TEXT & TALK SUBMISSION

**Authors**: Laura Hidalgo-Downing1, Paula Pérez-Sobrino2

**Main author**: Laura Hidalgo-Downing, Paula Pérez-Sobrino (joint main-authorship)

**Main author’s affiliation**: 1Universidad Autónoma de Madrid (Spain), Universidad de La Rioja (Spain)

**Institutional address**: 1Departamento de Filología Inglesa, Facultad de Filosofía y Letras, Universidad Autónoma de Madrid, Calle Tomás y Valiente 1, Madrid 28049, Spain

2Departamento de Filologías Modernas, Universidad de La Rioja, c/San José de Calasanz 33, 26004, Logroño, La Rioja. Email:

**Email**: 1[laura.hidalgo@uam.es](mailto:laura.hidalgo@uam.es); 2[paula.perezs@unirioja.es](mailto:paula.perezs@unirioja.es)

**Full title of article**: Developing an annotation protocol for evaluative stance and metaphor in discourse: theoretical and methodological considerations

**Short title of the Article (for running head):** An annotation protocol for evaluative stance

**Word count (all inclusive):** 8,267 words (excluding abstract and keywords, tables and references)

**Character count (with spaces):** 55,443

**Bionotes**

Laura Hidalgo Downing is Full Professor at Universidad Autónoma, Madrid. Her research interests include discourse analysis at the cross-roads between functional linguistics and cognitive linguistics, stance and evaluation, multimodality and metaphor. Her recent book-length publication is *Performing metaphoric creativity across modes and contexts* (2020, John Benjamins). Address for correspondence: Departamento de Filología Inglesa, Facultad de Filosofía y Letras, Universidad Autónoma de Madrid, 28049 Madrid, Spain. Email: [Laura.hidalgo@uam.es](mailto:Laura.hidalgo@uam.es)

Paula Pérez-Sobrino a lecturer in Modern Languages at the University of La Rioja. Her work deals with the ways in which metaphors and other types of figurative language help or hinder cross-cultural communication. She has published two research monographs titled *Multimodal Metaphor and Metonymy in Advertising* (2017, John Benjamins) and *Unpacking Creativity: The Power of Figurative Communication in Advertising* (2021, Cambridge University Press) and has a third monograph under contract for publication in 2022: *Emotion-Colour Associations across Languages and Genders* (Cambridge University Press). Address for correspondence: Departamento de Filologías Modernas, Universidad de La Rioja, c/San José de Calasanz 33, 26006, Logroño, La Rioja. Email: paula.perezs@unirioja.es

**Developing an annotation protocol for evaluative stance and metaphor in discourse: theoretical and methodological considerations**

**Abstract**

The process of identification and annotation of evaluation has received a lot of attention in recent years. However, given the complexity of the topic, the discussion of some of the central issues is still ongoing. The present article contributes to this debate by presenting an annotation scheme that is designed for the identification and annotation of evaluative stance in a corpus of four English genres, namely, newspaper discourse, political discourse, newspaper scientific popularization and fora. A 4862-word corpus was sampled from a larger 400,000-word corpus compiled within the research project STANCEDISC on the study of stance in discourse varieties. The scheme posits a series of ad hoc categories designed to optimise the transparency, reliability and replicability of the identification, annotation and analysis of evaluative stance. The categories are as follows: parts of speech (Noun Phrase, NP; Adjectival Phrase, AP; Adverbial Phrase, ADVP; Verbal Phrase, VP), function (classifying, predicational and attitude), metaphoricity (metaphoric and non-metaphoric), and value (positive and negative). The aim of this paper is to explain the scheme together with the theoretical justification of the categories and the methodological procedure adopted, and to illustrate the implementation of the scheme by discussing examples taken from different genres.

**Keywords**: evaluation, stance, stance annotation, inter-rater reliability, evaluative metaphor

1. **Introduction[[1]](#footnote-1)**

The study of evaluation has attracted a lot of attention in recent years but research on the methodological issues involved in the identification and annotation of evaluation is relatively recent (see Fuoli 2018). It is generally agreed that annotating and analysing evaluative language is a very challenging endeavour, since evaluation is often expressed cumulatively and implicitly throughout a text (Hunston 2011; Fuoli 2018; Simaki et al. 2018, 2020). Discussions of a methodological nature are “an important part of the process of theory building” (Fuoli 2018: 229) because they help to refine theoretical models and contribute to (re)formulate categories that better capture the expressions occurring in real discourse across various genres.

The present article aims to contribute to ongoing discussions on the challenges in the identification and annotation of evaluation by presenting a scheme for the annotation of evaluative stance together with the methodological process involved in its implementation, and the discussion of its application to the analysis of examples from a corpus of four different genres. The scheme has been developed to ensure the consistency, replicability, transparency and reliability of the annotation of a 400,000-word corpus consisting of four different genres in English which include political discourse, newspaper opinion discourse, science popularizations and fora. The corpus was compiled within a research project on the study of stance variation across genres from a critical perspective, in which the corpus was annotated for epistemic and deontic stance on the one hand, and evaluative stance, on the other.[[2]](#footnote-2)

The article is organized as follows: section 2 provides an overview of previous works on stance, evaluation and approaches to the identification and annotation of evaluation; //this will need rewriting – see comments later// section 3 presents the categories of the annotation scheme; section 4 explains the methodology, including the inter-rater reliability process and the steps carried out in the process of implementation of the scheme; and the article closes with conclusions in section 5.

**2 Literature review: Conceptual and methodological approaches towards the annotation of evaluative stance**

Our approach to evaluation envisages the function of language as one of the dimensions of stance, following Du Bois (2007: 163), who defines stance as follows:

“Stance is a public act by a social actor, achieved dialogically through overt communicative means, of simultaneously evaluating objects, positioning subjects (self and others), and aligning with other subjects, with respect to any salient dimension of the sociocultural field.”

This organization of stance into three dimensions – evaluation, positioning and alignment – is represented visually in the ‘stance triangle’ illustrated in Figure 1 below:

Diagrama

Descripción generada automáticamente

*Figure 1. The stance triangle (Du Bois 2007: 163)*

This conception allows for the understanding of stance as involving both subjectivity (positioning of subjects) and intersubjectivity (alignment with other subjects) and it highlights the context-dependent nature of the term by making explicit reference to the potential dimensions of the socio-cultural field that may be relevant for the stance-taking act, which are thus open to evaluation. The relation between stance and evaluation is mentioned explicitly in various frameworks (Biber et al. 1999; Hunston and Thompson 2000; Martin and White 2005; Englebretson 2007; Hunston 2011; Thompson and Alba-Juez 2014), each of which addresses stance from different angles. In our approach to evaluation, stance is a more general abstract phenomenon, while evaluation as a dimension of stance is manifested by specific linguistic choices within the semantic domains of “value judgements”, “assessments”, and “attitudes”. This domain is different from the two other domains which may express stance, namely affect or “personal feelings” and epistemicity or “commitment“ (Englebretson 2007: 17), which are not included in our annotation scheme.[[3]](#footnote-3) While some approaches to stance address the complexity of these three complementary dimensions within one framework (Biber et al. 1999; Hunston and Thompson 2000; Martin and White 2005; Alba-Juez and Thompson 2014), most work on stance focus either on the study of evaluation or epistemicity or affect, each of which have long standing traditions.

Turning to the definition of evaluation as a dimension of stance, we provide a preliminary characterization of evaluative stance as follows:

* The social act of assessing social actors, events and propositions by assigning positive or negative values which are grounded in socio-cultural systems of beliefs and opinions and which express a subjective positioning of the speaker and intersubjective alignment or disalignment with specific communities of speakers.
* From the perspective of the linguist’s task as a discourse analyst, evaluative stance is the set of lexical resources which express positive or negative assessment of entities, events and propositions, which are open to examination by using linguistic analysis and annotation methods.

Regarding methodological approaches to the identification, annotation, and analysis of evaluation, two trends can be identified: those that make use of corpus-assisted methods for the identification of evaluation patterns, often in large corpora (Hunston 2011, Hunston and Su 2019, Bednarek 2009), and those that carry out smaller-scale qualitative studies (see, for example, Bednarek 2014). Corpus-assisted studies often involve the analysis of adjectives as the prototypical form for the expression of evaluation, given that numerous evaluative adjectives contain positive or negative assessment as part of the stable meaning of the word (see White 2016: 81). However, numerous studies on evaluation are smaller scale studies, with a focus on the context-dependent nature of the evaluation.

**3 Data and methodological procedure**

In this section we describe the categories that make up the annotation scheme of evaluative stance and briefly explain the reasons which underlie the choice of each of the categories. We also address specific conceptual and methodological issues discussed in previous literature. Finally, we provide prototypical examples for each category by way of “gold standard”, with the aim of guiding the annotation process.

**3.1. The corpus**

In the project within which the present study was carried out, our aim was to provide an overview of the presence of evaluative stance expressions in our corpora, without, however, limiting this to a specific grammatical category such as the adjective or the adverb. The critical perspective underlying the corpus motivated the collection of two subsets of data within the corpora of opinion and political discourse, so that these corpora would allow for potential studies comparing evaluation strategies used by conservative and progressive politicians and newspapers. [[4]](#footnote-4)

The corpora are annotated manually and are used both for fine-grained analysis of evaluative expressions in smaller-scale case studies[[5]](#footnote-5) and for corpus-assisted studies of the distribution and frequency of the categories in the four genres.

**3.2. Establishing a scheme for the identification and annotation of evaluative stance in discourse: conceptual and methodological challenges**

*What counts as the unit of analysis?*

The first and most difficult step is determining what counts as an evaluative expression. In our annotation scheme, an evaluative expression must meet the following criteria:

1. It expresses positioning of the speaker/writer with regard to a topic
2. It evaluates an event or an entity, or a proposition
3. The conceptual domain is the expression of attitudes and opinions
4. It expresses positive or negative value
5. It expresses dis/alignment with communities of speakers or voices
6. It expresses first-hand stance (speaker/writer stance, that is Attribution is excluded)

Criteria 1, 2 and 5 (positioning, dis/alignment and evaluation) correspond with Du Bois’s (2007) notion of stance. Criteria 3, 4 and 6 restrict the notion of evaluation, narrowing it down to expressions concerning a speaker/writer’s attitudes or opinions and which manifest a positive or negative assessment of phrases or propositions (for a similar approach to evaluation see Bednarek 2009). This narrow perspective on evaluation focuses only on what has been referred to by some authors as moral evaluation (based on the opposition between positive and negative or good and bad) rather than on descriptive evaluation (for this distinction, see Alba-Juez and Makenzie 2014). Similarly, although we acknowledge the importance of second-degree stance (Attribution in Martin and White’s 2005 Appraisal model) in certain types of texts, the objective of our project in an initial stage was to analyze first-degree stance, leaving the expression of second-degree stance to a later stage.

For practical reasons, the unit of analysis adopted in our annotation scheme is the word (“hypocrisy”, “mendacious”, “damage”), and, when necessary, the phrase. Multi word units such as phrasal verbs (“storming out”), compound words (“stumbling blocks”), and hyphenated words (“life-destroying”) are considered as one word unit. The identification of metaphoric expressions, in some cases, required the identification of phrases as units (“close the door nicely”, “storm out in a huff”).

*Annotating part of speech*

Our annotation scheme consists of four levels of annotation: *part of speech, function, metaphoricity* and *value*. Part of Speech (PoS) and function are closely connected, since each function is prototypically performed by a related PoS, but this is not a one-to-one relation and for this reason we decided to maintain the two levels of analysis. This would enable us to carry out studies across genres for the purpose of determining specific tendencies regarding PoS. Within PoS, we consider *nouns*, *adjectives*, *verbs* and *adverbs*, leaving out *prepositions* and other units above the word and phrase levels.

In most cases, the label for the PoS will refer to a word category, and only exceptionally to phrases. The most prototypical category for the expression of evaluation, as confirmed by previous research (Biber et al. 1999: 512-513), is the category *adjective*, which has been the easiest to identify. Prototypical examples of evaluative adjectives are expressions such as “mendacious”, “hypocritical”, “fluent”, “reactionary”, “bad”, “best”, “worse”, “rank”, “dreadful”, “deplorable”. With regard to *adverbs*, we include two types of evaluative adverbs in our scheme. First, sentential adverbs which express stance towards a proposition such as, for example, “fortunately”, “sadly”, “luckily”. These correspond to Biber et al.’s (1999: 856) attitudinal adverbials. Second, manner adverbs express an assessment that takes scope over the verb phrase, for example “*desperately* clawing back their lost power”, “acting *dispassionately”*.

Some difficulties were encountered in the identification of evaluative adverbs, which had to do mostly with the differentiation between attitudinal sentential adverbs and style adverbials which do not express evaluation, such as “frankly” or “indeed*”* (Biber et al. 1999: 858).[[6]](#footnote-6) An additional challenge was the differentiation between evaluative manner adverbs and markers of *Graduation* (see Martin and White 2005), i.e., adverbs such as “very” or “extremely”; these are not considered as markers of evaluation in our scheme, since their semantic domain does not concern the expression of attitudes or opinions, but the degree of force or focus of an evaluative expression which they modify. Attention will have to be paid to expressions such as “terribly”, which may express *Graduation* in one context (terribly confused) but evaluation in another (you are playing terribly).

*Nouns* and *verbs* are particularly challenging when it comes to the identification and coding of evaluative stance. The reason for this concerns the distinction between descriptive and moral evaluation mentioned earlier and the fact that evaluative nouns and verbs perform a double function simultaneously, namely *categorizing or classifying* an entity or an event and at the same time *expressing an attitude* towards the entity or event in question. Van Dijk (1995: 29) provides the example of the word “thief”, which can be used to represent or categorize a person in terms of a fact as established, for example, by a court of law, or it can be used to express an opinion or attitude. The difference between these two uses can only be determined in context.

It can be argued that numerous categorizations form part of socially established norms of behaviour and beliefs, as can be observed in words such as “racism”, “truth”, “lie”, and even “white”. In many contexts, words such as these will be describing facts, but, in others, they may express opinions or attitudes. In order to determine whether words such as these are evaluative or not the crucial criterion we have adopted is the application of Du Bois’ (2007) definition of stance; that is, to check whether the use of the expression shows the speaker/writer’s positioning and dis/alignment with communities of speakers (section 2 above) in specific contexts of use. Prototypical examples of evaluative nouns in our scheme are words such as “bigotry”, “hypocrisy”, “cynicism”, “problem”, and prototypical examples of evaluative verbs are “legitimize” and “demonize*”*. Numerous examples of evaluative nouns and verbs will also be metaphoric, as discussed below.

*Annotating function*

The second level of annotation is termed *function* and includes three categories: *classifying, predicational* and *attitude*. *Classifying* and *predicational* categories relate to nomination and predicational strategies in critical discourse analysis (CDA) (Van Leeuwen 2008: 40-41; Reisigl and Wodak 2009: 95), although the categories in our present scheme do not correspond exactly with the categories proposed by CDA. The motivation for establishing this level of analysis lies in the critical discourse orientation of the project within which the annotation scheme is developed.

The classifying strategy in our scheme includes nouns and verbs used to represent social actors and events in a positive or negative light, while the predicational strategy includes adjectives and manner adverbs, which are used to assess social actors and events according to positive or negative values. The category *attitude* is inspired in Biber et al.’s (1999: 856) categorization of stance adverbials. While classifying and predicational expressions relate to phrases within the clause, evaluative expressions within attitude take scope refer to the whole proposition.

*Annotating metaphoricity*

The third level of annotation is concerned with *metaphoricity* and distinguishes between *metaphoric* and *non-metaphoric* evaluative expressions. The evaluative role of metaphor has been pointed out in previous approaches to evaluation, such as Martin and White (2005: 64-64), who argue that within Appraisal, lexical metaphors provoke rather than invoke “an evaluative response in readers”. In our annotation system we do not distinguish between *invoked* and *provoked* evaluation; however, we believe that evaluative metaphors deserve being examined as a separate category, given their important role in framing discourse topics (Deignan 2010; Semino et al. 2018) and in creativity (Fuoli, Littlemore and Turner 2021). Thus, including metaphoricity as a dimension of evaluation will allow researchers to determine the frequency and distribution of evaluative metaphors in different discourse genres and in texts with different ideological motivations.[[7]](#footnote-7)

In order to determine whether an evaluative expression is metaphoric in context we have applied two criteria: first, following MIP (Pragglejazz 2008) and MIPVU (Steen et al. 2010), the contextual meaning of the unit (that is, the one in the text) is compared to its basic meaning (i.e., the first lexical entry in the MacMillan dictionary). The second criterion is that it is possible to identify a mapping from a source domain which prototypically (though not always) will carry a positive or negative connotation onto a target domain. If these two criteria are met the expression is annotated as metaphoric; if the criteria are not met it is annotated as not metaphoric. Prototypical examples of evaluative metaphors are those which are not highly conventionalized, but rather involve high degrees of incongruity between source and target domains (for example, the “Brexit rainbow”) or involve processes of de-agentivization or de-humanization. For example, in a *Guardian* article on Theresa May, she is described as having been “rubbished” and “junked” by the conservatives in her own party. These verbs are good examples of evaluative metaphors in that they talk about a person in terms of a (negative) object, rubbish, thus de-humanizing the represented social actor and clearly conveying a negative attitude. Other clear examples are metaphors used to represent Brexit as a discourse topic, for example Brexit as a “cliff-edge” or as a “precipice”, which highlight the potential negative effects of Brexit.

*Annotating value*

The fourth level of analysis is *value*, which we restrict to *positive* and *negative* (see Bednarek 2009: 166). Although value is described in the literature as a cline which ranges from positive to negative including neutral positions (Martin and White 2005: 93), it is also recognized that the opposition between positive and negative is one of the basic parameters in the analysis of evaluation (see Biber and Finegan 1989). Identification of value involved some difficulties, since evaluation is totally context-dependent, as pointed out by numerous scholars. For example, White (2016: 81) shows how the word “fabricate” can be used in one context with no evaluative meaning and with a negative evaluative meaning in another.

The difficulties in the identification of value have not received sufficient attention in the literature. A key issue is the lack of fit between positive and negative connotation of words and the evaluative stance that may be adopted in specific contexts.[[8]](#footnote-8) In other words, the positive or negative connotation of a word such as, “war” or “peace”, does not guarantee that the choice of that word in a given context will express the positioning of a writer towards a topic. That is, the inscribed positive or negative connotation of a word does not necessarily imply an act of evaluative stance-taking, which has as a result a positive or negative evaluation of an entity or event. This is due to the fact that while word connotation takes place at the level of semantics, stance-taking is a discourse-pragmatic phenomenon. This problem poses a great challenge for the annotation of evaluation, since all potentially evaluative words need to be checked in context. Two criteria have been adopted to determine value: first, that a positive or negative connotation can be observed in the context of use motivated by the speaker/writer’s positioning towards a topic, and second, that the expression invokes relations of dis/alignment with other voices or communities of speakers. These criteria are discussed and illustrated in section 4 below.

**4 Testing the reliability of a step-wise method for the annotation of evaluative stance**

In this section, we report on the method we have followed for the identification and annotation of evaluative stance and on the inter-rater reliability process. We show the stages followed in the design and application of the scheme and the steps in the annotation process. We illustrate how the protocol is applied to the analysis of evaluative terms in an extract and we discuss the results of the inter-rater reliability process.

**4.1 Streamlining the annotation procedure in six stages**

The complete process was carried out by a team of seven researchers who worked together in the initial stages of the procedure (stages 1 to 4 in Figure 2 below) and then split into pairs to complete the annotation of the four genres. In this article, we report on the results of inter-rater reliability carried out by the two authors of this article. We discuss examples from the various genres analysed in stages 1 to 4 below, although our task, following the completion of these stages, was to annotate the press corpus. In order to reduce the analyst bias in the task of identification of evaluation, we followed the stages illustrated in *Figure 2*, which rely on previous studies of a similar nature (see Fuoli and Hommemberg 2015; and Fuoli, Littlemore and Turner 2021).

Diagrama

Descripción generada automáticamente

*Figure 2. Stages in the annotation of the corpus*

After compiling the corpus (stage 1), we developed an annotation manual that contained the explanation of the categories described in Section 3 with prototypical examples (stage 2).[[9]](#footnote-9) Due to the inherent subjective nature of the task, the team established a 4-round training stage (stage 3) to annotate random samples of the full corpus containing text extracts from press, political speeches, Internet fora and science popularization. The aim of the inter-rater reliability test was (1) to confirm the understanding of the categories by the annotators, (2) to test the versatility of the model to account for a myriad of evaluative stance examples in different genres, and (3) to revise and/or refine the guidelines and categories when necessary. Inter-rater reliability scores were reported after each round to motivate the need for further adjustments to the protocol (stage 4). Finally, once the inter-rater reliability scores confirmed that both researchers applied the protocol in a similar and consistent way, the remainder of the corpus was divided between the researchers and coded independently (stage 5). Ambiguous cases were resolved through joint discussion between the two researchers, and annotations were removed whenever agreement was not reached, although this was the exception rather than the rule.

**4.2 Stage 2: the coding scheme**

Our stepwise method for the identification and classification of expressions with potential for evaluation (designed in stage 2 and applied in stages 3 and 5) is illustrated in *Figure 3* and comprises 7 different steps (a-g), which address the dimensions of our scheme for the annotation of evaluative stance described in Section 3. In line with previous studies that have addressed the annotation of stance, the process involves two different tasks: identification of markable units, that is, selecting lexical units which have the potential to express evaluation (step c) and classifying such units along several dimensions (steps d-g).

Diagrama

Descripción generada automáticamente

*Figure 3. Steps in the annotation of evaluative stance*

Let us illustrate the application of the coding scheme with the following example (1) extracted from the corpus of press articles. For the sake of clarity, we underline the lexical units of interest in this extract, together with the annotation that was carried out in each case.

(1) The resolution of the backstop question has become a game <NP, CLA, MT, NG> of political machismo <NP, CLA, NMT, NG>. (…) Brexit is a headache <NP, CLA, MT, NG> not of Ireland’s making, in which its economic fortunes are the most heavily implicated of all EU countries: an unenviable <AP, PRE, NMT, NG> combination. (EOT-10)

After reading the full text (a), the second step (b) is to identify the main topic and/or social actors addressed in the example under consideration. In this case, the opinion article is entitled “Britain and Ireland are truly in this together”, and according to the subheading, the stance of the author is clear: “Ignore identity. Pragmatism and co-operation are key to the border issue”. Therefore, one should expect a positioning of the author in favour of a customs regime that does not regard any difference between Northern Ireland and the rest of Ireland, that is a pragmatic solution for the backstop problem within Brexit. The next step involves deciding whether (c) the author is expressing a positioning towards the events or social actors referred to in the potentially evaluative units underlined in the extract. We consider that the underlined terms in this context express a critical positioning of the author of the article towards Brexit and the way the backstop issue is being dealt with (*headache, game*), the politicians involved in this process and their actions (*machismo*) and the situation of Ireland in this process (*unenviable*), at least in the sense of recognizing the situation as very problematic.

The next steps seek to confirm how this evaluative potential is manifested in each lexical item. The items are first classified according to part of speech (step d) and function (step e) with regard to their sentence positions and functions. Thus *headache, game* and *machismo* are annotated as Nouns performing a Classifying function, while *unenviable* is annotated as Adjective performing a Predicational function. The next step (f) involves annotating metaphoricity. In order to determine whether the underlined expressions are metaphoric, we apply the two criteria established in our protocol, namely, to check whether the meaning of the expression coincides with the first lexical entry in the MacMillan dictionary and to observe whether there is a mapping of a positive or negative source domain onto a target domain. When applying these criteria, we find that the terms *game* and *headache* are metaphoric, while *machismo* and *unenviable* are not metaphoric and are annotated accordingly. The meanings in context of *machismo* and *unenviable* correspond to the first definitions in the dictionary. The meaning of the term *game* corresponds to lexical entry 6 in the MacMillan dictionary: ‘an activity or situation that someone seems to be treating less seriously than it should be treated’, while lexical entry 1 is: ‘an activity that you do for fun that has rules, and that you can win or lose’. Additionally, we can identify a mapping between a source domain of a (card or board) game as fun and not serious onto the target domain of a political activity (the backstop resolution). Similarly, the word *headache* in this extract has the meaning displayed in lexical entry 2 in the dictionary, ‘something that causes you a lot of problems’, and not the one in lexical entry 1 (a pain in your head). A mapping from a physical, more concrete, domain which has a negative connotation (pain in head) to a more abstract domain (the Ireland problem within Brexit) can be observed.

The final step (g) involves annotating value. In order to determine value, we apply the criteria of identifying positive or negative connotations of the words in context and the evocation of relations of dis/alignment with other voices. When considering the connotations, the words *machismo, headache* and *unenviable* carry explicit negative connotations, that is, they show inscribed evaluation. The word *game* acquires the negative connotation in context. The incongruity between games as fun activities and the seriousness of the backstop issue leads us to interpret that the word *game* here projects a negative criticism of the way in which the issue is being dealt with within Brexit negotiations.

We finally examine the expressions to determine whether they evoke relations of dis/alignment with communities of speakers, following Du Bois’s (2007: 162) claim that all evaluative expressions evoke different degrees of dis/alignment. We argue that the evaluative expressions considered together express different degrees of disalignment of the writer with other voices, specifically those of politicians responsible for Brexit in that they are not finding an acceptable solution for the Northern Ireland situation. This disalignment is explicitly manifested in the expressions ‘game’, which intertextually evokes the discourse of politicians involved in Brexit as not dealing with the issue seriously enough, and (political) ‘machismo’, an open criticism, and consequently a rejection, of their attitude. But disalignment is also implied in the negative construal of Brexit as a ‘headache’, that is, as something which is deeply problematic, and the representation of Ireland as being in an ‘unenviable’ situation. These words do not necessarily evoke other real voices or positions, but they seem to evoke alternative imagined situations as possible, that is Brexit as unproblematic, and Northern Ireland’s situation as enviable.

In this sense, all these choices contrast and resonate with the intertextually evoked positionings of politicians such as Boris Johnson and the Conservative party, whose slogans highlight Brexit as unproblematic or even as having positive effects, such as ‘let’s make Great Britain great again’ (positive construal of effects of Brexit) and ‘let’s get Brexit done’ (executive attitude to Brexit which downplays emotional and attitudinal dimensions). These examples show that negative and positive connotations in context are crucial for determining value, and that identifying dis/alignment can support this analysis and can provide a broader contextual perspective on how evaluation works in a text. However, while disalignment can be observed explicitly in some expressions, it is often either implicit or even not so relevant for determining value. It is possible that dis/alignment is manifested through a cumulative process throughout the text, while value needs to be assigned individually for each term.

**4.3 Stage 3: Inter-rater reliability testing**

Reliability testing (stage 3) was a key stage in the process of refining the method of evaluative stance annotation. Inter-rater reliability refers to the extent to which independent coders make similar annotations based on the same set of rules. High inter-rater reliability scores indicate that a set of guidelines is explicit and transparent enough for two independent annotators to produce a similar annotation and can therefore be taken as a proxy of the *reliability* of a protocol. Furthermore, if two annotators can produce similar results separately by relying on the same set of instructions, it can be expected that other researchers, external to this project, can replicate the analysis based on the same set of instructions. However, it should be acknowledged that a set of transparent and explicit instructions, by themselves, may not suffice; and that the researchers should also meet additional conditions to find converging annotations.

Identifying and annotating evaluative stance is a highly complex and subjective process. Despite having explicit guidelines, as well as the levels and granularity of the annotation, many decisions that are taken before starting the annotation are highly dependent on text-specific features, for example, the topic of the text or the genre to be annotated, as well as on the researchers’ own bias, expertise, or attention. These features may lead to slight deviations from the initial guidelines for annotation and may have an impact on the quantitative and qualitative analyses based on the protocol. In order to optimise the consistency in the application of the rules, as well as the transparency in the formulation of such guidelines, an inter-rater reliability test was conducted prior to the full annotation of the corpus.

Inspired by the fine-grained study carried out by Fuoli and Homemberg (2015), we adopted a double-coding approach where two researchers (the authors) annotated independently a random sample of the corpus in 4 subsequent rounds as explained in section 4.2 above. As shown in Table 1 below, the inter-rater reliability coding amounted to 4862 words (1215 words by round on average).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Press** | **Politics** | **Fora** | **Science** |
| *Round 1* | 443 | 378 | 357 | 505 |
| *Round 2* | 340 | 228 | 190 | 228 |
| *Round 3* | 269 | 231 | 214 | 258 |
| *Round 4* | 307 | 331 | 333 | 250 |
| *TOTAL* | 1359 | 1168 | 1094 | 1241 |

*Table 1. Words coded in each round of the inter-rater reliability test*

The annotation involved two tasks: selecting text units expressing evaluative stance, and annotating them in four dimensions: *PoS, function, metaphoricity* and *value*. After each round, inter-rater reliability scores were calculated, and a joint discussion of the diverging annotation was held to either (a) refine the protocol where there was an agreement on an aspect that needed clarification or (b) discard ambiguous cases where agreement was not reached. We are aware that inter-rater reliability as a method has its limitations, in the sense that it shows only the degree of agreement of the researchers involved, and, additionally, it means that those items on which there is disagreement are discarded, with the consequent potential damage for the identification of evaluative expressions. However, we considered that the benefits of the method surpassed its limitations.

In order to assess the robustness of our protocol, we have adopted a mixed-test approach for our inter-rater reliability study. For the first task, i.e., the identification of units of evaluative stance (step b), we report *precision, recall and F-score* to assess the overlap between the units identified by both researchers. For the classification task (steps d-g), we removed the units of analysis identified by only one of the researchers and performed a Cohen’s Kappa test on the ‘agreed’ units of analysis.

**4.4 Inter-rater reliability results**

Because of the reliability results after four rounds of training, the instructions were clarified, the prototypical examples were revised, and more transparent guidelines and instructions were provided to deal with ambiguous cases, complemented with a joint discussion at the end of the process to clarify the most controversial cases (stage 4). A refinement and elaboration of these instructions was carried out during the inter-rater reliability rounds and as a result the final version of the annotation protocol was confirmed.[[10]](#footnote-10)

***Task 1:* *Identification of units of evaluative stance (step b)***

*Table 2* reports the results of the reliability test performed on the first annotation task on 211 lexical units, where we identified the lexical units with potential for evaluation. *Precision* is calculated by taking the number of matches between both researchers divided by the total number of units identified by the second researcher (which might include some lexical units not identified by the first researcher, that is taken artificially as the ‘gold standard’). In turn, *recall* measures the success rate of the second researcher, i.e., the extent to which their annotation coincides with the first researcher. It is calculated by taking the matches between both authors divided by the total number of units coded by the first researcher (in other words, what is artificially considered as ‘correct’). Finally, F-score is the harmonic mean of precision and recall.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Author 1** | **Author 2** | **Matches** | **Precision (Matches/**  **Author 2)** | **Recall**  **(Matches/**  **Author 1)** | **F-score**  **(Precision +**  **Recall / 2)** |
| *Round 1* | 45 | 47 | 39 | 0,83 | 0,87 | 0,85 |
| *Round 2* | 38 | 41 | 36 | 0,88 | 0,95 | 0,91 |
| *Round 3* | 47 | 41 | 36 | 0,88 | 0,77 | 0,82 |
| *Round 4* | 57 | 55 | 49 | 0,89 | 0,86 | 0,88 |

*Table 2. Inter-rater reliability tests on the first annotation task (identification of lexical units with potential for evaluation)*

Even though there is no standard threshold of acceptability, our findings show high reliability compared with similar studies (Fuoli and Homemberg 2014; Read & Carroll 2012; Taboada & Carretero 2010). In particular, the F-score in the third round shows the greatest mismatch between both researchers: from a total of 53 of lexical units identified as having potential for evaluation, both coders agreed only on 36, meaning that they differed on 17 lexical units (11 coded exclusively by the first researcher, and 6 by the second researcher). For example, there was disagreement regarding the extent to which words with potentially inscribed evaluation (such as “new”) show the writer’s positioning to a topic. Consider (2):

(2) If you run an exporting business in Penarth, you need a Brexit that will open up new global markets.

It is unclear whether “new” in this context refers to a market that did not exist before and therefore conveys the idea of something that has been recently developed (in line with the first four entries in Macmillan) or whether it also evokes happiness and hope associated with new things (fifth entry in Macmillan). It might be that both interpretations are valid and impossible to separate from each other, thus reflecting the challenges and subtleties behind the identification of evaluative stance.

Another source of debate had to do with labels used to refer to groups of people with a potential for evaluation. Consider (3):

(3) This weekend it took the El Paso native and Democratic candidate Beto O’Rourke to draw (…) explicit link between the rhetoric of Donald Trump and the white nationalism that leads to shootings like this one, and he was of course right.

The researchers disagreed regarding the extent to which “white nationalism” expressed a positioning of the writer towards the topic and, consequently, an evaluation. It was annotated as evaluative by one researcher, who considered the expression implied a critical positioning towards a certain way of thinking in the US associated to Trumpism in this text, which also included explicitly negative evaluative terms such as ‘supremacist’. However, it can also be argued, as pointed out by the second researcher, that it is unclear whether there is a non-evaluative option to refer to white people with a far-right ideology, regardless of the stance of the writer with respect to this concept and what it represents, even if it is acknowledged that it expresses evaluation. In cases such as these, in which the disagreement was not resolved, we decided to be conservative and the expressions were disregarded. However, such disagreements are written down and will be taken into account for future refinements of the protocol.

***Task 2: classification task (steps d-g)***

For the second reliability test (classification of evaluative lexical units), we only retained the *matches* between both researchers in the first task, which left us with 160 units for the second task. *Table 3* shows the Cohen’s kappa scores for the classification of the lexical units in parts of speech (PoS) in general, and then for adverbs, adjectives, nouns and verbs.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **PoS** | **ADVP** | **AP** | **NP** | **VP** |
| *Round 1* | 0.96 | 1 | 0.95 | 0.92 | 1 |
| *Round 2* | 0.933 | 1 | 1 | 0.873 | 0.842 |
| *Round 3* | 1 |  | 1 | 1 | 1 |
| *Round 4* | 1 |  | 1 | 1 | 1 |

*Table 3. Cohen’s kappa scores for the classification of the evaluative lexical units in PoS*

As can be seen, despite the overall positive results (after all, PoS might be one of the most objective dimensions of annotation) mismatches in agreement can be observed in the second round. This had to do with the fact that the word as a unit had not been established as a criterion, and, consequently, there were mismatches in the identification of the PoS of evaluative expressions. For example, the sequence ‘far right’ was annotated as NP (far right) by one annotator while it was annotated as AP (far) by the other. The criterion was then adopted to annotate only the evaluative word or expression, in this case, the word ‘far’ modifying ‘right’.

*Table 4* reports the reliability scores for the annotation of functions (in general and in particular, for attitude, classifying and predicational). No major disagreements were encountered in this dimension.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Function** | **ATT** | **CLA** | **PRE** |
| *Round 1* | 0.96 | 0.84 | 1 | 0.95 |
| *Round 2* | 1 | 1 | 1 | 1 |
| *Round 3* | 1 |  | 1 | 1 |
| *Round 4* | 1 | 1 | 1 | 1 |

*Table 4. Cohen’s kappa scores for the classification of the evaluative lexical units in discourse functions*

We now turn to the reliability scores for the annotation of metaphors (*Table 5*).

|  |  |
| --- | --- |
|  | **Metaphor** |
| *Round 1* | 0.87 |
| *Round 2* | 0.59 |
| *Round 3* | 0.89 |
| *Round 4* | 1 |

*Table 5. Cohen’s kappa scores for the classification of the evaluative lexical units in terms of their potential for having a metaphorical interpretation*

In the first and second rounds, a mismatch was found in the annotation of conventional metaphors such as “far right” (4) and “enlightened” (5), where the evaluative expressions were annotated as metaphorical only by one researcher.

(4) The far right has already done the same.

(5) It was Disraeli’s Government that recognised improving public health depended on passing enlightened environmental legislation.

To resolve the disagreement, the extension of the Metaphor Identification Procedure developed at the Vrije Universiteit (MIPVU) was applied and this criterion revealed that “far” in “far right” is metaphoric, as there is a sharp contrast between its contextual meaning (how extreme someone’s actions are) and the more basic meaning in the dictionary (far used to talk about distance). In contrast, “enlightened” is not metaphoric, since the meaning of the expression corresponds with the first lexical entry in the dictionary: “someone who is enlightened has attitudes that are considered modern and reasonable”. Even if both expressions are evaluative, only the first one is annotated as metaphoric.

Cases such as “enlightened” are examples of highly conventional metaphors which have lost their metaphorical value and can be understood without having to evoke the source domain – this is known in the literature as “dead metaphors” (Müller 2009). In general terms, our goal was not to annotate all metaphoric expressions, but only to annotate metaphors that had a deliberate aim of expressing evaluation. This meant that evaluative metaphors typically turned out to be less conventional and to have more creative, ad hoc mappings.[[11]](#footnote-11)

Finally, we report the Cohen’s kappa scores for the fourth and last dimension of analysis, value (*Table 6*).

|  |  |
| --- | --- |
|  | **Value** |
| *Round 1* | 0.90 |
| *Round 2* | 0.88 |
| *Round 3* | 1 |
| *Round 4* | 1 |

*Table 6. Cohen’s kappa scores for the classification of the evaluative lexical units in terms of their potential positive or negative value*

Once again, we observe the patterns detected in the previous tables: the first two rounds seem to be the most challenging ones for the researchers, and it is only following the decisions adopted and the refinements made to the protocol after the second round that agreement improves.

An interesting case in the annotation of value occurred in the second round, in the annotation of the word “traditionalist”, which appears twice in (6):

(6) It's a traditionalist argument to say that marriage should be between two consenting adults in the same way it is a traditionalist argument to say that marriage should be between one man and one woman.

This example was retrieved from a forum on gay marriage. It was unclear from the thread of posts whether the overall positioning of the thread was for or against, as fora are characterized by their multi-voiced profile (see Sánchez-Moya and Maíz-Arévalo forthcoming //update//). Is “traditionalist” used here with a positive or negative value? The writer of the post later says “Just to clarify, I admit my faith in Christianity naturally leads me to a biased position against the notion of gay marriage”, which seems to suggest an awareness of the negative connotations of “traditionalist” as old-fashioned and unfair (and therefore negative, as the first researcher coded), but still the writer chooses to express alignment with those traditions. In this case, “traditionalist” would show positive evaluation, as coded by the second researcher.[[12]](#footnote-12) This example clearly reflects the ambivalence of some words with alleged “inscribed” evaluation that in some contexts might be interpreted differently by specific readers. Interestingly, the writer’s awareness of the potential confusion triggered by the choice of the word “traditionalist” is reflected in the mismatch between the annotations of the researchers. The criterion that prevailed in this case was that of the perceived “alignment” of the writer with positions in favour of marriage, be it heterosexual or gay.

**5. Conclusions**

This article has discussed some of the difficulties and challenges involved in the identification and annotation of evaluative stance in discourse by presenting a scheme for the annotation of evaluative stance and the methodological procedure we have followed in the implementation of the scheme. It can be argued that our scheme and method contribute to advancing research on the identification and annotation of evaluation for several reasons: first, we propose a set of clear criteria for the identification of evaluation based on the relation between the concepts of evaluation and stance as involving positioning and dis/alignment (du Bois 2007). Second, we present a replicable protocol which consists of four dimensions and we show how it has been applied to the annotation of a sample of texts as a preliminary step in the annotation of a larger corpus. Third, we test the degree of agreement between annotators, which has proved useful in determining the reliability of the annotation scheme.

However, the study also has some limitations, which need to be taken into account for further refinements of the protocol to inform future studies on evaluation. First, the design of the protocol was influenced by the nature and aims of our research project, in which the concept of evaluation is conceived as restricted regarding several important issues. These include the concept of evaluation itself, which focuses on opinion and attitude but excludes epistemicity and affect; the notion of value, which is restricted to positive and negative, thus excluding neutral value; and the use of the word and the phrase as unit of analysis, thus excluding longer stretches of discourse. Second, although the inter-rater reliability process proved extremely useful to debate and clarify the characteristics of evaluative expressions and to establish agreement among annotators, its application also meant that more difficult, and perhaps more interesting examples of evaluation were disregarded at this stage. All in all, our study contributes to showing the importance of developing annotation protocols and methods for the study of evaluation, an area of research which certainly deserves greater attention.

**References //follow T&T house style strictly//**

Alba-Juez, Laura and Geoff Thompson (eds.). 2014. *Evaluation in context*. Amsterdam: John Benjamins.

Alba-Juez, Laura & J. Lachlan Mackenzie. 2016. *Pragmatics: Cognition, context and culture*. Madrid: McGraw Hill.

Bednarek, Monika. 2009. Language patterns and Attitude. *Functions of Language*

16(2). 165-192.

Bednarek, Monika. 2014. “An astonishing season of destiny!” Evaluation in blurbs used for advertising TV series. In Laura Alba-Juez and Geoff Thompson (eds.) *Evaluation in context*. 197-220. Amsterdam: John Benjamins.

Bednarek, Monika and Helen Caple. 2014. Why do news values matter? Towards a new methodological framework for analysing news discourse in critical discourse analysis and beyond. *Discourse & Society* 25(2). 135-158.

Biber, Douglas and Edward Finegan. 1989. Styles of Stance in English. Lexical and Grammatical Marking of Evidentiality and Affect. *Text* 9. 93-124.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *The Longman grammar of spoken and written english*. London: Longman.

Deignan, Alice. 2010. The evaluative properties of metaphors. In Graham Low, Zazie Todd, Alice Deignan and Lynne Cameron (eds.) *Researching and applying metaphor in the real world*, 357-374. Amsterdam: John Benjamins.

Du Bois, Jean W. 2007. The stance triangle. In Robert Engelbreston (ed.) *Stancetaking in discourse*, 139-182. Amsterdam: John Benjamins.

Englebretson, Robert (ed.). 2007. *Stancetaking in discourse*. Amsterdam: John Benjamins.

Fuoli, Matteo. 2018. A stepwise method for annotating Appraisal. *Functions of Language* 25(2). pp. 229–258.

Fuoli, Matteo and Charlotte Hommerberg. 2015. Optimising transparency, reliability and replicability: Annotation principles and inter-coder agreement in the quantification of evaluative expressions. *Corpora* 10(3). 315–349.

Fuoli, Matteo, Jeannette Littlemore, and Sarah Turner, S. 2021. Sunken ships and screaming banshees: Metaphor and evaluation in film reviews. *English Language and Linguistics*. 1-29.

DOI: <https://doi.org/10.1017/S1360674321000046>

Hidalgo-Downing, Laura and Paula Pérez-Sobrino. Forthcoming. Pushing Britain off the precipice: A CDA approach to negative evaluative stance in opinion articles on Brexit. In Juana Marín Arrese, Laura Hidalgo-Downing and Juan Rafael Zamorano Mansilla (eds.), *Stance, Inter/subjectivity and identity in discourse*. Frankfurt: Peter Lang.

Hidalgo-Downing, Laura, Paula Pérez-Sobrino, Laura Filardo-Llamas, Carmen Maíz-Arévalo, Begoña Núñez-Perucha, Alfonso Sánchez-Moya, Julia T. Williams Camus. Submitted. A protocol for the annotation of evaluative stance and metaphor across four discourse genres. *Revista Española de Lingüística Aplicada*.

Hunston, Susan. 2011. *Corpus approaches to evaluation: Phraseology and Evaluative Language*. London: Routledge.

Hunston, Susan and Geoff Thompson. 2000. *Evaluation in text*. Oxford: Oxford University Press.

Hunston, Susan and Hang Su. 2019. ‘Patterns, constructions, and local grammar: A case study of evaluation’. *Applied Linguistics* 40(4). 567-593.

Martin, James R. and Peter R. R. White (2005) *The language of evaluation. Appraisal in English*. Palgrave, Macmillan.

Müller, C. 2009. *Metaphors Dead and Alive, Sleeping and Walking. A Dynamic View*. Chicago: University of Chicago Press.

Núñez-Perucha, Begoña and Laura Filardo-Llamas. Forthcoming. From "roaring lion" to "chlorinated chicken": evaluative stance and ideological positioning in a corpus of British political discourse. In Marín-Arrese, Juana, Laura Hidalgo Downing and Juan Rafael Zamorano Mansilla (eds.) *Stance, Inter/subjectivity and identity in discourse*. Frankfurt: Peter Lang.

Pragglejazz. 2007. A method for identifying metaphorically used words in discourse’, *Metaphor and Symbol* 22(1). 1-39.

Read, J., & Carroll, J. (2012). Annotating expressions of Appraisal in English, *Language Resources and Evaluation*, 46(3), 421–447. https://doi.org/10.1007/s10579-010-9135-7

Reisigl, Martin, and Ruth Wodak. 2009. The discourse historical approach. In Ruth Wodak and Michael Meyer (eds.). *Methods for critical discourse analysis*, 87-121. London: Sage.

Sánchez-Moya, Alfonso and Carmen Maíz-Arévalo. forthcoming. “Histrionic, appalling, a major turkey”: the expression of evaluative stance in the discourse of online forums. In Juana Marín Arrese, Author 1 and Juan Rafael Zamorano Mansilla (eds.), *Stance, Inter/subjectivity and identity in discourse*. Frankfurt: Peter Lang.

Semino, Elena, Zsofia Demjén, Sheila A. Payne and Paul Rayson. 2018. *Metaphor, cancer and the end of life: A corpus-based study*. London: Routledge.

Simaki, Vasiliki, Paradis, Carita, Kerren, Andreas. 2018. Evaluating stance-annotated sentences from the Brexit Blog Corpus: A quantitative linguistic analysis. *ICAME journal* 42(1). 133-166.

Simaki, Vasiliki, Paradis, Carita, Skeppstedt, Maria, Sahlgren, Magnus, Kucher, Kostiantyn, Kerren, Andreas. 2020. Annotating Speaker Stance in Discourse: The Brexit Blog Corpus. *Corpus linguistics and linguistic theory* 16(2). 215-248.

Steen, Gerard, Aletta G. Dorst, J. Berenike Herrmann, Anna A. Kaal, and Tina Krennmayr. 2010. Metaphor in usage. *Cognitive Linguistics* 21(4). 765-796.

https://doi.org/10.1515/cogl.2010.024

Taboada, M., & Carretero, M. (2010). Contrastive analyses of evaluation in text: Key issues in the design of an annotation system for attitude applicable to consumer reviews in English and Spanish, *Linguistics and the Human Sciences*, *6*(1–3), 275–295. https://doi.org/10.1558/lhs.v6i1-3.275

Thompson, Geoff. 2014. affect and emotion, target-value mismatches, and Russian dolls: Refining the appraisal model. In Laura Alba-Juez and Geoff Thompson (eds.) *Evaluation in context*, 47-66. Amsterdam: John Benjamins.

Van Dijk, Teun A. 1995. Power and the news media. In David L. Paletz (ed.) *Political communication and action*, 9-36. Cresskill, NJ: Hampton Press.

White, Peter R.R. 2016. Evaluative contents in verbal communication. In Andrea Rocci and Louis de Saussure (eds.) *Verbal communication, handbook of communication science*, 77-96. Berlin/Boston: Walter de Gruyter.

1. Hidalgo-Downing and Pérez-Sobrino have contributed equally to this research paper, and therefore should be both considered *joint first authors*. [↑](#footnote-ref-1)
2. The research project is entitled ‘STANCEDISC: Stance and Inter/Subjectivity in Discourse’ (PGC2018-095798), funded by the Spanish Ministry of Science, Innovation and Universities. The present study has also received funding from other research projects funded by the State Agency of Research - Spanish Ministry of Science, Innovation and Universities (FFI2017-82730-P, PID2020-118349GB-I00 and PID2021-123302NB-I00). [↑](#footnote-ref-2)
3. Within the stance project in which the present study was carried out, the annotation of stance was carried out by two complementary groups, one annotating epistemic and deontic stance and the other 3.2. sdsannotating evaluation. [↑](#footnote-ref-3)
4. See Hidalgo-Downing and Pérez-Sobrino (forthcoming) //update// and Núñez-Perucha and Filardo-Llamas (forthcoming) //update// for examples of such studies. In the former, attitudes to Brexit in opinion articles in *The Times* and *The Guardian* are compared, while in the latter, evaluative resources in Conservative and Labour politicians are compared. [↑](#footnote-ref-4)
5. See the following cases studies which have applied the protocol to the analysis of evaluation in different genres: Hidalgo-Downing and Pérez-Sobrino (forthcoming), Núñez-Perucha and Filardo-Llamas (forthcoming), Sánchez-Moya and Máiz-Arévalo (forthcoming) and Williams (Forthcoming). //update all refs// [↑](#footnote-ref-5)
6. In Biber at al. (1999: 857-858) attitude adverbials express evaluation, while style adverbials focus on style and comment on the manner of conveying the message. [↑](#footnote-ref-6)
7. Hidalgo-Downing and Pérez-Sobrino (forthcoming) //update// found that in *The Guardian* evaluative metaphors are more frequent than in *The Times*, and that they tend to be more critical towards Brexit and more creative. [↑](#footnote-ref-7)
8. See Bednarek and Caple (2014: 155) for this distinction. [↑](#footnote-ref-8)
9. Readers interested in the full annotation can download the final version of the protocol, the texts (clean and annotated), and R scripts from our public repository: https://osf.io/5jw7f/. [↑](#footnote-ref-9)
10. For a detailed explanation of the categories and the steps in the protocol, see Hidalgo-Downing et al. (submitted).//update// [↑](#footnote-ref-10)
11. See Fuoli, Littlemore and Turner (2021) for similar results. [↑](#footnote-ref-11)
12. For a discussion of how different interpretations of value may be analysed in terms of the reading positions of readers as compliant, resistant and tactical, see Martin and White (2005: 206). [↑](#footnote-ref-12)