

A Semantically-Based Computational Approach to the Annotation of Narrative Style

Rodolfo Delmonte
Ca' Foscari University of Venice
(delmont@unive.it)

Giulia Marchesini
Ca' Foscari University of Venice
(giuliamarches@gmail.com)

Abstract

This work describes the annotation of the novel "The Solid Mandala" (Patrick White, 1966), carried out combining sentiment and opinion mining on character level with the Appraisal Theory framework, here used to identify evaluative statements and their contribution to the social dimension of the text. Our approach was inspired by research on the correlation between White's style and the personality of his main characters. The annotation was manually executed by second author using an XML standard markup system and double-checked by first author. In this paper we comment on the selected features, focusing on the ones acquiring specialized meaning in the context of the novel, and provide results in terms of quantitative data. Comparing them, we are able to extract story units in which *special or significant events* take place, and to predict the presence or similar units in the narrative by detecting concentrations of features. Eventually collecting all annotations has made available a lexicon where all ambiguities are clearly identifiable and verifiable. The lexicon will be used in the future for the automatic annotation of other novels.

1 Introduction

In this paper we will demonstrate the use of deep semantic features in the annotation of a complex narrative text, the novel "The Solid Mandala", by White (1966), in an effort to prove the main tenet of White's style as purported by G. Collier in his "The Rocks and Sticks of Words" (Collier (1992) hence, GC1992). The theory states that White makes use of linguistic features to define his characters more accurately and at a deeper level, deliberately using specific words, phrases, syntactic structures and semantic features to profile each main character in the narrative and to highlight personality traits. Collier defines many interesting recurring features which seem to be prominent in the novel, features which we therefore incorporated in our annotation scheme.

"The Solid Mandala" is a particularly interesting sample for analysis because of the peculiarity of its structure and of its style. It is in fact divided into four distinct sections, each written as if through the eyes of one of the main characters. Using such a narrative as a starting point, we aim in the future to expand the limitations of the current annotation system, which is now partly tailored to the novel, to include many other kinds of narrative texts. This will be made possible by the collected annotations in a lexicon where head words are associated to their lemmata (see Delmonte and Marchesini (2017)).

In our case, White's style always maintains internal consistency, but it does so adapting its qualifying linguistic elements according to the lens of the narrating point of view – there are three main characters, and each of them is tasked with narrating a portion of the same story. Other elements impacting linguistic features are the nature of the relationships between the characters and the specificity of the events in their lives, and all of them are considered in the evaluation of the annotation.

As for the elements chosen for analysis, we started from semantic features connected to psychology and affection and included particularly meaningful syntactic structures, as we will see in more detail below. Predicates and verbs in general, for example, are often crucial to interpret many layers of textual analysis. Different aspects, tenses and modes all influence reader interpretation: an imperfective psychological verb might be related to a conscious, lasting mental process, while a perfective one might indicate unconscious realization and can come back later in the form of recognition.

For a general overview, we want to give a first quantitative outline of the novel. Its text has been divided into 131 **narremes**, defined as minimal independent story units as in GC1992 (p. 35-36) (see also Bonheim (2000)), each of independent size and connected to a specific event.¹ Table 1 indicates the total number of tokens considered and the number of annotations, both of single tokens (as in "thinks") and of compounds (as in "is thinking").

The manual annotation made on a text editor took around three months of work to complete, including revision and evaluation of the preliminary results. The contents of this paper refer to these evaluations. We are currently looking for trained annotators in order to provide data on inter-annotator agreement on at least a subset of the novel.

| Items | Total | Standard Deviation |
|-------------|---------|--------------------|
| Tokens | 120,249 | 1149.3 |
| Annotations | 8616 | 84.26 |

Table 1: Number of tokens (including punctuation), annotations, and their standard deviation

2 Semantic Features as Inherent Items of White's Style

As mentioned, we considered semantic features (see Bos and Delmonte (2008), Delmonte and Pallotta (2011)) as the main elements used to link the style and the personality of the three main characters. They are the twins Waldo and Arthur Brown, each narrating one of the two main sections of the book, and their neighbour Mrs Poulter, who instead narrates the shorter first and last sections corresponding to prologue and epilogue.

Both the fabula – the chronological order of events – and the sujet – their stylistic rearrangement, or plot, here a complex back and forth of memories alternated with actional present – were taken in consideration, mostly thanks to the fragmentation of the text in single events (narremes). After the annotation it is simpler to look at the narrative stream and to immediately spot evident concentrations of features, signaling highly relevant and symbolic events.

As for the annotation task itself, it was organized in three main meta-tags and a number of hierarchically related more specific ones, as shown in the table below. The annotation scheme is original and based on characteristics of the XML markup standard (elements, attributes and values). In this section we will discuss the reasoning behind the chosen categories.

2.1 Commenting Features and Their Relevance to the Storyline

Our three high level features, the meta-tags, are *uncertainty*, *subjectivity* and *judgement*. Additionally, we annotated with the element *negative* all negative forms in the novel.

With "uncertainty" we mark all the parts of the text which carry in their syntactic, semantic or pragmatic value a sense of interpretation of the storyworld by the characters (e.g. "it seems"). This interpretation may express the more or less conscious doubts in the minds of the protagonists. At the opposite end, it can also signal a judgement of certainty, which – ironically – generates in turn insecurity in the reader and raises the question: 'Is the narrator reliable?'. In this analysis the element of uncertainty, abbreviated

¹By *narreme* we (and he) mean here a basic story unit and a microstructure covering one independent event, as conveyed via a single or multiple points of view. As Collier notes (Collier (1992) pp. 37-43), it is important to remember that these narremes are merely heuristic abstractions, albeit very useful ones when it comes to psychological narrative. Not so is the opinion of other narratologists, who claim a structural causal relation intervening between each narreme (Wittmann (1975) pp. 19-28). In his analysis, Collier only takes action-oriented narremes, or "process statements", as opposed to "stasis statements". The latter passages of the novel are mostly a-temporal, descriptive, and purely psychological; in any case, they are never tied to a specific event or series of events, unlike the former. Since we needed to cover the whole text of the novel, and not only the process statements, an expansion of the original list was needed. So eventually we passed from 124 narremes of Collier to 131

| Uncertainty | Subjectivity | Judgement |
|--------------------|----------------------|------------------------|
| <i>Non-factual</i> | <i>Psychology</i> | <i>Social-esteem</i> |
| Seeming | Perception | Positive/Negative |
| Gnomic | Precognition | <i>Social-sanction</i> |
| Concessive | Cognition | Positive/Negative |
| Conditional | PerformWill | |
| DefDesire | <i>Affect-emot</i> | |
| Will | Positive/Negative | |
| Possibility | <i>Affect-inclin</i> | |
| Ability | Positive/Negative | |
| Obligation | <i>Affect-secur</i> | |
| Assumption | Positive/Negative | |
| | <i>Affect-satisf</i> | |
| | Positive/Negative | |

Table 2: Hierarchy of deep semantic features used in the annotation

as *uncertnty* in the annotation, has only one obligatory attribute: non-factual (see Sauri and Pustejovsky (2012)). It is in fact crucial in this case to establish the annotated expression as non-real, non-factual, as something that is only going on in the character's mind and which does not have an equivalent in the 'real world' of the story.

"Subjectivity" (see Taboada et al. (2011), Pang and Lee (2004)) focuses instead on various facets of character psychology. The main difference between *uncertainty* and *subjectivity* lies in the fact that the first category is about how the protagonists rationalize their reality, while the second marks the modalities in which they actively and subjectively contribute to the narrative. The former is non-factual in nature, but the latter always has consequences. *Subjectivity* includes active psychological processes both conscious and unconscious as studied by the cognitive sciences, as well as expressions of emotion and different kinds of feelings. These sub-categories are grouped into five attributes: *psychology*, *affect-emotivity*, *affect-inclination*, *affect-security* and *affect-satisfaction*. While psychology can have different values, the other four can only have two: positive or negative.

The last element is "judgement", abbreviated as *judgmnt* in the annotation, which marks all evaluative expressions and aims at highlighting social and personal reactions to the storyworld. The theoretical basis of this category can be found in the so-called "Appraisal Theory" (see Martin and White (2005) (hence M&W2005), Read and Carrol (2012), Taboada and Grieve (2004)). This theory emphasizes the relevance of impressions and judgements in the formation of feelings, emotions, and complex thoughts. Environment and psychology are here understood as being in a relation of mutual dependency, with the reactions of each individual to stimuli evoking different responses. Since we are talking about a novel, this system is of course artificial and created by the author. From a general standpoint we can say that *judgement* and the above-mentioned *affect* have a lot in common, both dealing with indices of emotion and sentiment (see Kim and Hovy (2004), Kao and Jurafsky (2012)). In this specific study, however, it was decided to annotate *judgement* as an independent element for two main reasons: firstly in order to stress our interest in evaluative language, and secondly in order to allow a more detailed internal differentiation between the attributes of *social-esteem* (personal dimension, e.g. appreciated/unappreciated) and *social-sanction* (social dimension, e.g. allowed/forbidden).

A more detailed description of attributes and values follows.

2.2 Features in Detail: Uncertainty

<uncertnty non-factual= "seeming">

Seeming is the probably the most representative of all the values of *non-factual*, and *uncertainty* in general. It is in fact the class that most refers to irrealia – indices referring to non-real elements of the storyworld. These figural and stylistic indices, meant to mark the internalization of experience by the characters of the novel, are defined as indices of *figural apperception* (see GC1992, pp. 140-141)². This value covers indicative verbs (e.g. "seemed"), modal statements (e.g. "must have"), adverbial clauses (e.g. "as though") and discourse markers (e.g. "perhaps"). In the case of modal verbs cases of ambiguity can arise, particularly with "must have" and the *assumption* class. To keep the distinction clear, modal verbs were only marked as *seeming* when their meaning was obviously indexical to character judgement.

<uncertainty non-factual= "gnomic">

Gnomic sentences are characterized by frequent use of the present tense, sententious tone, and generalization. Their use in White's style has been largely controversial in academic discussions and often interpreted as authorial intrusiveness, but in this analysis we choose to consider it as another element of apperception: through the formulation of generalizations and conclusions the characters rationalize their reality. In this interpretation gnomic expressions, either complete sentences or significant parts of larger ones, are read as the manifestation in the style of a psychological process.

<uncertainty non-factual= "concessive">

Fragments of text annotated as *concessive* always have a clear semantic value and structure. We made a distinction between 'true' concessives (e.g. "even if/though") and *if-concessives* or *pseudo-concessives* (e.g. "good, if bitter") (see GC1992; p. 187), not to be confused with the similar *if-conditional* type. Their occurrence in the novel is so high that we can talk about it as a pervasive characteristic of White's writing, at least in this particular novel.

<uncertainty non-factual= "conditional">

Conditional markers are another example of how a common syntactic structure can carry significant stylistic and narratological weight. In our case hypotheticals are particularly tied to character psychology and provide differentiations between the narrating characters. They see different applications in varying degrees of rationalization and justification, both real and putative, depending on the case.

<uncertainty non-factual= "defdesire">

Defdesire is here short for "desire defeated by grammar". It is a peculiar way of using conditionals, when a character's desire is negated by the structure of the sentence itself. Because of this high level of specialization and its importance for character psychology – it carries in fact negative semantic traits connected to pessimism and failure – we considered it as a separate value in the current annotation.

<uncertainty non-factual= "will">

Expressions of *will* represent the first of five groups of non-factual modal – or in this case, modal-like – values. We selected for this category all verbs of will and wish (including "like", "would like to" and "would have liked to" when synonyms of "want") and the noun "will". More difficult to solve were expressions of "be willing to", potentially overlapping with *inclination* or even *psychology* (see below). In order to disambiguate, we selected only the expressions stressing the verb and the result – action or inaction, highlighted by the presence of "to".

<uncertainty non-factual= "possibility">

Possibility is connected to prospects, odds, and opportunities. It consists of a variety of different elements, from modal verbs to nouns, from adjectives to adverbs, all of them playing a similar role in conveying the concepts of 'option' or 'likelihood' as seen from the point of view of the protagonists (e.g. "may", "might", "can", "could", "possibility", "possible", "possibly" ...).

<uncertainty non-factual= "ability">

Ability (e.g. "can", "be able to") is a class of modality which can sometimes be confused with the previous one, *possibility*. Its range of meaning is generally described as varying between general abilities (e.g. "I can swim" – always) and specific abilities (e.g. "You can do this" – now), and this distinction is maintained in the present research. Additionally, in "The Solid Mandala" we annotated as *ability* a specific "would", which seems to be a linguistic irregularity associated with Australian slang.

²*Seeming* and *uncertainty* in general can sometimes be judged as ironic statements, as well, but Collier chooses not to cover this possibility in his analysis and we in turn follow his suggestion. (Collier (1992) pp. 44-45, Reyes et al. (2012))

<uncertainty non-factual= "obligation">

It is dedicated to expressions of obligation and need, including commitments, duties, necessities, coercions, and sometimes even a sense of inevitability (e.g. "must", "ought to", "be forced to", "necessity", "necessary"...). Ambiguous occurrences were the expressions of "need": when followed by "to" they usually express obligation, and when on their own (e.g. "need something") they are not synonyms of "have to" and were marked as *inclination*. Another ambiguous case was the resolution of "should", which can also indicate *assumption*.

<uncertainty non-factual= "assumption">

Assumption (e.g. "should/should have", "must/must have") can have a twofold scope of meaning: the first is related to hypotheses, speculations, deductions, beliefs and abstract ideas; the second refers to suggestions or offers. The psychological value of assumption relates to two kinds of processes: rationalization for the first range of denotations, and preparations of future events for the second.

2.3 Features in Detail: Subjectivity

Four of the values of *subjectivity*, the ones connected to *affect* (inspired by Martin and White (2005), see also Wiebe et al. (2005), Turney and Littman (2003), Kao and Jurafsky (2012)) can be defined with a positive/negative polarity of values, while only one of them, *psychology* can take four different values. *Psychology* contains all cognitive indices, including terms indicating perception, memory, thought, imagination, metacognition, and so on.

<subjectivity psychology= "perception"> *Perception* can mark two groups of terms: those directly connected to the physical five senses (e.g. "hear", "see", "the sound"), and those related to the primary mental processes tasked with interpreting them (such as basic attention, e.g. "noticeably").

<subjectivity psychology= "precognition">

Unconscious or semi-conscious mental processes, represented by psychological verbs, belong to this value (e.g. "listen", "wonder"), and often tend to overlap with other classes, as is the case for instance with expressions of fear or doubt. In order to maintain the distinction as clear as possible, we have marked under *precognition* all generic expressions of feeling, wondering, wide-ranging ideas and immediate impressions, as well as interpretations of stimuli (e.g. forms of listen) and thoughts (when they take the form of feelings, intuitions, or impressions; e.g. thinking something of someone's attitude). Moreover, the concept of knowledge is here considered as a cognitive function instead of precognitive.

<subjectivity psychology= "cognition">

This value mostly describes psychological verbs connected to consciousness and awareness (e.g. "remember", "realize", "understand"). As occurs in precognition, cognition is a category of psychological verbs only, despite their great variety. For this reason the concept of "knowledge" is considered to involve consciousness and to be more similar to the idea of awareness, and therefore as a part of the cognitive group, while perfective and "one-time" uses of verbs like "think" (when it indicates feelings etc.; see above) are classified as precognitive. The goal is to distinguish more clearly between the domain of the "spirit" – feelings, impressions and abstract thought, all precognitive – and the domain of the "mind" – with awareness, memory, decisional processes, hopes and knowledge (both theoretical and with a semantic substrate of "acquaintance") among its most important functions. Another reason for the distinction is, then, the implementation of the "emotional" categories that we shall witness in the following sections, which may cause overlapping and confusion between the groups.

<subjectivity psychology= "performwill">

Short for "performative cognition and will", *performwill* marks expressions of extreme self-control and imposition of personal will on reality (e.g. special use of "decide"). The storyworld isn't rationalized as it is, but as the character wills it. Expressions of performwill are not simply cognitive indices but represent an extreme form of self-control as for instance for Waldo and his perpetual attempt to control reality and the ways in which this influences how he perceives external stimuli and other people. This strong manifestation of will can be either positive or negative: it is positive when a character – usually Waldo – more or less consciously forces himself to say something or to act in a certain way, and it is

negative when the same willpower is directed at restraining oneself from doing something or at avoiding and negating some realization or other.

<subjectivity affect-emot= "positive/negative">

The attribute *affect-emot* (emotion) marks all the expressions indicating happiness or unhappiness in the novel, the former distinguishing between cheer and affection (e.g. "jovially", "love") and the latter between misery and antipathy (e.g. "sorry", "despise").

<subjectivity affect-inclin= "positive/negative">

The positive aspect of *affect-inclin* (inclination) expresses varying degrees of desire and longing, all with comparable active traits of optimism and eagerness (e.g. "desire", "waiting"). Its negative end is instead annotated when someone is shown as far from eager to do, experience, or say something. It involves feelings of suspicion, mistrust and concern, if not outright anxiety or even terror (e.g. "afraid", "had feared"). As with the other affect categories, *affect-inclin* (inclination) covers an emotional spectrum. Its positive end usually expresses varying degrees of desire and longing, all with comparable active traits of optimism and eagerness. In this respect, this category was mentioned earlier in the context of the annotation of uncertainty non-factual="will"; the two aspects are alike in their general acceptance of "desire" and "will", but while the category under uncertainty non-factual deals with the modal side of the semantic field, *affect-inclin* marks projections of eagerness and feelings of longing or craving.

<subjectivity affect-secur= "positive/negative">

Short for security, it refers in its positive value to expressions of self-assurance, calm, hope and reliance (e.g. "trust", "confidence", "relief") and in its negative one to the reverse: disquiet, insecurity, anxiety, lack of confidence, and sometimes indices of fear (e.g. "surprise", "sweat"). A surprise is not always negative per se, but in this case always shows some degree of lack of security for a character, and is therefore part of the negative value. *Affect-secur* (security) marks yet another aspect of the varied overview of sensations. Its positive value is referred to in expressions of self-assurance, calm, hope and/or reliance, while the negative indicates the reverse: insecurity, anxiety, lack of confidence, and sometimes indices of fear. It is at this point necessary, focusing on fear, to establish a difference between this "security fear" and the "inclination fear" presented in the previous section, which are part of the same semantic field and can sometimes overlap – in a similar way to "non-factual will" and "inclination will". The main divergence between the two classes is in the semantic traits associated with each category: while "inclination fear", as we saw, expresses "not wanting to do something", often a concern for the future or inward rebellion against an idea, "security fear" indicates a sense of alarm and fright. It is not apprehension, but a nervousness caused by a lack of security, be it self-confidence or danger coming from the outside world.

<subjectivity affect-satisf= "positive/negative">

Affect-satisf (satisfaction) deals with the last aspect of the emotive sphere as it is considered in the current analysis. The positive value of this attribute marks all indices of approval, pleasure, agreement and curiosity, (e.g. "appeal") all deeply connected with the idea of "happiness" explored above and sometimes slightly overlapping with positive inclinations. At the opposite pole we find the negative connotations of dissatisfaction, including bother, sense of monotony, disapproval, impatience and anger (e.g. "yawned", "furious"). Expressions based on the notion of "satisfaction", in fact, are deeply connected with self-fulfillment and the possible lack thereof, which are among the most relevant themes of the whole novel. We can distinguish among four self-explanatory subcategories, Interest (satisfaction); Pleasure (satisfaction); Ennui (dissatisfaction); Displeasure (dissatisfaction).

2.4 Features in detail: judgement

The last content-related element introduced in the annotation is, as was the case with the affect attributes, an addition to the original theory formulated by Collier. In this annotation, *judgmnt* (judgement) marks all evaluative expressions relating to the characters of the novel, aimed at highlighting both social and personal reactions to the storyworld and particularly to the other characters and their behavior. The theoretical basis of this category can be found in the so-called "appraisal theory", (see Martin and White

(2005) p.111) which underscores the relevance of impressions and judgments in the formation of feelings, emotions, and complex thoughts. Environment and psychology are here understood as standing in a relation of mutual dependency, with the reactions of each individual to events and stimuli evoking different responses. Speaking of the characters of a novel, in this case, we can say that the “artificially created” psychology of every character reacts in substantially different ways to what happens in the storyworld. From a general point of view we can say that judgment and affect as categories have a lot in common, both dealing as they do with indices of emotion and sentiment. In this specific study, however, it was decided to annotate judgment as an independent element, for two main reasons: the first was to emphasize our interest in the evaluative language used by the author, and the second was to allow a more detailed internal differentiation between the categories of social-esteem and social-sanction and their respective polarities. For this very reason the element has two attributes with two values each. Both social-esteem and social-sanction are characterized by a social and a personal component. The social one is predominant, since in both attributes something is evaluated as positive when it is socially acceptable, and is considered negative when it does not respect social norms. The personal approach to the evaluation is as important, however, particularly in a novel like *The Solid Mandala* in which the protagonists are rarely socially accepted themselves, and were raised in a rather closed and secluded environment.

<judgment social-esteem= "positive/negative">

This category is used to mark adjectives and a few adverbs indicating respect and admiration (e.g. "healthy-looking", "reliable") or, at the negative pole, criticism and denigration (e.g. "peculiar", "capricious"). We identified three subcategories of this attribute, each with the internal polarity positive/negative: Normality (“how special?”), including lucky, predictable, fashionable / unlucky, eccentric, obscure, etc.; Capacity (“how capable?”), including robust, experienced, competent / weak, stupid, ignorant, etc.; Tenacity (“how dependable?”), including brave, thorough, faithful / timid, reckless, inconstant, etc.

<judgment social-sanction= "positive/negative">

In the case of *social-sanction*, the annotation focuses on particularly strong expressions of admiration or social denounce (see Martin and White (2005)). Similar in nature to social-esteem instances, these indices tend to be more general and often deal with moral and ethical issues, labeling adjectives and adverbs which refer to universal appraisal (e.g. "kindly", "honest") or universal condemnation (e.g. "indecent", "dishonest"). Two subclasses distinguished by semantic traits can be found under this attribute: Veracity (“how honest?”), deals with issues of truth and general trust, including truthful, candid, tactful / deceitful, manipulative, blunt, etc.; Propriety (“how far beyond reproach?”), deals with ethical issues. It includes moral, law-abiding, humble / evil, mean, irreverent, etc.

<negative= "/negative">

Negative is the last element in this annotation, and its characteristics lead it to be considered as separate from the others. It does not represent, in fact, a textual tag (like *narreme*, *p* and *s*) but a content-related one, and at the same time it is much “simpler” than the other elements explained above (uncertainty, subjectivity and judgment). It marks all the negations in the novel, in all their possible forms - taking care of double negations which are not regarded as such. The annotation often includes the complete form of the verb being negated (e.g. <negative>did not say</negative>), but not necessarily whole expressions, especially when dealing with modals (e.g. <negative>did not want</negative> to go). While in this particular study the frequency of negative versus positive verbs is particularly relevant because of the numerous verbal tags, these are not the only elements to be annotated. The negative tag was applied to many adjectives, pronouns and adverbs when appropriate, as well as to colloquial forms of “no” – both in direct and in indirect speech. A last comment must be made on the annotation of negative as applied to verbs. In the case of never (see the second example) and in some cases of not (fourth example), it was decided not to include the verbs they depended on, in order to highlight the negation in itself. This decision relied on the fact that “never” and “not”, the latter when followed by an adjective, mostly occur independently. Tagging them by themselves is the easiest way to make quantitative evaluations about their distribution in the text.

3 Discussion of Experimental Data

The first thing to notice is the still-high level of ambiguity in our extracted lexicon. All our extracted information was collected in a lexicon with each occurrence (e.g. "here") followed by a compound feature label, or series of labels, (e.g. "subject-psychology-cognition") and the narreme index. From this general lexicon we derived another one, this time only composed of unique forms. High ambiguity was then apparent. As seen in Table 1, the total number of extracted features amounts to 8646, but the non-repeated count only amounts to 1584. This means that there are forms which are present in the lexicon more than once and within different categories. An example is the form "hear", which can belong to *cognition*, *perception* or *precognition* depending on the case. We calculated a total of 121 entries having double or triple meaning.³ This final lexicon, made of word-forms associated with lemmata, potentially makes ambiguity a resource: it allows in fact to check for possible associated meaning in other meaning repositories. One of our goals is to use this principle to try and automatically tag other texts by the same author, checking the level of coverage and eventually moving on to other kinds of texts.

Back to the current analysis, extracting information from the annotated text allowed us to confirm our starting hypothesis: the style exhibits different characteristics depending both on the profile of the narrating character and on the nature and context of the specific event in the narrative. In the extraction of data and subsequent discussion the categories employed in the annotation system are referred to as *traits*. Their names (seeming, cognition, affect-secur...) stay the same, but each and every one of them is hereby envisioned as a psychological trait to be investigated.

We applied quantitative information about trait distribution to every main character – Waldo Brown, Arthur Brown and Mrs Poulter – using the above-mentioned narremes (story units), which provide us with the narrator for each event. With three distinct portions of text, each related to a protagonist, we were able to study which traits prevailed and therefore to draw conclusions about character profiling.

For each trait we calculated a ratio using the formula:

$$[\text{annotations of trait } X \text{ for given character} / \text{total of annotations of trait } X]$$

We then considered the length of the three sections, with Waldo's being the longest by far and Mrs Poulter's the shortest, and established as *significant* for our analysis every trait with a ratio of more than 0.65 for Waldo, 0.3 for Arthur, and 0.11 for Mrs Poulter. You can see a list of these *significant traits* for each character in Table 3.

The distribution of traits gives us valuable data on character profiling. We can see, for example, that Waldo Brown almost has exclusive access to *performative will* (with a ratio of more than 0.9) and that many of his significant traits are connected to appraisal with a negative polarity – manifesting a tendency to judge himself and/or others, and a feeling of being judged. It is not by chance if we find the negative declinations of *affect-inclination* and *affect-security*, as well, since they both deal with indices of apathy, anxiety, insecurity and fear. Moreover, the high frequency of *precognition* and *seeming* suggests that the character uses indices of interpreting phenomena before interiorizing them, raising the chance that we are dealing with a particularly unreliable narrator. Even for someone who has not read the novel, the general profile of the character is at this point established.

The same principle is valid for the other main characters, even with less significant traits. Arthur Brown's highest value is assigned to *defeated desire* which, along with the positive value of *inclination*, *ability*, *assumption* and both values of *satisfaction*, suggests a character striving to do something, pondering his abilities and those of others, and mostly exhibiting a pessimistic note about final impossibility

³A few more examples, as they appear in our Prolog-compiled lexicon:

lx(confused,confuse,[subject-affect-secur-negative,subject-psychology-cognition]).

lx(contemplate,.,contemplate,[subject-psychology-perception,subject-psychology-precognition]).

lx(could-not-have-been,be,[uncrtn-non-factual-ability,uncrtn-non-factual-possibility]).

lx(embarrassing,embarass,[subject-affect-emot-negative,subject-affect-secur-negative]).

lx(frustration,frustrate,[subject-affect-emot-negative,subject-affect-satisf-negative]).

lx(intended,intend,[subject-affect-inclin-positive,subject-psychology-cognition,-psychology-performwill]).

| Waldo | Arthur | Mrs Poulter |
|-------------------|---------------|--------------------|
| 1. PerformWill | 1. DefDesire | 1. S-sanction (p) |
| 2. Possibility | 2. Inclin (p) | 2. Gnostic |
| 3. S-sanction (p) | 3. Ability | 3. S-esteem (p) |
| 4. Precognition | 4. Gnostic | 4. Possibility |
| 5. S-esteem (n) | 5. Satisf (n) | 5. Inclin (p) |
| 6. Seeming | 6. Satisf (p) | |
| 7. S-sanction (n) | 7. Assumption | |
| 8. Inclin (n) | | |
| 9. Secur (n) | | |
| 10. S-esteem (p) | | |

Table 3: Significant traits for each character

to reach the goal. Interestingly, the sad and more realistic meaning of *defeated desire* is almost diametrically opposite to the wishful thinking of Waldo's *performative will*. Experience in Arthur's section is mostly expressed by positive *inclination* – opposed to Waldo's negative – towards life and others, and by a more generalized *gnomic* approach. These traits are present in Mrs Poulter, as well, indicating an important similarity between the two characters, but the differences are even more interesting. Mrs Poulter's section lacks all the dynamism and the interpretation of the world seen in Arthur's, using more basic *gnomic* rationalization and the positive appraisal values of *social sanction* and *social esteem*.

Additionally, distribution of traits was extremely useful to identify *significant events* (i.e. narremes with a high concentration of traits) and to investigate the nature of different relationships between main and secondary characters of the story. All these pieces of information contributed to an even more detailed psychological profile of the protagonists.

You can see the distribution of the occurrences in Table 4.

4 Conclusion

We presented the annotation of the whole text of "The Solid Mandala", by Patrick White, which was carried out by second author and double-checked by first author. The research was inspired by Gordon Collier's book about the same novel, "The Rocks and Sticks of Words", and it is based on many features suggested in it, with many changes and additions due to the specificity of our task. Suggestions from Martin & White's Appraisal Theory were included as well, introducing new features related to judgement and affect. The final results are here summarised with the help of tables and diagrams, clearly showing the effectiveness of this approach in highlighting psychological profiles of characters. The connection between profiles, authorial style and narratological level of analysis was also proved to be very strong. We intend to use the lexicon derived from the XML annotation to try and automatically annotate another novel by the same author, in order to evaluate the results in terms of coverage and precision. A second related goal is to use this analysis as a starting point to expand the annotation scheme to a larger target of narrative texts.

| Tags/Chars | Waldo | Arthur | Mrs Poulter |
|--------------|-------|--------|-------------|
| Perception | 674 | 303 | 109 |
| Precognition | 379 | 141 | 39 |
| Cognition | 953 | 460 | 140 |
| PerformW | 39 | 3 | 0 |
| Seeming | 512 | 198 | 56 |
| Will | 74 | 33 | 8 |
| Possibility | 224 | 54 | 42 |
| Gnomic | 32 | 20 | 9 |
| Ability | 347 | 195 | 51 |
| Obligation | 178 | 79 | 28 |
| Concessive | 150 | 75 | 26 |
| Conditional | 264 | 117 | 32 |
| DefDesire | 49 | 35 | 2 |
| Assumption | 73 | 36 | 9 |
| Emot-Pos | 400 | 183 | 43 |
| Emot-Neg | 247 | 107 | 30 |
| Inclin-Pos | 83 | 50 | 17 |
| Inclin-Neg | 75 | 26 | 12 |
| Secur-Pos | 133 | 65 | 19 |
| Secur-Neg | 297 | 116 | 40 |
| Satis-Pos | 136 | 66 | 10 |
| Satis-Neg | 215 | 112 | 25 |
| Esteem-Pos | 223 | 71 | 47 |
| Esteem-Neg | 346 | 129 | 39 |
| Sanct-Pos | 80 | 19 | 18 |
| Sanct-Neg | 127 | 47 | 16 |

Table 4: Distribution of semantic linguistic features for the three main characters

References

- Bonheim, H. (2000). Shakespeare's narremes. *Shakespeare Survey: Shakespeare and narrative* 53.
- Bos, J. and R. Delmonte (2008). *STEP '08 - Proceedings of the 2008 Conference on Semantics in Text Processing*. Semantics. London: College Publications.
- Collier, G. (1992). *The Rocks and Sticks of Words Style, Discourse and Narrative Structure in the Fiction of Patrick White*. Amsterdam Atlanta: Editions Rodopi B. V.
- Delmonte, R. and G. Marchesini (2017). A semantically based computational approach to narrative structure. In *Proceedings of IWCS, International Workshop on Computational Semantics*, Stroudsburg, PA, USA, pp. 8–17. ACL.
- Delmonte, R. and V. Pallotta (2011). *Opinion Mining and Sentiment Analysis Need Text Understanding*, pp. 81–95. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Kao, J. and D. Jurafsky (2012). A computational analysis of style, affect, and imagery in contemporary poetry. In *Proceedings of NAACL Workshop on Computational Linguistics for Literature*, Stroudsburg, PA, USA, pp. 8–17. ACL.
- Kim, S.-M. and E. Hovy (2004). Determining the sentiment of opinions. In *Proceedings of the 20th*

international conference on computational linguistics - COLING, Stroudsburg, PA, USA, pp. 1367–1373. ACL.

Martin, J. and P. White (2005). *Language of Evaluation, Appraisal in English*. London and New York: Palgrave Macmillan.

Pang, B. and L. Lee (2004). A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd annual meeting of the Association for Computational Linguistics*, Stroudsburg, PA, USA, pp. 271–278. ACL.

Read, J. and J. Carrol (2012). Annotating expressions of appraisal in english. *Language Resources and Evaluation* 46, 421–447.

Reyes, A., P. Rosso, and D. Buscaldi (2012). From humor recognition to irony detection: The figurative language of social media. *Data & Knowledge Engineering*.

Sauri, R. and J. Pustejovsky (2012). Are you sure that this happened? assessing the factuality degree of events in text. *Computational Linguistics* 38(2), 261–299.

Taboada, M., J. Brooke, M. Tofiloski, K. Voll, and M. Stede (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics* 37(2), 267–307.

Taboada, M. and J. Grieve (2004). Analyzing appraisal automatically. In *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*, pp. 158–161. AAAI Press.

Turney, P. D. and M. L. Littman (2003). Measuring praise and criticism: Inference of semantic orientation from association. *ACM Trans. Inf. Syst.* 21(4), 315–346.

White, P. (1966). *The Solid Mandala*. London: Eyre & Spottiswood.

Wiebe, J., T. Wilson, and C. Cardie (2005). Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation* 39(2), 165–210.

Wittmann, H. (1975). Theorie des narremes et algorithmes narratifs. *Poetics* 4(1), 19–28.

APPENDIX

Figures of Deep Semantic Features and their Distribution in Narremes

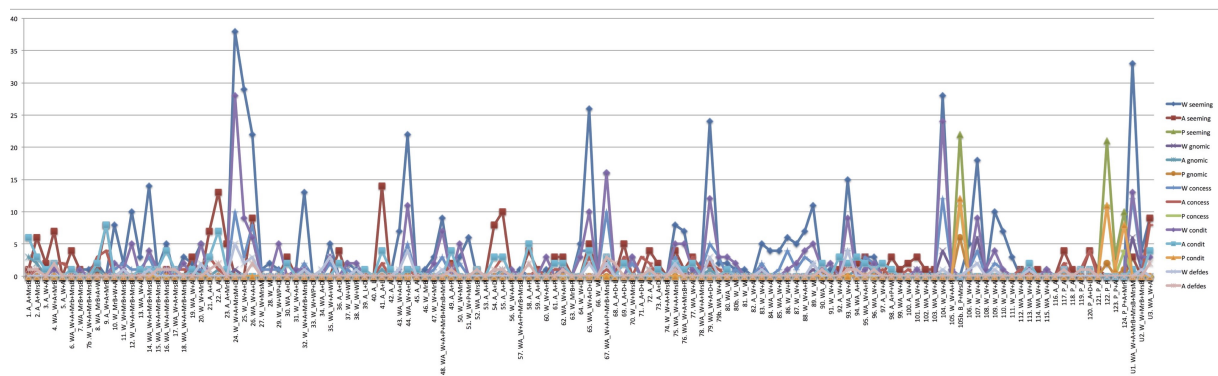


Figure 1: Distribution of uncertainty in 131 narremes

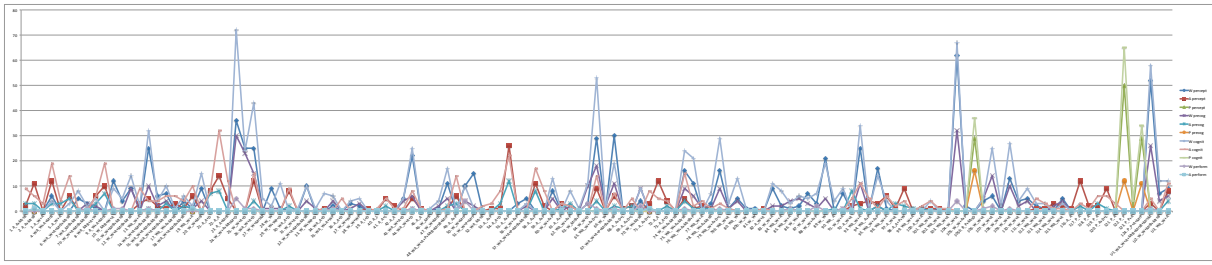


Figure 2: Distribution of psychology in 131 narremes

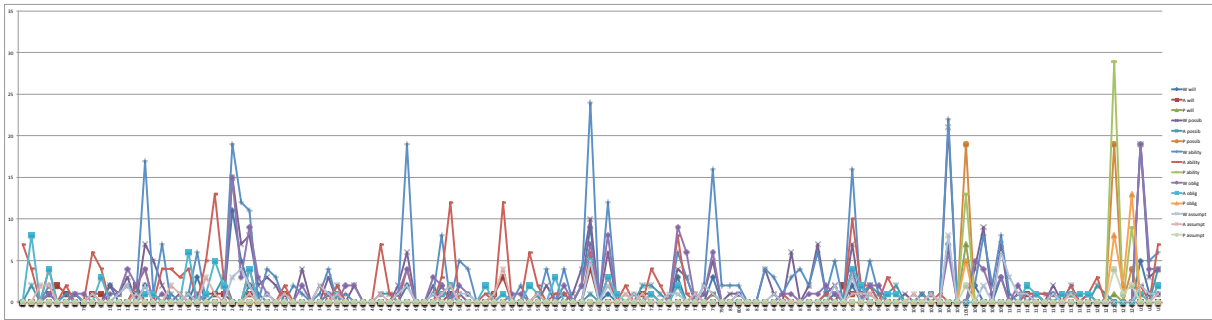


Figure 3: Distribution of modality in 131 narremes

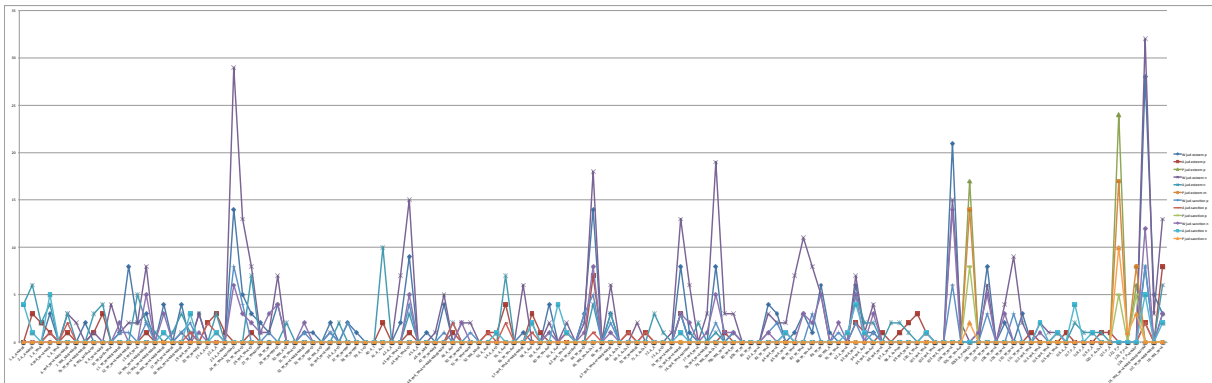


Figure 4: Distribution of judgement in 131 narremes