Persuasion Technique Taxonomy used in the Shared Task on the Dectection and Classification of Persuasion Techniques in Texts for Slavic Languages

Abstract

This documents provides essential information on the persuasion technique taxonomy used in the task.

1 Taxonomy

In this task we exploit the taxonomy from SemEval 2023 Task 3 (Piskorski et al., 2023), which is extended by two new persuasion techniques, namely: *false equivalence*[1], and *appeal to pity*[2]. The extended taxonomy consisting of 6 main categories subdivided into 25 fine-grained categories is shown in Figure 1.

2 Definitions and Examples

Below we provide the definitions of the persuasion techniques accompanied with some examples.

2.1 Attack on Reputation

Name Calling or Labelling: a form of argument in which loaded labels are directed at an individual or a group, typically in an insulting or demeaning way. Labelling an object as either something the target audience fears, hates, or on the contrary finds desirable or loves. This technique calls for a qualitative judgement that disregards facts and focuses solely on the essence of the subject being characterized. This technique is in a way also a manipulative wording, as it is used at the level of the nominal group rather than being a full-fledged argument with a premise and a conclusion. For example, in the political discourse, typically one is using adjectives and nouns as labels that refer to political orientation, opinions, personal characteristics, and association to some organisations, as well as insults. What distinguishes it from the *Loaded Language* technique (see 2.6), is that it is only concerned with the characterization of the subject.
Example: *'Fascist' Anti-Vax Riot Sparks COVID Outbreak in Australia.*

Guilt by Association: Attacking the opponent or an activity by associating it with another group, activity, or concept that has sharp negative connotations for the target audience. The most common example, which has given its name in the literature (i.e. *Reduction ad Hitlerum*) to that technique is making comparisons to Hitler and the Nazi regime. However, it is important to emphasize, that this technique is not restricted to comparisons to that group only. More precisely, this can be done by claiming a link or an equivalence between the target of the technique to any individual, group, or event in the presence or in the past, which has or had an unquestionable negative perception (e.g., was considered a failure), or is depicted in such way.
Example: *Manohar is a big supporter for equal pay for equal work. This is the same policy that all those extreme feminist groups support. Extremists like Manohar should not be taken seriously.*

Casting Doubt: Casting doubt on the character or the personal attributes of someone or something in order to question their general credibility or quality, instead of using a proper argument related to the topic. This can be done for instance, by speaking about the target's professional background, as a way to discredit their argument. Casting doubt can also be done by referring to some actions or events carried out or planned by some entity that are/were not successful or appear as (probably) resulting in not achieving the planned goals.
Example: *This task is quite complex. Is his professional background, experience and the time left sufficient to accomplish the task at hand?*

Appeal to Hypocrisy: The target of the technique is attacked on its reputation by charging them with hypocrisy or inconsistency. This can be done explicitly by calling out hypocrisy directly, or more implicitly by underlying the contradictions between different positions that were held or actions that were done in the past. A special way of calling out hypocrisy is by telling that someone who criticizes you for something you did, also did it in the past.
Example: *How can you demand that I eat less meat to reduce my carbon footprint if you yourself drive a big SUV and fly for holidays to Bali?*

Questioning the Reputation: This technique is used to attack the reputation of the target by making strong negative claims about it, focusing specially on undermining its character and moral stature rather than relying on an argument about the topic. Whether the claims are true or false is irrelevant for the effective use of this technique. Smears can be used at any point in a discussion. One particular way of using this technique is to preemptively call into question the reputa-

---

[1] https://www.logicallyfallacious.com/logicalfallacies/False-Equivalence
[2] https://www.logicallyfallacious.com/logicalfallacies/Appeal-to-Pity

```
ATTACK ON REPUTATION
- Name Calling or Labelling
- Guilt by Association
- Casting Doubt
- Appeal to Hypocrisy
- Questioning the Reputation

JUSTIFICATION
- Flag Waiving
- Appeal to Authority
- Appeal to Popularity
- Appeal to Fear, Prejudice
- Appeal to Values

DISTRACTION
- Strawman
- Whataboutism
- Red Herring
- Appeal to Pity

SIMPLIFICATION
- Causal Oversimplification
- False Dilemma or No Choice
- Consequential Oversimplification
- False Equivalence

CALL
- Slogans
- Conversation Killer
- Appeal to Time

MANIPULATIVE WORDING
- Loaded Language
- Obfuscation, Intentional Vagueness, Confusion
- Exaggeration or Minimisation
- Repetition
```

Figure 1: Two-tier Persuasion Technique taxonomy.

tion/credibility of an opponent, before he had any chance to express himself, therefore biasing the audience perception. Hence, one of the name of that technique is "poisoning the well."

The main difference between *Casting Doubt* (introduced earlier) and *Questioning the reputation technique* is that the former focuses on questioning the capacity, the capabilities, and the credibility of the target, while the latter targets undermining the overall reputation, moral qualities, behaviour, etc.

Example: *I hope I presented my argument clearly. Now, **my opponent will attempt to refute my argument by his own fallacious, incoherent, illogical version of history***

2.2 Justification

Flag Waving: Justifying or promoting an idea by exhaling the pride of a group or highlighting the benefits for that specific group. The stereotypical example would be national pride, and hence the name of the technique; however, the target group it applies to might be any group, e.g., related to race, gender, political preference, etc. The connection to nationalism, patriotism, or benefit for an idea, group, or country might be fully undue and is usually based on the presumption that the recipients already have certain beliefs, biases, and prejudices about the given issue. It can be seen as an appeal to emotions instead to logic of the audience aiming to manipulate them to win an argument. As such, this technique can also appear outside the form of well constructed argument, by simply making mentions that resonate with the feeling of a particular group and as such setting up a context for further arguments.

Example: **We should make America great again, and restrict the immigration laws.**

Appeal to Authority: a weight is given to an argument, an idea or information by simply stating that a particular entity considered as an authority is the source of the information. The entity mentioned as an authority may, but does not need to be, an actual valid authority in the domain-specific field to discuss a particular topic or to be considered and serve as an expert. What is important, and makes it different from simply sourcing information, is that the tone of the text indicates that it capitalizes on the weight of an alleged authority in order to justify some information, claim, or conclusion. Referencing a valid authority is not a logical fallacy, while referencing an invalid authority is a logical fallacy, and both are captured within this label. In particular, a self-reference as an authority falls under this technique as well.

Example: **Since the Pope said that this aspect of the doctrine is true we should add it to the creed.**

Appeal to Popularity: This technique gives weight to an argument or idea by justifying it on the basis that allegedly "*everybody*" (or the vast majority) agrees with it or "*nobody*" disagrees with it. As such, the target audience is encouraged to gregariously adopt the same idea by considering "*everyone else*" as an authority, and to join in and take the course of the same action. Here, "*everyone else*" might refer to the general public, key entities and actors in a certain domain, countries, etc. Analogously, an attempt to persuade the audience not to do something because "*nobody else is taking the same action*" falls under our definition of Appeal to Popularity.

Example: **Because everyone else goes away to college, it must be the right thing to do.**

Appeal to Values: This technique gives weight to an idea by linking it to values seen by the target audience as positive. These values are presented as an authoritative reference in order to support or to reject an argument. Examples of such values are, for instance: tradition, religion, ethics, age, fairness, liberty, democracy, peace, transparency, etc. When such values are mentioned outside the context of a proper argument by simply using certain adjectives or nouns as a way of characterizing something or someone, such references fall under another label, namely, *Loaded Language*, which is a form of *Manipulative Wording* (see 2.6).

Example: **It's standard practice to pay men more than women so we'll continue adhering to the same standards this company has always followed.**

Appeal to Fear, Prejudice: This technique aims at promoting or rejecting an idea through the repulsion or fear of the audience towards this idea (e.g., via exploiting some preconceived judgements) or towards its alternative. The alternative could be the status quo, in which case the current situation is described in a scary way with *Loaded Language*. If the fear is linked to the consequences of a decision, it is often the case that this technique is used simultaneously with *Appeal to Consequences* (see Simplification techniques in 2.4), and if there are only two alternatives that are stated explicitly, then it is used simultaneously with the *False Dilemma* technique (see 2.4).

Example: *It is a great disservice to the Church to maintain the pretense that there is nothing problematical about Amoris laetitia. **A moral catastrophe is self-evidently underway** and it is not possible honestly to deny its cause.*

2.3 Distraction

Strawman: This technique consists in making an impression of refuting the argument of the opponent's proposition, whereas the real subject of the argument was not addressed or refuted, but instead replaced with a false one. Often, this technique is

referred to as misrepresentation of the argument. First, a new argument is created via the covert replacement of the original argument with something that appears somewhat related, but is actually a different, a distorted, an exaggerated, or a misrepresented version of the original proposition, which is referred to as "*standing up a straw man.*" Subsequently, the newly created '*false* argument (the strawman) is refuted, which is referred to as "*knocking down a straw man.*" Often, the strawman argument is created in such a way that it is easier to refute, and thus, creating an illusion of having defeated an opponent's real proposition. Fighting a strawman is easier than fighting against a real person, which explains the origin of the name of this technique. In practice, it appears often as an abusive reformulation or explanation of what the opponent *actually*' means or wants.

Example: *Referring to your claim that providing medicare for all citizens would be costly and a danger to the free market, I infer* **that you don't care if people die from not having healthcare, so we are not going to support your endeavour.**

Red Herring: This technique consists in diverting the attention of the audience from the main topic being discussed, by introducing another topic. The aim of attempting to redirect the argument to another issue is to focus on something the person doing the redirecting can better respond to or to leave the original topic unaddressed. The name of that technique comes from the idea that a fish with a strong smell (like a herring) can be used to divert dogs from the scent of someone they are following. A strawman (defined earlier) is also a specific type of a red herring in the way that it distracts from the main issue by painting the opponent's argument in an inaccurate light.

Example: *Lately, there has been a lot of criticism regarding the quality of our product.* **We've decided to have a new sale in response, so you can buy more at a lower cost!**.

Whataboutism: A technique that attempts to discredit an opponent's position by charging them with hypocrisy without directly disproving their argument. Instead of answering a critical question or argument, an attempt is made to retort with a critical counter-question that expresses a counter-accusation, e.g., mentioning double standards, etc. The intent is to distract from the content of a topic and to switch the topic actually. There is a fine distinction between this technique and *Appeal to Hypocrisy*, introduced earlier, where the former is an attack on the argument and introduces irrelevant information to the main topic, while the latter is an attack on reputation and highlights the hypocrisy of double standards on the same or a very related topic.

Example: **A nation deflects criticism of its recent human rights violations by pointing to the history of slavery in the United States.**

Appeal to Pity: A technique that evokes feelings of pity, sympathy, compassion or guilt in audience to distract it from focusing on evidence, rational analysis and logical reasoning, so that it accepts the speaker's conclusion as truthful solely based on soliciting the aforementioned emotions. It is an attempt to sway opinions and fully substitute logical evidence in an argument with a claim intended to elicit pity or guilt.

Example: **If this person is found guilty of this crime, his ten children will be left without a parent at home, therefore the jury must submit a verdict of innocence.**

2.4 Simplification

Causal Oversimplification: Assuming a single cause or reason when there are actually multiple causes for an issue. This technique has the following logical form(s): (a) *Y occurred after X; therefore, X was the only cause of Y*, or (b) *X caused Y; therefore, X was the only cause of Y+ (although A, B, C...etc. also contributed to Y.)*

Example: *School violence has gone up and academic performance has gone down since video games featuring violence were introduced.* **Therefore, video games with violence should be banned, resulting in school improvement.**

False Dilemma or No Choice: Sometimes called the *either-or* fallacy, a false dilemma is a logical fallacy that presents only two options or sides when there actually are many. One of the alternatives is depicted as a *no-go* option, and hence the only choice is the other option. In extreme cases, the author tells the audience exactly what actions to take, eliminating any other possible choices (also referred to as *Dictatorship*).

Example: **There is no alternative to Pfizer Covid-19 vaccine. Either one takes it or one dies.**

Consequential Oversimplification: An argument or an idea is rejected and instead of discussing whether it makes sense and/or is valid, the argument affirms, without proof, that accepting the proposition would imply accepting other propositions that are considered negative. This technique has the following logical form: *if A will happen then B, C, D, ... will happen.* The core essence behind this fallacy is an assertion one is making of some '*first*' event/action leading to a domino-like chain of events that have some significant negative effects and consequences that appear to be ludicrous. This technique is characterized by **ignoring and/or understating the likelihood of the sequence of events from the first event leading to the end point** (last event). In order

to take into account symmetric cases, i.e., using *Consequential Oversimplification* to promote or to support certain action in a similar way, we also consider cases when the sequence of events leads to positive outcomes (i.e., encouraging people to undertake a certain course of action(s), with the promise of a major positive event in the end).

Example: ***If we begin to restrict freedom of speech, this will encourage the government to infringe upon other fundamental rights, and eventually this will result in a totalitarian state where citizens have little to no control of their lives and decisions they make***.

False Equivalence: A technique that attempts to treat scenarios that are significantly different as if they had equal merit or significance. In particular, an emphasis is being made on one specific shared characteristic between the items of comparison in the argument that is way off in the order of magnitude, oversimplified, or just that important additional factors have been ignored. The introduction of the certain shared characteristics of the scenarios is then used to consider them equal. This technique has the following logical form: *A and B share some characteristic X. Therefore, A and B are equal.*

Example: ***The introduction or restrictive hours of alcohol sales boosted the black market industry, and analogously, one can expect that the introduction of too restrictive anti-abortion regulations will lead to growth of the illegal abortion business***.

2.5 Call

Slogans: A brief and striking phrase that may include labeling and stereotyping. Slogans tend to act as emotional appeals.

Example: ***Immigrants welcome, racist not!***

Conversation Killer: This includes words or phrases that discourage critical thought and meaningful discussion about a given topic. They are a form of *Loaded Language*, often passing as folk wisdom, intended to end an argument and quell cognitive dissonance.

Example: *I'm not so naïve or simplistic to believe we can eliminate wars.* ***You can't change human nature.***

Appeal to Time: The argument is centered around the idea that time has come for a particular action. The very timeliness of the idea is part of the argument.

Example: *This is no time to engage in the luxury of cooling off or to take the tranquilizing drug of gradualism.* ***Now is the time to make real the promises of democracy. Now is the time to rise from the dark and desolate valley of segregation to the sunlit path of racial justice.***

2.6 Manipulative Wording

Loaded Language: use of specific words and phrases with strong emotional implications (either positive or negative) to influence and to convince the audience that an argument is valid. It is also known as *Appeal to Argument from Emotive Language*.

Example: *They keep feeding these people with trash. They should stop.*

Obfuscation, Intentional Vagueness, Confusion: This fallacy uses words that are deliberately not clear, so that the audience may have its own interpretations. For example, an unclear phrase with multiple or unclear definitions is used within the argument and, therefore, does not support the conclusion. Statements that are imprecise and intentionally do not fully or vaguely answer the question posed fall under this category too.

Example: ***Feathers cannot be dark, because all feathers are light!***

Exaggeration or Minimisation: This technique consists of either representing something in an excessive manner – by making things larger, better, worse (e.g., *the best of the best*, *quality guaranteed*) – or by making something seem less important or smaller than it really is (e.g., saying that an insult was just a joke), downplaying the statements and ignoring the arguments and the accusations made by an opponent.

Example: *From the seminaries, to the clergy, to the bishops, to the cardinals,* ***homosexuals are present at all levels, by the thousand***.

Repetition: The speaker uses the same word, phrase, story, or imagery repeatedly with the hope that the repetition will lead to persuade the audience.

Example: ***Hurtlocker deserves an Oscar****. Other films have potential, but they do not **deserve an Oscar like Hurtlocker does**. The other movies may deserve an honorable mention but **Hurtlocker deserves the Oscar***.

References

Jakub Piskorski, Nicolas Stefanovitch, Nikolaos Nikolaidis, Giovanni Da San Martino, and Preslav Nakov. 2023. Multilingual multifaceted understanding of online news in terms of genre, framing, and persuasion techniques. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3001–3022, Toronto, Canada. Association for Computational Linguistics.