electronics

Article

Emotion Recognition Based on the Structure of Narratives

Tibor Pólya [1,2,*] and István Csertő [2]

[1] Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, 1117 Budapest, Hungary
[2] Institute of Psychology, Károli Gáspár University of the Reformed Church in Hungary, 1037 Budapest, Hungary
* Correspondence: polya.tibor@ttk.hu

Abstract: One important application of natural language processing (NLP) is the recognition of emotions in text. Most current emotion analyzers use a set of linguistic features such as emotion lexicons, *n*-grams, word embeddings, and emoticons. This study proposes a new strategy to perform emotion recognition, which is based on the homologous structure of emotions and narratives. It is argued that emotions and narratives share both a goal-based structure and an evaluation structure. The new strategy was tested in an empirical study with 117 participants who recounted two narratives about their past emotional experiences, including one positive and one negative episode. Immediately after narrating each episode, the participants reported their current affective state using the Affect Grid. The goal-based structure and evaluation structure of the narratives were analyzed with a hybrid method. First, a linguistic analysis of the texts was carried out, including tokenization, lemmatization, part-of-speech tagging, and morphological analysis. Second, an extensive set of rule-based algorithms was used to analyze the goal-based structure of, and evaluations in, the narratives. Third, the output was fed into machine learning classifiers of narrative structural features that previously proved to be effective predictors of the narrator's current affective state. This hybrid procedure yielded a high average F1 score (0.72). The results are discussed in terms of the benefits of employing narrative structure analysis in NLP-based emotion recognition.

Keywords: emotion recognition; emotion lexicon; narrative structure; machine learning

check for updates

Citation: Pólya, T.; Csertő, I. Emotion Recognition Based on the Structure of Narratives. *Electronics* **2023**, *12*, 919. https://doi.org/10.3390/electronics12040919

Academic Editors: Oscar Araque, Lorenzo Gatti, Álvaro Carrera Barroso and Kyriaki Kalimeri

Received: 28 December 2022
Revised: 8 February 2023
Accepted: 8 February 2023
Published: 11 February 2023

1. Introduction

Emotion recognition in text is a highly important research field in Natural Language Processing (NLP). There are various reasons why emotion recognition is an important task. One of them is that emotion is a significant factor in mental functioning, and, consequently, it has a role in cognitive functioning (e.g., decision making), political commitment, intergroup relations, communicating with a chatbot, and forming sentiments. While NLP-based emotion analysis in these fields would generally require an assessment of text producers' self-rated emotional experiences to ensure adequate validity of the findings, the related empirical studies mostly use external criteria. For example, Bostan and Klinger [1] recently reviewed 14 emotion corpora, only one of which was found to have emotion annotations obtained from the text producers themselves [2], while self-labeling (also called distant supervision) was used to compile another corpus [3], in which tweets were annotated with the associated hashtags describing an emotion category (e.g., #joy). In the remaining 12 cases, emotions were annotated by experts or crowdsourced readers. Importantly, however, Buechel and Hahn [3] conducted a comparative quality assessment of text producers' vs. readers' annotations of the same texts, which revealed that the former had better quality than the latter in terms of both inter-annotator agreement and distribution of annotations. The present study aimed to contribute to the apparently less popular line of research focused on a valid NLP-based assessment of text producers' *subjective* emotional experiences. In this study, participants were asked to recount two emotional episodes, one positive and one negative. The aim of this study was to test whether participants' emotions could be recognized from the structure of their narratives on those emotional experiences.

2. Related Work

Emotion recognition in text is commonly defined as a task of text classification. The existing methods are based either on lexicons or on machine learning.

2.1. Lexicon-Based Methods

Lexicon-based classification typically relies on dictionaries of emotion labels and emotionally charged words. Illustrative examples are the Linguistic Inquiry and Word Count (LIWC) [4] and the NRC (National Research Council Canada) Valence, Arousal, and Dominance Lexicon (Mohammad 2018a) [5]. The LIWC has dictionaries of positive and negative emotions (337 and 618 words, respectively). Negative emotions comprise three subcategories such as anxiety, anger, and sadness. The latest version of the LIWC [6] also has dictionaries of positively and negatively toned words (102 and 1530 words, respectively). The LIWC was compiled by expert annotators, while the NRC Valence, Arousal, and Dominance Lexicon by crowdsourcing. This latter has around 20,000 words, whose annotation was based on the three-dimensional model of emotions including arousal, valence, and dominance [5].

Lexicon-based methods are widely used in emotion recognition due to their transparency and straightforward applicability. Since the compilation of emotion lexicons is based on theoretical considerations concerning separate words, such lexicons have the advantage of versatile application across various text types and domains. However, lexicon-based methods sometimes have limited coverage of relevant words and thus limited applicability to large datasets [7].

2.2. Machine Learning Methods

Machine learning methods improve the performance of lexicon-based methods by widening the feature set of linguistic analysis (e.g., word n-grams, character n-grams, word embeddings, affect lexicons, negation, punctuation, emoticons, or hashtags [1]), and by their capacity to infer decision rules on which emotion recognition is based. While previous studies experimented with various models of emotion recognition such as Random Forest [8] and Support Vector Machine [9], the state-of-the-art models of emotion classification are based on deep learning methods, including Convolutional Neural Network [10], Long-Short Term Memory [11], and BERT [12].

Due to the expanded set of linguistic features, machine learning methods usually have high coverage. However, since machine learning methods are trained with one particular corpus, and thus they are adjusted to the specificities of that corpus, their performance is less generalizable across different corpora [1]. Bostan and Klinger [1], for example, compared the quantitative similarity across the 14 emotion corpora by selecting the 5000 most common words from each corpus and calculating their cosine similarity measure. They found that it varied between 0.48 and 0.96. It indicates that there are domains that are similar to each other (e.g., blogs and tweets). However, there are also dissimilar domains (e.g., news headlines and tweets). Furthermore, recent research also indicates that there could be significant differences in styles among texts coming from the same domain. Recently, there has been active research on how style differences can be transformed by NLP methods [13–15].

It is an observable general trend in NLP-based emotion recognition that machine learning methods outperform lexicon-based methods in terms of F1 score, even though the related studies use different corpora and different emotion category systems. For example, one study [7] used the LIWC emotion dictionaries to analyze about 664 thousand tweets for eight basic emotion categories (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust), obtaining a relatively low performance level (F1 = 0.35). By comparison, Potapova and Gordeev [8] analyzed 10,000 sentences from movie reviews coded as either positive or negative. They compiled a set of words and phrases expressing aggression and, using the Random Forest method, obtained a substantially higher performance level (F1 = 0.58). Mohammad [9] developed a lexicon for Ekman's six basic emotion categories (anger, disgust, happiness, sadness, fear, and surprise) based on the occurrences of emotion category words as tweet hashtags. To evaluate the lexicon, he analyzed 1000 headlines and used the Support Vector Machine model. He observed somewhat lower performance level (F1 = 0.44). The state-of-the-art machine learning models usually achieve even higher performance. Deriu et al., [16] for example, used the Convolutional Neural Network model to analyze 300 million tweets coded for valence (positive, negative, or neutral), reaching higher performance (F1 = 0.62). Felbo et al. [11] analyzed 1.2 billion tweets for 64 categories using the Long-Short Term Memory distant supervision method, also observing slightly higher performance levels (F1s = 0.69 to 0.75). Finally, Cortiz [17] classified 58 thousand Reddit comments under 27 emotion categories with the BERT model and obtained a low performance level (F1 = 0.46).

2.3. Study: Emotion Recognition Based on Narrative Structure Analysis

The present study tested the effectiveness of narrative structure analysis in NLP-based emotion recognition, which is the first attempt in the field, to the authors' knowledge. The rationale behind this innovative approach is that narrative is a text genre commonly used to describe and express emotions [18]. A narrative comprises a set of temporally and logically connected actions [19]. Although most approaches focus on the role of narrative in knowledge representation (e.g., [20]), it can be argued that narrative is also related to emotions. Some authors even argue that emotions and narratives have a homologous structure [21–23]. The present study focuses on two homologous structural features of narratives and emotions, which are defined in the context of a popular distinction between the three levels of a narrative, including the story, which concerns the events recounted, the discourse, which concerns how the events are recounted, and narration, which concerns the situation in which the events are recounted [24].

The first homologous structural feature is a goal-based structure, which can be described at the level of discourse. According to the appraisal theory of emotion (e.g. [25]), an emotional response is elicited by appraisal of an event. Appraisal is based, to a considerable extent, on how the event influences goal attainment. Events supporting vs. hindering goal attainment are appraised as positive vs. negative events, respectively. Similarly, the actions comprising a narrative are selected, connected, and evaluated according to their relationship with a goal pursued by the protagonist. In this perspective, a narrative begins with an event that corrupts or contaminates the status quo. In response to the altered state of affairs, the protagonist sets a new goal, makes plans for goal achievement, and takes action to achieve the goal. The chain of actions terminates when the goal is achieved or changed. Arguing for the common goal-based structure of emotions and narratives, Stein and Hernandez [23] suggest that mapping a goal-based narrative structure onto emotional processes provides a testable model of the relationship between cognitive appraisal and emotional responses.

The second homologous structural feature is evaluative structure, which can be described at the level of narration. As it was mentioned, narratives may contain descriptions of past appraisal processes. However, when past events are recounted verbally, the person may reappraise the past events. Correspondingly, the narrator evaluates the narrative events during narration. Evaluation is carried out by adding an evaluative structure to the events included in the narrative [18]. The successful sharing of a narrative requires the narrator to use linguistic means of evaluation, by which they clarify the significance of the events for their audience. By telling a narrative lacking in evaluation, the narrator risks the loss of their audience's interest. However, recent research shows that narratives elicit affect not only in their audience [26], but also in their narrators [22].

2.4. Natural Language Processing and Narrative Analysis

Recent research in Natural Language Processing (NLP) focuses increasing attention on the analysis of narrative structure [for a review, see 21]. The related NLP developments are targeted at diverse objectives including the detection of various narrative features such as agents and their relationships [27], plotline [28], and temporality [29], to only mention a few examples.

However, emotion recognition methods are rarely designed specifically for narrative texts. One of the few exceptions is the NARCODER [20], which reveals the goal-based structure of narratives by coding keywords and multiword expressions. Other studies focus on the distribution of emotion words in narratives. Alm and Sproat [30] analyzed 23 Grimms' fairy tales. The researchers manually coded a small set of emotions (angry, disgusted, fearful, happy, sad, positive and negative surprise) in each sentence of the texts. They found that the frequency of positive, negative, and neutral emotions followed a wave-shaped trajectory in the tales, each segmented into five units of equal length. The trajectory can be explained by considering the goal-based structure of narratives. In the first unit, there is low emotion intensity as the narrator describes the initial status quo. In the second unit, there is a peak of negative emotions, probably due to the precipitating event. Finally, in the last unit, there is a peak of positive emotions, probably due to goal achievement. In a similar vein, Boyd and his colleagues [31] analyzed a large and diverse set of narratives. They found a universal wave-shaped pattern of cognitive tension in the narratives, which were also segmented into five units each. The use of cognitive process words increased in the middle of the narratives (between the second and fourth units). This pattern can also be explained by the goal-based structure of narratives, since the increased frequency of cognitive process words is consistent with the process of working through the difficulties of goal achievement.

However, it is important to note that studies on the automatic generation of narratives, which focus on the suspense elicited in the audience, consider narratives and emotions as closely related to each other [32].

The present study proposes a novel approach to emotion recognition, which utilizes measurable basic properties of overall narrative structure to assess text producers' affective states, in contrast with existing approaches focused on the distribution of specific emotion terms (as opposed to non-emotion terms making up the bulk of the analyzed text). The structural properties utilized in the present study are elements of goal-based structure and evaluative structure, which are discussed in detail below.

2.4.1. Analysis of Goal-Based Structure: Narrative Transformation

Narratives having a goal-based structure provide ample insight into the mental realm of the protagonist and other characters. This is because a narrative having a goal-based structure describes not only the goal of the protagonist but its subjective reactions to the chain of narrative actions as well. This aspect of the narrative is aptly grasped by Jerome Bruner, who notes that "narrative deals with the vicissitudes of human intentions" [33] (p. 16). One way to detect the mental realm in a narrative is based on the differentiation between the landscape of action and the landscape of consciousness [30]. While events described objectively pertain to the landscape of action, the same events can be described subjectively, that is, in relation to the consciousness of a character. An action can be transferred from the landscape of action to the landscape of consciousness by using narrative transformations. An early study [34] describes 12 categories of narrative transformations (see Table 1 for definitions and examples). Narrative transformation can be achieved by several linguistic operations, some of them using single words, while others adding a new clause to the description of an action.

Table 1. Definitions and examples for narrative structural features (Hungarian lexicon examples are added in square brackets).

Narrative Structural Features	Definition	Example (Lexicon and Sentence)
Narrative transformation		
Mode	Expression of possibility, impossibility, necessity, or prohibition of an action.	will, must, used to [fog, kell, szokott] I will open the book.
Intention	The intention to perform an action.	want, decide, goal [akar, dönt, cél] I wanted to go there.
Result	Action presented as already accomplished.	manage, achieve, prevent [sikerül, elér, megakadályoz] I managed to get to the station on time.
Manner	Specification of the manner in which an action occurs, or expression of its intensity.	adverb He firmly pointed out the flaws.
Aspect	Expression of the temporal contour of an action.	start, finish, bring up [kezd, befejez, megállít] I started to get into running.
Status	Negation of the action	not, never, without [nem, soha, nélkül] I did not make the mess.
Appearance	Indication of the replacement of one event by another.	seem, bewilder, pretend [tűnik, megtéveszt, színlel] It seemed to stop raining.
Knowledge	Description of awareness of the action.	understand, contrive, feel [megért, kitalál, érez] I understood what he was doing.
Description	Description of an act of communication.	call, describe, chat [hív, leír, beszélget] I called my sister.
Supposition	Description of anticipation of a future action.	expect, tomorrow, then [vár, holnap, majd] I expected something different.
Subjectivation	Attribution of the action, as an object of observation, to a subject.	remember, consider, doubt [emlékszik, fontol, kételkedik] I remember vividly.
Attitude	Description of the state elicited in the subject by the action.	wonder, laugh, enjoy [csodálkozik, kinevet, élvez] I wonder why you are here.
Narrative evaluation		
Comparative	Comparison between any aspects of two narrative events.	Comparative and superlative adjective, conjunction than [mint] I was happier this morning than yesterday.
Quantifier	Expression of the quantity of any aspect of an event included in a narrative.	all, some, enough [összes, egész, elég] All my hopes are gone.
Qualification	Emphasis added to the description of any aspect of a narrative event.	Primary adjective It was a happy day.
Explanation	Insertion of unknown information in the narrative.	so, consequently, because [ezért, következésképp, mert] So now they have come to us.

2.4.2. Analysis of Evaluative Structure: Narrative Evaluation

Another early study [18] provides a detailed description of those linguistic features that can be used to evaluate narrative events. These linguistic features increase the complexity of the otherwise simple syntax of narrative clauses. This study deals with four subtypes of narrative evaluation as follows. *Quantifiers* are a subtype of intensifiers, which express the quantity of any aspect of an event included in a narrative. *Comparatives* are a subtype of comparators, which make a comparison between any aspects of two narrative events. Finally, *qualifications* and *explanations* are two subtypes of explicatives. Qualifications locate events on the dimension of valence defined by negative and positive endpoints, while explanations insert some information in the narrative that helps the appropriate interpretation of any aspect of a narrative event unfamiliar to the audience (see Table 1 for definitions and examples).

The present study tested a novel strategy of NLP-based emotion recognition based on linguistic features of the goal-based structure and evaluative structure of narratives. Considering these linguistic features as the most basic and thus ubiquitous building blocks of narrative texts, as opposed to emotion lexicons showing sporadic occurrences and highly variable frequencies within and across narratives, the proposed structural measures were expected to have higher coverage than emotion lexicons.

On the basis of the theoretical considerations suggesting that emotions and narratives have a homologous goal-based structure and evaluative structure, it was hypothesized that NLP-based measures of these two narrative structural features would reveal the narrator's current affective state, thereby contributing to NLP-based emotion recognition.

The hypotheses were tested in an empirical study.

3. Materials and Methods

Participants

The study involved 117 university students (85 females and 32 males) aged 18 to 28 years ($M = 24.2$, $SD = 2.8$). All participants were native Hungarian speakers, and they took part in the study voluntarily.

Measures

The participants' affective state during narration was assessed with the Affect Grid [35], which is a brief yet sensitive measure of short-term changes in one's affective state. The Affect Grid has a single item presented in a 9×9 grid format, which represents affective states in the two-dimensional space of Russell's [35] model. The respondent marks the cell that best indicates their current affective state. Affect descriptors placed at each corner and at the midpoint on each side support the respondent's orientation in the grid. The psychometric properties of the measure indicate good convergent validity and adequate interrater reliability.

Procedure

Every participant wrote an autobiographical narrative about each of two topics, including losing some property and finding someone else's lost property. The order of presentation of the narrative writing tasks was counterbalanced. Immediately after completing each narrative (i.e., twice during the procedure), the participants indicated their affective state using the Affect Grid.

Analysis of Narratives

First, the magyarlanc linguistic parser [36] was used to preprocess the text of the narratives. The magyarlanc annotates Hungarian texts for token boundaries, part-of-speech (POS) categories, lemmata, and morphological features. Stopwords were not excluded from the analysis.

Second, the narratives were analyzed with a set of rule-based algorithms for the 12 narrative transformations and four types of narrative evaluation described above. The analysis was based on customized lexicons and occasionally part-of-speech categories (see Table 1 for examples of lexicon and sentence context).

The rule-based algorithms were tested for reliability by comparing the output of machine analysis with that of human coding obtained for a sample of 23 narratives. The F1 scores computed for the 16 algorithms indicated that each algorithm was adequately reliable (F1 > 0.80 in all cases, which is a commonly accepted criterion level in psychological text analysis; (see, e.g., [37])).

4. Results

Each of the 117 participants wrote two narratives, thus the overall corpus comprised 234 narratives (total sentence count = 1583; total word count = 22,019). The word count of each narrative ranged from 22 to 404 ($M = 94.1$, $SD = 64.9$).

Frequency data for the 12 narrative transformations, four types of narrative evaluations and emotion lexicons are presented in Table 2. While each structural feature showed zero frequency in one or more narratives, a repeated measures ANOVA test indicated a significant difference between the average frequencies of the three features ($F(1.139,265.296) = 324.1$, $p < 0.001$). Follow-up paired *t*-tests indicated that the participants used narrative transformations significantly more frequently than narrative evaluations ($t(233) = 16.7$; $p < 0.001$), and, furthermore, they used narrative evaluations significantly more frequently than emotion lexicons ($t(233) = 16.9$; $p < 0.001$).

Table 2. Frequency of emotion lexicons, narrative transformations, and evaluations.

Variables	Frequency			
	Minimum	**Maximum**	***M***	***SD***
Emotion lexicons	0	10	2.0	1.8
Narrative transformations	0	141	24.7	19.7
Narrative evaluations	0	54	11.3	9.3

To assess the potential of the narrative transformation and narrative evaluation measures for emotion recognition, the absolute frequencies obtained for each structural feature were normalized and then fed into a machine learning procedure. Regarding the narrator's affective state, the dimensional data obtained with the two nine-point scales of the Affect Grid (i.e., valence and arousal) were transformed into grouping variables for the purpose of the classification task. On the one hand, the overall corpus of 234 narratives was divided into low vs. high arousal groups using the median as the cutpoint. On the other hand, each narrative was assigned to one of the following four groups: high positive, low positive, low negative, and high negative valence. The narratives eliciting positive vs. negative affective states were used separately due to known differences in the structure of narratives about positive vs. negative experiences [38]. Of all narratives, 80% were used to train the machine learning algorithm, and 20% for testing. The *k*-nearest neighbors (*k*-NN) method was used for machine learning. Distances were calculated with the overlap metric. Testing different values for K shows that the best value for K in our application is 5. For validation, a *k*-fold cross-validation method was used where $k = 10$. The entire analytic process was repeated on the TD-IDF vectorized texts. The contribution of narrative features to emotion recognition was assessed with the performance measures of recall, precision, accuracy and F1 score. Table 3 presents the obtained performance measures, which indicate the power of the algorithms for predicting the narrator's arousal level and the valence of their positive or negative affective state.

As baseline, we present a model that employs emotion lexicons as the predictors of the narrator's affective state. Lexical emotion categories were analyzed with the emotion dictionaries of the Narrative Categorical Content Analysis system (NarrCat) [39], which contain words and expressions marking positive and negative emotions. We used these lexicons because psychological text analysis generally relies on lexicon-based methods (see, e.g., LIWC [4]). As presented in Table 3, the baseline model showed moderate and fluctuating power for predicting the narrator's affective state. The baseline model had the highest predictive power for positive valence, followed by that for arousal and for negative valence. The employed *k*-NN method used with normalized data produced somewhat higher performance levels than the *k*-NN method used with TD-IDF vectorized data.

Table 3. Performance measures of the models (R: Recall, P: Precision, A: Accuracy, and F1 score).

Classifier	Predictor Variables	Arousal				Positive Valence				Negative Valence			
		R %	P %	A %	**F1 %**	R %	P %	A %	**F1 %**	R %	P %	A %	**F1 %**
k-NN	Emotion lexicons	54.7	54.6	55.1	**54.6**	66.5	66.5	66.5	**66.7**	50.0	29.3	58.5	**36.9**
	Narrative transformations	73.7	75.2	75.4	**74.4**	70.8	72.2	70.3	**71.5**	70.9	71.9	69.2	**71.4**
	Narrative evaluations	61.4	63.7	63.2	**62.5**	58.8	59.9	58.4	**59.3**	62.5	64.8	62.2	**63.6**
$k\text{-NN}_{\text{TF-IDF}}$	Emotion lexicons	55.2	55.4	52.6	**55.3**	64.3	65.3	63.2	**64.8**	50.0	29.3	58.5	**36.9**
	Narrative transformations	67.2	67.7	67.6	**67.4**	63.6	64.3	64.1	**63.9**	65.6	65.1	66.2	**65.3**
	Narrative evaluations	52.8	52.3	54.1	**52.5**	52.2	53.7	51.8	**52.9**	51.3	52.1	58.2	**51.7**

The model with narrative transformations as the predictor showed the best performance in emotion recognition, having equally high predictive power for all three measures of affective state (see Table 3).

Finally, the model with narrative evaluations as the predictor showed performance levels comparable to those of the baseline model. The predictive power of narrative evaluations was equal for negative valence and arousal and slightly lower for positive valence.

5. Discussion

The present study tested a new strategy of NLP-based emotion recognition utilizing narrative structure rather than emotion lexicons. The results confirmed the success of the proposed strategy. Frequency data for narrative structural features produced higher coverage compared to those for emotion lexicons. This difference was due to treating the tested linguistic features as structural dimensions of narratives.

The study also found that narrative transformations performed better than emotion lexicons in predicting the narrator's affective state. Furthermore, the model with overall frequency of narrative transformations as the predictor is comparable to several alternative predictive models [8,9], and it is very close to the performance level of the state-of-the-art deep learning methods [10–12]. That is, the model meets the general trend in NLP-based emotion recognition that machine learning methods, and especially state-of-the-art models, outperform lexicon-based methods in terms of F1 score. The predictive power of narrative evaluations was comparable to that of emotion lexicons. One possible reason for the relatively low performance levels obtained for narrative evaluations lies in the communicative context of the study. Specifically, the participants were not presented with any information about the prospective reader or readers of their autobiographical narratives, thus they recounted their memories in an impersonal rather than interactive setting, which might result in less elaborate reappraisals of past events.

This latter finding is in line with the view that emotions and narratives have a homologous structure [21–23], which has several implications for future research. First, it significantly expands the range of those linguistic features that are potentially useful in emotion recognition, which is an important advantage of the proposed approach. Since narrative structure is rather complex, many narrative structural features can be tested for their value in emotion recognition. Besides goal-based structure [23] and narrative evaluation, Habermas [21] also enumerates actions and normality. The inclusion of these homologous structural features may further improve the effectiveness of NLP-based emotion recognition.

Second, it suggests that narrative structural features have promising potential for assessing text producers' affective states without using external criteria (e.g., self-report measures, independent observers' judgments), which may be especially useful in cases when emotion recognition is motivated by an interest in psychological processes.

One important aspect of the evaluation of an algorithm considers its generalizability across different domains and styles [13–15]. The tested narrative-based emotion recognition procedure can be expected to have high generalizability for at least two reasons. First, although not all of the diverse text types previously sampled in test corpora are typically characterized by a proper narrative structure (e.g., blogs, tweets, social media posts and comments, news articles, fairy tales), they all contain some of the narrative transformations and evaluations on which the analytical procedure employed in the present study was focused. Second, the present study adopted a dimensional conceptualization of affect proposed by Russell and his colleagues [35], according to which the two dimensions of affect are arousal–sleepiness and pleasure–displeasure. This dimensional approach is more consistent with a narrative-structure-based strategy of emotion recognition compared to a categorical approach, since narratives usually contain occurrences of diverse emotion categories. A dimensional approach offers better manageability of emotional diversity and better generalizability of findings obtained for narratives about different events.

Finally, the obtained findings broaden the scope of utilizing narrative structure in NLP research, in which field narratives provide an important subject of empirical research on knowledge representation [20], while the present study demonstrates their value in the investigation of emotions, which is another important advantage of the proposed approach.

The findings demonstrate that a hybrid lexicon- and machine-learning-based analysis of narrative structure is feasible. The procedure employed in the present study combined lexicon-based analysis with machine learning algorithms to recognize emotions. These two levels of analysis were integrated into an intermediate-level procedure where lexicons and other relevant linguistic features such as part-of-speech categories and morphological properties were used to implement the linguistic operationalization of narrative structural features. Due to the complexity of these features, their linguistic operationalization was implemented by a set of rule-based algorithms.

Furthermore, the results contribute to a better understanding of the emotionality of narratives. According to the appraisal theory [25], emotions are dependent on appraisal processes. Emotional appraisal is indicated by evaluations in narratives [23]. However, the argument for the homologous structure of emotions and narratives implies that there are two layers of appraisal when a person recounts past events: the past appraisal of the events and the reappraisal of the events during narration. These two layers of appraisal can be related to the protagonist and the narrator, respectively. The narrative approach adopted in the present study points out that autobiographical narratives integrate these two layers of emotions during narration. Studies of narrative generation also distinguish between these two layers of emotions by making a distinction between suspense elicited by the protagonist and suspense elicited by the narrative [32].

The present study has several limitations. One of them is that the findings were obtained for a relatively small and homogeneous sample of young adults. The small sample size calls for caution when interpreting the obtained performance measures, since it might result in increased classification accuracy [40] and biased F1 scores as well. Nonetheless, a factor mitigating the potential effects of the small sample size is the low dimensionality of the data, which was ensured by a hybrid method that enabled us to use a feature set limited to only 16 dimensions (including 12 types of narrative transformations and four types of narrative evaluation). As a result, a low feature-to-sample ratio was obtained, which indicates, according to Vabalas et al. [40], that the tested models are unlikely to be substantially overfitted. Furthermore, a notable merit of the dataset against its small size is that it is a result of a high-cost data collection process. The reported findings concern the predictive relationship between structural features of personal narratives and self-reported data on the narrators' affective states at the time of narration, with which no comparable findings, obtained with a hybrid method, are available to our knowledge. Another limitation is posed by the topical homogeneity of the analyzed corpus, which is due to collecting narratives in response to two closely related topics. A further limitation is that while the proposed high generalizability of the obtained findings relies on the assumption of a universal narrative structure of diverse text genres, the study provided no related empirical data. Future research should examine whether and to what extent the results of the study could be generalized across different text types and domains frequently used in studies of NLP-based emotion recognition.

6. Conclusions

The present study demonstrates the utility of narrative structure analysis in NLP-based emotion recognition. Furthermore, the obtained findings underline the importance of further empirical studies on the homologous structural properties of emotions and narratives, which may foster progress in both theoretical and applied research on emotion recognition, especially regarding studies motivated by an interest in text producers' psychological processes.

Further research efforts are planned to be made in the following directions: (1) using larger sample of narratives; (2) direct comparisons between the proposed narrative-structure-based model and other machine learning methods on the same corpus; (3) assessment of the generalizability of the proposed model across different text types (e.g., blogs, public communications via social media and social networking sites, news articles, online costumer reviews), thereby possibly gaining important insights into the methodological implications of the diversity of domain and language style [13–15]; (4) development of an English version of the model; (5) employing the model in the exploration of emotional contagion processes in personal and internet-mediated communication, which may provide valuable empirical insights into group formation processes and the dynamics of diverse groups (e.g., of customers, voters, or citizens).

Author Contributions: Conceptualization, T.P. and I.C.; methodology, T.P.; software, T.P.; formal analysis, T.P.; writing—original draft preparation, T.P.; writing—review and editing, I.C.; funding acquisition, T.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Research, Development and Innovation Office, grant number K124206.

Data Availability Statement: The data presented in this study are openly available in FigShare at https://doi.org/10.6084/m9.figshare.21788342.v1.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Bostan, L.A.N.; Klinger, R. An Analysis of Annotated Corpora for Emotion Classification in Text. In Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe, NM, USA, 20–26 August 2018; Association for Computational Linguistics: Santa Fe, NM, USA, 2018; pp. 2104–2119.
2. Scherer, K.R.; Wallbott., H.G. Evidence for universality and cultural variation of differential emotion response patterning. *J. Personal. Soc. Psychol.* **1994**, *66*, 310–318. [CrossRef] [PubMed]
3. Buechel, S.; Hahn, U. Readers vs. writers vs. texts: Coping with different perspectives of text understanding in emotion annotation. In *Proceedings of the 11th Linguistic Annotation Workshop*; Association for Computational Linguistics: Valencia, Spain, 2017; pp. 1–12.
4. Pennebaker, J.W.; Francis, M.E.; Booth, R.J. *Linguistic Inquiry and Word Count: LIWC 2001*; Lawrence Erlbaum Associates: Mahway, NJ, USA, 2001; p. 71.
5. Mohammad, S.M. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. In Proceedings of the Annual Conference of the Association for Computational Linguistics (ACL), Melbourne, Australia, 15–20 July 2018.

6. Boyd, R.L.; Ashokkumar, A.; Seraj, S.; Pennebaker, J.W. *The Development and Psychometric Properties of LIWC-22*; University of Texas at Austin: Austin, TX, USA, 2022.
7. Saravia, E.; Liu, H.C.T.; Huang, Y.H.; Wu, J.; Chen, Y.S. CARER: Contextualized affect representations for emotion recognition. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; Association for Computational Linguistics: Brussels, Belgium, 2018; pp. 3687–3697.
8. Potapova, R.; Gordeev, D. Detecting state of aggression in sentences using CNN. In *Speech and Computer: 18th International Conference, SPECOM 2016, Budapest, Hungary, August 23-27, 2016, Proceedings 18*; Springer: Cham, Switzerland, 2016; pp. 240–245.
9. Mohammad, S. #emotional tweets. In Proceedings of the *SEM 2012: The First Joint Conference on Lexical and Computational Semantics, Montréal, Canada, 7–8 June 2012; Association for Computational Linguistics: Montreal, Canada, 2012; pp. 246–255.
10. Abdul-Mageed, M.; Ungar, L. Emonet: Fine-grained emotion detection with gated recurrent neural networks. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Vancouver, BC, Canada, 30 July–4 August 2017; pp. 718–728.
11. Felbo, B.; Mislove, A.; Søgaard, A.; Rahwan, I.; Lehmann, S. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. *arXiv* **2017**, arXiv:1708.00524.
12. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
13. Che, W.; Zhang, Y. *Deep Learning in Lexical Analysis and Parsing. Deep Learning in Natural Language Processing*; Deng, L., Liu, Y., Eds.; Springer Nature: Singapore, 2018; pp. 79–116.
14. Li, D.; Zhang, Y.; Gan, Z.; Cheng, Y.; Brockett, C.; Sun, M.; Dolan, W.B. Domain adaptive text style transfer. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Hong Kong, China, 3–7 November 2019.
15. Savchenko, E.; Lazebnik, T. Computer aided functional style identification and correction in modern russian texts. *J. Data Inf. Manag.* **2022**, *4*, 25–32. [CrossRef]
16. Deriu, J.; Lucchi, A.; De Luca, V.; Severyn, A.; Müller, S.; Cieliebak, M.; Hofmann, T.; Jaggi, M. Leveraging large amounts of weakly supervised data for multi-language sentiment classification. In Proceedings of the 26th International Conference on World Wide Web, Perth, Australia, 3–7 April 2017; pp. 1045–1052.
17. Cortiz, D. Exploring transformers in emotion recognition: A comparison of bert, distillbert, roberta, xlnet and electra. *arxiv* **2021**, arXiv:2104.0204. Available online: https://www.semanticscholar.org/paper/Exploring-Transformers-in-Emotion-Recognition%3A-a-of-Cortiz/5d63495e669a410770373c5017a6738feec74141 (accessed on 10 February 2023).
18. Labov, W. The transformation of experience in narrative syntax. In *Language in the Inner City*; Labov, W., Ed.; Blackwell: Oxford, UK, 1972; pp. 354–396.
19. László, J. *The Science of Stories. An Introduction to Narrative Psychology*; Routledge: London, UK, 2008.
20. Akimoto, T. Cogmic space for narrative-based world representation. *Cogn. Syst. Res.* **2021**, *65*, 167–183. [CrossRef]
21. Habermas, T. *Emotion and Narrative. Perspectives in Autobiographical Storytelling*; Cambridge University Press: New York, NY, USA, 2019.
22. Pólya, T. Temporal structure of narratives reveals the intensity of the narrator's current affective state. *Curr. Psych.* **2021**, *40*, 281–291. [CrossRef]
23. Stein, N.L.; Hernandez, M.V. Assessing understanding and appraisals during emotional experience. In *Handbook of Emotion Elicitation and Assessment*; Coan, J., Allen, J., Eds.; Oxford University Press: Oxford, UK, 2007; pp. 298–317.
24. Piper, A.; So, R.J.; Bamman, D. Narrative theory for computational narrative understanding. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Punta Cana, Dominican Republic, 7–11 November 2021; Association for Computational Linguistics: Punta Cana, Dominican Republic, 2021; pp. 298–311.
25. Scherer, K.R. The dynamic architecture of emotion: Evidence for the component process model. *Cogn. Emot.* **2009**, *23*, 1307–1351. [CrossRef]
26. Oatley, K. Why fiction may be twice as true as fact: Fiction as cognitive and emotional simulation. *Rev. Gen. Psychol.* **1999**, *3*, 101–117. [CrossRef]
27. Labatut, V.; Bost, X. Extraction and analysis of fictional character networks: A survey. *ACM Comput. Surv.* **2019**, *52*, 1–40. [CrossRef]
28. Elsner, M. Abstract representations of plot structure. *Linguist. Issues Lang. Technol.* **2015**, *52*, 89. [CrossRef]
29. Yauney, G.; Underwood, T.; Mimno, D. Computational prediction of elapsed narrative time. In Proceedings of the Workshop on Narrative Understanding, Minneapolis, Minnesota, 2–7 June 2019.
30. Alm, C.; Sproat, R. Emotional sequencing and development in fairy tales. In *International Conference on Affective Computing and Intelligent Interaction*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 668–674.
31. Boyd, R.L.; Blackburn, K.G.; Pennebaker, J.W. The narrative arc: Revealing core narrative structures through text analysis. *Sci. Adv.* **2020**, *6*, eaba2196. [CrossRef] [PubMed]
32. Doust, R.; Piwek, P. A model of suspense for narrative generation. In Proceedings of the 10th International Conference on Natural Language Generation, Santiago de Compostela, Spain, 4–7 September 2017; Association for Computational Linguistics: Santiago de Compostela, Spain, 2017; pp. 178–187.
33. Bruner, J.S. *Actual Minds, Possible Worlds*; Cambridge University Press: New York, NY, USA, 1986.
34. Todorov, T. Narrative transformations. In *The Poetics of Prose*; Cornell University Press: Ithaca, NY, USA, 1977; pp. 218–233.

35. Russell, J.A.; Weiss, A.; Mendelsohn, G.A. Affect Grid: A single-item scale of pleasure and arousal. *J. Personal. Soc. Psychol.* **1989**, *57*, 493–502. [CrossRef]
36. Zsibrita, J.; Vincze, V.; Farkas, R. Magyarlanc: A toolkit for morphological and dependency parsing of Hungarian. In Proceedings of the International Conference Recent Advances in Natural Language Processing, Hissar, Bulgaria, 9–11 September 2013.
37. Gottschalk, L.A.; Bechtel, R.J. *Computerized Content Analysis of Speech and Verbal Texts and Its Many Applications*; Nova Science Publisher: New York, NY, USA, 2008.
38. Gobbo, C.; Raccanello, D. How children narrate happy and sad events: Does affective state count? *Appl. Cogn. Psychol.* **2007**, *21*, 1173–1190. [CrossRef]
39. László, J.; Csertő, I.; Fülöp, É.; Ferenczhalmy, R.; Hargitai, R.; Lendvai, P.; Péley, B.; Pólya, T.; Szalai, K.; Vincze, O.; et al. Narrative language as an expression of individual and group identity. *SAGE Open* **2013**, *3*, 1–12. [CrossRef]
40. Vabalas, A.; Gowen, E.; Poliakoff, E.; Casson, A.J. Machine learning algorithm validation with a limited sample size. *PLoS ONE* **2019**, *14*, e0224365. [CrossRef] [PubMed]