Unveiling subjectivity in press discourse: a statistical and qualitative study of manually annotated articles

Révéler la subjectivité dans le discours de presse : une analyse statistique et qualitative d'articles annotés manuellement

Escouflaire Louis

Descampe Antonin

Fairon Cédrick

UCLouvain, IL&C (Institute for Language & Communication)

Résumé (French abstract)

Cet article présente les résultats d'une expérience d'annotation impliquant 36 participants chargés d'évaluer la subjectivité de 150 extraits d'articles de presse belges francophones et d'identifier les indicateurs linguistiques de subjectivité dans ces extraits. Nous explorons d'abord l'accord inter-annotateurs et les corrélations entre les annotations et les variables associées aux articles, puis nous analysons qualitativement un échantillon de textes. Nous introduisons la « cartographie d'attention textuelle », une méthode pratique pour visualiser au niveau des tokens les résultats des annotations. L'étude révèle que la perception de la subjectivité dans le texte varie d'un lecteur à l'autre, mais que certains tokens sont considérés par beaucoup comme des indicateurs de subjectivité. Des indicateurs attendus, tels que les mots de sentiment et la ponctuation expressive, ont été identifiés, des marqueurs moins établis tels que les marqueurs de discours et les verbes modaux ont également joué un rôle. L'utilisation du pronom « on » a aussi été identifiée comme un indicateur significatif de la subjectivité, malgré son utilisation courante pour dissimuler la présence de l'auteur dans l'écriture journalistique. Ces résultats offrent des perspectives sur la subjectivité dans le discours journalistique et ont des implications significatives pour le domaine du journalisme et pour l'amélioration des modèles de détection automatique de la subjectivité.

Mots-clés (French keywords): subjectivité, journalisme, annotation, analyse de discours, discours de presse

Abstract

This article presents the results of an annotation experiment involving 36 participants tasked with rating the subjectivity of 150 excerpts from Belgian French press articles and with identifying linguistic indicators of subjectivity in the excerpts. We first explore inter-annotator agreement and correlations between the annotation results and the variables associated with the articles, then perform qualitative analysis of a sample of annotated texts. We introduce "textual heat maps", a convenient method to visualise the results of the token-level annotations. The

study reveals that the perception of subjectivity in text varies among readers, but that some tokens are widely considered as indicators of subjectivity. While expected indicators such as sentiment words and expressive punctuation were found to influence interpretations, less established markers such as discourse markers and modal verbs also played a role. The use of the pronoun “on” was identified as a significant subjectivity indicator, despite its common use to conceal the author's presence in journalistic writing. These findings offer insights into subjectivity in journalistic discourse and have significant implications for the field of journalism and for the improvement of automated models for subjectivity detection.

Keywords: subjectivity, journalism, annotation, discourse analysis, press discourse

1. Introduction

Objectivity has long been a fundamental guideline in the landscape of journalism, yet its realization remains a perpetual challenge at the centre of heated debates. Many journalists strive to present news stories in an impartial and neutral manner, aiming to deliver information free from personal bias (Wallace, 2020). However, the very nature of the journalistic process and the intricacies of human perception turn objectivity (or the complete absence of subjectivity) into an unattainable ideal (Steensen, 2017).

In this digital era where the lines between fact and opinion are fading, with more and more users gathering information from the biased and polarizing recommendation algorithms of social media platforms (Levy, 2021), understanding to what extent a press article is influenced by its author's personal opinions is an important matter. The presence of subjectivity in press discourse not only impacts the credibility and trustworthiness of news sources but also has far-reaching implications for media literacy, shaping how audiences interpret and engage with the information they encounter (Ku et al., 2019).

To shed light on the linguistic mechanisms of subjectivity in press discourse, we conducted an annotation experiment involving 36 participants tasked with evaluating the subjectivity of 150 Belgian French-written journalistic texts and identifying linguistic indicators of subjectivity in the texts. We define indicators of subjectivity as "words, groups of words or punctuation marks that signal the presence of the author's opinion in the text." This study aims to identify patterns and tendencies in annotators' perceptions, unveiling the textual elements triggering their interpretations of subjectivity. We also introduce "textual heat maps", a convenient method to visualise and analyse, at a very granular level, the annotated indicators of subjectivity. Through this exploration, we aim to contribute valuable insights to the field of subjectivity analysis in press discourse.

In section 2, we present the state of the art of the question of subjectivity in the fields of linguistics, natural language processing, and journalism studies. We also give an overview of previous studies about indicators of subjectivity in press discourse. In section 3, we present our corpus and detail the design of the experiment and the profile of the participants. In section 4, we examine the results of the study, first looking at statistical findings, then analysing qualitatively a sample of selected annotated texts. Section 5 is dedicated to the limitations of the experiment, followed by our conclusions in section 6.

2. State of the art

2.1. Objectivity and subjectivity in journalism

In journalism studies, the issue of objectivity in press discourse has always ignited fervent debates. In American journalism, particularly since the beginning of the 20th century, the objectivity norm has held a pivotal position as one of the guiding principles of journalistic writing (Schudson, 2001). It is widely regarded by most journalists as a "structural ideal" that they must try to pursue, although many acknowledge its paradoxical unattainability (Steensen, 2017). The inherent

subjectivity of reporting is primarily due to the inevitable processes of decision-making that guide all journalistic steps: story selection, choice of format, prioritization of some articles over others, and more (Tong, Zuo, 2021). Subjective judgments also come into play during the writing phase, for example when framing topics, ordering citations, or simply choosing adequate words. The depiction of facts is invariably influenced by the author's personal interpretation of these facts, shaped by their unique perspective and life experiences, which makes complete objectivity impossible in news reporting (Muñoz-Torres, 2012).

To mitigate the inherent subjectivity of the news-making process, journalists use several techniques, in accordance with what Tuchman (1972) refers to as the "strategic ritual of objectivity". This is realised through a range of neutralizing mechanisms designed to mask the journalist's personal opinions on the content of the text (Koren, 2004). Such guidelines, disseminated through journalism instructional materials or within newsrooms, include practices such as citing information sources, using impersonal expressions, and opting for a neutral lexicon (Charaudeau, 2006).

2.2. Linguistic subjectivity

Following Benveniste's Theory of Enunciation (1966), which rejects the classical conception of language as a simple tool of communication, declaring subjectivity as inherent in language, several linguists turned their attention for the first time to the question of the place occupied in language by the subject. Kerbrat-Orecchioni (1980) extended Benveniste's problem of enunciation by seeking to build the first inventory of enunciative traces in discourse, the "linguistic processes by which the speaker makes his mark on the utterance, inscribes himself in the message (implicitly or explicitly) and situates himself in relation to it" (Kirakossian, 2015).

In the field of natural language processing (NLP), opinion mining is a complex task consisting in the automatic extraction of the points of view that emerge from a text (Birjali et al., 2021). It has become a central task in several real-world domains, such as customer review surveys, reputation management, or press discourse analysis. A preliminary task in opinion mining consists in subjective text classification, which is very impactful in the opinion mining pipeline (Mihalcea et al., 2007), and is considered as one of its most complex steps (Ravi, Ravi, 2015). Subjective text classification can be carried out at sentence, paragraph, or document level (Marchand, 2012), and often involves the identification of subjective units in the text (Eensoo-Ramdani et al., 2011). In this paper, we refer to these subjective units as "indicators of subjectivity", which can be found at the morphosyntactic, the lexicosemantic or the textual levels of discourse.

2.3. Indicators of subjectivity in journalistic discourse

Several studies have attempted to assess the effectiveness of various types of indicators of subjectivity in press discourse, mostly in English, but also in other languages such as French.

Regarding morphosyntactic elements, the frequency of deictics, especially of first and second-person personal pronouns (singular and plural), is considered as a relevant indicator of the author's presence (De Cock, 2016). In French, the use of the indefinite pronoun *on* (*we*, *one*) is also sometimes considered as a good indicator of subjectivity,

since it places both the author and the reader in the context of enunciation (Todirascu, 2019). Analysing multiple indicators across different Belgian French media, Ho-Dac and Küppers (2011) found that modal adverbs are particularly frequent in subjective articles. Todirascu (2019), seeking to automatically classify *news* (mostly objective) and *opinion* (mostly subjective) articles based on linguistic descriptors, observed that the presence of relative pronouns and adjectives is more prevalent in *opinion* articles. In English discourse, Regmi and Bal (2015) also found an unequal distribution of parts-of-speech between journalistic genres: *opinion* articles had proportionally more adjectives, while *news* articles contained more verbs. The presence of negation words also seems to be one of the best predictors for *opinion* classification in English (Krüger et al., 2017).

At the lexicosemantic level, the use of sentiment lexicons (resources seeking to gather all the subjective words of a given language) is an effective way of analysing subjectivity in press discourse (Vis et al., 2012). Such lexicons include the English NRC EmoLex lexicon (Mohammad, Turney, 2013), and the French emotional valence lexicon of the Lexique3 sentiment database (Gobin et al., 2017). Regarding indicators of subjectivity at the textual level, the average length of words in the article can contribute to the performance of subjective text classification models (Alhindi et al., 2020). Various punctuation marks have also been identified as relevant markers of subjectivity. French *opinion* articles contain more question marks, exclamation marks and semicolons (Todirascu, 2019). Chaput (2019) has also shown the importance of four specific punctuation marks for analysing subjectivity in journalistic texts in French: ellipses, parentheses, dashes, and quotation marks. Based on a corpus of American articles, Krüger et al. (2017) found that the number of digits used (dates, amounts, percentages) was proportionally greater in *news* articles, and that lexical complexity was higher in *opinion* articles. Escouflaire (2022) evaluated the effectiveness of 30 state-of-the-art indicators of subjectivity for French *news* vs. *opinion* genre classification, identifying the number of negation words, relative pronouns, and adjectives as the most reliable indicators of subjectivity.

2.4. Annotation of subjective texts

Although the journalistic genre to which the article belongs according to the journalist or the media may be used as a ground-truth for research on *opinion* vs. *news* articles classification (Krüger et al., 2017), manual annotation is generally considered a particularly reliable way of obtaining a compliant reference dataset for training subjective text classification models (Ravi, Ravi, 2015). Subjectivity in text can be annotated in various ways: using a binary label (objective vs. subjective), a graded axis (from completely objective to very subjective), or by asking annotators to identify subjective sentences, words, or groups of words in the text. Janyce Wiebe and her team carried out several manual annotation experiments of subjectivity in articles from the American press (Wiebe et al., 2003; Wiebe, Riloff, 2005). The MPQA Opinion Corpus (Multi-Perspective Question Answering) contains more than 500 articles in which the "potentially subjective elements" have been annotated manually by three individuals (Wilson et al., 2017). Aker et al. (2019) had the subjectivity of 250 press articles assessed by annotators on a scale between 1 (totally objective) and 100 (totally subjective), with each article being read and rated by at least 10 people.

In French, Lark et al. (2015) grouped 10,000 tweets manually annotated for subjectivity at token level.

To minimise the influence of external factors on the annotations, one must carefully create precise annotation instructions and compile them in an annotation guide that can be consulted by the annotators at any time during the experiment (Ide, Pustejovsky, 2017). If needed, these instructions can be made deliberately vague so as not to overly influence the decisions of the participants (Wiebe et al., 2005; Ho-Dac et al., 2012).

When analysing the annotations produced during the experiment, inter-annotator agreement (IAA), also known as "inter-rater agreement" or "inter-coder reliability", must be computed to statistically evaluate the homogeneity of the choices made by the different annotators on all the documents (Paun et al., 2022). The higher the IAA, the more we can consider that the observations are shared between the annotators (Mathet, Widlöcher, 2019). Spooren and Degand (2010) insist, however, that IAA cannot be considered a measure of outcome validity in tasks involving linguistic analyses, given the inevitable involvement of many subjective factors in annotators' interpretive processes (Stede, Peldszus, 2012). Wiebe et al. (2005) obtained low to medium agreement scores when asking annotators to identify subjective elements. Balahur et al. (2013) found that the subjective evaluation of press discourse is a difficult task for human readers: the personal convictions and backgrounds of readers may influence their interpretation of the texts, and the lack of context or knowledge about a situation might also lead to differing evaluations of a text's subjectivity.

In this context, our work aims to identify known and new patterns in annotators' perceptions of subjectivity in French press articles. We also introduce a convenient method to visualise and analyse indicators of subjectivity annotated at token level, which we call "textual heat maps". The results of our analyses may be used to better understand how subjectivity is interpreted by readers in French press discourse, and to train and evaluate automated models for *opinion* vs. *news* articles classification.

3. Methodology

3.1. Data

The dataset used in this experiment consists of 150 excerpts of articles from the RTBF Corpus[1] (Escouflaire et al., 2023), an open-source collection of web articles published between 2008 and 2021 by the RTBF (*Radio-Télévision Belge Francophone*), the public service media of the French-speaking region of Belgium. We extracted a random sample of 5.000 articles from the "opinions" feed of the RTBF website (such as editorials and reviews), and 5.000 articles tagged as "news" by the RTBF (news, press agency dispatches), which were taken from the "Belgium", "World" and "Society" feeds (Bogaert et al., 2023). Based solely on the categorization made by the media, *opinion* articles are assumed to be more subjective in essence than *news* articles, which are expected to reflect the author's opinions less.

[1] https://dataverse.uclouvain.be/dataset.xhtml?persistentId=doi:10.14428/DVN/PEVSSI

We used the logistic regression classifier introduced by Escouflaire (2022), based on 18 state-of-the-art linguistic features of subjectivity in discourse, such as the presence of expressive punctuation signs in the text, the ratio of sentiment words or the use of relative pronouns. We trained this classifier on a random sample of 4.000 *opinion* and 4.000 *news* articles from the corpus, then used it to attribute a subjectivity rating (between 0 and 1) to the remaining 1.000 *opinion* and 1.000 *news* articles[2]. Those 2.000 articles were then ranked from lowest to highest subjectivity score (i.e., from least to most subjective). We use the term Automated Subjectivity Rating (henceforth ASR) to refer to the subjectivity score obtained through the logistic regression classifier.

Then, we randomly selected three subsets of articles from each genre (*opinion* and *news*): a subset of articles from the 10% with the lowest ASR, another subset of articles from those with an ASR between 45% and 55%, and articles from the top-10% ASR. This enabled us to produce a dataset with a higher diversity in terms of subjectivity, relying not only on the binary genre associated to each article by the media. The number of articles per subset for each genre is presented in Table 1. With 75 *opinion* articles and 75 *news* articles across three levels of ASR, we obtained a subset of 150 articles displaying different levels of linguistic subjectivity.

Finally, following the assumption that most of the subjective content of a press article is found at the start and at the end of the text (Chenlo, Losada, 2014), we reduced each article to either its first or last paragraph only. This choice was also made to avoid too much lack of context in the excerpts, which could hinder the annotators' interpretation. We randomly selected (with a balanced distribution) whether the start or the end of the article was used as the excerpt.

Table 1. Distribution of excerpts in the final dataset following Automated Subjectivity Rating (ASR) and genre (tagged by the media source).

	10% highest ASR	**45-55% ASR**	**10% lowest ASR**	**Total**
***Opinion* articles**	20	20	35	75
***News* articles**	35	20	20	75
Total	55	40	55	150

We also decided to provide the annotators with short texts, containing between two and five sentences, to increase the readers' focus on the linguistic and stylistic aspects of the excerpts and to reduce the influence of textual context on their interpretation. This sentence segmentation was manually operated: we decided at which sentence of the article the excerpt should start or end based on textual (e.g. end of paragraph) or thematic boundaries (e.g. change of topic).

[2] The classifier is presented in more detail in Appendix 1.

After this step, the 150 excerpts in the dataset contain between 50 and 150 words, with the mean token count across the sample being 92 tokens. Excerpts from the two genres (*opinion* and *news*) do not differ significantly in terms of token count.

3.2. Participants

A group of 36 native speakers of French participated actively in the experiment. All participants were first-year master's students at the Louvain School of Journalism (EJL), at UCLouvain, Belgium. At the time of the experiment, participants were between 20 and 26 years old, with the mean age being 23. Regarding gender distribution, 16 participants (44%) identified as female and 20 as male (56%). This selective sample of participants allows us to control the influence of age and educational background on subjectivity perception within the annotator pool, though it limits the generalizability of the experiment to a specific socio-cultural segment of the population.

Out of all 36 participants, 32 (88.9%) said that they read press articles at least once a week, and 19 participants (52.7%) declared reading the press every day. More specifically, 14 participants (38.8%) indicated that they read opinion pieces (e.g. editorials, reviews) at least once a week. Almost all participants (34 out of 36) consumed information regularly through web articles (from websites such as *www.rtbf.be*), making this the most popular source of information consumption among the group. Other popular media for gathering information include television (77.8% of all participants), as well as radio, smartphone apps of media sources, Twitter, Facebook and Instagram (all above 50% use).

When asked to evaluate their personal experience and skills in journalistic writing on a scale from 1 to 7 (1 being "not skilled at all" and 7 being "professional skills"), most participants evaluated their skills between 3 and 6 (mean = 4.3). Only 3 participants rated themselves below 3. A third of all participants evaluated their skills as 5 out of 7.

3.3. Annotation guide

The annotation experiment conducted for this study took place over four weeks, from October 10th to November 4th, 2022. All participants were students following the course "Journalism and information literacy". They were given the choice between either participating in this experiment or writing an essay about journalistic objectivity in the digital era. Serious participation in the experiment and respect of deadlines were explicit criteria for the final evaluation of the course.

The annotation environment used for the experiment was Label Studio (Tkachenko et al., 2020), an open-source annotation software suited for various types of data, and which allows users to freely highlight character spans in textual items[3].

Only two definitions were provided in the annotation guide:

- "The more the content and structure of a text are influenced by the author's opinion, the more that text is subjective."

[3] A screenshot of the Label Studio annotation interface is presented in Appendix 2.

- “Indicators of subjectivity are words, groups of words or punctuation marks that signal the presence of the author’s opinion in the text.”

During the experiment, participants were asked to read and annotate 50 journalistic excerpts.

For each excerpt, 4 steps were required from the annotator, in the following order:

A) Carefully *read* the excerpt below.
B) According to you, how *subjective* is this excerpt?
    1. Totally objective
    2. Slightly subjective
    3. Quite subjective
    4. Very subjective
    5. Totally subjective
C) How *confident* are you with your response to the previous question?
    1. Not confident at all
    2. Slightly confident
    3. Quite confident
    4. Very confident
    5. Totally confident
D) *Highlight* the “indicators of subjectivity” in the excerpt.

Questions B and C are Likert-scale questions: the annotators were asked to rate, from 1 to 5, the overall subjectivity of the excerpt, and the confidence with which they evaluated this subjectivity.

A preliminary two-hour session was conducted with all participants before the start of the experiment. During this session, an annotation test was first conducted on paper: all annotators received the same three excerpts to read and annotate following the four steps described above. After this test, they were able to ask questions about the experiment. A significant part of the session was then used to read through the annotation guide with the participants and to clarify some instructions with them. Participants were then shown how to install and use Label Studio on their personal computers. At the end of the session, a second annotation test was conducted in real conditions, in the Label Studio environment, using a different sample of three excerpts. Again, participants were then able to ask any methodological questions that this final test raised.

During this session, participants did not receive a detailed overview of research about subjectivity in discourse or about journalistic objectivity. They were simply told that the aim of the experiment was to identify the textual elements in which subjectivity can be found in journalistic discourse, from the reader's point of view. They were told that they were annotating excerpts from the start or the end of articles but were not told whether each specific excerpt was found at the start or the end of its corresponding article. It was explicitly said that they would not receive precise criteria for identifying subjectivity, as we did not know what these criteria were yet. We also added that we wanted them to use their experience as readers and their human intuition.

Some more practical recommendations were made to the participants before the start of the experiment. In case they met any methodological hesitations, the following recommendations were available at the end of the annotation guide:

- Read and annotate excerpts on your own and without consulting anyone.
- When annotating, be as consistent with yourself as possible, from one excerpt to the next, from one day to the next.
- Interpret words and information in the context of the sentence and excerpt in which they are written. For example, the adjective *cold* does not carry the same meaning in sentence X as in sentence Y.

  (X) This is a *cold* drink.

  (Y) This is a *cold* conversation.

- Try to set aside your personal opinions. Remain as impartial as possible when reading and evaluating the excerpts.
- Focus on subjectivity at text level, not at editorial level. Editorial subjectivity concerns the choices made by the media or the journalist before writing the article: e.g. choice of subject, choice of illustration, choice of participants to quote.
- The subjectivity you need to identify is that of the author of the excerpt, not that of other sources/interactors mentioned in the excerpt.
- When highlighting indicators of subjectivity, do not highlight whole sentences that you believe are subjective, but rather highlight the specific words or groups of words that signal the subjectivity.
- Avoid including unwanted characters when highlighting inside text spans (e.g. whitespaces, punctuation marks that you do not deem relevant, etc.).
- There is no correct answer. Trust your reader instinct.

3.4. Annotation design

Each participant was assigned 50 excerpts to read and annotate (12 to 13 excerpts per week, for 4 weeks). The 150 excerpts in the dataset were attributed to participants using a distribution algorithm designed to follow two main rules:

- Each excerpt was annotated by at least 10 different participants, of whom at least 3 were female and at least 3 were male.
- Each participant received 25 excerpts marked as *opinion* and 25 excerpts marked as *news* (this information remained hidden from the participants). The distribution of these genres was balanced every week. Participants were not aware of this balanced distribution.

In total, 145 excerpts were read and annotated by 10 to 14 participants. Because 3 students did not complete the experiment to the end, 5 excerpts out of the 150 did not follow these rules as they were annotated by only 8 to 9 participants. 36 out of the 39 initial participants completed the experiment.

In the end, 1762 full annotations were produced over the course of the experiment. All annotations were anonymised.

4. Results and discussion

4.1. Subjectivity and confidence ratings

The mean subjectivity rating (HSR) attributed by the 36 participants in question B ("According to you, how subjective is this excerpt?") on a scale from 1 to 5 to each of the 150 excerpts was 2.46 ($\sigma = 0.96$). This rating will be referred to as Human Subjectivity Rating (henceforth "HSR").

Then, we investigated the influence of different variables on the mean HSR. These variables are the genre of the excerpt (*news* or *opinion*), the Automated Subjectivity Rating (ASR) that was attributed to the excerpt by the classification model used for selecting the samples (divided into *low, medium,* or *high* subjectivity), and the mean Human Confidence Score ("HCR") given by the annotators to their own subjectivity ratings.

Table 3. Results of Welch's ANOVA and Pearson tests on the answers to questions 2 and 3 based on genre of excerpt and Automated Subjectivity Rating (ASR).

	Genre				ASR			
	W's ANOVA		Pearson's R		W's ANOVA		Pearson's R	
	W	*p*	R	*p*	W	*p*	R	*p*
Mean HSR	121.63	< 0.001	0.671	< 0.001	22.18	< 0.001	0.481	< 0.001
Mean HCR	1.641	0.20	- 0.104	0.20	1.201	0.30	-0.105	0.20

Using Levene's test, we found that our subjectivity data violated the assumption of homogeneity of variances for both genre and HSR. We therefore performed Welch's ANOVA tests for analysis of variance and Pearson's correlation tests (see Table 3), and found that both the genre of the excerpt and the ASR were significantly correlated ($p < 0.001$) with the mean subjectivity rankings given by the annotators to each excerpt. In addition, according to both Welch's ANOVA and Pearson's R tests, the correlation between mean subjectivity and genre (represented in Figure 1) was stronger than the correlation between mean HSR and ASR (represented in Figure 2). *Opinion* excerpts were annotated as more subjective than *news* excerpts, and excerpts with a lower ASR were considered less subjective than excerpts with a higher ASR. These results suggest that human readers are in average able to correctly identify the genre of the article, and that their interpretations of excerpt subjectivity are in par with the ASR attributed by the classification model which relies on state-of-the-art linguistic features of subjectivity. A thorough analysis of some outlier excerpts in the boxplots will be carried out in Section 4.3.2.

Then, we measured the correlation between these two categorical variables and the mean HCR of the annotators attributed to each excerpt by the annotators in question C ("How confident are you with your response to the previous question?"). Overall, the mean HCR of the annotators was 3.76 ($\sigma = 0.35$). As shown in Table 3, we found that neither the genre of the excerpt nor its ASR was statistically correlated with the mean HCR.

Figures 1 (left) and 2 (right). Boxplots of the mean Human Subjectivity Rating (HSR) by genre (1) and by Automated Subjectivity Rating (ASR) (2).

Insérer ici les Figures escouflaire_figure1.png et escouflaire_figure2.png

The Pearson test revealed that the mean HSR was not significantly correlated with the mean Human Confidence Rating (R = 0.102, $p = 0.215$). However, when looking at Figure 3, a scatterplot representing the distributions of both values for all excerpts, it appears that excerpts which were given either a very low (< 1.5) or a very high mean HSR (> 4) were also given a higher mean HCR in average than the other excerpts. This suggests that excerpts which appeared as either very subjective or very objective to the annotators were also those for which the annotators were the most confident, and that they were more hesitant with excerpts that appeared more or less subjective.

Because our annotation design allowed us to include gender as a balanced variable in the analysis, we also statistically tested the hypothesis that the gender of the participants might impact their judgments of excerpt subjectivity. We computed correlation tests using annotation results and gender data of the participants. We found no statistical correlation between the gender of the annotators and their HSR, according to a t-test ($T = 0.518$, $p = 0.6$). However, gender was negatively and significantly correlated with the HCR ($T = -3.33$, $p < 0.001$), indicating that female annotators were more likely to show less confidence in their annotations than male annotators.

To evaluate the consistency of subjectivity ratings given by the different annotators for an excerpt, we measured the variance of HSR in each excerpt, i.e. how much the ratings of the excerpt deviated from the mean HSR of the excerpt (if all annotators gave the same HSR to the excerpt, variance = 0). The mean variance in the dataset was 0.66 ($\sigma = 0.41$; max = 1.84). In the next section, we look deeper into consistency over the entire dataset, by evaluating inter-annotator agreement.

Figure 3. Scatterplot of mean Human Subjectivity Rating (HSR) by mean Human Confidence Rating (HCR) of all 150 excerpts, separated based on their genre.

Insérer ici la Figure escouflaire_figure3.png

4.2. Agreement

Inter-annotator agreement (IAA) was measured on the answers for each excerpt to question B (Human Subjectivity Rating). We expected agreement to be quite low, as evaluating the subjectivity of an excerpt is in itself a very subjective question (Balahur et al., 2013). When interpreting a journalistic excerpt, however factual, readers are always influenced by their personal beliefs, experiences, and views of the world (Muñoz-Torres, 2012). However, the main reason why we computed IAA was to

scrutinize the consistency of the annotations rather than to assess the reliability of our results.

We measured IAA across all 1762 answers made to question B. Each annotator ranked the subjectivity level of the excerpt on a Likert scale from 1 ("totally objective") to 5 ("totally subjective"). For example, excerpt [1] is the text with the highest variance in terms of HSR (1.84), to which all five different scales of HSR were assigned by the ten annotators who read it. This example is the illustration that the subjectivity of a single excerpt can be perceived very differently by different individuals. The words are highlighted following the "textual heat map" method presented in Section 4.3.2.

> [1] A Fukushima l'homme passe du contrôle au combat. On ne dirige plus la machine on l'affronte. La bataille, incertaine, fait des victimes. Comme dans une tragédie antique. Celle-ci commence par le mythe de Prométhée et la domestication du feu nucléaire. Après la colère des dieux, nous basculons dans la chute d'Icare. Comme dans un tableau de Brueghel. On prétendait dompter l'énergie, mais la mer nous engloutit.
>
> *At Fukushima, man has moved from control to combat. We no longer control the machine, we confront it. The battle, uncertain, claims victims. As in an ancient tragedy. It begins with the myth of Prometheus and the domestication of nuclear fire. After the wrath of the gods, we tumble into the fall of Icarus. As in a painting by Brueghel. We claimed to tame energy, but the sea engulfs us.*
>
> Example 1 (piece of excerpt); mean HSR = 3.4; mean HCR = 3.5; *opinion* genre; high ASR.

Measuring IAA with Likert scale data can be achieved in different ways: either by treating each level on the scale as a separate class that the annotator can select, or by aggregating some of the levels together to reduce the granularity of the results (Vogel et al., 2020). We computed IAA on the HSR results with both setups. First, using all 5 Likert levels as 5 separate classes; second, aggregating levels 1 and 2 together as one class, "low subjectivity", level 3 as "medium subjectivity", and levels 4 and 5 together under a "high subjectivity" class, keeping only 3 distinct classes. In addition, we applied filters consisting of "confidence thresholds" (Tanguy et al., 2018), based on the results of the annotators' answers to question C (HCR). The scatterplot in Figure 3 shows that the distribution in mean HCR is quite diffuse (between 3 and 4.6) across the 150 articles. We selected five different HCR thresholds: 3, 3.5, 3.75, 4 and 4.25. These thresholds allow us to include in the computation of the agreement for HSR only the excerpts for which the mean HCR is higher than a certain value (between 1 and 5).

We used Krippendorff's alpha (α) to measure IAA, because of its robustness in experiments in which not all documents were annotated by all participants and in which the number of annotations per document is variable (Hayes, Krippendorff, 2007), and because it has been adopted as a standard in discourse analysis.

Table 2. Results of the inter-annotator agreement study on the answers to question B (Human Subjectivity Rating) using Krippendorff's alpha (α), filtered by Human Confidence Rating (HCR).

Mean HCR threshold (*c*)	*> 3*	*> 3.5*	*> 3.75*	*> 4*	*> 4.25*
Excerpts (*N*)	*150*	*119*	*74*	*42*	*20*
IAA (*α*) with 5 classes	0.218	0.233	0.280	0.306	0.363
IAA (*α*) with 3 classes	0.350	0.393	0.485	0.598	0.755

When using all five ratings of HSR as five classes, the results presented in Table 2 show that α is between 0.2 and 0.4, i.e. "quite low" (Krippendorff, 2018), even when applying more restrictive confidence thresholds. If we filter only the 20 excerpts with the highest mean HCR (above 4.25), α remains quite low. This low agreement may suggest that the question asking participants to rate the subjectivity of an excerpt is difficult, or rather that it is a highly subjective question and that the results are strongly influenced by the personal interpretations of the readers. This confirms the observations made by Wiebe et al. (2005), although our work focuses on excerpts and not on complete texts. However, when the categories are aggregated into three distinct classes instead of five, α rises considerably, reaching 0.755 ("high agreement") when isolating the 20 excerpts with the highest mean HCR among annotators. These higher agreement results show that, even though there is an overall large degree of variability in the HSR assigned by the different annotators across the 150 excerpts, some consistency is found in a small portion of the dataset. This led us to believe that taking a closer look at the excerpts with the lowest and highest mean HCR, and at those with the least and the most consistent HSR in our results, could provide interesting insights. We also examined excerpts presenting the most variance in HSR among annotators, to find where disagreement may come from.

4.3. Subjective text spans analysis

4.3.1. Statistical analysis

In this section, we analyse the answers provided by annotators to question D ("Highlight the indicators of subjectivity in the excerpt.") The results of this question take the form of text spans containing at least one token.

In total, 4722 text spans were highlighted by the annotators during the whole annotation process, for a mean of 2.68 text spans per annotation (over 1762 annotated excerpts in total). The average text span was 12 characters long, with the lowest being only 1 character (punctuation signs) and the longest being 140 characters long (a whole sentence was highlighted by one annotator). When examining the 150 excerpts in terms of text span annotations, the mean number of text spans per article was 31.5 spans ($\sigma = 24.26$). Of the 150 excerpts, four excerpts were not highlighted at all by any annotator. These are all excerpts belonging to the *news* genre and that were given a low ASR by the automated model, and excerpts for which the mean HSR was 1 and mean HCR was above 4. Example [2] is one of these excerpts which all annotators rated 1 in terms of subjectivity (variance = 0), with a high mean HCR (4.36), and did not highlight any single text span as subjective.

[2] Les ministres des Finances de l'UE ont donné leur feu vert mardi à une augmentation de la capacité et de la durée du "plan Juncker", destiné à encourager croissance et emploi, portant les investissements à au moins 500 milliards d'euros à l'horizon 2020.

EU finance ministers gave the go-ahead on Tuesday for an increase in the capacity and duration of the "Juncker Plan", designed to encourage growth and employment, raising investment to at least €500 billion by 2020.

Example 2 (piece of excerpt); mean HSR = 1; mean HCR = 4.36; *news* genre; low ASR.

The Pearson test showed that there was a strong significant correlation between the mean number of spans identified in an excerpt and the mean HSR attributed to it ($R = 0.924$, $p < 0.001$). This correlation is visualised on Figure 4: the higher the mean HSR of an excerpt, the more text spans are highlighted by the annotators in this excerpt. The scatterplot also illustrates that significantly more text spans were highlighted in *opinion* articles than in *news* articles, as confirmed by a Welch's ANOVA test ($W = 89.99$; $p < 0.001$). A similar test showed that the number of highlighted spans was significantly correlated with the ASR associated to the excerpt ($W = 16.51$; $p < 0.001$): excerpts that were considered more subjective by the automated model were highlighted more often by the annotators.

In addition, we found that 2127 different types of text spans were highlighted. Of these, 1345 were highlighted only once. We take a closer look at the most frequently highlighted text spans in section 4.4.

Figure 4. Scatterplot of mean Human Subjectivity Rating (HSR) by number of highlighted subjective text spans, separated based on their genre.

Insérer ici la Figure escouflaire_figure4.png

4.3.2. Heat map visualization

We developed "textual heat maps" in order to visualise in HTML format which text spans were the most frequently identified by the annotators in a given excerpt. The colours of the highlighting correspond to the number of times each text span was highlighted by annotators, from light blue (highlighted once) to dark red (highlighted by six or more annotators), as shown in Figure 4. These heat maps can be used to easily understand which characters and words were the most widely considered by annotators as indicators of subjectivity.

A total of twelve textual heat maps are presented in this paper. The excerpts that will be qualitatively analysed using these heat maps were selected because they corresponded to outliers identified in the quantitative insights in the previous sections, or because they include clear examples of indicators discussed in Section 4.4. The textual heat maps are presented alongside the excerpt's English translation, its mean

HSR and HCR, its genre (*news* or *opinion*), and the ASR level it was assigned (low, medium or high). All English translations of the excerpts presented in this paper are ours.

Figure 4. Colours assigned to the number of annotators who highlighted a given text span.

Insérer ici la Figure escouflaire_figure5.png

4.3.3. Qualitative analysis of sample texts

In this section, we selected five excerpts (represented as textual heat maps), according to different criteria, to further examine the results presented earlier. We will come back to some of these excerpts in section 4.4.

> [3] Le gouvernement de Charles Michel est divisé avant un budget, rien d'étonnant en fait... Ce qui se passe est d'une banalité affligeante. On retrouve dans la séquence qui se déroule en ce moment une bonne partie des maux qui frappent la politique belge depuis au moins 15 ans, depuis le dernier gouvernement Dehaene. On retrouve des négociations marathons où telle taxe, telle coupe dans les soins de santé est décidée au bout de la nuit parce qu'il faut bien avoir quelque chose à livrer aux médias et au parlement. On retrouve le même empressement, le même amateurisme qui conduit des mesures importantes à s'écraser en plein vol faute de préparation.
>
> *Charles Michel's government is divided before a budget, which is hardly surprising... What is happening is distressingly banal. The sequence of events unfolding at the moment reflects many of the problems that have plagued Belgian politics for at least 15 years, since the last Dehaene government. There are marathon negotiations in which a particular tax or cut in healthcare is decided at the end of the night because there has to be something to deliver to the media and parliament. There is the same haste, the same amateurism that leads important measures to crash in mid-air due to lack of preparation.*
>
> Example 3; *opinion* genre; mean HSR = 4.6; mean HCR = 4.5; high ASR.

Example [3] is the excerpt with the highest mean HSR (4.6) among all 150 excerpts: all annotators rated it as 4 or 5 (on a scale from 1 to 5). It is also the excerpt for which participants showed the highest mean HCR (4.5) in the entire corpus, even higher than example [2] (which is the sentence with the lowest mean HSR), suggesting that its subjectivity is particularly obvious to human readers. Among the words that were highlighted the most consistently by annotators in this excerpt (in bright red), we find sentiment-related adjectives (*étonnant*, *affligeante*) and nouns carrying a strong axiological value (*banalité*, *amateurisme*), as well as the argumentative locution *il faut bien*.

One of the articles with the lowest mean Human Confidence Rating (HCR) is example [4], which was not considered very subjective by annotators (mean HSR = 1.75), but for which they reported an average HCR of only 3.15 out of 5. Of the eight participants who annotated this excerpt, five rated its subjectivity as 1, one participant rated it as 2, one rated it as 3 and one gave it 4. None of the annotators who assigned it a rating of 1 highlighted a single word in the excerpt, and only the word *pourtant* was highlighted by more than one annotator. A possible explanation of the annotators' unsureness in evaluating the subjectivity of this excerpt and of the disparity in highlighted text spans may be related to its topic, namely the Israeli-Palestinian conflict, which is known to be a deeply contentious and multifaceted issue. Example [5] also suggests that the topic of the articles influences the perception of its subjectivity by human readers. This assumption should be tested in greater detail in further research.

> [4] Les Palestiniens de Cisjordanie vivent toujours sous occupation militaire et sont privés de tout droit politique. Ceux de Jérusalem-Est, pourtant annexée par Israël depuis 1980, n'ont pas obtenu la citoyenneté israélienne et ne peuvent participer qu'aux élections municipales.
>
> *Palestinians in the West Bank still live under military occupation and are deprived of all political rights. Those living in East Jerusalem, which has been annexed by Israel since 1980, have not been granted Israeli citizenship and can only take part in municipal elections.*
>
> Example 4 (piece of excerpt); mean HSR = 1.75; mean HCR = 3.12; *opinions* genre; low ASR.

We then searched for articles with the most variance (or lowest consistency) among all Human Subjectivity Ratings (HSR). One of the articles with the least consistent HSR (variance = 1.24) is example [5]: two readers considered it "totally objective", one found it "slightly subjective", three "quite subjective" and the majority (six annotators) rated it as "very subjective". This disparity may again be related to the topic of the excerpt, or to the words used: the most frequently highlighted subjective spans in the excerpt are not adjectives and nouns, as in example [3], but adverbials (*sérieusement*, *très peu*) and verbs (*peut*, *pointent*). Still, it is interesting to note that some words were highlighted by many annotators (*à juste titre*), while others were only highlighted by one (*concept*). This observation suggests that some indicators of subjectivity may be more subtle, or not as clear as other types of markers, such as axiological adjectives.

> [5] Si la plupart des analyses pointent à juste titre la stigmatisation volontaire de cette "communauté musulmane" dont on peut sérieusement se demander si elle existe autrement que comme concept à usage politique, elles ne pointent que très peu le processus par lequel le ministre fonde son argumentation.
>
> *While most of the analyses rightly point to the deliberate stigmatisation of this 'Muslim community', which we can seriously question exists as*

anything other than a concept for political use, they do very little to highlight the process by which the Minister bases his argument.

Example 5 (piece of excerpt); mean HSR = 3.08; mean HCR = 3.58; *opinion* genre; low ASR.

Example [6] is a very interesting excerpt in the dataset. First, it is one of the few excerpts to which all scales of subjectivity were assigned by annotators (from "totally objective" to "totally subjective"), displaying a very high variance in HSR (1.27). Second, it is the excerpt with the least consistent mean HSR between female and male annotators: while female readers gave it a mean HSR of 2.6, males rated it 3.33 in average. Although we found no statistical correlation between gender and HSR in our results (see Section 4.1), the gap in this excerpt may be related to the topic of the article (i.e. women's rights). It suggests that the human perception of subjectivity is at least somewhat influenced by the personal subjectivity of the person reading the excerpt, by their sociological profile and experiences.

[6] En cette journée internationale de lutte, nous rappelons qu'il est essentiel de protéger et garantir les droits des femmes à disposer de leurs corps et de leurs vies. D'autant plus dans ce contexte de crise prolongée. C'est une question de santé publique, de justice sociale et d'égalité des chances : l'État doit créer, en toutes circonstances, des conditions favorables à l'accès à la santé reproductive et sexuelle et à la réalisation du droit de toutes et tous à l'autodétermination.

On this international day of struggle, we reiterate that it is essential to protect and guarantee women's rights to control their bodies and their lives. Even more so in this context of prolonged crisis. It is a question of public health, social justice, and equal opportunities: the State must create, in all circumstances, conditions favourable to access to reproductive and sexual health and to the realisation of the right of all to self-determination.

Example 6; mean HSR = 3; mean HCR = 3.64; *opinion* genre; high ASR.

Finally, example [7] is one of the outliers in the *news* boxplot in Figure 1: an article tagged as belonging to the *news* class by RTBF, but which was considered as very subjective by the annotators (mean HSR = 3.75). The ASR given by the classifier to this excerpt was also high. This discrepancy between the categorisation made by the media and the subjectivity ratings of both the human annotators and the automated model is very interesting, as it may lead us to question or at least nuance the relationship between journalistic genre and linguistic subjectivity. This perspective should be examined more closely in further research. Several punctuation marks were highlighted in the excerpt (!, ", ...), which are analysed in more detail in section 4.4.

[7] Le président américain Donald Trump a trouvé la parade ultime pour contrer toutes les "fake news" dont sont truffés les médias selon lui... du moins quand elles le concernent (et ce alors qu'un des consultants de Fox News accuse la chaîne d'avoir publié, à dessein, de fausses informations défavorables aux démocrates à la demande du président des Etats-Unis...)

Comme on n'est jamais si bien servi que par soi-même, c'est via sa page Facebook qu'il a mis en ligne un programme court, intitulé... “The Real News”, et présenté par sa belle-fille!

US President Donald Trump has found the ultimate way of countering all the "fake news" he believes the media are full of... at least when it concerns him (and this at a time when one of the Fox News consultants is accusing the channel of having deliberately published false information unfavourable to the Democrats at the request of the President of the United States...) And since, if you want a job done, you'd better do it yourself, he has posted a short program on his Facebook page, entitled... "The Real News", presented by his daughter-in-law!

Example 7; mean HSR = 3.75; mean HCR = 3.75; *news* article; high ASR.

4.4. Indicators of subjectivity

To find out which textual elements were considered most as indicators of subjectivity throughout the 150 texts by the annotators (question D), we counted the frequency of all the different text spans that were highlighted during the experiment.

The most frequently highlighted spans are presented in Table 4. We included in the table text spans that were highlighted at least 7 times, and that occurred more than once in the dataset. This allows us to avoid including textual elements that may have been present in the table only due to the small size of the dataset, mitigating bias that could be linked to the specificity of the 150 excerpts. For example, the word *terrifiant* (‘terrifying’) was highlighted 8 times but was only present once in the dataset, and was therefore not included in Table 4.

As presented in section 4.3.1, 4722 text spans were highlighted by the annotators. We first cleaned them by removing starting and ending whitespaces. We then counted together occurrences of the same text spans to establish frequency measures. We decided not to tokenise text spans containing more than one token (word or punctuation sign) when measuring text span frequencies, in order to shed light on multiword expressions that may be considered as indicators of subjectivity by the annotators and to avoid including stop words or non-meaningful tokens in the analysis. This decision may however hide some tokens that may have been highlighted both as part of multi-token text spans and on their own. Regarding text spans containing only one token, we grouped together under a single entry tokens belonging to the same lemma (*doit/doivent*, *spectaculaire(s)*), but only when more than one declension or conjugation of this lemma was highlighted by the annotators. For example, the feminine *petite* was the only form of the adjective *petit* (‘small’) that was highlighted in our experiment, therefore we only included the conjugated form in the table.

In this section, we analyse the text spans presented in Table 4 to find patterns in the textual elements highlighted as subjective elements by the annotators. We will first focus on text spans related to linguistic features previously identified by the literature on subjectivity in discourse, then on newly identified indicators of subjectivity.

Table 4. The most frequently highlighted text spans across the 150 texts, ranked by number of times they were highlighted (*H*), total occurrences in the dataset (*N*) and proportion of the number of times they were highlighted on their total occurrences (*H/N*).

Text span	*H*	*N*	*H/N*	Text span	*H*	*N*	*H/N*
on	98	57	1.72	*austère*	12	2	6
...	96	25	3.84	*émouvant(es)*	12	2	6
!	80	28	2.86	*bon(nes)*	12	8	1.5
nous	79	41	1.93	*semble(nt)*	11	4	2.75
"	74	215	0.34	*important*	11	4	2.75
?	38	40	0.95	*pourtant*	11	6	1.83
je/j'	34	12	2.83	*certain(es)*	11	8	1.37
doit/doivent	28	19	1.47	*finalement*	10	5	2
évidemment	27	4	6.75	*étonnant*	10	3	3.33
dramatique	23	3	7.66	*peut-être*	9	4	2.25
déjà	21	18	1.17	*chez nous*	9	5	1.8
que	21	134	0.16	*vraiment*	9	7	1.29
spectaculaire(s)	20	4	5	*il faut*	9	11	0.82
en fait	20	7	2.86	*alors*	9	17	0.53
très	19	14	1.36	*bien*	9	26	0.35
donc	19	21	0.90	*rapidement*	8	4	2
grand(es)	18	22	0.81	*encore*	8	16	0.5
mais	18	81	0.22	*comme*	8	50	0.16
hypocrisie	16	2	8	*pire*	7	2	3.5
bref	16	3	5.33	*tout de même*	7	2	3.5
sans doute	16	4	4	*petite*	7	2	3.5
beau/belle	16	5	3.2	*drastiques*	7	2	3.5
mon/ma/mes	15	6	2.5	*moi*	7	3	2.33
mauvais(es)	14	9	1.55	*surtout*	7	5	1.4
si	14	28	0.5	*rien*	7	11	0.64
même	14	30	0.47	*toujours*	7	15	0.47
nombreux(ses)	13	12	1.08	*plus*	7	91	0.08

4.4.1. Established indicators

The main observation that can be made when examining Table 4 is that many of the spans of text that were most highlighted by annotators correspond to indicators of subjectivity found in the state of the art and that were included in the feature-based classification model which was used to select articles (Escouflaire, 2022): the indefinite pronoun *on*, expressive punctuation (*..., ?, !*), first person pronouns and determiners (*nous*, *je, mon/ma/mes*), adjectives (*dramatique, grand(es), mauvais(es)*), and modal adverbs (*évidemment*, *sans doute*, *peut-être*). These text spans are also those that display the highest ratio between highlighted text spans and total occurrences in the dataset (highlighted-occurrences ratio; H/N ratio). These results suggest that the textual elements on which our annotators based their interpretation of subjectivity are very similar to indicators that have been identified in the state of the art, as those were the indicators used in the model. This assumption is also reinforced

by the significant correlation that was found in section 4.2. between the ASR given by the model and the mean HSR attributed by the annotators to each text.

We took a closer look at the main indicators of subjectivity that were highlighted by many annotators throughout the sample.

The overall most frequently highlighted token was the third-person singular pronoun *on*, which was highlighted 98 times, with a highlighted-occurrence (H/N) ratio of 1.72. In French, *on* is a very particular word. As a personal pronoun, it can be used as a more informal alternative to the first-person plural pronoun *nous* ("we", "us"), referring to the subject of the enunciation and possibly including the interlocutor ("you") and/or third-party agents ("they"). As an indefinite pronoun, *on* can be used by the subject to show that a source of information is anonymous or unknown, to avoid mentioning the said source of information, or to shift responsibility away from the subject (Rabatel, 2001).

In French journalistic metadiscourse on objective writing, *on* is often presented as a typical example of "enunciative erasure" (Koren, 2004): relying on the ambiguity of the pronoun *on* can be seen as a stylistic strategy used by journalists to distance themselves from a particular statement, or simply to hide their own subjectivity (Marnette, 2004; Aouda, 2019). In our corpus, *on* appears 41 times in *opinion* excerpts, and 16 times in *news* excerpts.

> [8] Si on se donne la peine d'écouter le texte, on comprend pourtant vite qui s'exprime ici : si des hommes (en particulier mineurs) peuvent aussi être victimes d'agressions sexuelles, ils ne vivent pas la culpabilisation dénoncée dans la chanson.
>
> *If **you** take the trouble to listen to the lyrics, **you**'ll soon realise who's talking: while men (especially minors) can also be victims of sexual assault, they don't experience the guilt-tripping denounced in the song.*
>
> Example 8 (piece of excerpt); mean HSR = 3.72; mean HCR = 3.27; *opinion* genre; medium ASR.

We examined several instances of *on* that were highlighted by at least 4 annotators in the text in which they appeared. We found that the personal and indefinite uses of *on* coexist and are both considered by the readers as indicators of subjectivity. As shown by example [8], the journalist may use *on* to create a sense of collective identity or to establish a connection with their audience, presenting information from a shared perspective. This use is highlighted as subjective by the annotators.

Another use can be found in example [9], where the anaphoric repetition of *on*, here used to refer to anonymous victims, was systematically highlighted as subjective by the readers. This subjective feeling could derive from the sense of identification with the victims that the multiple *on* may aim to convey in this text.

> [9] En Syrie, depuis 5 ans, on meurt sous les balles, on meurt sous les bombes. Mais en plus, aujourd'hui, si on survit au conflit, on meurt de faim. On meurt de faim parce que des dizaines de villes sont assiégées. La

plupart par l'armée, mais aussi par les rebelles ou par l'organisation terroriste Etat Islamique.

*In Syria, for the last 5 years, **people** have been dying from bullets, **people** have been dying under bombs. And today, if **people** survive the conflict, **they** die of hunger. **They** are starving to death because dozens of towns are under siege. Most of them by the army, but also by the rebels or by the terrorist organisation Islamic State.*

Example 9 (piece of excerpt); mean HSR = 2.91; mean HCR = 3.25; *opinion* genre; high ASR.

In example [5] (Section 4.3.3), the single *on* was highlighted multiple times. While it may be interpreted as a way for the journalist to distance themselves from their comment made about the existence of the "Muslim community", and thus to mask their own subjectivity, it is however the most consistent indicator of subjectivity identified in this text by its readers. Some texts in the corpus also use *on* as a purely indefinite pronoun: in example [10], *on* is simply used to refer to unspecified individuals. In such cases, annotators do not appear to identify *on* as an indicator of subjectivity as often as with its other uses.

[10] Quelques spécimens ont en effet élu domicile sur la côte belge. On les trouve parfois sur les bancs de sable de l'Escaut. Ils ne se sont, à ce jour, pas attaqués à l'homme. Mais la prudence doit rester de mise face à ces mammifères marins aux canines acérées.

*A few specimens have indeed taken up residence on the Belgian coast. **One** may sometimes find them on the sandbanks of the Scheldt. To date, they have not attacked humans. But caution is still called for when confronted with these sharp-toothed marine mammals.*

Example 10 (piece of excerpt); mean HSR = 1.71; mean HCR = 3.79; *news* article; medium ASR.

These observations suggest that while *on* is often used in journalistic discourse as a strategy of enunciative erasure, its presence does not seem to hide the journalist's subjectivity, as it rather appears mostly to human readers as a marker of the author's presence in the text. The highlighted-occurrence ratio of *on* (see Table 4) is quite low, which shows that some occurrences of *on* are not systematically considered as indicators of subjectivity. Our experiment also suggests that readers are capable of understanding and disambiguating the different uses of *on*.

Other elements high on the list in Table 4 include punctuation marks, in particular ellipses, quotation, question marks, and exclamation marks.

Ellipsis (...) is the second most frequently highlighted indicator of subjectivity in the corpus: it was identified 96 times by the annotators, with a higher highlighted-occurrences ratio (3.84) than *on*, showing a stronger consensus in subjectivity among occurrences and annotators. Ellipses can convey a sense of suspense or expectation (Dalhet, 2003), as in example [7] in which they are used to highlight the irony of the program's name. In journalistic discourse, ellipsis can also be used to signal the journalist's choice not to reveal all his thoughts, leaving it up to the reader to interpret

the implied message (Chaput, 2019), as in example [3]. Ellipses often reflect the author's subjectivity or the presence of their implicit opinion, without it being necessary to reveal its content.

Exclamation marks (!) are also identified as very good cues of subjectivity, due to the unambiguous effect of emphasis or surprise that they carry. Question marks (?) are typical of the *opinion* genre as well, in which interrogative sentences are used to involve the reader in the text and rhetorical questions may reinforce the point made by the author (Todirascu, 2019), as in example [11]. However, they show a lower highlighted-occurrences ratio (0.95) than other text spans, suggesting that they are not as consensual as other indicators of subjectivity.

> [11] Marine Le Pen n'a jamais été aussi haute dans les sondages, elle se qualifierait sans problème pour le deuxième tour de cette présidentielle. C'est une question souvent évoquée : le Front national a-t-il changé ? En France, la plupart des observateurs ne qualifient plus le FN de parti d'extrême droite, mais de parti populiste ou de droite nationaliste. Est-ce vraiment le cas ? Marc Sirlereau s'est rendu dans le sud de la France à la rencontre d'élus et de militants.
>
> *Marine Le Pen has never been so high in the polls, and she would have no problem qualifying for the second round of this presidential election. It's a question often asked: has the Front National changed? In France, most observers no longer describe the FN as a far-right party, but as a populist or nationalist right-wing party. Is this really the case? Marc Sirlereau travelled to the south of France to meet elected representatives and activists.*
>
> Example 11; mean HSR = 2.33; mean HCR = 3.33; *news* genre; high ASR.

Another punctuation mark that ranks high in the list is the quotation mark ("). Although its H/N ratio is even lower than that of the question mark, suggesting that many quotation marks in the dataset are not considered indicators of subjectivity by annotators, it was still highlighted 35 times. While it is characteristic of the *news* genre, in which it is mostly used to cite sources, signaling direct reported speech, quotation marks are also sometimes used by journalists to distance themselves from a particular word or expression (as in example [7] with *fake news*; this usage is sometimes called “scare quotes”) or to add emphasis to it (Gutzmann, Stei, 2011), as in example [12].

> [12] En Flandre, ces "assouplissements" suscitent énormément de débats et de discussions, que ce soit dans les médias, ou dans la sphère privée. Ce qui est étonnant, c’est que ces "erreurs de grammaire tolérées" existaient en fait déjà dans les anciennes versions de l’ANS, mais qu’elles semblent aujourd’hui susciter plus d’émoi qu’avant.
>
> *In Flanders, these "relaxations" are generating a great deal of debate and discussion, both in the media and in private. What's surprising is that these "tolerated grammatical errors" already existed in previous versions of the SLA, but now seem to be causing more of a stir than before.*

Example 12; mean HSR = 2.83; mean HCR = 3.42; *opinion* genre; low ASR.

4.4.2. New indicators

Some text spans displayed in Table 4 are not part of the indicators of subjectivity previously identified in the literature and were not used in the ASR of the model used in (Escouflaire, 2022). We grouped them into different categories. First, we found many adverbs and adverbial phrases that are typically used as discourse markers in the sequential domain (Crible, Degand, 2019), signaling discourse continuity: *en fait* ("in fact"), *bref* ("in short") and *alors* ("then") are part of this category. They are mostly used with discursive functions of addition, consequence or specification (Crible, Degand, 2019). As illustrated in example [13] (here with a function of consequence), *bref* was almost unanimously considered as an indicator of subjectivity in the text by the annotators, with a ratio of highlighted forms to total occurrences of 5.33. Some sequential occurrences of *en fait* show a concessive function, as in example [3]. Sequential discourse markers thus seem to be productive indicators of subjectivity, even more so when they are concessive.

[13] Il y a donc bien eu de multiples infractions à la loi, et pas simplement l'expression d'opinion. Sinon, cela revient d'ailleurs à alimenter la thèse qu'ils défendent, et aussi, en Belgique, les propos de la NVA et de ses différents leaders. Bref, si on suit cette idée, De Wever et les autres dirigeants de la NVA seraient déjà poursuivis pour avoir simplement exprimé une opinion indépendantiste…

So there were multiple breaches of the law, and not just the expression of opinion. Otherwise, it would be tantamount to fueling the thesis they are defending, and also, in Belgium, the statements made by the NVA and its various leaders. In short, if we follow this line of thinking, De Wever and the other leaders of the NVA are already being prosecuted for simply expressing a pro-independence opinion...

Example 13; mean HSR = 3.27; mean HCR = 3.55; *opinion* genre; medium ASR

We also found that other discourse markers belonging to the grammatical categories of subordinating and coordinating conjunctions, i.e. *donc* ("therefore"), *si* ("if") and *comme* ("like", "because"), were often highlighted in the texts. This is likely due to the fact that argumentative writing is very common in the *opinion* genre of journalism (Alhindi et al., 2020) and that those words are often used as markers of argumentation, as can be seen in examples [7] with *comme*, in [8] with *si* and in [13] with *donc*, *si* and *sinon* ("otherwise"). However, it should be noted that these discourse markers display a low H/N ratio, which may indicate that their subjectivity is dependent on the context or on the annotators' interpretations.

Furthermore, intensifiers and mitigators also represent another category of words found in the list in Table 4: *plus* ("more"), *très* ("very"), *presque* ("almost"), *peut-être* ("maybe").

Finally, an interesting finding concerns the presence of the verbs *doit* and *doivent* among the most frequently highlighted indicators of subjectivity (with an H/N ratio of 1.47), which are the singular and plural forms of the third person present indicative of *devoir* ("to have to"), a modal verb that can be used to express obligation. Our hypothesis is that the sense of necessity associated with this verb can be seen as a strong marker of the author's opinion in the text. Example [14] is an excerpt from an *opinion* article that received a low ASR from the model based on linguistic features, but that was considered as quite subjective by human readers. The annotations at word-level show that the presence of *doit* and *doivent* in the text, alongside nouns such as *nécessité* and *importance*, highly influenced their perception of the text as subjective. The same observation can be made in example [3], where the locution *il faut bien* reflecting the same deontic modality is highlighted by several annotators. This suggests that modal verbs, especially those related to obligation, should be considered when analysing subjectivity in journalistic discourse. Further work should be dedicated to putting into perspective discrepancies between the automated model and the subjectivity analyses made by human readers.

[14] La réaction face au groupuscule djihadiste doit être globale. Sur le terrain, les opérations militaires doivent aller de pair avec des gestes politiques et humanitaires. Chez nous, les attaques de Bruxelles ont montré la nécessité d'une plus grande collaboration entre les forces de police et les services de renseignements. Mais aussi l'importance d'un véritable travail de fond face à la radicalisation des esprits.

The response to the jihadist group must be comprehensive. On the ground, military operations must go hand in hand with political and humanitarian gestures. At home, the Brussels attacks demonstrated the need for greater collaboration between police forces and intelligence services. But also the importance of real in-depth work to counter the radicalisation of minds.

Example 14; mean HSR = 3.33; mean HCR = 3.67; *opinion* article; low ASR.

5. Limitations

It is important to acknowledge the limitations inherent to such a research project, as they may both help to contextualise our findings and to suggest avenues for future studies. First, the participants are all young adult Master students in journalism, which may limit the generalizability of the results. In addition, the annotations may be influenced by the opinions of the participants and by their personal knowledge of the topics treated in the corpus. Broadening this study to a larger group of participants would help to mitigate the impact of annotator subjectivity. Second, the sample of journalistic texts chosen for this study may not be representative of the wide spectrum of French-written press articles in terms of topics or writing styles. It was decided to limit the corpus to one media and one language to limit the influence of non-textual variables on the results. Again, a further study with a larger and more diverse corpus of texts would provide more generalizable findings and may provide interesting comparative inter-linguistic or inter-source insights. Finally, by focusing on

punctuation, words or groups of words, the design of our annotation made it difficult or impossible to analyse long-term dependencies and stylistic features that are not reducible to token level, such as sentence length or lexical complexity, which may be important for subjectivity interpretation by human readers.

6. Conclusions

We conducted a one-month experiment in which 36 participants evaluated the subjectivity of 150 excerpts of French press articles and highlighted the textual elements that they considered as indicators of subjectivity, i.e. words, groups of words or punctuation marks that signal the presence of the author's opinion in the text. A method for visualising multi-annotated texts in HTML format at token level was developed in the form of textual heat maps. Our study showed that the perception of a text as subjective may vary vastly between different readers, especially regarding texts discussing topics on which they have personal opinions. We confirmed that human readers tend to base their interpretations of subjectivity on elements that had already been identified in the literature, such as sentiment words and adjectives, and expressive punctuation. Our results also showed that less expected indicators, such as quotation marks, discourse markers, and modal verbs, also play a part in evaluating subjectivity in press discourse for human readers. We analysed the presence of the pronoun *on* as one of the most important indicators of subjectivity highlighted by annotators throughout our corpus, despite its use being typically encouraged in journalistic writing to conceal the author’s presence in the text.

These findings not only contribute to our understanding of subjectivity in journalistic discourse but also hold implications for the field of journalism. Recognizing the linguistic indicators that can sway reader perceptions towards subjectivity may prompt news organizations to refine their editorial processes and enhance transparency, ultimately enhancing reader trust. The new indicators of subjectivity identified in this study may be used to improve NLP models for automated subjective text classification and opinion mining, such as the model that was used in this paper to select the sample of articles.

Additional avenues for research could involve conducting a qualitative examination of participants' responses to the post-experiment questionnaire, delving deeper into their conceptions of objectivity and their impressions on the annotation task. A more comprehensive analysis of the influence of more independent variables such as gender and age of participants, or article topics, on subjectivity ratings, would likely be very informative. Furthermore, expanding the scope of the study by encompassing diverse media sources or languages, in conjunction with a more diverse pool of annotators, could yield valuable insights and enhance the comprehensiveness of this research project.

7. References

AKER, A., GRAVENKAMP, H., MAYER, S. J., HAMACHER, M., SMETS, A., NTI, A., ERDMANN, J., SERONG, J., WELPINGHUS, A. and MARCHI, F. 2019. Corpus of News

Articles Annotated with Article Level Subjectivity. *Workshop on Reducing Online Misinformation Exposure-ROME.*

ALHINDI, T., MURESAN, S. and PREOTIUC-PIETRO, D. 2020. Fact vs. Opinion: The Role of Argumentation Features in News Classification. *Proceedings of the 28th International Conference on Computational Linguistics*: 6139-6149.

AOUDA, M. 2019. Le pronom On dans la presse écrite: une marque d'effacement ou d'engagement énonciatifs? *Revue Ihalat* 228.

BALAHUR, A., STEINBERGER, R., KABADJOV, M., ZAVARELLA, V. VAN DER GOOT, E., HALKIA, M., POULIQUEN, B. and BELYAEVA, J. 2013. Sentiment Analysis in the News. *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC'2010)*: 2216-2220.

BENVENISTE, E. 1966. *Problèmes de linguistique générale* 1, Paris : Gallimard.

BIRJALI, M., KASRI, M. AND BENI-HSSANE, A. 2021. A Comprehensive Survey on Sentiment Analysis: Approaches, Challenges and Trends. *Knowledge-Based Systems* 226: 107134.

BOGAERT, J., ESCOUFLAIRE, L., DE MARNEFFE, M. C., DESCAMPE, A., STANDAERT, F. X. and FAIRON, C. 2023. TIPECS: A corpus cleaning method using machine learning and qualitative analysis. *International Conference on Corpus Linguistics (JLC).*

CHAPUT, L. 2019. Sur quelques marques de subjectivité dans le journalisme d'information politique de 1945 à 2015 au Québec. *Mots. Les langages du politique* : 151-168.

CHARAUDEAU, P. 2006. Discours journalistique et positionnements énonciatifs. Frontières et dérives. *Semen* 22.

CHENLO, J. M. and LOSADA, D. E. 2014. An Empirical Study of Sentence Features for Subjectivity and Polarity Classification. *Information Sciences* 280: 275-288.

CRIBLE, L. and DEGAND, L. 2019. Domains and functions: A Two-dimensional Account of Discourse Markers. *Discours. A Journal of Linguistics, Psycholinguistics and Computational Linguistics* 24.

DALHET, V. 2003. *Ponctuation et énonciation*. Matoury : Ibis rouge éditions.

DE COCK, B. 2016. Register, genre and referential ambiguity of personal pronouns: A cross-linguistic analysis. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)* 26(3): 361-378.

EENSOO-RAMDANI, E., BOURION, E., SLODZIAN, M. and VALETTE, M. 2011. De la fouille de données à la fabrique de l'opinion: Enjeux épistémologiques et propositions. *Les Cahiers du numérique* 7(2) : 15-40.

ESCOUFLAIRE, L. 2022. Identification des indicateurs linguistiques de la subjectivité les plus efficaces pour la classification d'articles de presse en français. *Actes de la Conférence JEP-TALN-RECITAL 2022* : 69-82.

ESCOUFLAIRE, L., BOGAERT, J., DESCAMPE, A. and FAIRON, C. 2023. The RTBF Corpus: A Dataset of 750,000 Belgian French News Articles Published Between 2008 and 2021. *International Conference on Corpus Linguistics (JLC).*

GOBIN, P., CAMBLATS, A. M., FAUROUS, W. and MATHEY, S. 2017. Une base de l'émotionalité (valence, arousal, catégories) de 1286 mots français selon l'âge (EMA). *European Review of Applied Psychology* 67(1): 25-42.

GUTZMANN, D. and STEI, E. 2011. How Quotation Marks What People Do with Words. *Journal of Pragmatics 43(10)*: 2650-2663.

HAYES, A. F. and KRIPPENDORFF, K. 2007. Answering the Call for a Standard Reliability Measure for Coding Data. *Communication Methods and Measures* 1(1): 77-89.

HO-DAC, L. M. and KÜPPERS, A. 2011. La subjectivité à travers les médias: étude comparée des médias participatifs et de la presse traditionnelle. *Corpus* 10 : 179-199.

HO-DAC, L. M., FABRE, C., PÉRY-WOODLEY, M. P., REBEYROLLE, J. AND TANGUY, L. 2012. An empirical approach to the signalling of enumerative structures. *Discours. A journal of linguistics, psycholinguistics and computational linguistics* 10.

IDE, N. and PUSTEJOVSKY, J. 2017. *Handbook of Linguistic Annotation*. New York: Springer.

KERBRAT-ORECCHIONI, C. 1980. *L'énonciation : de la subjectivité dans le langage.* Malakoff : Armand Colin.

KIRAKOSSIAN, A. 2015. *La subjectivité linguistique dans l'acceptation de Catherine Kerbrat-Orecchioni*. Université d'État d'Erevan.

KOREN, R. 2004. Argumentation, enjeux et pratique de l'« engagement neutre » : Le cas de l'écriture de presse. *Semen* 17.

KRIPPENDORFF, K. 2018. *Content Analysis: An Introduction to Its Methodology*. Newbury Park: Sage Publications.

KRÜGER, K. R., LUKOWIAK, A., SONNTAG, J., WARZECHA, S. and STEDE, M. 2017. Classifying News Versus Opinions in Newspapers: Linguistic Features for Domain Independence. *Natural Language Engineering* 23(5): 687-707.

KU, K. Y., KONG, Q., SONG, Y., DENG, L., KANG, Y. and HU, A. 2019. What Predicts Adolescents' Critical Thinking about Real-life News? The Roles of Social Media News Consumption and News Media Literacy. *Thinking Skills and Creativity* 33: 100570.

LARK, J., MORIN, E. and SALDARRIAGA, S. P. 2015. Canéphore : un corpus français pour la fouille d'opinion ciblée. *Actes de la Conférence JEP-TALN-RECITAL 2015.*

LEVY, R. E. 2021. Social Media, News Consumption, and Polarization: Evidence from a Field Experiment. *American Economic Review* 111(3): 831-870.

MARCHAND, M. 2012. État de l'art: L'influence du domaine sur la classification de l'opinion. *Actes de la Conférence JEP-TALN-RECITAL 2012* : 177-190.

MARNETTE, S. 2004. L'effacement énonciatif dans la presse contemporaine. *Langages* 4 : 51-64.

MATHET, Y. and WIDLÖCHER, A. 2019. Annotation, évaluation et mesure d'accord en linguistique de corpus. *Revue française de linguistique appliquée* 24(1) : 111-128.

MIHALCEA, R., BANEA, C. and WIEBE, J. 2007. Learning Multilingual Subjective Language via Cross-lingual Projections. *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*: 976-983.

MOHAMMAD, S. M. and TURNEY, P. D. 2013. Crowdsourcing a Word–Emotion Association Lexicon. *Computational Intelligence* 29(3): 436-465.

MUÑOZ-TORRES, J. R. 2012. Truth and Objectivity in Journalism: Anatomy of an Endless Misunderstanding. *Journalism Studies* 13(4): 566-582.

PAUN, S., ARTSTEIN, R., and POESIO, M. 2022. *Statistical Methods for Annotation Analysis.* San Rafael: Morgan and Claypool.

RABATEL, A. 2001. La valeur de « on » pronom indéfini/pronom personnel dans les perceptions représentées. *L'Information Grammaticale* 88(1) : 28-32.

RAVI, K. and RAVI, V. 2015. A Survey on Opinion Mining and Sentiment Analysis: Tasks, Approaches and Applications. *Knowledge-Based Systems* 89: 14-46.

REGMI, S. and BAL, B. K. 2015. What Make Facts Stand Out from Opinions? Fistinguishing Facts from Opinions in News Media. *Creativity in Intelligent, Technologies and Data Science*: 655-662.

SCHUDSON, M. 2001. The Objectivity Norm in American Journalism. *Journalism* 2(2): 149-170.

SPOOREN, W. P. and L. DEGAND 2010. Coding Coherence Relations: Reliability and Validity. *Corpus Linguistics and Linguistic Theory* 6(2): 241–266.

STEDE, M. and PELDSZUS, A. 2012. The Role of Illocutionary Status in the Usage Conditions of Causal Connectives and in Coherence Relations. *Journal of Pragmatics* 44(2): 214–229.

STEENSEN, S. 2017. Subjectivity as a Journalistic Ideal. In B. K. FONN, H. HORNMOEN and N. HYDE-CLARKE, et al. (eds.), *Putting a Face on It: Individual Expose and Subjectivity in Journalism*. Oslo: Cappelen Damm Academic Press: 25–47.

TANGUY, L., FABRE, C. and RIVIÈRE, L. 2018. Extending the Gold Standard for a Lexical Substitution Task: is it worth it? *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC).*

TKACHENKO, M., MALYUK, M., HOLMANYUK, A. and LIUBIMOV, N. 2020. *Label Studio: Data Labeling Software.* Available online: https://github.com/heartexlabs/label-studio.

TODIRASCU, A. 2019. Genre et classification automatique en TAL: le cas de genres journalistiques. *Linx. Revue des linguistes de l'université Paris X Nanterre* 78.

TONG, J. and ZUO, L. 2021. The Inapplicability of Objectivity: Understanding the Work of Data Journalism. *Journalism Practice* 15(2) : 153-169.

TUCHMAN, G. 1972. Objectivity as Strategic Ritual: An Examination of Newsmen's Notions of Objectivity. *American Journal of Sociology* 77(4): 660-679.

VIS, K., SANDERS, J. and SPOOREN, W. 2012. Diachronic changes in subjectivity and stance – A corpus linguistic study of Dutch news texts. *Discourse, Context and Media* 1(2-3): 95-102.

VOGEL, C., KOUTSOMBOGERA, M. and COSTELLO, R. 2020. Analyzing Likert Scale Inter-Annotator Disagreement. *Neural Approaches to Dynamics of Signal Exchanges*: 383-393.

WALLACE, L. R. 2020. *The View from Somewhere: Undoing the Myth of Journalistic Objectivity.* University of Chicago Press.

WIEBE, J., BRECK, E., BUCKLEY, C., CARDIE, C., DAVIS, P., FRASER, B., LITMAN, D., PIERCE, D., RILOFF, E., WILSON, T., DAY, D. and MAYBURY, M. T. 2003. Recognizing

and Organizing Opinions Expressed in the World Press. *New Directions in Question Answering*: 12-19.

WIEBE, J. and RILOFF, E. 2005. Creating Subjective and Objective Sentence Classifiers from Unannotated Texts. *Computational Linguistics and Intelligent Text Processing: 6th International Conference*. Berlin: Springer Berlin Heidelberg: 486-497.

WIEBE, J., WILSON, T. and CARDIE, C. 2005. Annotating Expressions of Opinions and Emotions in Language. *Language Resources and Evaluation* 39: 165-210.

WILSON, T., WIEBE, J. and CARDIE, C. 2017. *MPQA Opinion Corpus.* Handbook of Linguistic Annotation: 813-832.

Appendix 1

Table 5. List of all features of the logistic regression classification model used for selecting the sample articles, along with the class associated to each feature (Escouflaire, 2022).

Opinion	News
nb_on	*nb_digits*
pron_rel	*pron_1*
nb_neg	*nb_vb*
nb_adj	*length_words*
lexique3_sentiment	*nb_citations*
nrc_sentiment	
nb_exclam	
nb_interrog	
nb_pointvirg	
nb_susp	
nb_deuxpoints	
textblob_sent	
cttr	

Table 6. The description of each feature presented in Table 5.

Feature	Description
pron_1	Ratio of first-person pronouns and determiners
pron_rel	Ratio of relative pronouns
nb_vb	Ratio of verbs
nb_adj	Ratio of adjectives
nb_neg	Ratio of negation words
lexique3_sentiment	Ratio of words with positive/negative sentiment (Lexique3 lexicon).
nrc_sentiment	Ratio of sentiment words (NRC lexicon)
textblob_sent	Mean sentiment score (*TextBlob*)
length_words	Mean length of words
nb_interrog	Ratio of question marks
nb_on	Ratio of *on* tokens
nb_exclam	Ratio of exclamation marks
nb_pointvirg	Ratio of semicolons
nb_deuxpoints	Ratio of colons
nb_susp	Ratio of ellipses
nb_citations	Ratio of text between quotes
nb_digits	Ratio of digits
cttr	Corrected type-token ratio

Appendix 2

Screenshot of the Label Studio annotation interface. Participants were first asked to carefully read the excerpt. Then, they had to rate the excerpt's overall subjectivity on a scale from 1 to 5, and to evaluate (also on a scale from 1 to 5) their confidence towards their own subjectivity rating of the excerpt. Then, they were told to use the 'IDS' button to manually highlight, directly on the excerpt above, the indicators of subjectivity.

Insérer ici la Figure escouflaire_figure6.png

Appendix 3

An article which received a low automated subjectivity score but that was identified as highly subjective by human annotators:

> [16] Les conséquences de la crise en Europe sont que les Etats doivent réduire leur "train de vie", mais dans des conditions totalement absurdes. Ce sont ceux qui sont les faibles, ceux qui gagnent le moins qui doivent le payer le plus. Ce sont ceux qui ont besoin de relancer leur économie qui doivent serrer la ceinture de leur population. Ce sont ceux qui ont déjà des difficultés à qui on impose des taux d'intérêts et donc des niveaux de remboursement des emprunts encore plus drastiques. Le cercle vertueux s'est transformé dans la plus infernale et vicieuse spirale, entraînant salariés et allocataires vers l'abîme.
>
> *The consequences of the crisis in Europe are that governments are having to cut back on their "way of life", but under totally absurd conditions. It is those who are the weakest, those who earn the least, who have to pay the most. It is those who need to revive their economies who have to tighten the belts of their populations. It is those who are already in difficulty who are being forced to pay even more drastically in terms of interest rates and loan repayments. The virtuous circle has been transformed into the most infernal and vicious spiral, dragging employees and beneficiaries towards the abyss.*
>
> Example 16; mean HSR = 4.18; mean HCR = 4.27; *opinion* genre; low ASR.