

A Semantically-Based Computational Approach to Narrative Structure

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

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

Abstract

In this paper we will define narrative structure as characterized by a basic element, the *narreme*, which is here described as the basic unit of narrative structure, the smallest possible unit of the story. We annotated a full novel – "The Solid Mandala" (1966) by Nobel laureate Patrick White – combining two approaches: one is related to sentiment and opinion mining, including deeper aspects connected to event factuality and subjectivity, and the other focuses on evaluative features derived from the Appraisal Theory framework. After characterizing the style, we will show the main significant events of the plot as they emerge from the distribution of deep semantic features. Narreme boundaries will be identified by presence of specific speech acts, change of point of view, and movement in spatio-temporal coordinates through flashbacks. An experiment with our system of text understanding, GETARUNS, has been carried out to test its ability to automatically identify narremes.

1 Introduction

According to Collier (1992) (hence GC92) Patrick White characterizes the plot of his books through the use of deep linguistic features: in particular, an accurate choice of words, syntactic structures, and semantic features is used to highlight specific portions of the role of each character in the narrative depending strictly on his/her personality traits and on the structure of the story. White's style remains always the same for all characters and sequences of events, but it varies its qualifying linguistic elements according to the point of view, to the events in the lives of the characters, and to the relationships in the storyline portrayed in the novel.

In order to make our narratological approach more transparent, we will here provide a brief summary of the main theme of "The Solid Mandala". Our novel tells the story of the life of twins Arthur and Waldo Brown, of their family, and of their neighbour, Mrs Poulter, all living in a suburb of Sydney. Waldo Brown is an appropriate example of many other important characters of Patrick White's novels, as well. He is the representative of the intellectual who failed to become an artist or even to accomplish anything of significance, ending up being a simple clerk in a municipal library. His life is empty of events and positive emotions. He is educated and despises his community, which he considers too uninteresting and uncultured for him to be a part of. This voluntary isolation translates in a general growing resentment and in open hostility towards his twin brother, who is completely different from him and yet always a constant in his life. Arthur and Waldo could not be more diametrically opposed. This opposition is specifically crafted to portray and therefore study two basic drives: intellect and intuition. We find in Waldo every characteristic of the academic individual driven by intellect, as we said before; in Arthur, instead, there is a more "feminine" intuition which is often painted as direct result of his weak wits. Contrary to his brother, Arthur is far from studious and clever: it is often difficult to understand him even from his point of view, and this does not seem to hold particular meaning for him. He has difficulty speaking and expressing himself, even though some of his thoughts are deeper and more significant than Waldo's. He loves others, even and mostly the brother who despises him and considers him an handicap. Most of all, he is completely, almost unbelievably good, always humble and helpful in his simple way of living.

This complex story was manually divided by Collier into *narremes*¹ and chronologically reorganized in order to reconstruct what in narratology is defined as *fabula*. This was limited to those units that have “event character”(GC92,p.36), “or which are processes or event-sequences of clearly demarcated duration”. Narremes as entities are also defined as follows: “these narrative units might equally well be termed functions (at the level of the story, fabula, or recite) and motifs (at the levels of narrative, sujet, or discourse)”. Each of them covers one independent event as narrated by a single point of view or by more than one²

For our study we created a new subdivision of the novel into narremes, partly following the approach suggested by Collier, but implementing 131 narremes to account for all events in the story, contrary to his total of 124. Four temporal blocks of narremes were selected: Childhood, Adulthood, Old Age and After Waldo’s Death – this one only in Arthur’s point of view.

As a first quantitative outline of the novel, we present here a short table indicating the main entities as they have been derived from the counts of tokens in each of the 131 narremes – or story units.

Quant.Data	Toks.Arthur	Annots.Arthur	Toks.Waldo	Annots.Waldo	Toks.Others	Annots.Others
SUM	54312	4060	66365	4894	32771	2208
MEAN	848.625	63.437	1106.083	81.567	1092.367	73.6
ST. DEV.	1040.998	82.528	1700.591	106.066	653.367	37.477

Table 1: Distribution per main characters of tokens, annotations, and their standard deviation

We also report in a second table general overall data to evaluate differences. As can be noticed, values of standard deviation are higher than the mean indicating a great irregularity in data distribution. However values for non protagonists, under "Others" are more regular thus marking a neat difference from the rest.

Items	Total	Mean	St.Deviation
Tokens	120,249	917.9	1149.3
Annotations	8616	65.78	84.26

Table 2: Number of tokens (including punctuation), annotations, and their standard deviation

Without considering punctuation, the total number of tokens is 106,935, the total of the types 9,419, and the type/token ratio is 0.0881. The repetition rate is very high, as can be gathered from the so-called Vocabulary Richness index. Coverage of 50% of the total text tokens is reached after first 70 entries in the Rank list. The lower portion of the Rank List, made up of Rare Words appearing only 3/2/1 times, is made of 6848 types. In the novel, Waldo is explicitly mentioned 1063 times, while Arthur 985, slightly less. This difference is also apparent in the distribution of tokens and annotations, as shown above in Table 1. The only case in which st.dev. is lower than the mean is in the distribution of annotations for other less important characters. As to presence of female and male characters, the pronoun “he” is used 2641 times, and “she” 1225 times. The two most prominent female characters of the story, Dulcie Feinstein and Mrs Poulter, are mentioned 282 and 277 times respectively. “Mrs”, however, is mentioned 640 times: if compared to “Mr”, appearing only 213 times, which makes women’s role in the story highly important.

¹As has been defined in Bonheim (2000) (pp. 1-11) and in Dorfman (1969). See also Wittmann (1975).

²Quoting Collier, “Some of the narremes could have been divided further, others perhaps merged where they are adjacent. The reader-response principles upon which the selection of narremes was made are at once simple and complex. They derive from the relative salience of given events to the reader’s eye (after at least a second reading), and from the less overt pressure exerted by the presence of manifold details, leitmotifs and echoes or the microstructure level.” Collier (1992), p. 37

In the second part of the paper we will describe the algorithm used in the experiment that we carried out to try and find narremes boundaries. The output was derived from a system for text understanding called GETARUNS, which has been used lately for event discovery and semantic similarity challenges. The novel is made up of sentences which by way of an homology (see Bal (1987)) are assumed to correspond to the linguistic content of the novel itself. Event units are expressed by eventive words – verbs, deverbal nouns, non-stative adjectives – and in our system they represent an action, a transition from one state to another. The system has been used to automatically detect and annotate event structures of news articles for the Event Workshop 2013 ³.

2 Highlighting Events in the Narrative Structure

Investigating the distribution of annotations throughout "The Solid Mandala" can also mean taking a closer look at differences and similarities in style amongst the narremes. The original structure of the novel, in fact, makes it difficult for the readers to judge the impact of earlier events on later occurrences at a first glance, and complicates the *sujet* – how events are presented in order to produce an emotive reaction from the reader - with a complex system of memories and triggers.

On the contrary, starting from the *fabula* – the logical sequence of events semantically related in space and time - is an easier way to compare events between themselves. Two premises on the peculiar characteristics of the narremes are in order. It is essential to keep in mind that there are two main kinds of narremes: some are tied to the actional present, and others are connected to memories, which can be either presented as such or directly connected to other narremes without explicit markers. Another relevant aspect is that not all the narremes have the same length, but they depend on the matching event. Some are consistent parts of the narration and cover many pages, while others are only made up of one or two sentences, implying disparity in word counts. At a first glance this difference could seem a problem for the analysis, but it really only shows how the author favored some events in the narrative, giving them a higher level of importance.

As anticipated, it is possible to identify three main stages of life for the twins Waldo and Arthur, each precisely represented by a certain number of narremes. They are:

- Childhood and youth (28 narremes)
- Adulthood (69 narremes)
- Old age (19 narremes)

This count only excludes narremes in which Mrs Poulter represents the point of view, and Arthur's narremes after Waldo's death. This arbitrary distinction aims to give better insight on the lifestyle of the two brothers and on how it changes throughout the years and decades.

2.1 The Density of Semantic Features Highlights Significant Events

We considered semantic features (see Bos and Delmonte (2008)) 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. 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). 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. Table 3 illustrates the hierarchy of the various features, with the attributes represented in italics and the values in normal text. ⁴.

³The 1st Workshop on EVENTS: Definition, Detection, Coreference, and Representation, HLT-NAACL, Atlanta

⁴We omit comments on the features because all detailed information about semantic features annotation are reported in a companion paper by the same authors in a workshop in this conference (see Delmonte and Marchesini (2017))

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 3: Hierarchy of deep semantic features used in the annotation

Narreme length and number of annotations are both markers signaling importance. Even without a complex analysis, it is apparent that some events are fundamental for the story: we call these narremes “significant events”. Significant events are characterized by a dense co-occurrence of stylistic features in more general terms, as well. In the two following Figures, we show the inter-relation between number of words annotated as a ratio of the total number of tokens. The areas of main concentration of the features are coincident with the ones that have been defined as “significant events” and are respectively: 1st area: 23-27; 2nd area: 42-45; 3rd area: 53-55; 4th area: 63-68; 5th area: 74-79b; 6th area: 92-94 + U2-U3; 7th area: 103-110; 8th area 121-125. As can be easily noticed, the peaks coincide with significant events, detected through individual assignment of relevant traits. The two diagrams below show how in some narremes the number of annotations is higher than in others. It was mentioned before that the length of the units plays an important role in many cases: to longer narremes are normally, but not consistently, associated more occurrences of each semantic trait.

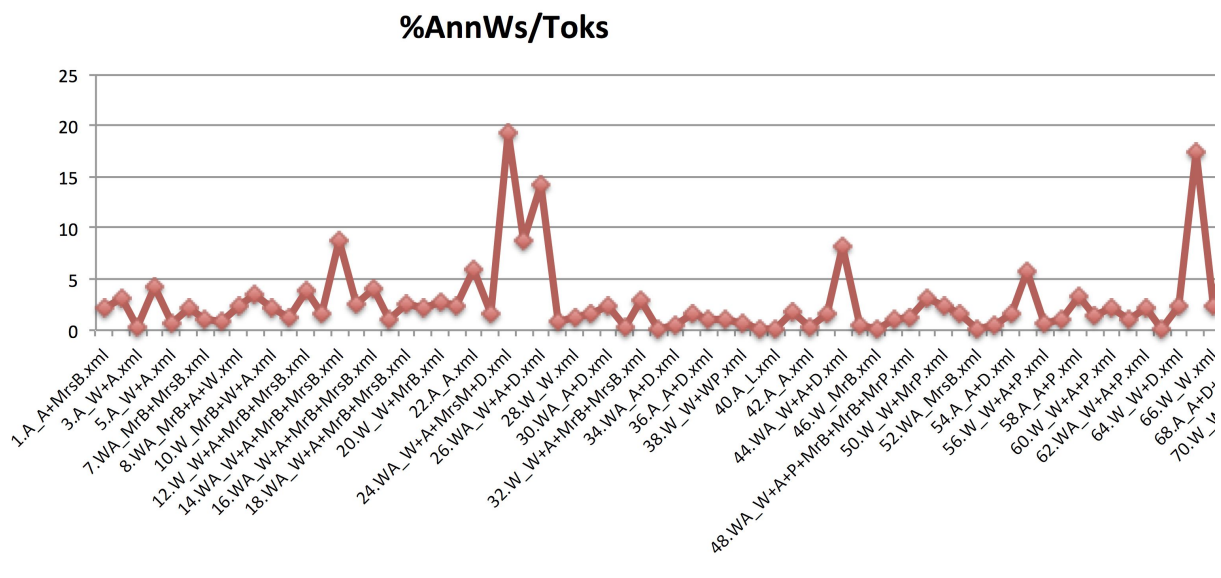


Figure 1: Distribution of words annotated in each narreme as a ratio of total tokens - first half of the novel

Narreme 104 - belonging to Waldo’s section - represents the beginning of the actional present in

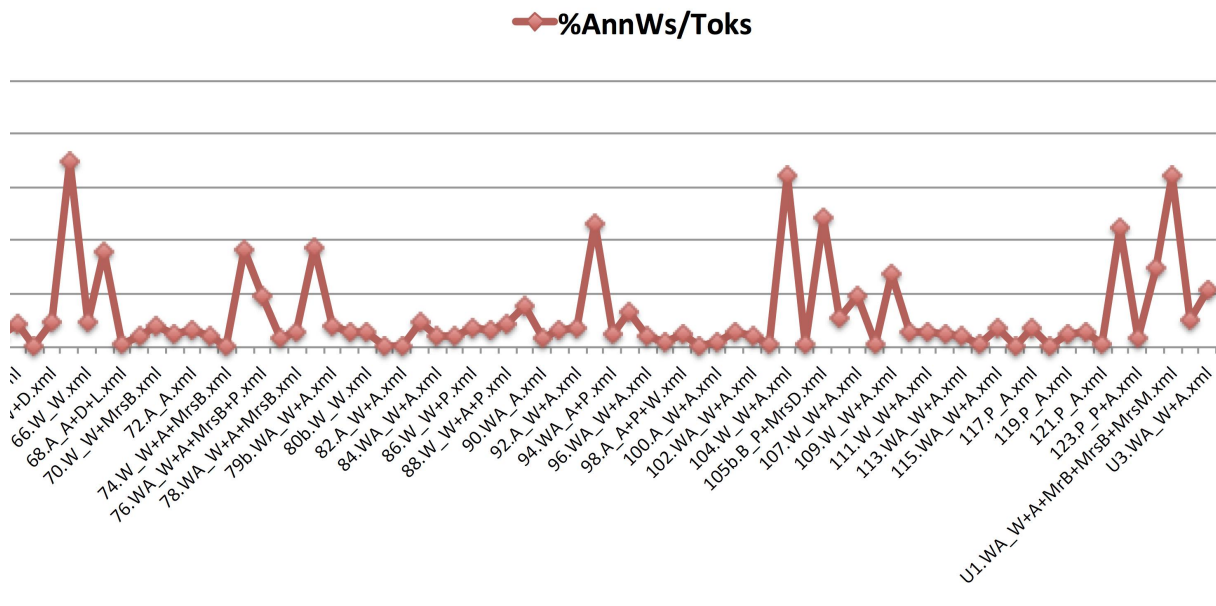


Figure 2: Distribution of words annotated in each narreme as a ratio of total tokens - second half of the novel

the *fabula*. Everything before this point is in the past and is constituted by memories. From here on, the events of the *fabula* follow the twins through Waldo's last days and, later, in Arthur's disorganized escape to Sydney. If we examine this section from a perspective involving the *sujet*, on the other hand, this is the first time in the novel that we meet Waldo as a narrator. The beginning of narreme 104 is in fact the beginning of the second section of the novel, Waldo's, immediately after the prologue-section "In the Bus". It covers roughly sixteen pages between 23 and 61, all in the first part of Waldo's section, and functions as a starting point for the stream of memories that comes afterwards. Interestingly, if we take into consideration the first segment of Arthur's narremes according to the *sujet* – first part of narreme 2, narreme 1, and then second part of narreme 2 – we notice something similar. In this case, strictly limiting the analysis at Arthur's section, the *sujet* almost coincides with the *fabula*, and we find an unusually high number of annotations if compared to most of the other narremes.

3 GETARUNS: a System for Text Understanding and Narreme Encoding

GETARUNS⁵, the system for text understanding developed at the University of Venice, is organized as a pipeline which includes two versions of the system: what we call the Partial and the Deep GETARUNS. They work in a backoff policy: there are in fact three interconnected parsers, and they are activated in order to prevent failure. The system has a middle module for semantic interpretation and discourse model construction cast into Situation Semantics, and a higher module where reasoning and generation take place.

The system is based on the LFG theoretical framework and has a highly interconnected modular structure. The output of grammatical modules is fed onto the Binding Module, which activates an algorithm for anaphoric binding. Antecedents for pronouns are ranked according to grammatical function, semantic role, inherent features and their position at f-structure. Eventually, this information is added into the original f-structure graph and then passed on to the Discourse Module (hence DM).

⁵The system has been tested in STEP competition (see Bos and Delmonte (2008)), and can be downloaded in two separate places. The partial system called VENSES in its stand-alone version is available at http://www.aclweb.org/aclwiki/index.php?title=Textual_Entailment_Resource_Pool. The complete deep system is available both at http://www.sigsem.org/wiki/STEP_2008_shared_task:_comparing_semantic_representations, and at <http://project.cgm.unive.it/html/sharedtask/>

GETARUNS has a linguistically based semantic module which is used to build up the DM. Semantic processing is strongly modularized and distributed amongst a number of different submodules. They in turn take care of Spatio-Temporal Reasoning, Discourse Level Anaphora Resolution, and other subsidiary processes, like Topic Hierarchy, which cooperate to find the most probable antecedent of coreferring and cospecifying referential expressions when creating semantic individuals. These are then asserted in the DM, which is at this point the only knowledge representation here used to solve nominal coreference. The system employs two resolution submodules working in sequence: the first one is run whenever a free sentence external pronoun is spotted; the second one takes the results of the first submodule and checks for nominal anaphora. They have access to all data structures contemporarily and pass the resolved pair of anaphor-antecedent to the following modules. Semantic Mapping is performed in two steps: at first is created a Logical Form, which is a structural mapping from DAGs onto unscoped well-formed formulas. These are then turned into situational semantics informational units, infons, which may become facts or sits. Each unit has a relation, a list of arguments, which in our case receive their semantic roles from lower processing – a polarity, a temporal and a spatial location index.

All entities and their properties are asserted in the DM with the relations in which they are involved; in turn the relations may have modifiers - sentence-level adjuncts – and the entities may have modifiers and attributes, as well. Each entity has a polarity and a couple of spatio-temporal indices linked to main temporal and spatial locations (if any exists). Otherwise they are linked to a presumed time reference derived from tense and aspect computation. On the second occurrence of the same nominal head the semantic index is recovered from the history list and the system checks whether it is the same referring expression and has non-conflicting attributes or properties. In all other cases a new entity is asserted in the DM, being however computed as included in (a superset of) or by (a subset of) the previous entity.

3.1 Coreference Links

In this section we briefly present the addition to the GETARUNS system added for the purpose of this task. The annotation of each narreme is shown in a XML file obtained with the following steps:

1. The GETARUNS system provides a deep analysis of each narreme on a sentence-by-sentence basis. At the end of this preliminary analysis, markables are collected and all semantic information is attached to each word in the sentence. With this system we collected all verbs, eventive nominals, and possible eventive modifiers. This is done in two steps:

- 1a. At the end of parsing, each word of the sentence is associated to its lemma. General semantic categories are collected from the analysis, as well.

- 1b. The system provides the steps required for building the Discourse Model, i.e. where entities, relations and properties are asserted along with their attributes. Semantic indices are assigned to each new entity, and previous mentions receive the previously assigned indices. At this point the contents of the Discourse Model are associated to each word in the sentence.

2. At the end of the analysis of the narreme the system collects all markables, which are internally made of four elements: markable index, word, lemma, semantic index (extracted from the Discourse Model) or a generic indicator of eventuality for all verbs.

3. The complete Discourse Model is searched to create a list of all entities, relations and properties, this time with spatio-temporal coordinates, relations and polarity – as documented in situation semantics. Additional information is derived in this phase from WordNet, FrameNet or SumoMilo, and is made available to the coreference algorithm. Another component activated at this point is sentiment analysis, which computes affective labels associated to each markable – if possible – and classifies each markable into three different classes: positive, negative or neutral.

4. The coreference algorithm works as follows: for each markable it checks all possible coreference links, at first only on the basis of inherent semantic features – wordform and lemma identity; then semantic similarity is measured on the basis of a number of lexically-based similarity criteria, without statistical measures. Searching WordNet synsets, we assign a score according to whether the markables are directly contained in the same synset or not. A different score is assigned if their relation can be inferred from the hierarchy.

5. After collecting all possible coreferential relations between semantically validated markables, we proceed to filter out all inconsistent or incompatible links according to three criteria: - first criterion: diverse sentiment polarity - second criterion: different argument structure - third criterion: non-related spatio-temporal relations Both argument structure and spatio-temporal relations are collected in the discourse structure, which also contains dependence relations expressed by discourse relations in discourse structure, temporal logical relations – as computed from an adaptation of Allen’s algorithm –, and a point of view computed on the basis of presence of "reportive" verbs – direct speech, reported speech, reported indirect speech. Another selected criterion relates to the nature of semantic similarities as computed by the system: values below a certain threshold indicate that the coreference was chosen on the basis of weak similarity. This last point is based on thesauri classification.

4 Computing Narreme Boundaries

As said above, narreme boundaries are identified by the presence or absence of specific speech acts, by changes in the point of view, and by movements in spatio-temporal coordinates through flashbacks. The module looking for narremes tries to individuate their boundaries, taking as input the analysis of the text at a syntactic and semantic level. To evaluate results, it makes use of the ranges of all narremes as they were manually identified through start and sentence numbers. More in detail, the input is made of the following semantic information: - the list of entities evaluated as main, secondary and potential topics of discourse by the specialized centering algorithm with Topic Hierarchy, which works on the output of the anaphora resolution modules (see Grosz and Sidner (1986) Grosz et al. (1995)); - the list of discourse structures computed one for each clause, containing information on discourse moves, relations and main predicates in each clause.

At first the algorithm creates clusters of clauses around each entity in a sequence; it then keeps only the ones constituting frequently mentioned entities, or characters of the story. It eventually extracts another subset of clauses, this time characterized by the fact that the main predicate is included in a list of speech acts – verbs like "decide", "discover" etc., and those introducing direct discourse like "say", "ask", "tell" etc. Clusters are a function of four variables: Topic linguistic description - which may be subject to different types of mentions identified and unified under the same Semantic Identifier in the Discourse Model; Topic Type, in a hierarchy including Expected, Potential, Secondary and Main; Discourse Move - one among three (Up, Down, Level) - which are computed by the Discourse Structure module and assigned to a Clause identifier; Clause Relevance, one of two possibilities, Foreground or Background - computed on the basis of Tense, Aspect and Factuality again associated to a Clause Identifier.

At this point the algorithm has all the information it needs to decide whether two adjacent clusters may be unified into a single one or be left separate. If the second case is true, the boundary is individuated in the clause following the last one of each cluster, by matching the list of speech-act-related clauses or examining the distance between the two adjacent clusters ⁶.

4.1 A Close Look at the Narreme Clustering Algorithm

We report here below the higher portion of the algorithm for narreme identification written in Prolog. In order to make the algorithm testable, we split the novel into 4 separate sections obeying criteria indicated above. The algorithm starts by loading system output for a given section of the novel and then the sentence range associated to each narreme as computed manually in order to make evaluation possible. Then it loads the structures of each clause computed at discourse level which are made up

⁶Our method is different from previous attempts at text segmentation, but is also similar in some aspects. It is different from plot-units segmentation proposed by Lehnert (1995) and lately by Goyal et al. (2010) which is solely based on affect and mental states of characters. It is different from Topic based segmentation purported by Hearst (1997) in its TextTiling system, and lately attempted in literary texts by Kazantseva and Szpakowicz (2012), because topics alone are not sufficient cue for our novel. It is also different from the approach proposed in Swanson et al. (2014) where the authors elaborate an annotation scheme at clause-level with an analysis based on Labov (1997) which distinguishes three types of clauses for narrative labeling: Action, Orientation and Evaluation. These are then used to create plot units

by a SentenceNumber, a ClauseNumber, the main verbal predicate, the Relevance computed and the Discourse Move. As a second element of the computation, the algorithm loads all topics as they have been computed by the coreference module and the topic hierarchy module. Topics are collected by their Semantic unique Identifier in the Discourse Model, the Topic Head, and the Clause number.

`findallnarremes:-`

```

consult(systemoutput1),
consult(sents_range),
findall(Nsen-Cl-Pred-Relv-DsMv, (
    sd_structure(Nsen, Cl, Sbj, DsRel, Tens, Pred, Relv, DsMv, Move)
), AllMovess),
reverse(AllMovess, AllMoves),
findall(SemId-Head-Cl, (
    topic(Cl, Type, SemId),
    topps(Ty, Cls, Head, SemId, [P,G,N, Sf, Sr])
), AllCharss),
sort(AllCharss, AllChars),
unifyallChars(AllChars, Uniques),
eliminacharsminors(Uniques, MainChars),
elaborate_output(AllMoves, MainChars).

```

As a first step, the system unifies all Topics or Characters of the novel by their frequency - in number of clauses in which they are present-, choosing the most frequent ones, thus indirectly selecting the protagonists. Minor characters are filtered and deleted from the list. The net result is a list of protagonists with their Semantic identifier and all clauses in which they appear, which is accompanied by all moves - in the variable AllMoves - in discourse structure and passed to the call "elaborate output".

`elaborate_output(AllMoves, MainChars):-`

```

clusteringchars(MainChars, Clusters),
createclusters(AllMoves, Clusters, Created),
extractculminated(AllMoves, Culminated),
countranges(1, Created, Outputs),
searchadjacents(Culminated, Outputs, Output),
tell(allranges),
writeoutputall(Output),
told,
!.

```

The final call of the algorithm does the following actions. It finds clusters of events associated to the main character/s. The first call "clusteringchars" decomposes the single list of clauses related to a single character into a set of ClauseNo-TopicHead-SemIdentifier triples, in order to reconstruct the "fabula"-based sequence of events per each protagonist. Then "createclusters" associates discourse level information and sentence number to each TopicHead thus make a quadruple, SentNumb-ClauseNo-TopicHead-Move. Then it creates a cluster by unifying in the same sublist all quadruples which have a Clause Number smaller than a threshold computed on the basis of the difference between the current clause number and the new one. This difference has been set to 7. In this way clusters may contain clauses referable to same or different character/s as long as they are close enough. The following call, "extractculminated" chooses those discourse structures that have "culminated" verb predicates associated, i.e. special events that can mark the beginning of a new narreme and that will be used to individuate narreme boundaries. "Countranges" filters those clusters which are too small, i.e. smaller than 10 elements. Eventually the last call "searchadjacents" glues clusters which are close enough into one single "narreme" also checking whether the beginning and end are suitable.

Results from the analysis of the output of the whole of Waldo's and Arthur's chapters indicate an accuracy of 48%, which is the best result obtained with various attempts using different weights on the

data available. In particular, we tried at first to make the choice of the ending boundary dependent on the presence of a new entity being asserted as Potential Topic, i.e. having entered the Entity List for the first time. Then we modified the algorithm to take into account presence of a clause with a predicate included in the list of speech acts. We eventually combined the two choices and got the best result, which is still however below 50%.

5 Conclusion

In this paper we focused on the analysis of a novel, "The Solid Mandala" (1966) by Nobel laureate Patrick White, annotating its whole text with semantic features. In addition, we manually subdivided the text into narremes, basing our work on the reconstruction of the *fabula* seen in Collier (1992). The problematic aspect of White's style is that it merges plot characteristics – connected to the *fabula* – with character traits. As a consequence, feelings, emotions and actions of the narrating characters are all deeply intertwined with the plot and the narrated events.

We tried an automatic extraction of narremes from the output of a system for text understanding, which is strongly semantically driven. Results are still however below the 50% threshold, suggesting that the level of difficulty of the task is extremely high.

References

- Bal, M. (1987). *Narratology*. Toronto: University of Toronto Press.
- 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 the annotation of narrative style. In *Proceedings of IWCS, International Workshop on Computational Semantics*, Stroudsburg, PA, USA, pp. 1–12. ACL.
- Dorfman, E. (1969). *The narreme in the medieval romance epic: An introduction to narrative structures*. University of Toronto Press.
- Goyal, A., E. Riloff, and H. D. III (2010). Automatically producing plot unit representations for narrative text. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing - EMNLP 2010*, Stroudsburg, PA, USA. ACL.
- Grosz, B., A. K. Joshi, and S. Weinstein (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* 21(2).
- Grosz, B. and C. Sidner (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics* 12(3).
- Hearst, M. A. (1997). Texttiling: Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics* 23.
- Kazantseva, A. and S. Szpakowicz (2012). Topical segmentation: a study of human performance. In *Proceedings of Human Language Technologies: The 2012 Annual Conference of the NAACL*, Stroudsburg, PA, USA, pp. 211–220. ACL.

- Labov, W. (1997). Some further steps in narrative analysis. *Journal of narrative and life history* 7, 395–415.
- Lehnert, W. G. (1995). Plot units and narrative summarization. *Cognitive Science* 5(4), 293–331.
- Swanson, R., E. Rahimtoroghi, T. Corcoran, and M. A. Walker (2014). Identifying narrative clause types in personal stories. In *Proceedings of the SIGDIAL 2014 Conference*, Stroudsburg, PA, USA, pp. 171–180. ACL.
- Wittmann, H. (1975). Thorie des narrèmes et algorithmes narratifs. *Poetics* 4(1), 19–28.