Exploring the Relationship between the Progressive, Sentiment, and Subjectivity

A Computational Study of Reddit Posts

Jan Michel

Bachelor Thesis
Submitted to the Institute for English and American Studies

RWTH Aachen University
Institute for English and American Studies
English Linguistics

Contents

1 | Introduction

The progressive aspect and the use of the progressive constructions (BE -ing) have long been subject to extensive linguistic research because of their multiple functions, ranging from the strictly temporal to more nuanced uses such as expressing subjectivity. The main focus of early research has been on the temporal and aspectual use of the progressive, i.e. referring to a current state of affairs at reference time. This is the usual grammatical description of the progressive aspect. While subjectification in grammaticalisation, a shift in the meaning of grammatical constructions, such as the progressive, has long been observed, the diachronic evaluation of such a shift in the use of the progressive has also been documented. Additionally, comparative synchronic corpus studies of native speakers of inner and outer circle Englishes use of progressive show differences in use, suggesting a possible non-temporal use of the progressive. These suggestions have subsequently been taken up in research approaches able to show the modal use of the progressive in a "*non-aspecto-temporal*" (Killie 2004; de Wit and Brisard 2014) sense. While this research provides general insights into the tension between the use of the progressive and its aspectual description, it also suggests that the progressive affords to functions as a marker of epistemic contingency, i.e. an ongoing, inconclusive subjective evaluative process or subjectivity overall. Irritation and surprise are often associated with epistemic contingency, which could also be interpreted as subjectivity. Another link to subjectivity has been discussed in de Wit and Brisard (2014). According to further research by de Wit and Brisard, the progressive is used to express and emphasise personal accounts of certain states of affairs. Research suggests the different use of the progressive can primarily be observed in discoursive settings, such as conversations, but also social media posts.

Since the progressive serves as a marker of epistemic contingency, which in itself is directly related to subjectivity, a relationship with the expression of subjectivity through sentiment in a text seems plausible. An ongoing evaluative process, expressed through the emotions of irritation and surprise, among others, at least seems likely to translate directly into sentiment in a text.

Recent advances in computational approaches to language, particularly in machine learning algorithms and natural language processing, provide new methods for advanced research on this particular relationship. These methods aim to facilitate the analysis of large textual data sets. With the growing interest in probabilistic approaches to language in linguistics and the analysis of patterns in language, algorithms can use such perspectives to understand context in textual data. The consideration of context goes beyond the capabilities of established corpus approaches. The context of words is understood in terms of probabilities, with the algorithms transforming words into representational vectors and using topographical approaches to compute the relationship between words. One of the possible applications of these representation-based language models is sentiment analysis, which uses the context of the text to make predic-

tions about the underlying sentiment in the given instance of text.

Using the aforementioned theoretical framework of the progressive as a marker of subjectivity and computational sentiment analysis, this analysis aims to shed light on the relationship between sentiment and the use of the progressive. To achieve this, the approach presented here aims to answer the question of whether there is a correlation between sentiment and the use of the progressive. Specifically, this quantitative analysis attempts to test two main hypotheses:

H1 There is a correlation between sentiment (intensity) in text and the use of the progressive.

H2 Negative sentiment correlates with an increased use of progressive constructions.

The main data will be provided in the form of Reddit posts and comments, representing the discoursive setting in written text. The analysis of these texts will then follow a mixed methods approach, combining computational techniques and statistical modelling. A fine-tuned machine learning based BERT model will be used for sentiment classification, predicting both negative and positive sentiment intensity as well as a combined compound score. This model is especially qualified for this task, as it considers the context for its predictions, which goes beyond dictionary-based approaches to sentiment. The occurrence of progressive constructions within each sentence of the corpus will be marked, making use of a part-of-speech annotation of the dataset after sentiment intensity analysis and annotation. This data will be used to build a comprehensive data framework for analysis. Statistic modelling in the form of logistic regression analyses will then be used to explore potential correlations between the occurrence of progressive constructions (BE -ing), and sentiment strength.

A comparable approach was previously presented by Viola (2023), analysing the correlation between sentiment polarity and the use of the progressive in Italian mixed-genre texts. This research was able to show a small correlation between negative sentiment polarity and the use of progressive. However, this research was based only on a classification of the textual data into positive and negative sentiment, while leaving the sentiment strength out of the picture. Additionally, the correlation analysis was not facilitated in the form of statistical modelling, possibly leaving out other significant predictors for progressive use.

To shed light on the potential links between the progressive aspect, sentiment, and subjectivity, this approach employs a comprehensive, interdisciplinary investigation utilizing both linguistic and computational linguistic methods. First, a detailed review of existing research on the progressive aspect and computational methods will provide a theoretical foundation. This is followed by the methodology and data section, which outlines the data set, computational model development, and statistical methods. The results of the analysis are then presented, highlighting significant correlations and potential linguistic patterns. Finally, a qualitative examination of selected text samples will provide context and nuance, demonstrating the interplay of sentiment and use of the progressive and its relationship to subjectivity and epistemic contingency in real communicative settings.

2 | Academic discussion of the progressive and sentiment analysis

To provide a theoretical basis for the links of progressive constructions (BE -ing) and sentiment, it is necessary to consider the complex accounts of the English progressive and its use. The term progressive construction will be used throughout this account, to account for the aspectual, and non-aspectual use of the progressive. While traditional research on the progressive has mainly focused on its temporal and aspectual qualities, signalling a "durative aspect" (Hatcher 1951, 254; also Dennis 1940) of an event or action or ongoing activity (Biber et al. 1999, 470), later studies revealed more non-standard uses of the progressive as in marking subjectivity. As it is arguable, whether 'subjective' progressives occur within the boundaries of the aspectual description of the construction, the term progressive construction will be most frequently used to account for all uses of the progressive.

2.1 The progressive and its use

The general, grammatical account of the progressive aspect is focused on its properties in signalling an ongoing activity or event, which usually is temporally limited, mostly occurring in conversation and fiction (470–71). The progressive aspect is frequently used with verbs from a wide variety of semantic domains, including action, stative and mental verbs (Biber et al. 1999, 471; also Dennis 1940; Hatcher 1951). With the sole focus on its aspectual qualities, this description of the progressive construction has limitations and does not account for low-frequency uses identified in several later analyses. Hatcher's account suggests that conveying emotional intensity can be an effect of using the progressive (Hatcher 1951, 272–73), which overall results in signalling the "involvement of the subject" (Hatcher 1951, 279).

A stronger focus on such uses of the progressive is set in Bland (1988), suggesting that when used instead of the simple form with stative verbs, the progressive can slightly alter the meaning of the stative verbs by imposing dynamics (Bland 1988, 60). It is arguable, whether this could already suggest the involvement of subjectivity in the conveyed meaning. However, in these instances, the progressive, in effect, carries additional meaning that could be described as the speaker's personal interpretation or evaluation regarding the referred state. Additionally, Bland suggests the progressive can make an utterance "more personal" when the progressive is used instead of the simple form (Bland 1988, 61). While this does not yet fully account for non-standard uses of the progressive construction, together with Hatcher (1951), it provides a basis for further research on such cases.

The shift from the purely temporal understanding of the progressive to recognising its less frequent non-standard uses represents a key development in research on the progressive con-

struction. As described by Bland (1988), especially in discourse, the progressive is used in deviation from its grammatical account, especially concerning its use with stative verbs. This deviation shows both highlighting of subject's involvement, which has also been noted in Hatcher (1951), as well as when used with stative verbs, also to convey a slightly different meaning of the stative verb (Bland 1988, 60), hinting at the speaker's evaluation of the referenced state.

In empirical research, mostly diachronic perspectives, a shift in the use of the progressive has been visible from the sixteenth century onward. This shift accounts for both the general rise in the use of the progressive construction and its increasing frequency to occur with stative verbs (Collins 2008; Killie 2004). This change in the use of the progressive can be described in terms of subjectification, a process that can occur within the grammaticalisation of the progressive and embeds its meaning more and more in the speaker's personal evaluation and attitude (Wright 1994, 470). Traugott (1995) offers a more detailed description of this ongoing pragmatic-semantic change. While Hatcher suggests that the progressive's use with stative verbs simply forms an extension of its aspectual qualities, Bland and Wright both argue for a more nuanced description, including further notions of subjective evaluations and attitudes. Wright also introduces the term 'modal' progressives, referring to progressive constructions that mainly convey the speaker's subjective stance or attitude (Wright 1994, 470). This additionally suggests the progressive's use in a non-aspectual manner, especially when the progressive is used in situations where the simple form would be more intuitive. Generally, Wright puts forward a set of diagnostics for the identification of subjective progressives:

a. syntactic environment: **main** vs. subordinate clause;
b. tense: **present** vs. past;
c. verb type: **private (cognitive)** vs. activity
d. identity of subject: **first**, second, third person

(Wright 1994, 472, *own emphasis*)

Additionally, the subjectification with the progressive is believed to causally contribute to the rise in the use of the progressive with stative verbs, showing a "general change in function/meaning" (Killie 2004, 27; cf. Traugott 1995). Killie evaluated the diagnostics provided by Wright in a mostly qualitative corpus-based enquiry. Therefore, Killie analysed a large corpus of literary works from the sixteenth and seventeenth centuries, as this is believed to mark the beginning of the process of subjectification in the progressive. As suggested by Traugott (1995) and Killie (2004), modifying adverbs of the 'always'-type (i.e. always and closely related synonyms) occurring in the progressive construction have the largest influence on the subjectification of the progressive construction. Hence, Killie focuses on such progressive constructions to evaluate Wright's diagnostics. Since most progressive forms in her corpus are situated in subordinate clauses, and this finding also holds true for progressives occurring with an 'always'-type adverb, Killie suggests this diagnostic cannot be empirically supported. Likewise, her findings on the tense of the 'always'-type progressives suggest that those instances

of the progressive frequently occur in past tense (Killie 2004, 38). Also with regard to the semantic categories of the verbs, Killie states that correlations between specific groups such as cognitive/private verbs do not seem intuitive and cannot be supported by her findings (Killie 2004, 40-1). This is generally in line with the findings presented by Paulasto, who mentions the attitudinal use of the progressive, often occurring with a temporal adjunct of the always and is often connected to expressing a negative subjective attitude (Collins 2008, 239).

Killie's evaluation of Wright's diagnostics is complicated. In "The mystery of the modal progressive," Wright explicitly states that the diagnostics are not ordered and it generally is not suggested that an 'always'-type adverb is a necessary requirement for subjective progressives. It is however stated that such temporal adjuncts in the progressive construction alone significantly contribute to a subjective meaning of the construction (Wright 1994, 478). This especially holds true for the instances in which the progressive is used to convey an unfavourable judgment as implicit criticism about a recurring habitual situation (Collins 2008, 239). For a more extensive evaluation of Wright's diagnostics, an analysis of progressive constructions that do not explicitly collocate with a temporal adjunct would have possibly provided deeper insights into such collocational features. However, with regard to the modality of such instances of the progressive construction, both Killie and Wright seem to accept the notion that these instances indeed occur in a non-aspectual way (Killie 2004, 29; Wright 1994, 469).

More generally, these uses demonstrate a shift in the progressive form from a purely apsectotemporal quality to signalling the speaker's subjective evaluation. This phenomenon of subjectification of grammatical constructions has been discussed in many languages in linguistic research. Regarding the English language, Traugott (1995, 1989) describes subjectification as a pragmatic-semantic process where "meanings become increasingly based in the speaker's subjective belief state/attitude toward the proposition" (Traugott 1989, 31). This process accounts for the change in the use of the English progressive, where it is now more frequently used to mark the speaker's subjective attitude. Concerning grammaticalisation, it refers to the process by which grammatical forms or constructions, such as the progressive, shift towards expressing increasingly abstract, pragmatic, interpersonal, and speaker-based functions (Traugott 1995, 32). According to Killie (2004), this may account for the rise and shift in the use of the progressive that is observed in corpus-based research.

While these findings already provide a basis for the relation of the progressive and subjectivity, more recent research focused on such uses of the progressive. While it has already been noted by Killie (2004) and Paulasto (2014), the progressive is often used in a "non-aspectotemproal" sense (de Wit and Brisard 2014, 1) when expressing a subjective evaluation. In addition to instances where the progressive is used instead of the more usual simple form (1a,b), the research provided by de Wit and Brisard focuses more on modal progressives (de Wit and Brisard 2014; Brisard and de Wit 2014; Koss, de Wit, and Auwera 2022). According to de Wit et al., these uses of the progressive can act as a marker for epistemic contingency, an ongoing subjective evaluative process. Two of the more prominent expressions of this process are

irritation and surprise, which are generally supported by previous accounts on the progressive and subjectivity. Here the progressive form references the ongoing process of evaluation rather than the referred action. Especially with regard to subjectification, this highlights the change in meaning, with the focus shifting to the speaker's internal processes rather than the referenced situation (cf. Traugott (1995) and Killie (2004))

This overall account of the progressive form highlights its shift from a primarily temporal function to its increasingly frequent role in expressing subjective perspective and the speaker's attitude. This link between subjectivity and the use of the progressive provides the basis for an analysis of how sentiment, a central concept in understanding the emotional valence of texts, might interact with progressive constructions.

2.2 Sentiment and the progressive

While the link between the progressive aspect and subjectivity is well-documented, it remains to be fully explored whether more intense sentiment (either positive or negative) correlates with an increased use of progressive forms. Given that sentiment in text arises from subjective evaluations, a link between the progressive and sentiment generally seems plausible (Paulasto 2014; Liu et al. 2010). However, there is little research on this specific connection in the English language.

With the general rise in computational approaches to language, more opportunities for large-scale linguistic research become available. One such application is sentiment analysis (SA), which encompasses predicting the sentiment polarity or valence of a text sequence. Prior studies on this relationship have often been limited in scope or relied on manual annotation. Computational approaches, particularly SA, offer the ability to examine this potential correlation across large-scale corpora. While limited, existing research provides a point of reference. Notably, a study by Viola (2023) used the novel approach of SA to examine the connection between progressives and sentiment in Italian non-genre-specific texts from the EVALITA dataset. The article therefore provides a methodological framework for using SA in linguistic research. As research on this phenomenon suggests, subjectification of the progressive form is present in several languages (Kranich 2013, 4). Therefore the general methodological framework of Viola's account could provide a feasible basis for research on this phenomenon in the English language.

With an identical research goal of finding possible correlations between sentiment and progressive, Viola (2023) emphasised the study aimed to provide a methodological basis for these approaches. For the specific approach, the FEEL-IT sentiment classifier is used which is based on an Italian version of the *BERT* language model (UmBERTo) that is fine-tuned for SA (Bianchi, Nozza, and Hovy 2021). Input is classified into underlying emotions, (i.e. anger, fear, sadness, joy, etc). That model, however, is not specifically trained on the genre of text which is known to improve accuracy significantly (Viola 2023, 2; cf. e.g. Hutto and Gilbert 2014). Apart from genre, specific training and fine-tuning, the categorical classification of texts

into sentiment polarity (i.e. negative, positive) is based on collapsing the identified emotions into their respective groups. Not taking the valence or 'strength' into account, the explanatory power is limited. These limitations are also pointed out by Viola, while the general methodology still provides useful insights into possible applications of those computational approaches in linguistic research.

Results show that the majority of sentences containing progressive forms are classified as negative sentiment, which generally aligns with Collins (2008) and the description of attitudinal progressives. A small but statistically significant correlation between negative sentiment and the use of the progressive could be demonstrated. However, Viola discusses several limitations to this approach with regard to the limited explanatory power of categorical sentiment polarity and accuracy, as the sentiment analysis was performed on separate sentences rather than a full text, significantly limiting the context and hence the accuracy of the analysis. In addition, the non-genre-specific corpus is discussed as a limitation as well (Viola 2023, 5). Another critique is the limited scope of the discussed account, as additional factors, such as the progressive verb category, are not included in the analysis.

Overall, Viola provides a strong and adaptable methodological foundation for linguistic research on relations between sentiment and the use of the progressive. The discussed limitations can be overcome with slight adaptations to the machine-learning model for sentiment analysis, and incorporating a broader scope for the overall analysis of the progressives.

2.3 Natural language processing and sentiment analysis

As used in Viola (2023), sentiment analysis and natural language processing in general offer a large set of approaches towards language. While some approaches, such as part of speech annotation and dependency parsing are already well established in linguistics, especially with regard to corpus construction, other approaches such as sentiment analysis, are novel to linguistic research (Viola 2023, 1).

Generally, the term natural language processing refers to the computational processing and understanding of language. This very broad term therefore describes various applications from part of speech tagging to more complex approaches and systems such as large language models. Many of the algorithms in natural language processing utilise dictionaries and sets of rules to process text (cf. Hutto and Gilbert 2014; Taboada 2016). Usually, these algorithms are genre-specific as the language in the data set needs to be represented within the dictionary and rules and therefore bear certain limitations.

Recent advancements in natural language processing, especially with regard to neural networks and machine learning have introduced several language models that can be trained for the same purpose. Instead of dictionaries, these models use computational understanding to make predictions on textual data. As text cannot be directly computed and calculated with, tokens, usually words or 'sub-words', are transformed to representational high-dimensional vectors

with the values representing the syntactic, semantic, grammatical and lexical relations of the token (Mikolov et al. 2013). Another major advancement was the introduction of attention to transformer-based neural networks for language processing. Attention heads enable the algorithm to assign an importance marker to other tokens in a sequence that influences the meaning of a token in question. This enables the models to take more context into account and react to shifts in meaning due to semantic or lexical constructions (cf. Devlin et al. (2018)).

Sentiment analysis, as an application of natural language processing, was originally developed for consumer research with special regard to qualitatively analysing product reviews (Viola 2023, 2). Leveraging the state of the art of natural language processing, sentiment analysis itself also developed from highly genre-specific dictionary-based approaches to machine learning approaches, capturing the complex interplay of context and structure of the text. This development has broadened the field of applications for sentiment analysis, as machine-learning models can be fine-tuned to perform tasks on a specific genre of text.

While there are several language models available that are capable of performing sentiment analysis tasks, there are several factors to consider. Earlier language models were limited by their unidirectional nature (Devlin et al. 2018, 1), which BERT addresses through its bidirectional approach to contextual understanding. Using bidirectional self-attention (cf. Vaswani et al. 2017), BERT calculates the contextual relevance of each token in relation to the entire surrounding sequence. This feature distinguishes BERT from other large language models (LLMs) such as OpenAI's GPT (Devlin et al. 2018, 3), which only consider tokens to the left of the relevant token. With these mechanisms, the importance and influence of tokens on the meaning of the overall text sequence can be calculated. BERT is pre-trained on a massive data set using two methods: masked language modelling and next-sentence prediction. Due to this pre-training, the model is already familiar with most words that could occur during fine-tuning and application for sentiment analysis.

For classification fine-tuning, the algorithm is tuned with an annotated data set containing a text sequence of one or multiple relevant classifications. These fine-tuning data sets are usually classified by humans so that the ML algorithm can adapt to these predictions. Another option is using an automated algorithmic annotation using rule-based systems for instance. Even with this approach, the predictions of an ML algorithm often are more precise, capturing significantly more context than rule- or dictionary-based systems. The text sequences are encoded as described above, and an additional layer added to the base model is used to predict the classification of the text sequence (cf. Devlin et al. 2018, 9). While this approach can be used to predict categorical classifications as FEEL-IT (cf. Bianchi, Nozza, and Hovy 2021), predictions can also be continuous to account for sentiment valence with regard to SA.

Overall, the utilisation of tools such as sentiment analysis provides interesting methodological approaches for the automated semi-qualitative analysis of large corpora. Using such methods, links between sentiment and choice for the progressive aspect can be empirically analysed.

3 | Data and methodology

As the overall aim of the quantitative analysis is to show a correlation between sentiment and use of the progressive, the methodology is an adapted version of the methodology presented in Viola (2023), specifically addressing the limitations discussed in 2.2. Therefore additional linguistic features, such as verb category and intervening words are captured within the process of identifying progressive constructions. Similar data has been analysed in Killie (2004) and Collins (2008) and will present more nuanced results aimed at a better understanding of how sentiment influences the use of progressive constructions in text. The overall methodological setup is a Python-based pipeline, in which the corpus and a training subset are constructed, pre-processed, and used for training the machine learning algorithm and for the final statistical analysis.

3.1 Corpus construction

As argued in previous research, especially Bland (1988), the more flexible use of the progressive in uncommon uses can be observed especially in discourse. As the language used in conversation-based social networks represents this discoursive character, an analysis of such texts seems appropriate. However, in contrast to Viola (2023), this approach focuses on a single-genre corpus of social media posts. Therefore, the corpus consists of Reddit posts and comments, as they offer a specific insight into discourse use of language, as opposed to texts from multiple genres. Since Reddit posts are not limited in characters, such as Twitter, Reddit posts provide a solid basis for analysis (Proferes et al. 2021). The dataset is part of the `convokit` (Chang et al. 2020) toolkit for natural language processing (NLP) developed at Cornell University. It generally contains 948,169 subreddits with their respective posts and comments until the data cutoff in October 2018. For the corpus creation, three popular subreddits (*r/amitheasshole*, *r/idiotsincars*, *r/confessions*) provided the population from which a subset was sampled randomly. These popular subreddits were chosen based on the type of posts, namely personal stories which bear an increased likelihood of present subjectivity or sentiment.

Since previous research has already pointed out that subjective or attitudinal progressives are a low-frequency use of the progressive, and largely different text lengths ranging from just a single word to multiple paragraphs, a corpus of 20,000 texts should provide enough data to analyze the less common uses of the progressive (Killie 2004; Viola 2023). While this corpus size seems enormous, it is necessary to capture a large number of texts, as it is to be expected that the majority of sentences does not contain progressive constructions, and the frequency of subjective or attitudinal progressives (cf. Killie 2004; Collins 2008) is expected to be relatively small among those sentences that contain progressive forms. All these factors make it necessary for the corpus to be large in size, as this is required to detect less frequent patterns in language.

Table 3.1: Corpus composition, randomly sampled from three subreddits

Subreddit	No. of texts	Sentences
r/amitheasshole	8,219	-
r/idiotsincars	8,114	-
r/confessions	3,667	-
Total	20,000	64,214

For this analysis, cloud computing with Google Colab was utilised to overcome computational limitations due to the size of the corpus and the fine-tuning of a machine learning algorithm.

3.2 Sentiment analysis

The data obtained from `convokit` is then used for both fine-tuning the machine-learning model for sentiment valence classification and the analysis. The sentiment analysis on the corpus will be performed using a specifically fine-tuned BERT model. For the fine-tuning training data set, a randomised subset of 250,000 texts was pre-classified using the dictionary-based VADER model (Hutto and Gilbert 2014) and manually corrected for ambiguously classified texts. Since VADER is a dictionary-based model for sentiment analysis, the text needs to be pre-processed for it to efficiently yield meaningful annotations. Therefore, stop words, i.e. those words that do not carry significant semantic meaning, are removed from the original text. The new text without stopwords is then stored in a separate column of the data frame. This duplication of text is necessary, as the full text of the posts is necessary for fine-tuning of the BERT model, as machine learning models are not impacted by the inclusion of stopwords. For the sentiment analysis, in contrast to Viola (2023), the full text of the post is used as this captures more context and is, therefore, more reliable than analysing the sentiment in single sentences. Especially with regard to exaggerations and sarcasm, the larger context reduces the risk of false classification. This also accounts for the specific genre of the text as the model is specifically trained to classify Reddit posts, using this data set. Based on this data, a pre-trained BERT model for the English language was fine-tuned to predict three sentiment valence values: negative, positive, and a compound score, which all are generally based on the outputs offered by the VADER pre-classification. For this, BERT for sequence classification was trained in three epochs, having an early stopping mechanism in place to avoid overfitting (i.e. the machine learning model fitting too much to the training dataset). The fine-tuned model's performance is measured using a mean square error loss function on a validation subset, generally comparing the pre-classification values to the predicted values. The mean error and mean squared error are tested for improvement after each training epoch. Overall the model demonstrates good performance for sentiment valence prediction for Reddit posts with $ME = 0.093$ for the compound score and $ME = 0.002$ for both negative and positive (Table 3.2). These scores are not directly

comparable to the recall and F1 scores in nominal classification models focusing on sentiment polarity (Viola 2023; Pota et al. 2020), but demonstrate a reliable sentiment valence classification with negligible error regarding the focus of this research approach. A further qualitative analysis of a randomly chosen subset for classification accuracy further emphasised the model's precision with the predictions closely resembling the pre-classification values.

Table 3.2: Fine-tuned BERT sentiment valence classification performance

score	mean error (ME)	mean error squared (MES)
negative	0.020	0.038
positive	0.021	0.041
compound	0.070	0.175

The fine-tuned model was then used to classify the previously collected corpus consisting of 20,000 texts, which were not used for fine-tuning the BERT model. The three predicted sentiment scores (i.e. negative, positive, compound) are stored within the data frame. A qualitative analysis of a randomly selected subset of the text supported the model's performance with comprehensible sentiment valence classifications. The overall distribution, based on ranges according to the estimated possible error of the compound score shows an uneven sentiment distribution, which, based on the size of the corpus, should not pose a problem in the process of the analysis.

3.3 Extraction of linguistic features

As stated in Viola (2023), one of the major limitations of the approach lies in the limited data that was gathered on the progressives. As corpus research demonstrates that in some instances, the progressive in combination with stative and/or cognitive verbs was identified to be frequently associated with subjectification (Wright 1994). Therefore, while still basing the general methodology on Viola, the scope of the analysis is increased by additionally extracting the semantic category of the present participle in the progressive construction, as well as tense and clause localisation. This additional data both benefits the quantitative and qualitative analysis of the texts. With regard to Wright (1994), this additional information on the progressives, will make it possible to test the other diagnostics for subjective progressives, offered by Wright.

3.3.1 Identification of progressives

Before identifying the progressive constructions, the texts are split into sentences, while keeping reference to the original text. For this, the textual data is tokenized using the `nltk` tokenizer (Loper and Bird 2002). All sentences consisting of less than three words are dropped from the data set, as they cannot contain a full progressive construction and would only create noise in the statistical testing.

Algorithm 1 Identify Progressive Verb Forms in Sentences

```
Setup:
    Import SpaCy, Pandas, and NLTK                ▷ SpaCy and NLTK for NLP tasks, Pandas for working with the data frame
    Download NLTK data
    Load SpaCy model
function CATEGORIZEVERB(verb)
    Find and return the verb category             ▷ using WordNet lexnames as shown in table 3.3
end function
function IDENTIFYCLAUSE(sentence, index)
    Determine and return clause type and content  ▷ using the SpaCy dependency tree
end function
function GETTENSE(form of "(to) be")
    Return tense of the verb                      ▷ using a dictionary with the respective forms of (to) be
end function
function IDENTIFYPROGRESSIVES(df)
    Add new columns to store results
    for each sentence in df do
        Convert sentence to words and tag parts of speech
        Check for forms of "to be" (e.g., "is", "was")
        if form of "to be" found then
            Check neighbouring words for present participle (VBG)
            if not a "going-to" future then
                Record verb details, category, clause type, tense
            end if
        end if
    end for
    return updated df
end function
```

An algorithm searches the sentences in the text-level sentiment valence annotated sentences corpus for progressive constructions. For this `nltk` is used again for its word tokenizer and part of speech tagger. In doing so, the Penn Treebank tagset (Taylor, Marcus, and Santorini 2003) is utilized for 'querying' the corpus. While other, more precise taggers are available, the `nltk` tagger performs well on straightforward constructions such as the progressive. Additional limitations in the tagset can be overcome by directly utilising the token, as is done with the form of BE. Using the words and corresponding part of speech tags, the algorithm iterates through the sentences and identifies progressives consisting of a form of BE followed by up to three intervening words and a present participle (VBG) form (cf. 0). Intervening adverbs and modals are also identified and temporally stored.

Since going-to futures are not of interest for this analysis, as they have become more of an auxiliary, rather than a proper progressive construction through grammaticalisation (Perez 2023, 4). Therefore, an additional check takes place within the algorithm to identify instances of going-to futures, so that these are excluded from the identification process. Therefore all potential progressive constructions with "going" as the present participle are checked whether "going" is followed by "to" and a base form verb (VB). If the progressive construction is not an instance of the going-to future, the construction (form of BE, intervening words, present participle) are stored in the data frame. A binary variable is used to flag sentences containing progressive constructions, so that the data qualifies for logistic regression analysis for correlation testing.

Table 3.3: WordNet lexname utilisation for verb categorisation

	action	**stative**	**private**
lexnames	verb.motion verb.competition verb.change verb.communication	verb.stative	verb.cognition verb.perception verb.emotion

3.3.2 Progressive verb category

As Wright (1994) suggests there could be a correlation between the use of a specific category of verbs (i.e. private/cognitive verbs) in the progressive form and sentiment. Additionally, Viola (2023) discusses the lack of testing for correlations across verb categories as a limitation of her research. To efficiently capture verb categories in this large corpus of 20,000 texts, the progressive identification algorithm uses an additional function to retrieve the verb category of the present participle. This is done by leveraging the lexname feature of WordNet, which contains a semantic description of the word in question. With the implementation of WordNet in `nltk`, each present participle identified by the algorithm is reduced to its lemma and then categorised by its corresponding WordNet lexname. For this, the lexnames were grouped as presented in table 3.3. While this strategy has minor limitations in accuracy, this classification could still yield valuable insights into more features of subjective progressives.

3.3.3 Clause localisation

An additional function utilises the SpaCy library to identify the clauses in which the progressive constructions occur. To achieve this, the index of the token is passed to the function. The clause identification function then localises the respective token and iterates through the dependency tree until the token's head. The head's dependency relations are then used to classify the respective clause. If this relation is a clausal complement, open clausal complement, adverbial clause modifier, relative clause, or conjunct, the function classifies the localisation as sub-clause. Otherwise, the function returns "main clause". In addition to information retrieval, this function is also implemented in the identification of progressives, to ensure that all components of the progressive construction (i.e. the auxiliary 'be' and present participle) are within the same clause. This step increases the reliability and specificity of the progressive identification algorithm.

3.3.4 Tense

As suggested in Wright (1994), the tense of the auxiliary (BE) in the construction is also extracted. This is implemented in the form of a dictionary, from which the tense for the auxiliary is retrieved and stored in the dataset. This information is used to further enhance findings by testing for a correlation between sentiment and the tense of the progressive auxiliary verb.

3.4 Statistical testing

As stated in Viola (2023, 5), additional linguistic features of the progressives should be considered for potential correlations with sentiment. While the data to be extracted mainly followed the diagnostics presented in Wright (1994), the statistical testing is mainly focused on the correlation between sentiment and use of the progressive, as well as additional linguistic features such as verb category and clause subordination. For this, statistical modelling in the form of logistic regression will be used to test for possible correlations.

The approach in general aims to show correlations between use of the progressive and sentiment intensity. Therefore, a univariate logistic regression model, with is used to test for correlations between the sentiment intensities as the independent, and occurrence of the progressive as a binary dependent variables. A univariate analysis seems most appropriate, as this avoids problems of multicollinearity, which could arise using a multivariate analysis. For accuracy, the independent variable will also be used as the constant/intercept to predict a baseline for neutral sentiment for the respective intensity values. Since binary non-continuous variables are used for the progressive identification and additional features, logistic regression suits this analysis best.

Additionally, a logistic regression model is used to test for correlations between sentiment intensity and progressive verb category (i.e. action, stative, cognitive, and behavioural verbs). For this, dummy variables for each semantic verb category are created. Creating dummy variables is a commonly used, reliable method to perform regression analysis with categorical variables using mean comparison (Hardy 1993). This is implemented by a Python script that transforms the verb category into three categorical variables: *stative, cognitive, behavioural*. A comparable approach is used to test for possible correlations between sentiment and clause-type (*main clause, subordinate clause*). This transforms the multi-category variable into several binary variables. For the clause types, it is only necessary to test for correlations for one of both dummy variables. This analysis could help add a quantitative dimension to the notion of subjective progressives mostly occurring in main clauses (cf. Wright 1994).

An additional advantage of using univariate models for correlation testing is the assumption that the progressive construction mostly correlates with negative sentiment, as discussed in Viola (2023) and Collins (2008). As the sentiment analysis model predicts also a non-zero positive sentiment intensity for mainly negative texts especially with stronger overall sentiment intensity, this approach seems more reliable in detecting potential significant correlations.

4 | Results

With the methodological setup in place, the texts and sentences were processed to yield quantitative results on the correlation between sentiment and the progressives. Against the background of previous research on this phenomenon in Viola (2023), and descriptions of attitudinal and subjective uses of the progressive aspect, it is to be expected to find a correlation between the use of the progressive and an expression of negative attitude.

With the overall corpus consisting of 20,000 texts, there are 60,460 sentences that were tested for the occurrence of progressives. A total of 4,150 sentences within this data set contain one or more progressives, which makes up roughly 6.9% of all sentences. This is to be expected, as the corpus contains both top-level posts and comments that significantly vary in length so that some utterances might only contain a single word.

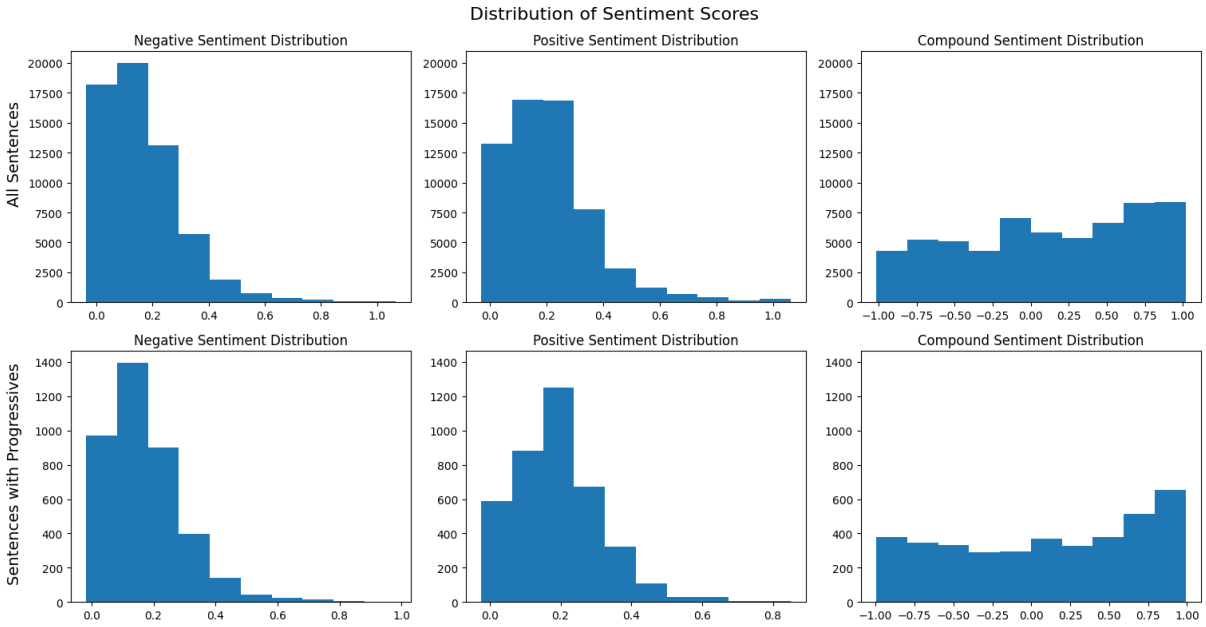


Figure 4.1: Distribution of sentiment valence scores in the sentences dataset

As suggested in Viola (2023, 5) and mentioned in 3.2, the sentiment analysis was performed on a text level before splitting the data set into single sentences. However, the distribution of sentiment suggests that most of the texts are close to neutral, with low scores for both positive and negative sentiment intensity in the majority of texts and texts with stronger sentiment intensity increasingly less frequent. This is further supported by the combined compound score. The subset consisting of only those sentences containing progressive forms could already hint at sentiment overall influencing the occurrence of progressive forms, as for all scores, there is an observable shift towards slightly increased scores with the scores close to zero (neutral) slightly decreased in proportion (cf. Fig. 4.1).

With regard to possible correlations between sentiment valence and intensity, and the use of the progressive, separate logistic regression models for each sentiment intensity score as the

Table 4.1: Logistic Regression Results for Sentiment Scores

	occurrence of progressives		
Score	**Negative**	**Positive**	**Compound**
Intercept	-2.7327***	-2.5459***	-2.6721***
Standard Error (SE)	(0.025)	(0.026)	(0.017)
Coefficient	0.3374**	-0.6641***	-0.0478
Standard Error (SE)	(0.110)	(0.107)	(0.028)
Observations	60460	60460	60460
Pseudo R-squared	0.00033	0.00137	0.00012

Note: * p < 0.05, ** p < 0.01, *** p < 0.001

independent variable and the progressive sentence classification as the dependent variable show expected correlations between negative sentiment and the use of the progressive. Not only is the correlation between negative sentiment intensity and the occurrence of the progressive statistically significant (p<0.01), it is also supported by a statistically significant negative correlation for positive sentiment intensity. This correlation confirms the hypothesis, as it suggests that with increasing negative sentiment, the probability of progressives to occur increases. It is to be noted, that although statistically significant, these results only account for a small proportion of variance in the occurrence of the progressive, as in Viola (2023). The coefficient for negative sentiment intensity is approximately 0.34 (95% CI: [0.121, 0.553]), (cf. Table 4.1). While not being of large effect, this shows that for a certain proportion, there is a distinct correlation between sentiment and use of the progressive. However, the small effect size aligns with the subjective or "attitudinal progressives" (Collins 2008) ('always'-type progressives respectively) being a low-frequency use of the progressive aspect.

Another hint for this same correlation is visible for the negative correlation between positive sentiment intensity and use of the progressive, with a pseudo R^2 of approximately 0.00137. The coefficient for positive sentiment intensity is approximately -0.664 (95% CI: [-0.874, -0.454]). This negative correlation further supports the correlation between the negative sentiment intensity and the occurrence of the progressive, as it suggests that a decrease in the positive sentiment intensity correlates with an increase in the likelihood of occurrence of progressive constructions, again hinting at the correlation between negative sentiment intensity and occurrence of progressive constructions. This generally supports the result of negative sentiment. While not statistically significant, the correlation in the compound score also hints at negative sentiment intensity correlating with an increased likelihood of progressive constructions to occur.

This finding already allows for a rejection of the alternative hypotheses, and accept the hypothesis that there is indeed a correlation between sentiment and use of the progressive, more specifically negative sentiment and use of the progressive. Further information on the specific uses of the progressive is necessary and can help evaluate other predictors for the occurrence of progressive constructions, as suggested in Viola (2023).

Additional information on the correlations is provided by the verb categories obtained in the progressive classification algorithm. As to be expected, the majority of verbs (56%) in the present participle form are action verbs that are most commonly associated with the progressive aspect. The other categories are less prominent, with stative verbs accounting for about 20% and private verbs for 15%. A small proportion (9%) could not be categorised.

As suggested in Viola (2023), logistic regression was used to test for possible correlations between certain verb categories and sentiment, when the progressive is used. However, the regressions did not yield statistically significant results with sufficient explanatory power to draw any conclusions from this data. The reason for this could be the limited size of the data set, as the amount of sentences per verb group is relatively small. The relative amounts of verbs, especially from the action and cognitive/private category however align with existing research (Collins 2008, 237). The small deviations in percentages could be a result of a slightly different categorisation (cf. Table 3.3).

Additional data is provided by the identification of the clause type in which the progressive construction occurs. Aligning with Killie (2004) on the progressive, the majority of progressive constructions are found in subordinate clauses (60%). A logistic regression test for a potential correlation between sentiment intensity and the progressive occurring in main clauses shows a positive correlation between negative sentiment intensity and occurrence of progressives in main clauses (cf. Table 4.2). As with the general correlation between sentiment intensity and occurrence of progressive constructions, this finding is supported by all three sentiment values with statistically significant correlations, although with small expected effect sizes. This finding generally aligns with previous accounts on subjective and attitudinal progressives predominantly occurring in main clauses (Wright 1994).

The same shows for the correlation of tenses and sentiment intensity. The statistically significant correlations all point towards a tendency of negative sentiment intensity increasing the likelihood of progressives to occur in the respective utterances. This correlation further supports the diagnostics offered in Wright (1994), showing that subjective progressives seem to tend to occur in present tense as opposed to past tense.

Table 4.2: Logistic regression results for correlation between sentiment intensity and further diagnostics based on Wright (1994)

	main clause			**present tense**		
Score	**Neg**	**Pos**	**Comp**	**Neg**	**Pos**	**Comp**
Int.	-3.75***	-3.47***	-3.62***	-3.43***	-3.20***	-3.30***
SE	(0.039)	(0.041)	(0.026)	(0.033)	(0.035)	(0.022)
Coeff.	0.688***	-0.828***	-0.181***	0.712***	-0.569***	-0.104*
SE	(0.164)	(0.169)	(0.044)	(0.141)	(0.142)	(0.038)

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5 | Qualitative analysis

As the quantitative results suggest a correlation between negative sentiment intensity and the use of the progressive, a qualitative analysis of selected sentences from the corpus will provide additional context on this correlation. With the theoretical foundation of subjective progressives (Killie 2004; Wright 1994) and attitudinal progressives (Collins 2008), the focus of this analysis will be directed at examining whether this correlation might align with such descriptions of non-standard uses of the progressive aspect. To achieve this, a selection of 'always'-type progressives from the corpus will be analysed against the diagnostics provided in Wright (1994, 472). Additionally, some instances will be analysed that go along with strong sentiment intensity.

5.1 On the *'always'*-type progressives

As the function for identification of the progressives also stores intervening words, the corpus was queried for 'always' progressives as described in Wright (1994) and Killie (2004). As stated in (2.1), this progressive construction contains an instance of 'always' in between the auxiliary verb and the present participle. In previous research, this type of progressive has often been brought up with regard to the subjectification of the progressive. In the overall dataset, there are 11 instances of the 'always' progressive. However, these instances all have rather low to medium sentiment intensity for all scores. Therefore, all these examples stem from rather neutral utterances in the corpus.

(1) a. she bites other children in kindergarten , she doesn't talk , she **is always screaming** when something doesn't gobher way.

(r/AmItheAsshole [*sic*], Utt.-ID: 88vxzp)

b. Is it really possible that their lives are THAT busy that they **are always doing** important things 24/7?

(r/AmItheAsshole, Utt.-ID: 8gv9rg)

c. Me and my SO both smoke weed, I smoke more than her and she **is constantly getting** on at me about how long it's lasting/smoking it without her.

(r/confessions, Utt.-ID: 9hpm3h)

As seen in (1a), the progressive aspect is used instead of the more common simple form, referring to a situation with a habitual character. The sentiment score for the post (compound: -0.23) does not suggest strong sentiment. This finding generally aligns with Killie (2004), as the use of the always progressive in that utterance could be considered part of an exaggerated statement. However, the verb category does not match the description in Wright (1994), which

suggests that in instances where the always progressive is used, the present participle often stems from the category of cognitive/mental verbs, as previously demonstrated in Killie (2004).

The same can be said for (1b), as the referred situation also has a habitual character. It is especially noteworthy that the always-type progressive can occur in interrogative utterances. As in (1a), the simple form, i.e. "... that they always do important things", would be more common to use with regard to the habitual character. Again, the content of the utterance suggests that the 'always'-type progressive is used to highlight the exaggerated statement of working around the clock. Especially with regard to Wright (1994), several criteria for subjective progressives are fulfilled as this instance marks an exaggeration, expresses disbelief or a negative attitude of the speaker and uses the adverb "always".

The progressive in (1c) differs in the intervening modifying adverb, where a close synonym of 'always' is used. This kind of deviation from the prototypical definition of the subjective 'always'-type progressive was previously discussed in Wright (1994) and Traugott (1995). Apart from the use of a close synonym, the other criteria are mostly fulfilled in this example, namely the sentence being in the present tense with a first or third-person pronoun.

With regard to overall verbal categories, both sentences can be classified with action verbs, i.e. those that reference externally visible processes or actions. This is the most common group of verbs in progressive constructions (cf. Collins 2008; Killie 2004; Paulasto 2014). This holds true for all identified 'always'-type progressives in the overall corpus. Another point made in Killie (2004, 31) is the potential that 'always'-type progressives are commonly associated with a figurative meaning. For the instances present in this corpus, this does not seem to be the case.

With regard to the clause type, the findings differ from the overlapping notions of attitudinal or subjective progressives in Collins (2008), Wright (1994), and Killie (2004). While there is a total of 11 'always'-type progressives, 8 of these occur in the sub-clause of the respective sentence. According to the aforementioned research, the majority of such progressives would occur in main clauses which cannot be confirmed with the findings from this corpus.

These instances are almost prototypical for the theoretical descriptions of the modal or subjective progressives, having the 'always'-type adverb which has been identified as the main characteristic of these types of progressive (cf. Traugott 1989, 1995; Wright 1994; Killie 2004; Collins 2008). However, because of the rather low to medium sentiment intensity for these few instances in the corpus, these cannot have led to the statistically significant correlation between negative sentiment intensity and the use of the progressive.

5.2 Fringe cases

In addition to the clear-cut examples in (1), there are other instances in which the progressive is similarly used to convey and highlight a subjective and evaluative character of an utterance. These instances do not occur with an intervening adverb but similarly use the progressive in instances where, in grammatical terms, the simple form would have been appropriate. Also, these

examples often feature strong sentiment intensity, which could therefore provide a qualitative basis for the existing correlation.

(2) To me it seems that the Jeep **was being** the idiot and the silver car was tired of their bullshit and went around but you're right it's hard to say what exactly is going on
(r/IdiotsInCars, Utt.-ID: dw7u3je)

In (2), the main characteristic of subjective progressives is missing. The construction however seems to serve a subjective purpose, rather than an aspecto-temporal one. While the subjective stance is already directly addressed at the start of the utterance, the use of the progressive seems to refer to the temporary internal evaluative process rather than directly to the behaviour of the jeep which is directly referenced in the utterance. This aligns with findings in Brisard and de Wit (2014, 218), hinting at epistemic contingency. This is further supported by the use of a private or cognitive verb as the present participle. Therefore, (2) clearly visualises the subjectification in the grammaticalisation of the English progressive while not fitting the previous definition of subjective progressives.

(3) But I **was just wanting** to see the approximate speed which in either case doesn't seem crazy enough, showing how dangerous an idiot can be even at low-ish speeds
(r/IdiotsInCars, Utt.-ID: dw7u3je)

As Traugott (1995) and Wright (1994) pointed out, the presence of the adverb ('always' or close synonyms) is the key characteristic of subjective or modal progressives. However, instances such as (2) suggest that this criterion might not be necessary for all subjective progressives. This is further demonstrated in (3), in which a private verb occurs in the present participle form. However, the progressive in past tense still takes on the role of conveying a layer of subjectivity in the utterance. In this instance especially, the epistemic contingency (Brisard and de Wit 2014, 218) seems to be a fitting explanation for the use of the progressive. In this instance, the progressive aspect highlights the temporary evaluative state expressed in the present participle "wanting", which is said to collocate very rarely with progressive constructions in their aspectual sense (Biber et al. 1999, 472).

These two examples might illustrate, how the findings by de Wit and Brisard might extend the diagnostics offered in Wright (1994), as this seems to capture the instances that go beyond the syntactic and grammatical markers identified by Wright.

6 | Discussion

Besides examining the suitability of computational methods, such as sentiment analysis, this analysis aimed to explore the relationship between sentiment and the use of the progressive aspect. Thus focusing on the progressive aspect's role in conveying subjectivity and sentiment. The statistical findings confirmed the initial hypothesis that there is indeed a correlation between the occurrence of progressive constructions (BE -ing) and negative sentiment intensity. This correlation was found in all analysed sentiment values (negative, positive, and compound) with negative sentiment showing an expected small but meaningful correlation.

As initially explained, a theoretical foundation is offered by previous research on subjectification with regard to the grammaticalisation of the progressive aspect. With this background, the findings presented before offer a perspective supporting a more nuanced interpretation of the use of the progressive aspect in discourse, that integrates with previous research on the use of the progressive with relation to subjectivity.

With the results from the quantitative and qualitative analyses, it is possible to gain further insights into the diagnostics provided in Wright (1994), which provided a theoretical base for the examination of the relationship between sentiment and the occurrence of progressive constructions. Her diagnostics suggest that certain features associated with the progressive afford to signal subjective and/or attitudinal meanings. With Killie (2004) partially supporting these diagnostics, the findings from the quantitative analysis provide further evidence for some of the diagnostics.

Additionally, the findings offer broader linguistic implications for understanding the shift in the meaning of grammatical constructions, such as the progressive, with regard to subjectification and grammaticalisation (Traugott 1995). Especially with regard to Bland (1988), this includes the implications for language teaching, where understanding more nuanced uses of the grammatical forms can be challenging. However, a clear understanding of the low-frequency uses can benefit language teaching, as described in Bland (1988).

Through this comprehensive analysis, the discussion seeks to bridge empirical data with theoretical insights, providing a robust platform for future research and practical applications in linguistics and related fields.

6.1 Theoretical implications

The findings on the correlation between sentiment and the use of the progressive bear several implications for well-established theoretical accounts on low-frequency uses of the progressive aspect, generally aligning well with the notion of subjective progressives.

This research supports linguistic theories that describe the progressive not only as a temporal marker but also as a device for expressing subjectivity and modality. Theories discussed by

Traugott (1995) and de Wit and Brisard (2014) have highlighted the potential for grammatical forms to evolve in usage, adapting to modal and subjective functions as part of their grammaticalisation process. The current findings lend empirical support to these theories, demonstrating how the progressive can extend beyond its primary temporal function to convey nuanced layers of subjective meaning, particularly in conveying sentiment.

The findings presented here support the theoretical accounts describing the progressive not only as a temporal marker but also as a means for conveying subjectivity or modality (cf. Wright 1994; Brisard and de Wit 2014). Especially subjectification theory has emphasised the potential for grammatical constructions to undergo semantic change to serve more modal and subjective functions in the process of grammaticalisation (Traugott 1995). The quantitative findings presented here empirically support these theories, demonstrating the progressive's use beyond its primary temporal function, expressing subjective meaning and sentiment.

The traditional aspect theory primarily positions the progressive as a marker of ongoing action, focusing on its temporal boundaries (Dennis 1940; Hatcher 1951). Hatcher suggests the use of the progressive with stative verbs to be an extension of its aspectual qualities. While both acknowledge the uncommon uses, they fall short of an explanation of the progressive's modal, non-aspecto-temporal uses. However, the empirical findings suggest a broader functionality, aligning well with findings in Wright (1994), suggesting the progressive's use to express subjectivity. The progressive aspect's role in signalling epistemic contingency, as seen in its correlation with negative sentiment, underscores a dual functionality that has been underexplored in aspectual analysis, as explained in Brisard and de Wit (2014).

Moreover, this study aligns with the shifts in aspect theory that embrace a more integrative approach to understanding grammatical aspects. For instance, Traugott (1995) argues that due to the process of grammaticalisation and subjectification the meaning of grammatical constructions, such as the progressive, shift from their pure aspectual meaning towards more nuanced meanings, also involving notions of subjectivity. This adaptation in theoretical stance allows for a more holistic view of how grammatical aspects operate within language, accounting for both traditional temporal uses and the emergent modal uses that characterize modern discourse.

The static view of grammatical categories is increasingly being challenged by empirical data suggesting that grammatical markers are capable of acquiring new functions over time. For the progressive in particular, this has long been observed in the rise of stative verbs in the progressive form, which have been pointed out in early research by Hatcher and Dennis, and have since been supported by empirical research such as Collins (2008) and Paulasto (2014). This examination of correlations between sentiment and the occurrence of the progressive contributes to this re-evaluation by providing concrete evidence and examples of the progressive aspect functioning in ways that traditional aspect theory does not fully account for.

Overall, the findings from this study challenge and extend traditional aspect theory by demonstrating the progressive aspect's capability to express not only temporal continuity but also subjective evaluations and sentiment. This observation can be interpreted in terms of gram-

maticalisation and subjectification of the progressive construction. This dual functionality could be considered in future theoretical and descriptive works on the aspect, ensuring that linguistic descriptions and pedagogical approaches align with the actual usage of language in varied communicative contexts, as mentioned by Bland (1988).

The diagnostics proposed by Wright (Wright 1994, 472) provide a valuable framework for understanding the subjective and modal uses of the progressive aspect. While Wright directly addresses subjectivity as the conveyed notion in such low-frequency uses of the progressive aspect, the quantitative findings presented before suggest that sentiment can indeed be used to detect subjectivity in textual data. This has also been suggested by Liu et al. (2010), stating that one of the main assumptions of sentiment analysis is that sentiment in the text is the expression of a subjective attitude, or, in alignment with Brisard and de Wit (2014), an evaluation.

In "The mystery of the modal progressive," Wright offers diagnostics for the identification of subjective progressives. Namely, their syntactic environment (occurrence in main clauses), tense (present), verb type (private), and the identity of the subject (mostly first person), (Wright 1994, 472). Killie (2004) evaluated the diagnostics empirically with a large corpus from 16th and 17th century English literature. In her findings, she was not able to support Wright's diagnostics besides the collocation with temporal adjuncts of the 'always'-type.

The findings from the quantitative analysis indicate a significant correlation between the use of the progressive in main clauses and negative sentiment, which aligns with Wright's diagnostic that subjective progressives often appear in main clauses. The correlation with negative sentiment fits the description of attitudinal progressives (Collins 2008, 239). However, this finding also suggests that Killie's description does not fully account for the subjective progressives when accepting the use of sentiment as an indicator of subjectivity.

Consistent with Wright's diagnostics and findings by Killie and Collins, the statistical findings also demonstrated a positive correlation between negative sentiment intensity and the occurrence of present tense progressive constructions. These findings reinforce the notion that the present tense may be more conducive to expressing subjective and attitudinal meanings through the use of the progressive. This finding also underscores the notion of attitudinal progressives in Collins (2008, 239), suggesting these instances of progressives often convey negative subjective attitudes.

Wright's suggestion with regard to the semantic categories of verbs used in the progressive could neither be supported nor refuted. As Wright suggests, progressive constructions involving a private verb as the present participle are more likely to fall into the category of subjective progressives. While this diagnostic could not be supported by empirical research, with findings suggesting no distinct involvement of private verbs in such uses of the progressive (Killie 2004, 40–41), an analysis for correlation between sentiment intensity semantic verb categories, as suggested in Viola (2023), did not yield meaningful results.

While the identity of the subject has not been examined within the quantitative analysis, the findings in the qualitative analysis show that especially for the 'always'-type progressives, the

diagnostics provided in Wright (1994) generally seem to align with the findings in the Reddit corpus. This diagnostic has, however, been previously examined in empirical corpus research (Killie 2004, 39). Her findings do not support Wright's notion of preferred identities of the subject for subjective progressives. Killie points out, that there is no evidence for subjective progressives' tendency to predominantly occur with first- or second-person subjects, as described by Wright.

The evaluation of Wright's diagnostics in light of these findings suggests that while some aspects of her framework hold true, others may need revision to accommodate changes in the findings presented here and in previous corpus research such as Killie (2004) and Collins (2008). Particularly, the broadening of verb types and the diminished relevance of the subject's identity point to inaccuracies in Wright's description, or towards the ongoing process of subjectification, that asks for a more dynamic interpretation of the progressive aspect with regard to such low-frequency uses.

Overall, the insights gained from this evaluation not only validate certain aspects of Wright's diagnostics but also highlight areas for further necessary research, particularly in expanding the theoretical framework to include a wider range of linguistic markers, more accurately capturing the ongoing subjectification of the progressive.

6.2 Sentiment and subjectivity

The relationship between sentiment intensity and the use of the progressive aspect is a central focus of this thesis. Especially with regard to empirical investigations on subjectification and grammaticalisation in the progressive aspect, the findings from the quantitative analysis point towards sentiment being a reasonably reliable indicator of subjectivity in texts. As mentioned before, this is supported by both the basic assumptions of sentiment analysis (Liu et al. 2010, 1), and previous examinations on the correlation between sentiment and use of the progressive (Viola 2023). As discussed in Kranich (2013), the process of subjectification can be observed in several "(loosely related and completely unrelated)" languages (Kranich 2013, 4). Therefore the findings from Viola's account provide a valuable foundation

The analysis has demonstrated a significant correlation between negative sentiment intensity and the use of the progressive, supporting the notion that the progressive aspect is often utilized to express negative evaluations and attitudes. This finding is consistent with theories that associate the progressive with subjectivity, such as those discussed Brisard and de Wit (2014), suggesting the progressive is frequently used to convey epistemic stances, such as the speaker's subjective evaluation. This is particularly noteworthy in instances where the progressive underscores an unexpected or contingent nature of an action, which is often associated with negative evaluations (Brisard and de Wit 2014, 202). This aligns with the findings from the quantitative analysis, hinting at a correlation between negative sentiment intensity and the occurrence of progressive constructions. With Liu's account, that sentiment is a subjective, evaluative stance

in text, the quantitative findings, therefore, seem to represent the subjective progressive, that has been previously discussed in Wright (1994) and Killie (2004).

The qualitative analysis of the 'always'-type progressives has provided deeper insights into how these forms contribute to the expression of sentiment in discourse. The examples from the Reddit corpus illustrate that these progressives are used not merely for signalling temporal duration but to convey nuanced, often negative sentiments about habitual actions or states, as highlighted by Killie (2004) and Wright (1994). Though it is debatable, whether these progressives occur in a "non-aspecto-temporal" manner, as argued for by Killie and Wright. This usage aligns with the diagnostic criteria set forth by Wright, emphasizing the expressive capacity of progressives in marking subjectivity and attitude.

The influence of intervening words, especially temporal adverbs like 'always', in progressive constructions has been notable in the corpus. Such modifiers intensify the subjective or evaluative nature of the progressive, as they often underscore the speaker's frustration or irritation with ongoing actions. This observation aligns with the findings from Killie (2004), who noted the subjectification of the progressive form in historical and contemporary English.

In summary, the analysis of sentiment in relation to the use of the progressive has underscored the aspect's flexibility in expressing more than mere temporal boundaries. It has shown that the progressive can effectively convey a wide range of emotional tones and attitudes, which are crucial in subjective and evaluative contexts. This underscores the need for aspect theories to consider these non-temporal uses more prominently, as suggested by recent linguistic research (Viola 2023).

Examining these dynamics, the findings presented here not only confirm the progressive's role in expressing subjectivity and sentiment but also enrich the understanding of how grammatical forms evolve to meet communicative needs in diverse discursive settings.

6.3 Broader linguistic implications

While the findings contribute to the thorough discussion of pre-existing research on the relation of subjectivity and the occurrence or use of the progressive, there are more implications brought up by these results.

The findings from this study suggest that the progressive aspect, traditionally taught primarily as a marker of ongoing action, also plays a significant role in expressing sentiment and subjectivity. This expanded understanding could be incorporated into language teaching, especially for ESL learners who may benefit from a more nuanced grasp of how grammatical forms can convey subtle nuances of meaning beyond their conventional uses. Specifically in the instances where the progressive is argued to be used in a non-aspecto-temporal or modal manner, as opposed to its aspectual use. Educators can develop teaching materials and exercises that highlight the multifunctional uses of the progressive, potentially enhancing learners' communicative competence and sensitivity to language nuances.

With Bland (1988) stating that such uses are well observable in discourse, it would be interesting to examine this phenomenon in spoken discourse. This could provide more insights into the specific use in such instances and could yield interesting findings, especially with regard to other possibly collocating markers, such as filled pauses and other prosodic features.

Apart from this, the study opens several directions for future research. One promising direction is the exploration of the progressive aspect across different genres or types of discourse to examine if the observed patterns hold consistently or vary significantly. By loosening the genre-specificity, the results then could be more generalized than those that heavily rely on genre-specific corpora, such as the Reddit corpus. This could involve analyzing the progressive in professional, academic, or literary texts to compare with the findings from online discourse used in this study.

Additionally, cross-linguistic comparisons could provide valuable insights into whether the patterns observed in English are unique or part of broader cross-linguistic trends in the use of grammatical aspects. Especially with regard to the findings presented in Viola (2023), which align well with the results from the quantitative analysis provided here. Such studies could help determine the universality of the progressive's role in expressing sentiment and subjectivity, contributing to a deeper understanding of aspects in a global context. However, Kranich (2013) already mentions, that such uses and the process of subjectification in the progressive aspect can be considered a cross-linguistic phenomenon, that is observed across a variety of languages.

With modern technologies and communication platforms, discourse use of written language becomes more prominent overall. With Biber et al. (1999) and Bland (1988) among others, stating that the progressive overall has the tendency to occur most frequently in discourse, the analysis of the impact of such digital communication platforms on the process of subjectification and grammaticalisation of the progressive aspect could be meaningful. This could extend the findings presented here and offer a comprehensive view of how language evolves with regard to new communicative environments.

Lastly, an analysis of the perception of such instances of the progressive could provide meaningful insights into how these low-frequency uses contribute to the overall meaning and processing of the sentences. Studies suggest that in instances where the progressive is used instead of the simple form, the processing of the sentence differs (cf. Madden-Lombardi, Dominey, and Ventre-Dominey 2017). An analysis of how these instances are processed could contribute to a more nuanced understanding of how such low-frequency uses shape and convey attitude in discourse.

In summary, the implications of this study extend well beyond theoretical linguistics, influencing practical applications in language teaching and setting the stage for future inquiries. By integrating the findings into further research, the understanding of the progressive aspect's role in language, especially with regard to low-frequency uses can be enhanced.

6.4 Reflection of the methodology

This section assesses the statistical significance and effect sizes of the correlations found in the study, offering insights into the practical implications of these results.

The statistical models employed in this study have confirmed significant correlations between the use of the progressive and sentiment, particularly negative sentiment. While these correlations are statistically significant, as evidenced by low p-values, the effect sizes, represented by pseudo R^2 values, are relatively small. This suggests that while the progressive's use is indeed associated with sentiment, it is one of many factors influencing its usage. These findings align with the nuanced uses of the progressive aspect as discussed by de Wit and Brisard (2014) and Brisard and de Wit (2014), where the progressive serves multiple communicative functions beyond mere temporal progression. The effect of sentiment intensity on the occurrence of the progressive can therefore be considered relatively small. This was already expected as this use type of the progressive was already identified to be a low-frequency use in Wright (1994), Killie (2004), and Collins (2008).

While logistic regression has proven useful for exploring the relationships between sentiment and progressive usage, there are some limitations to this approach. Generally, logistic regression assumes a linear relationship between the independent and dependent variables, which may not fully capture the complexities of language use and sentiment expression with other potential predictors such as personal writing style, which the statistical model does not account for. Additionally, the pseudo R^2 values, while helpful, do not provide a measure of explained variance as robust as R^2 in linear regression, potentially underestimating the impact of other unmeasured variables.

In summary, the statistical analyses performed in this study highlight a meaningful, if modest, correlation between sentiment intensity and the usage of the progressive aspect as a low-frequency use. These results reinforce the importance of considering multiple linguistic and contextual factors when interpreting the role of grammatical constructions in conveying sentiment. Future studies may benefit from employing mixed-methods approaches to better account for the layered nature of language use and sentiment expression.

In addition to the explanatory power of the statistical examination, the methodology provides a basis for further reflection. As the foundation for this methodology is provided by Viola (2023), several adaptations have been made to overcome limitations in Viola's approach. The breadth of the statistical analysis was drastically increased by incorporating further analyses of Wright's diagnostics. Also, as opposed to sentiment polarity, sentiment intensity was used to classify the texts from the corpus.

Overall, the methodological setup of this study, involving a comprehensive corpus analysis combined with advanced statistical modeling, allowed for a nuanced exploration of the progressive aspect's usage in relation to sentiment. Utilizing a large corpus derived from Reddit posts provided a rich data set reflecting language use in a specific communicative context.

With more than 60,000 sentences, the corpus size is already immense, which is necessary to meaningfully analyse the low-frequency use of the progressive in expressing subjectivity. However, the corpus size might not be sufficient to analyse correlations between sentiment and specific semantic categories of verbs. As previously discussed, no meaningful, statistically significant results could be yielded using logistic regression. This might change with more possible observations per semantic verb category in an even larger corpus.

The use of machine learning approaches, such as the BERT-based sentiment analysis provides helpful methods for annotating and coding large corpora. This approach bears some limitations. Even when understanding the algorithm's process in predicting sentiment, the values still represent predictions that might diverge from human annotations and classifications. While the risk for wrong predictions is relatively small, there is still a loss in quality and explicability compared to human classifications. Additionally, there is some instability involved in fine-tuning machine learning models, such as BERT (cf. e.g. Mosbach, Andriushchenko, and Klakow 2020).

Besides possible instability, which against the model's evaluation performance (cf. table 3.2) seems minimal, an additional limitation of this approach is due to the computational requirements for the training and fine-tuning of large language models. While genre specificity is considered to be a limitation in sentiment analysis tasks, fine-tuning such models is often necessary to account for the domain-specific features of language, such as the language used in Reddit discourse. The computational requirements for fine-tuning often exceed the computational power of a usual computer, as special AI accelerators are necessary for a reasonably fast fine-tuning process. To mitigate this limitation, the model used in this analysis was fine-tuned using Google Colab's cloud computing environment, which largely accelerated fine-tuning. Using the fine-tuned model for generating predictions for the corpus data requires significantly less computing power.

Overall, a mixed-methods approach including a thorough machine learning-assisted statistical analysis of the available data combined with a qualitative analysis seems to provide a solid basis for research on non-standard uses of the progressive. However, with this use being a low-frequency variant of the progressive a massive data set is required for full explanatory power. The general framework of the methodology based on computational and machine learning approaches, can easily be adapted for other linguistic research endeavors and significantly broadens the possibilities for empirical research, especially regarding nuanced uses of language.

6.5 Conclusion

This study has explored the intricate relationship between sentiment intensity and the use of the progressive aspect, highlighting the multifaceted role of the progressive in conveying not just temporal, but also subjective and evaluative meanings. The quantitative analysis provided strong evidence that negative sentiment intensity correlates significantly with the use of the

progressive, supporting theories by Brisard and de Wit (2014) and the foundational work by Wright (1994) that associate the progressive with subjective evaluations.

The qualitative examination of 'always'-type progressives further illustrated how these constructions are employed to express nuanced sentiments, particularly negative evaluations of habitual actions or states. These findings align with the diagnostics proposed by Wright (1994) and supported by Killie (2004), demonstrating that subjective progressives frequently occur in main clauses and are often modified by temporal adverbs. This underlines the expressive capacity of the progressive aspect in marking the speaker's attitude and subjectivity.

Moreover, the broader linguistic implications of this research suggest that the progressive aspect's evolution from a purely temporal marker to a device for expressing subjectivity and modality is part of a larger process of grammaticalisation and subjectification. This aligns with the theoretical perspectives of Traugott (1995), who described the shift in grammatical forms towards more abstract and speaker-oriented functions. The integration of computational methods, such as sentiment analysis, has proven invaluable in quantifying these correlations, providing a robust methodological framework for future linguistic research.

The findings from this study also have practical applications, for instance in language teaching. Understanding the progressive aspect's multiple functionalities can enhance teaching materials and exercises, enabling learners to grasp the subtle nuances of meaning conveyed by grammatical forms. Additionally, the results can be used to improve the accuracy of sentiment analysis models by using such examples in training data.

In conclusion, this research has contributed to a deeper understanding of the progressive aspect's role in language. It has shown that beyond its traditional aspectual function, the progressive is a significant marker of sentiment and subjectivity. Future research should continue to explore these dimensions, employing mixed-methods approaches to fully capture the layered nature of language use. This will ensure that theoretical descriptions and pedagogical practices remain aligned with actual linguistic phenomena, ultimately enriching both academic inquiry and practical language education.

References

Bianchi, Federico, Debora Nozza, and Dirk Hovy. 2021. "FEEL-IT: Emotion and Sentiment Classification for the Italian Language." In *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis,* edited by Orphee De Clercq, Alexandra Balahur, Joao Sedoc, Valentin Barriere, Shabnam Tafreshi, Sven Buechel, and Veronique Hoste, 76–83. Online: Association for Computational Linguistics. https://aclanthology.org/2021.wassa-1.8.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. 1999. *Longman Grammar of Spoken & Written English.* Grammar Reference. London, England: Longman.

Bland, Susan Kesner. 1988. "The Present Progressive in Discourse: Grammar Versus Usage Revisited." *TESOL Quarterly* 22 (1): 53–68. https://doi.org/10.2307/3587061. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2307/3587061.

Brisard, Frank, and Astrid de Wit. 2014. "Modal uses of the English present progressive." In *Core, Periphery and Evidentiality,* edited by Juana I. Marín-Arrese, Marta Carretero, Jorge Arús Hita, and Johan van der Auwera, 201–220. Berlin, Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110286328.201.

Chang, Jonathan P., Caleb Chiam, Liye Fu, Andrew Z. Wang, Justine Zhang, and Cristian Danescu-Niculescu-Mizil. 2020. "ConvoKit: A Toolkit for the Analysis of Conversations," https://doi.org/10.48550/ARXIV.2005.04246.

Collins, Peter. 2008. "The Progressive Aspect in World Englishes: A Corpus-Based Study." *Australian Journal of Linguistics: Journal of the Australian Linguistic Society* 28 (2): 225–249. https://search.ebscohost.com/login.aspx?direct=true&db=mlf&AN=2011933477&site=ehost-live.

de Wit, Astrid, and Frank Brisard. 2014. "A Cognitive Grammar account of the semantics of the English present progressive." *Journal of Linguistics* 50 (1): 49–90. http://www.jstor.org/stable/24583342.

Dennis, Leah. 1940. "The Progressive Tense: Frequency of Its Use in English." *PMLA/Publications of the Modern Language Association of America* 55 (3): 855–865. https://doi.org/10.2307/458746.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." *Preprint: arXiv,* https://doi.org/10.48550/ARXIV.1810.04805.

Hardy, Melissa. 1993. *Regression with Dummy Variables.* SAGE Publications, Inc. https://doi.org/10.4135/9781412985628.

Hatcher, Anna Granville. 1951. "The Use of the Progressive Form in English: A New Approach." *Language* 27 (3): 254–280. http://www.jstor.org/stable/409755.

Hutto, C., and Eric Gilbert. 2014. "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text." *Proceedings of the International AAAI Conference on Web and Social Media* 8 (1): 216–225. https://doi.org/10.1609/icwsm.v8i1.14550.

Killie, Kristin. 2004. "Subjectivity and the English progressive." *English Language and Linguistics* 8 (1): 25–46. https://doi.org/10.1017/s1360674304001236.

Koss, Tom, Astrid de Wit, and Johan van der Auwera. 2022. "The Aspectual Meaning of Non-Aspectual Constructions." *Languages* 7 (2). https://doi.org/10.3390/languages7020143.

Kranich, Svenja. 2013. "Functional layering and the English progressive." *Linguistics* 51 (1): 1–32. https://doi.org/10.1515/ling-2013-0001.

Liu, Bing, et al. 2010. "Sentiment analysis and subjectivity." *Handbook of natural language processing* 2 (2010): 627–666.

Loper, Edward, and Steven Bird. 2002. *NLTK: The Natural Language Toolkit.* https://doi.org/10.48550/ARXIV.CS/0205028.

Madden-Lombardi, Carol, Peter Ford Dominey, and Jocelyne Ventre-Dominey. 2017. "Grammatical verb aspect and event roles in sentence processing." Edited by Sonja Kotz. *PLOS ONE* 12 (12): e0189919. https://doi.org/10.1371/journal.pone.0189919.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. *Efficient Estimation of Word Representations in Vector Space.* https://doi.org/10.48550/ARXIV.1301.3781.

Mosbach, Marius, Maksym Andriushchenko, and Dietrich Klakow. 2020. *On the Stability of Fine-tuning BERT: Misconceptions, Explanations, and Strong Baselines.* https://doi.org/10.48550/ARXIV.2006.04884.

Paulasto, Heli. 2014. "Extended Uses of the Progressive Form in L1 and L2 Englishes." *English World-Wide: A Journal of Varieties of English* 35 (3): 247–276. https://search.ebscohost.com/login.aspx?direct=true&db=mlf&AN=2014303583&site=ehost-live.

Perez, Aveline. 2023. "Time in motion: grammaticalisation of the be going to construction in English," https://doi.org/10.26181/22202026.V1.

Pota, Marco, Mirko Ventura, Rosario Catelli, and Massimo Esposito. 2020. "An Effective BERT-Based Pipeline for Twitter Sentiment Analysis: A Case Study in Italian." *Sensors* 21 (1): 133. https://doi.org/10.3390/s21010133.

Proferes, Nicholas, Naiyan Jones, Sarah Gilbert, Casey Fiesler, and Michael Zimmer. 2021. "Studying Reddit: A Systematic Overview of Disciplines, Approaches, Methods, and Ethics." *Social Media + Society* 7 (2): 205630512110190. https://doi.org/10.1177/20563051211019004.

Taboada, Maite. 2016. "Sentiment Analysis: An Overview from Linguistics." *Annual Review of Linguistics* 2 (1): 325–347. https://doi.org/10.1146/annurev-linguistics-011415-040518.

Taylor, Ann, Mitchell Marcus, and Beatrice Santorini. 2003. "The Penn Treebank: An Overview." In *Text, Speech and Language Technology,* 5–22. Springer Netherlands. https://doi.org/10.1007/978-94-010-0201-1_1.

Traugott, Elizabeth Closs. 1989. "On the Rise of Epistemic Meanings in English: An Example of Subjectification in Semantic Change." *Language* 65 (1): 31. https://doi.org/10.2307/414841.

———. 1995. "Subjectification in grammaticalisation." In *Subjectivity and Subjectivisation,* 31–54. Cambridge University Press. https://doi.org/10.1017/cbo9780511554469.003.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. *Attention Is All You Need.* https://doi.org/10.48550/ARXIV.1706.03762.

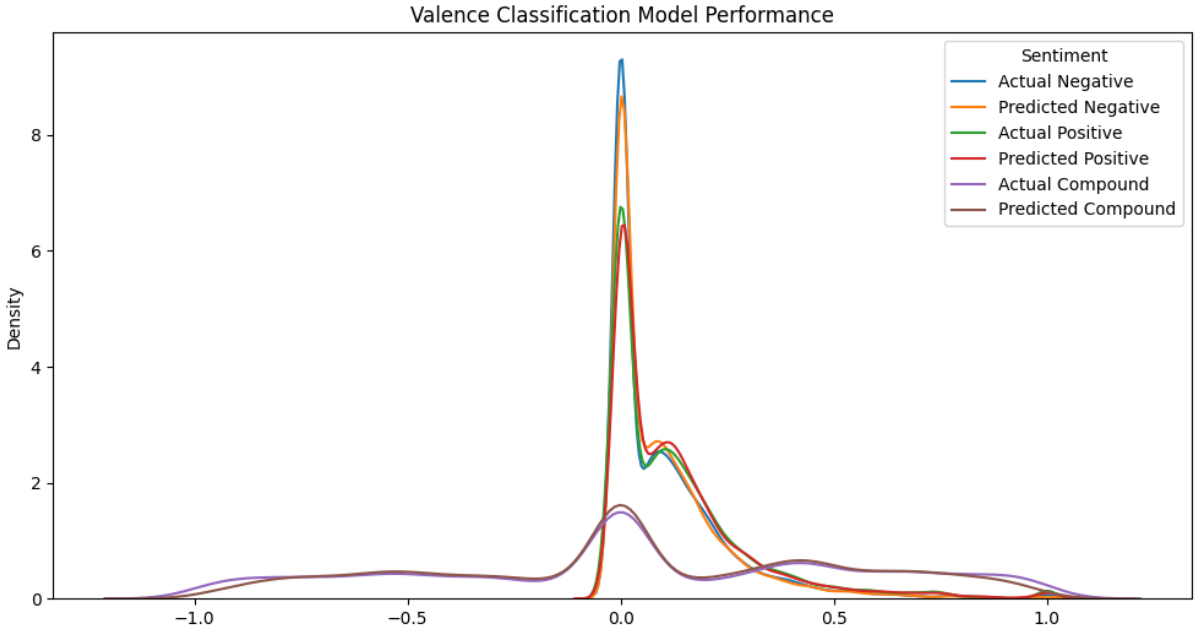
Viola, Lorella. 2023. "On the use of sentiment analysis for linguistics research. Observations on sentiment polarity and the use of the progressive in Italian." *Frontiers in Artificial Intelligence* 6. https://doi.org/10.3389/frai.2023.1101364.

Wright, Susan. 1994. "The mystery of the modal progressive." In *Studies in Early Modern English,* 467–486. DE GRUYTER. https://doi.org/10.1515/9783110879599.467.

List of Tables

List of Figures

Statistical data



KDE density graph showing BERT model performance against adapted VADER scores

Semantic verb categories of present participles

Semantic verb category	***N***	**proportion**	**mean sentiment**
action verbs	2163	56%	0.085
stative verbs	760	20%	0.152
private verbs	602	15%	0.135
uncategorised	364	9%	0.029

compound sentiment value

Regression results for sentiment and clause type / tense

	Clause Type (cl_main_clause)			Tense (t_present)		
Score	**Negative**	**Positive**	**Compound**	**Negative**	**Positive**	**Compound**
Intercept (Int.)	-3.7522***	-3.4746***	-3.6207***	-3.4318***	-3.1990***	-3.3016***
SE	(0.039)	(0.041)	(0.026)	(0.033)	(0.035)	(0.022)
Coefficient (Coeff.)	0.6884***	-0.8281***	-0.1812***	0.7124***	-0.5688***	-0.1036*
SE	(0.164)	(0.169)	(0.044)	(0.141)	(0.142)	(0.038)
Log-Likelihood (LL)	-7207.6	-7203.4	-7207.5	-9193.3	-9197.1	-9201.7
Pseudo R-squared	0.001163	0.001750	0.001178	0.001326	0.000909	0.000408
LLR p-value	4.179e-05	5.016e-07	3.739e-05	7.788e-07	4.300e-05	0.006160

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Regression results for sentiment and semantic verb categories

	Category Action			Category Private			Category Stative		
Score	**Negative**	**Positive**	**Compound**	**Negative**	**Positive**	**Compound**	**Negative**	**Positive**	**Compound**
Intercept (Int.)	-3.3436***	-3.1284***	-3.2846***	-4.6108***	-4.5652***	-4.6064***	-4.4031***	-4.3124***	-4.3785***
SE	(0.033)	(0.035)	(0.022)	(0.061)	(0.065)	(0.042)	(0.055)	(0.058)	(0.038)
Coefficient (Coeff.)	0.3022*	-0.8478***	-0.0935*	0.0698	-0.1688	0.0556	0.2407	-0.2538	0.1099
SE	(0.146)	(0.144)	(0.037)	(0.279)	(0.253)	(0.071)	(0.244)	(0.227)	(0.063)
Log-Likelihood (LL)	-9325.6	-9309.4	-9324.6	-3373.9	-3373.7	-3373.6	-4080.8	-4080.6	-4079.8
Pseudo R-squared	0.0002260	0.001956	0.0003331	9.241e-06	6.682e-05	9.191e-05	0.0001171	0.0001554	0.0003718
LLR p-value	0.04002	1.537e-09	0.01267	0.8028	0.5019	0.4310	0.3282	0.2600	0.08149

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$