajcay, T. Jajcayová, F. Mráz, D. Pardubská, M. Plátek (Eds.), Proceedings of the 22nd Conference Information Technologies – Applications and Theory (ITAT 2022), volume 3226 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2022, pp. 111–121. URL: https://ceur-ws.org/Vol-3226/paper13.pdf.

[4] F. Mráz, M. Plátek, D. Pardubská, D. Průša, One-side pumping and two-side pumping by complete CF(¢, $)-grammars, in: B. Brejová, L. Ciencialová, M. Holeňa, R. Jajcay, T. Jajcayová, M. Lexa, F. Mráz, D. Pardubská, M. Plátek (Eds.), Proceedings of the 23rd Conference Information Technologies – Applications and Theory (ITAT 2023), volume 3498 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2023, pp. 110–120. URL: https://ceur-ws.org/Vol-3498/paper14.pdf.

[5] P. Jančar, J. Šíma, The simplest non-regular deterministic context-free language, in: F. Bonchi, S. J. Puglisi (Eds.), 46th International Symposium on Mathematical Foundations of Computer Science, MFCS 2021, August 23–27, 2021, Tallinn, Estonia, volume 202 of *LIPIcs*, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021, pp. 63:1–63:18. doi:10.4230/LIPIcs.MFCS.2021.63.

[6] J. Šíma, M. Plátek, One analog neuron cannot recognize deterministic context-free languages, in: T. Gedeon, K. W. Wong, M. Lee (Eds.), Neural Information Processing – 26th International Conference, ICONIP 2019, Sydney, NSW, Australia, December 12–15, 2019, Proceedings, Part III, volume 11955 of *Lecture Notes in Computer Science*, Springer, 2019, pp. 77–89. doi:10.1007/978-3-030-36718-3_7.

[7] M. Lopatková, M. Plátek, P. Sgall, Towards a formal model for functional generative description: Analysis by reduction and restarting automata, Prague Bull. Math. Linguistics 87 (2007) 7–26. URL: http://ufal.mff.cuni.cz/pbml/87/lopatkova-et-al.pdf.

[8] P. Jančar, F. Mráz, M. Plátek, J. Vogel, Restarting automata, in: H. Reichel (Ed.), Fundamentals of Computation Theory, FCT '95, volume 965 of *Lecture Notes in Computer Science*, Springer, 1995, pp. 283–292. doi:10.1007/3-540-60249-6_60.

[9] P. Jančar, F. Mráz, M. Plátek, J. Vogel, On monotonic automata with a restart operation, J. Autom. Lang. Comb. 4 (1999) 287–311. doi:10.25596/jalc-1999-287.

[10] F. Otto, Restarting automata, in: Z. Ésik, C. Martín-Vide, V. Mitrana (Eds.), Recent Advances in Formal Languages and Applications, volume 25 of *Studies in Computational Intelligence*, Springer, 2006, pp. 269–303. URL: https://doi.org/10.1007/978-3-540-33461-3_11. doi:10.1007/978-3-540-33461-3_11.

[11] J. O. Shallit, A Second Course in Formal Languages and Automata Theory, Cambridge University Press, 2008.

[12] J. Hopcroft, J. Ullman, Introduction to Automata Theory, Languages, and Computation, Addison-Wesley, N. Reading, MA, 1980.