

1 Introduction

Media bias commonly refers to slanted or one-sided news coverage. Media bias is the perceived bias of journalists and news producers within the mass media in the selection of events and stories that are reported and how they are covered. The term media bias implies a pervasive or widespread bias contravening the standards of journalism, rather than the perspective of an individual journalist or article. The direction and degree of media bias in various countries is widely disputed.¹

The goal of this study is to annotate instances of bias by word choice and labeling. Therefore, we conduct a manual deductive content analysis. This codebook defines what needs to be annotated, provides annotation rules, and elaborates the annotation types with examples.

When referring to a semantic concept, such as an actor of the action in a news event, authors choose from multiple words or multi-word terms, such as Merkel, Angela Merkel or Bundeskanzlerin. While these three exemplary terms are considered neutral in this study, other terms may be not. If a chosen term may induce a different (emotional) judgment on the referred concept by a human reader, the term used by the author induces bias by word choice. Examples of bias by word choice include regime vs. government, dictator vs. chief of state.

Likewise, authors can label the same concept differently, i.e., different attributes added to the concept lead to a different perception by the human reader of the concept mentioned in the text. If a chosen attribute induces a different (emotional) judgment of its concept by a human reader, the attribute causes bias by labeling. Examples of bias by labeling include coalition forces vs. invasion forces, immigrants vs. undocumented immigrants (where the former does not contain any label), an anxious and uncertain President Trump.

While the previously given examples are only one of both forms of media bias, in practice, bias by word choice and bias by labeling can often not be distinguished clearly. Border cases, such as freedom fighters vs. terrorists, young undocumented immigrants vs. aliens, show that the manifestation of both bias forms is highly related. Moreover, their effect is identical, i.e., the perception of the referred (word choice) and attributed (label) semantic concept is changed in a specific way. Thus, in this study, we treat both forms as jointly, i.e., word choice and labeling (WCL).

2 Definitions and Examples

2.1 General remarks

All examples, links and referred resources provided in this document were valid at the time of writing but might not necessarily be when reading.

2.2 Term

We define a term as a single word or multi-word expression in a text.

¹ Source: https://en.wikipedia.org/wiki/Media_bias

2.3 Concept

We define a concept in the linguistic sense as a single, language independent, semantic unit. In a document, a concept is represented by a term. Synonyms, for instance, represent the same concept (or similar concepts). A concept, for instance, is the current German Chancellor, which might be represented in a document also as “Merkel,” “Bundeskanzlerin,” “Bundeskanzler,” “Angela Merkel,” “Dr. Merkel,” etc. In this study, concepts are usually entities, e.g., politicians, actions, such as “deport” or “overstep,” places, and times. Note that this list is not comprehensive, if you find a different concept type, please note the document and position where it occurs in your coding protocol.

2.4 Bias by Labeling (L)

- Bias by labeling is caused by words added to a concept, also called attributes, which change the perception of that concept. An **attribute** is “a quality or feature regarded as a characteristic or inherent part of someone or something”²
- Typically, labels are adjectives (modifying subjects or objects) or adverbs (modifying predicates), but can also be nouns (usually as part of compounds, modifying subjects or objects, e.g., freedom fighters, which could also be considered a word choice)
- Examples
 - an anxious and uncertain President Trump
 - Mr. Trump, exasperated, ...
 - his anguished deliberations
 - a difficult emotional decision for the president
 - Mr. Trump's hard-line policy adviser, Stephen Miller
 - less-educated, native-born Americans
 - cruel action

2.5 Bias by Word Choice (WC)

- Bias by word choice is caused by words that refer to a concept while inducing judgement at the reader.
- Note that often a label is added to a concept to describe it more precisely, e.g., freedom fighters, which in conjunction with the concept induces a judgement. In such cases, the resulting reference to the concept is a combined form of bias by word choice and labeling
- Examples
 - Subjects or objects
 - Coalition forces, invasion forces, illegal aliens, young immigrants, undocumented immigrants, immigrants who entered the US as children, young people,
 - Predicates
 - Mr. Obama had overstepped his authority
 - invade vs. cross border
 - the participants were taking jobs from U.S. citizens

² <https://en.oxforddictionaries.com/definition/attribute>

2.6 Bias by Word Choice and Labeling (WCL)

- This is the subject of our study, which needs to be annotated. Instances of WCL are bias by word choice, labeling, or a combination of both.

3 What to annotate and how

3.1 Definitions

There are two main units to be coded: frame properties and concepts.

FRAME PROPERTY: A frame property, or simply named PROPERTY, serves as a categorization of the judgement and or frame that phrase induces. All valid PROPERTYs are represented in a list in MAXQDA, each having its own definition, which you can find in the code memo (within the Code System view). Each PROPERTY targets a single (nearby) CONCEPT.

CONCEPT: A CONCEPT is any semantic concept mentioned in the article, including ACTIONS, ACTORS, ACTOR-Is, COUNTRYs, EVENTs, GROUPs, MISCs, and OBJECTs. Each CONCEPTs mention can be the target of one or more PROPERTY mentions.

A concept is one of the following types:

ACTION: An action that is performed by an ACTOR (or other CONCEPT) and/or that happens during an EVENT, e.g., the ACTION “Negotiate about the peace” happened during the EVENT “PRK USA Summit.” Consists typically mostly of verb phrases (VPs).

ACTOR: Someone or something that performs an action. Must be the subject of the sentence or a clause of the sentence. An individual that is part of or represents a specific group of individuals, e.g., DACA recipients, belongs to the same semantic actor. ACTORs are mostly single persons or single institutions, such as companies.

[CONCEPTNAME]-I (to be revised): If an individual is part of a GROUP, e.g., DACA recipients, but the document mentions also individuals of that group, i.e., a single person, e.g., Mr. Smith, who is part of the group, add another [CONCEPT]-I concept to the CONCEPT -TREE, using the suffix “-I”, e.g., “DACA recipients-I”, which you will use for all coreferences in the text that do not refer to the group but to individuals of that group, e.g., a DACA recipient interviewed at a demonstration. For coreferences referring to the whole group, use the “DACA recipients” code. If a CONCEPT A is a specific “form” of another CONCEPT B or located within CONCEPT B, add A as a sub-code to B in the ACTOR-TREE. For instance, McMaster (occurring frequently, hence being its own code) is a specific form of USA-I (which would be used for all non-frequently mentioned individuals that represent the US).

COUNTRY: Use the country’s three-letter country code, e.g., United State of America = USA, Russia = RUS, Germany = DEU, see: https://en.wikipedia.org/wiki/ISO_3166-1_alpha-3 Institutions or job positions that are part of a country/state (typically the government), need to be tagged as the country itself (not as a representative). For instance, The White House, Pentagon, White House Spokesman, are all permanently associated with / representing the USA, the US army; therefore, you must code all of them just as USA.

[COUNTRYCODE]/[Lastname of Head of state]: The head of state, e.g., Putin, Trump, and Merkel, is always its own sub-type. For instance, for RUS, Putin would be coded as shown here:



[COUNTRYNAME]-MISC: This includes any CONCEPT that is part of the country and has a state-function, but does not represent the country. For instance, we code the government and its members as [COUNTRYCODE], since they officially represent the country towards other countries, are eligible to sign contracts for the country, and make new laws, etc. The opposition, however, lacks the authority to do this and their views and opinions may often differ to those of the government. Thus, Russia's opposition would be coded as RUS-MISC.

EVENT: An event during which ACTIONS

The action of the event, e.g., an armed conflict between two parties, described by "humiliating defeat", "armed confrontation", "battle", "attack" etc.

GROUP: A group of individuals that act collectively, e.g., "immigrants," "DACA recipients," or "sports club." Thus, similar to ACTOR but is a group of individuals, whereas ACTOR is a single individual.

OBJECT: Something that is the object of one or more ACTIONS or the topic of EVENTS, e.g., "Steele dossier."

MISC: Any semantic concept that is mentioned frequently in the article and that does not fit into the other categories. Typically, broadly defined or abstract, e.g., "Denuclearization," "Reactions on deal."

CONCEPT-TREES: list of all CONCEPTs, specified in MAXQDA, e.g., the ACTOR-TREE contains all ACTORS.

When to tag [CONCEPTNAME] vs [CONCEPTNAME]-I? If an individual is officially *representing* the CONCEPT, e.g., the spokesman of the Pentagon (which itself is part of the US) or a Minister, you must tag the individual as [CONCEPTNAME]. If the individual is rather a "sample" of a group but does not have an official function within the group (particularly to represent the group towards other entities) then you must tag [CONCEPTNAME]-I.

3.2 Workflow of coding for a single news article

While the study is in the field of bias by WCL, for the coding of PROPERTIES, do not only consider instances or phrases of biased news coverage, but **any phrase that induces a judgment on a target concept**.

Note that in the following, we use the term ACTOR for ease of understanding and readability. However, all ACTOR rules and mentions also refer to any other CONCEPT.

- (1) read the article briefly, thereby scan for ACTORS involved in the event(s). is there any ACTOR that is not yet in the ACTOR-TREE but has at least four mentions (nominal, pronominal, and else) in the current news article or is a country?
if so, add that ACTOR to the ACTOR-LIST. To find how often an ACTOR occurs, search for its name in the search field (thereby look individually only for first name and family name, its job/function (prime minister), and pronominal mentions (he/she/...) before and after the nominal mentions.

If the ACTOR is a country, always also add as children to the country in the ACTOR-TREE the following: [Lastname of Head of state], [Countryname]-I, [Countryname]-MISC.

- (2) ACTOR-PHRASE: for each ACTOR from the ACTOR-TREE, scan for nominal occurrences in the article. code each nominal co-reference with its ACTOR (pronominal co-references must not be coded), where the PHRASE also must include any attribute, i.e., an adjective or adverb, close to the actual reference to the actor, e.g., code 'illegal aliens' but not only 'aliens' (for the DACA Dreamers) or 'Alleged Russian Fighters'.

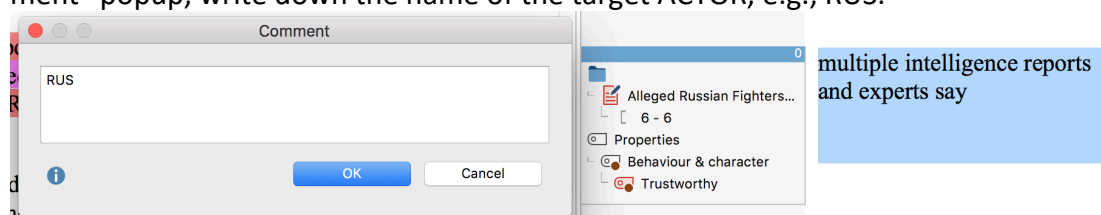
You must use 'Highlight coding', but not 'Code with a new code' (this way, the ACTORs will be visually apparent throughout the article). Code the job position, rather than the name, e.g., Pentagon spokesman Eric Pahon, since the name is not important here (except if the article is reporting on Eric Pahon, i.e., he is one of the main actors involved in the article's event). Note that MAXQDA only allows five different "Highlight coding" codes, so you will need to use 'Code with a new code' for additional ACTORs. In that case, change the colors of all Property Codes to black and use colored Codes for the ACTORs.

- (3) for each ACTOR-PHRASE coded in (2), check whether the ACTOR-PHRASE as a whole or individual words of it induces a judgement on the ACTOR. if so, add the corresponding code(s) of the type 'Properties' to whole phrase or individual tokens. If you haven't done so already, check the definition of each Property, by opening its memo in the "Code System" view. For instance, "Alleged Russian Fighters" is an ACTOR-PHRASE, but alleged=Misc@RUS and Fighters=Perpetrator@RUS.

- (4) for each ACTOR-PHRASE coded in (2) (do this for one ACTOR-PHRASE after the other, regardless of their color/ACTOR), check whether the paragraph where it occurs, the paragraph before and the one afterward contain PHRASES that induce a judgement on the ACTOR of the ACTOR-PHRASE (including non-nominal coreferences, e.g., pronominal co-references, such as 'he'). These PHRASES shall not include the target (this will be added as a "Comment", see below). Exception: If the target is neither at the beginning or end of a phrase but in the middle (and hence cannot be excluded from the phrase), you must include the target in the phrase, e.g., wished me well.

The PHRASE needs to be of the type VERB, VP, NP. hint: often, the verb itself does not induce the judgement, but only in conjunction with the object.

For each such PHRASE, add any code(s) of the type 'Properties' using 'Code with a new code'. For each code added this way to the PHRASE, add its target (the ACTOR on which the PHRASE induces the judgement) by double-clicking on the newly added code in the "Retrieved Segments" view (see <https://www.maxqda.com/help-max18/05-coding/creating-comments-for-coded-segments> for further information). In the "Comment" popup, write down the name of the target ACTOR, e.g., RUS.



add its ACTOR, i.e., the ACTOR on which the judgement is induced, and. If the document does not contain paragraphs, e.g., each sentence is in one line but there is no

paragraphing, check six sentences before the ACTOR-PHRASE's occurrence and afterward. TODO: add code memo instruction!!!!

3.3 Coding rules

A1: as short as possible: 'struggled' vs 'struggled to focus on his studies' (Difficulty->Difficult). because 'struggled' itself induces the judgement, only 'struggled' shall be coded. or 'hard to fully mentally prepare for it' -> only 'hard'. or 'the road may be longer and bumpier' -> only 'longer and bumpier'

A2: do not code the target of a biased VP/NP, e.g. 'The students' hostility to any compromise with Americans' -> code only 'hostility to any compromise with Americans'. **Exception:** if the target is necessary to induce a specific judgement, include it in the coded phrase, e.g., "has weakened the Russian economy", Negative economy@Russia (only "weakened" would not be sufficient, as it does not imply the economy).

B1: as long as necessary. never remove the immediate context of a phrase, e.g., (not a) balanced immigration policy (behavior->fair). phrases cannot be coded isolatedly but need to be coded as to their actual meaning in the text. for verbs: double check if you only want to code the verb (have been blessed vs have been blessed with all the opportunities). example for NP: 'hostility to any compromise with Americans'. in other words: the coded phrase by itself must induce the judgement, e.g., 'no' (behavior->unfair) does not do that, so do not code it. Also pay attention, if you want to code a noun that has an adjective before it; usually, the adjective is a label that describes the noun, and as such shall not be discarded but coded along with the noun. For instance, "suspected chemical weapons attack": would be perpetrator if it was only chemical weapons attack, but including the suspected word in front, it needs to be tagged as other-other

C1: an annotated phrase needs to decisively induce the coded judgement. if doubt regarding the induced judgement remains, think whether another, more general code fits better (the most general code is Other->Other that shall be used for phrases that induce judgement but neither positively nor negatively) or do not code it at all

D1: only code phrases that actually **induce judgement on a concept**, e.g., 'have free speech for everyone' (behavior->fair) or 'no' (behavior->unfair) does not induce judgement on a concept (full context: They have no respect for the First Amendment, they have no respect for our way of life, which is to have free speech for everyone.) in other words: do not code the meaning of a phrase, e.g., that 'free speech' or 'claiming rights' means behavior->fair. code 'free speech' if and only if it induces a judgement on a specific concept.

D2: only code phrases that actually **induce judgement**, e.g., 'When Breitbart News asked the illegals if new immigration limits were a fair trade for amnesty' (behavior->fair) & 'fair' (behavior->fair) does not induce judgement, since it is part of a question asked during an interview. Also, for instance, comparisons / metaphors do usually not induce a judgement, e.g. 'stepping up on my community' (context 'Oh, let's make a deal with the wall,' *** it is like *** I'm stepping up on my community,'). Do not simply search for terms that appear as they

induce a judgement! Rather read the text, and systematically find phrases that explicitly and specifically induce a judgement targeting a specific semantic concept (the ACTOR). So, special caution is required for hypothetical and fictional statements (even though, depending on the particular case, they might induce judgement!)

E1: use the most specific dimension, e.g., “she sobbed” -> Power-Weak (instead of, for instance, Affection-Empathy)

E2: If you would tag more than one dimension: Does one of categories (of one dimension) necessarily include a category (of another dimension)? For instance, if a concept is tagged as *perpetrator*, this will “always” include also *refusal* (no one likes bad people) In that case, only tag the independent category, i.e., perpetrator, but not the / none of the dependent dimensions, i.e., refusal.

E3: You must avoid personal interpretation. Think in terms of what the author of the text is likely to achieve with her choice of words. This is a rather soft criteria, but to get an idea what this is about, here are two examples. Obvious example: Just because you like Angela Merkel, do not code the phrase ‘Bundeskanzlerin Merkel’ as Affection-Empathy. Slightly more realistic but also vague example: If something arguable has happened in a news event, and the text describes that certain people ‘argue’ about that, do not code that as Other-Positive (or similar) because *you think* that it is a good idea that people argue about the issue.

E4: Consider context! As an extension to ‘Think in terms of what the author of the text is likely to achieve with her choice of words.’ from E3: Consider the context of a phrase to estimate its intended judgement. For instance, “boosting government spending” could be either good or bad, but if it is part of a listing where the author mentions only negative effects caused by DACA, e.g., “increase unemployment rate, boosting government spending, and increased number of crimes”, it is likely to be intended to induce a negative judgement, too.

Another example: “US warns Russia against backing Assad” Do not code “backing” as Affection@Assad, because the second clause of this sentence is just what the US warns Russia to not do.

E5: Code the judgement that the text (or a person/speaker who’s quoted within the text) shows towards the target. For instance, “I’m willing to take the risk for my family and for all the other DACA people out there” is twofold: “I’m willing to take the risk” -> Behavior-Confidence@DACA recipients-R; “take the risk” -> Power-Weak@DACA recipients.

E6: Do **not** consider the news outlet of the article when coding; only consider the appearance in text. For instance, “McMaster denounced Russia” could be both powerful@McMaster or Perpetrator@McMaster. Given that the article is published by Breitbart, it is likely that powerful would be more appropriate (since Breitbart is in favor of McMaster, usually) but you cannot use this information. Code only according to the text (here, both codes are desired).

F1: If you tag a code from the Other dimension, always write in the second line of the code comment the name of the dimension that would be most suitable for your annotation (if it existed). This way we can find whether certain dimensions are missing and will add them later. (The first line contains the target.)

4 Revisions

- 0.6: Full revision of Section 3 (FH)
- 0.8: Refined workflow and rules in Section 3 (FH)
- 0.9: Update rule (4) in Section 3.2 so that certain phrases can include the target (FH)
- 1.0: Distinguish between more fine granular concept types (FH)