Segmenting Narrative Synopses on Event Reporting Mode based on Heuristics on Constituency Parses

Pablo **Gervás**[1]

Facultad de Informática, Universidad Complutense de Madrid, Madrid, 28040 Spain

Abstract
A narrative of relative complexity will often include events beyond simple facts that happen in the story world–such as plans, orders, wishes of the characters– or references to events that happen in a different story world–stories or anecdotes being told by the characters. These correspond to different modes of reporting events. Identifying spans of text corresponding to different modes of reporting is a significant challenge. The present paper proposes a mechanism for segmenting the text of synopses of narrative plots into spans that correspond to different views of the storyworld in terms of temporal, spatial or modal coordinates. This is achieved by considering syntactic structure to identify cues for the start of embedded discourses and continuity over features such as tense, voice or mode to identify the points where the embedded discourse ends. This process can handle embedded discourses that span over several sentences and recursive nesting of discourses. The solution is tested against a corpus of synopses hand-annotated for start and end of embedded discourses. Acceptable precision and recall metrics are reported.

Keywords
modes of reporting events, embedded discourse, plot synopses, Stanford Core NLP

1. Introduction

The events included in a narrative do not always constitute a sequence of events that are equally true in the storworld in a steady succession. Sometimes the narrative refers to events that are not held to be true–but rather inform about wishes or plans of characters–or that were true at an earlier point in time. The start of the spans of text affected by such instances are often indicated linguistically by means of specific syntactic constructions, such as infinitive sentences, modal verbs or reported speech constructions. The end of the span of text that is affected may be indicated by changes in tense or voice. These phenomena constitute an important challenge for the interpretation of narrative text, because even assuming a certain ability to construct a conceptual representation of each event from the given text, the interpreter is still faced with the need to decide in what mode–as a truth or as a wish–and with what temporal and spatial coordinates to add it to the representation being constructed for the storyworld.

Examples of this type of sentence are: "the girl orders the gendarmes to destroy the garden" or "The forecaster warns the weather may be cold." These sentences describe a principal event–an order or a warning–which involve a secondary event–destroying the garden or the coming of cold weather. In each case, the meaning of the verb includes connotations about the distance between the secondary event and the principal event. Sometimes a sentence introduces an embedded discourse and then the reporting mode introduced by that embedded discourse continues over a number of sentences. Consider the following example:

> A tsar has his queen and their son Ivan. The groom predicts [a sister will be born who shall be a terrible witch. The sister shall devour her father and mother and all people under their command.] Ivan asks the tsar for permission to go out for a walk.

The second sentence introduces a prophecy. The third sentence continues describing parts of the prophecy. The fourth sentence returns to the telling of the main story, and no longer refers to the prophecy. In terms of changes in reporting mode, the span of text corresponding to the prophecy should be identified as the one marked between square brackets. This is the type of task we want to address. We work on the hypothesis that such changes in mode of reporting may be flagged by changes in the features of the clauses that determine the time/space coordinates for the event: tense, voice, mode. In the example, the events corresponding to the prophecy are all presented in future tense.

The present paper explores the viability of developing such a computational solution for identifying spans of text corresponding to different modes of reporting events based on the Stanford Core NLP tool for English [1].

2. Previous Work

The distinction between modes of reporting for a narrative that is considered in this paper is based on pragmatic criteria: it is intended to inform a process of representation of the information contained in the narrative, and it is concerned with clustering the events reported into sets that are attributed a similar degree of certainty in the same possible world.

From the point of view of linguistics, there are a wide range of phenomena that might influence this kind of classification/clustering task. The most relevant is *modality*, defined as the way in which statements in a language may be marked in terms of their relation with reality or truth [2]. In the particular case of English, modality is most often expressed in terms of auxiliaries–such as *may* or *can* but may also be expressed lexically–with verbs such as *want* or *need*.

Another relevant phenomenon is *reported discourse*, which appears when an agent reports discourse originally contributed by a different agent, and which may itself be an utterance or a belief [3]. Reported discourse may appears as *direct speech*–the reporting agent conveys the exact words of the reported agent–or as *indirect speech*–the reporting agent paraphrases what the reported agent said. Direct speech is typographically marked by presenting the reported discourse between quotes. Indirect speech is usually introduced by a complementizer– in English *that*.

Instances of reported discourse may be further characterised by their *attribution relations* [4]. The attribution relation for a statement involving reported discourse that explicitly encode the reported discourse, the speaker (or source) and some cue indicating the attitude attributed to the speaker. Attributions are marked in English by the use of specific prepositions, specific lexical phrases, specific reporting verbs and the verbs preceding *that*-clauses [5].

The work of [4] develops a system for the automatic extraction of attribution relations. The proposed system relies essentially on a k-NN classifier for identifying verbs that act as cues for attribution, a conditional random field labeller to identify the span of text that corresponds to the reported discourse, and a logistic regression classifier to identify the entity that is presented as the source. A number of additional components help to refine the specific entities involved in cases where they are conveyed in the text by complex expressions spanning several tokens.

A different approach addresses the task of segmenting plot synopses identifying turning points in the narrative [6]. This system relies on a corpus of synopses annotated with turning points–stages in the structure of the narrative–over which a neural network model is trained to identify the turning points. One important argument presented by these authors is that working on synopses presents significant advantages: (1) the shorter format makes annotation easier, so the effort is easier to scale, (2) interannotator agreement is likely to be higher for synopses, given that synopses are written at a higher level of abstraction.

When characters in a story report events by themselves telling a story, each such telling is considered to introduce a new *narrative level*. The task of annotating narrative levels over long texts has been addressed by [7]. The authors outline the difficulties presented by a lack of corpora annotated with the relevant information and propose a solution based on extending a small annotated sample with synthetically created data. In terms of the task itself, they propose a model of the task as one of segmentation of the text by identifying boundaries of narrative levels in the text.

The procedure described in this paper relies on the constituency parse provided by the Stanford Core NLP toolkit [1]. The constituency parse for a sentence is used as the main data structure to drive the process because it respects the relative order of appearance of the words in the input text. This is relevant for the procedure employed for identifying separate spans of text that correspond to different modes of reporting events, as they will be assigned to continuous spans of the text.

3. Segmenting Narrative Text into Spans with Different Reporting Mode

A faithful reconstruction of the meaning of a given story requires that the different modes of reporting be identified and correctly assigned to each and every one of the events mentioned. One possible way of achieving this is by identifying the spans of text corresponding to each reporting mode and the set of events mentioned in each span.

Table 1
Statistics of reporting mode spans in the manual annotation of the corpus of Propp tales.

Total number of sentences:	344
Total number of nestings:	66
Maximum depth of nesting:	3
Average depth of nesting:	1.05
Total number of words:	2607
Total number of words in embedded levels:	401
Percentage of the total number of words that appeared embedded:	15/100

Table 2
Statistics of reporting mode spans in the manual annotation of the corpus of Madamme d'Aulnoy tales.

Total number of sentences:	803
Total number of nestings:	178
Maximum depth of nesting:	4
Average depth of nesting:	1.11
Total number of words:	5390
Total number of words in embedded levels:	867
Percentage of the total number of words that appeared embedded:	16/100

3.1. A Corpus of Plot Synopses

Because the aim of this research initiative is to explore the application of text processing tools to the task of identifying the meaning of a narrative in terms of the storyworld it describes, it was considered more practical to operate on synopses of larger narratives rather than on the narratives themselves. This is based on the assumption that synopses attempt to condense the structure of the plot for a given narrative, leaving out the details that add value to the work but not necessarily to the general structure.
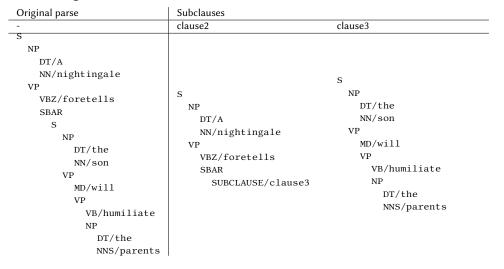
To inform this process, a small corpus of plot synopses has been considered. This corpus draws upon two different sources: the synopses of the set of plots for Russian folk tales originally analysed by Vladimir Propp [8] and the synopses of the set of French fairy tales by Madamme d'Aulnoy [9] as annotated in terms of Proppian character functions by [10]. This corpus has been manually annotated by hand to mark the start and end of all instances of embedded discourse which differ from other contributions to the text in the relative distance between the subject of the main clause and the reality of the embedded discourse. These distances sometimes involve modality, time or even a different storyworld.

The statistics of the spans of reporting mode that appear in the manual annotation are presented in Table 1 for the Propp tales and in Table 2 for the D'Aulnoy tales.

3.2. A Text Segmentation Process

The process of text segmentation needs to be informed by the following factors: the set of events reported by the text (often a single sentence in the text reports more than one event,

Table 3
Example of input sentence and breakdown into subclauses, showing relative order of presentation and relative co-indexing.

Original parse	Subclauses	
-	clause2	clause3
| ```
S
  NP
    DT/A
    NN/nightingale
  VP
    VBZ/foretells
    SBAR
      S
        NP
          DT/the
          NN/son
        VP
          MD/will
          VP
            VB/humiliate
            NP
              DT/the
              NNS/parents
``` | ```
S
  NP
    DT/A
    NN/nightingale
  VP
    VBZ/foretells
    SBAR
      SUBCLAUSE/clause3
``` | ```
S
  NP
    DT/the
    NN/son
  VP
    MD/will
    VP
      VB/humiliate
      NP
        DT/the
        NNS/parents
``` |

each represented by a particular subclause), the syntactic structure of sentences whose verbs indicate a specific reporting mode (such sentences are likely to imply a change in reporting mode between the event reported by the main clause and any reported in clauses subordinate to the main verb) and punctuation marks used to indicate relative groupings of subclauses (commas, brackets, or quotation marks).

An extraction procedure is proposed that builds a sequence of subclauses for each sentence–each corresponding to a subtree describing an event. We develop heuristics for identifying points in the narrative where the events being told differ significantly in the mode in which they are reported, and consider these points as candidates for breaks between spans for different reporting modes.

The different reporting modes that we have identified as being marked in this way are: events that occur in significantly different *temporal* coordinates (flashbacks, flashforwards), events that are narrated in a different *modality* (wishes, plans, prophecies, curses, orders), events that are specifically marked as being *narrated* by particular characters (indirect speech, reported speech).

3.2.1. Sub-Clause Extraction

A sentence with subordinate clauses will need to be broken down into its constituent parts. The relations between these parts are captured by assigning labels to identify subordinate clauses and adding these labels in argument positions in the main clause to represent the subordination relation.

The result of this process is a basic tree that represents the main clause as a simple structure with either noun phrases or subclause identification nodes as arguments. An example of this process is shown in Table 3.

Table 4
Example of subclause breakdown for past participle sentence.

Original parse	Subclauses	
-	clause5	clause6

```
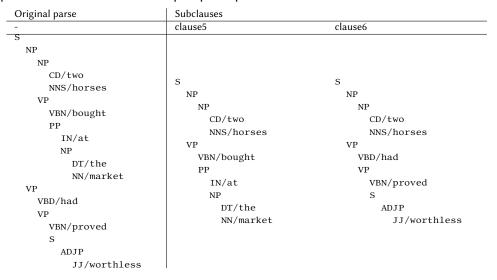S                      S                      S
  NP                     NP                     NP
    NP                     NP                     NP
      CD/two                 CD/two                 CD/two
      NNS/horses             NNS/horses             NNS/horses
    VP                     VP                     VP
      VBN/bought             VBN/bought             VBD/had
      PP                     PP                     VP
        IN/at                  IN/at                  VBN/proved
        NP                     NP                     S
          DT/the                 DT/the                 ADJP
          NN/market              NN/market                JJ/worthless
  VP
    VBD/had
    VP
      VBN/proved
      S
        ADJP
          JJ/worthless
```

A recursive search for subclauses is applied. Recursion allows consideration of subclauses that appear as adjuncts to noun phrases in other clauses. The type of subclauses that may appear nested within a noun phrase in the constituency trees produced by the Stanford parser are: relative clauses ("[The boy who was hungry] asked for a snack", past participle sentences ("[The horses bought at the market] proved worthless") and gerund sentences ("He found [a ring belonging to his mother]"). In each example, the part of the sentence corresponding to the subclause attached to the noun phrase has been marked in brackets. When extracting this type of subclauses, duplication of the subject may be required. An example of this process is shown in Table 4.

3.3. Identifying Embedded Discourse Boundaries

For each sentence, the set of subclauses extracted from it are ordered so that relations of embedding are presented in the same order as in the discourse: main clause followed by embedded clause. The procedure is applied in succession to the resulting sequence of subclauses.

For each clause identified in the manner described above, the system extracts the following information: voice, tense and main verb. The system also considers the following additional aspects that play a role on boundary identification: whether the clause includes a subordinated clause, whether the main verb of the clause has connotations of change in reporting mode, whether the clause includes a clause-grouping punctuation sign (opening or closing). Finally, given the recursive nature of embedded discourse, the system maintains stack data structures that allow it to keep track at any given point of: the depth of levels of quotation and the depth of levels of embedding.

At each point, a look ahead stage has been added to check whether: the following clause starts with quotation marks, or the following clause differs in tense/voice/mode from the current one.

Table 5

Metrics on precision and recall over extracted features for the set of Propp tales. Values are shown for span starts (S), span ends (E) and specific spans as annotated in the corpus (sp). In each case, precision is indicated by a P suffix and recall by a R suffix.

SP	SR	EP	ER	spP	spR
0.84	0.62	0.84	0.62	0.95	0.92

Table 6

Metrics on precision and recall over extracted features for the set of D'Aulnoy tales. Values are shown for span starts (S), span ends (E) and specific spans as annotated in the corpus (sp). In each case, precision is indicated by a P suffix and recall by a R suffix.

SP	SR	EP	ER	spP	spR
0.90	0.40	0.85	0.37	0.80	0.89

The heuristic that decides at each point of the discourse whether to consider the start of a span with a different reporting mode for events relies on the following conditions: whether the syntactic structure of the given clause includes embedded subclauses, whether the main verb of the clause is considered to change the mode of reporting–as defined by a resource file that lists verbs of this type–and whether the clause start with opening quotation marks. The decision on the end of a span of discourse relies on the following conditions: whether the clause ahead involves changes in tense, voice and mood; whether the current clause already opened a span of reported discourse–which would make it a main clause and necessarily followed by at least one subordinate subclause– and whether closing quotation marks appear at the end of the subclause.

4. Discussion

The described solution can be discussed from two points of view: in terms of recall and precision metrics of the outcomes of the solution with respect to the manually annotated reference corpus and in terms of how it compares with existing solutions for similar tasks.

4.1. Metrics over Reporting Mode Spans

To provide a quantitative evaluation of the quality of the results, we consider: precision and recall over positions of embedded story starts, precision and recall over positions of embedded story ends, and precision and recall over bags of words identified for specific spans as annotated in the reference corpus. The current outputs for the Propp tales are shown in Table 5 and current outputs for the D'Aulnoy tales are shown in Table 6.

Results for the two different data sets are shown separately due to the significant differences that exist between them. The synopses for Propp tales are telegraphic transcriptions of the formal analysis described in shorthand in Propp's book. The synopses for D'Aulnoy's tales are more elaborate descriptions provided by Williams in her thesis [10]. The differences in the complexity of the language employed give rise to the differences in scores. Nevertheless, results

are comparable across the two sets.

Reasonably high (0.84 and 0.90) precision on starts means that most of the starts of embedded stories found by the parse are indeed starts that appear marked in the reference. The lower values (0.62 and 0.40) for recall on starts means that a certain proportion of the starts marked are not being located correctly. This suggests that there are further features that identify the start of these spans, beyond verbs that indicate changes in reporting mode, that would need to be considered. Reasonably high (0.84 and 0.85) precision on ends means that most of the ends of embedded stories found by the parse are indeed ends that appear marked in the reference. Low (0.62 and 0.37) recall on ends means that a large proportion of the ends marked are not being located correctly. However, it is important to note that, because the procedure treats the location of starts and ends separately, the values for location of ends will be considerably affected by any errors in the identification of the corresponding starts. Relatively high values on precision (0.95 and 0.80) and recall (0.92 and 0.89) for the identification of words belonging to specific spans suggest that, once the values for locating starts and ends are improved, the procedure would work well.

It is also important to consider that some of the incorrect diagnoses leading to low values in precision and recall may be tracked back to error rates in the parser. The results of the parser chosen for the task very often misrepresent the relations between the constituents of the sentence. For instance, the parses of sentences such as "the girl orders the gendarmes to destroy the garden" consider that "the gendarmes" as subject of the clause "to destroy" and not as direct object of the clause "orders". These may explain why precision and recall for retrieving words for a specific span do not reach maximal values. A solution enhanced with the results of a dependency parse will be considered as further work and should help improve results.

4.2. Comparison with Prior Work

The fundamental challenge for the development of adequate machine learning solutions to tasks of segmenting/classifying reported discourse in text has been identified as the lack of training resources of appropriate size and coverage, both in the context of attribution relations [4] and narrative levels [7]. The corpus developed by [4] constitutes a valuable resources for the specific task of identifying attribution relations, but it does not provide coverage for phenomena beyond that. The work reported in [7] proposes a solution that relies on the generation of synthetic data to augment automatically a small data set used as seed. The resulting data set is valuable for the annotation of narrative levels but does not consider other phenomena.

As long as there are no data sets large enough to train machine learning solutions to the task, the proposed solution may provide a simple baseline that obtains results with acceptable values of precision and recall over basic features. Whereas the segmentations generated are clearly far from perfect, they can provide a valuable starting point for any efforts on developing corpora with the assistance of human annotators. At worst, they can be considered as a way to help create the seed data sets required for automatic augmentation via data synthesis.

5. Conclusions

A procedure for detecting spans of text that correspond to different modes of reporting events has been proposed. The proposed procedure operates directly on a syntactic parse of a story, without requiring any special depths of semantic analysis. It is also designed to operate on inputs based on their syntactic structure and a lexical resource that identifies verbs which indicate a change in reporting mode. These characteristics allow it to operate as a baseline solution for obtaining an initial segmentation in the absence of sufficient volumes of annotated data to train machine learning solutions.

Acknowledgments

References

[1] C. D. Manning, M. Surdeanu, J. Bauer, J. R. Finkel, S. Bethard, D. McClosky, The Stanford CoreNLP Natural Language Processing Toolkit., in: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014, June 22-27, 2014, Baltimore, MD, USA, System Demonstrations, The Association for Computer Linguistics, 2014, pp. 55–60.

[2] P. Portner, Modality, Modality, OUP Oxford, 2009.

[3] D. Y. Oshima, Perspectives in reported discourse, Ph.D. thesis, Stanford University Stanford, 2006.

[4] S. Pareti, Attribution: a computational approach, Ph.D. thesis, The University of Edinburgh, 2015.

[5] A. C. Murphy, Markers of attribution in english and italian opinion articles: A comparative corpus-based study, ICAME journal 29 (2005) 131–150.

[6] P. Papalampidi, F. Keller, M. Lapata, Movie plot analysis via turning point identification, arXiv preprint arXiv:1908.10328 (2019).

[7] N. Reiter, J. Sieker, S. Guhr, E. Gius, S. Zarrieß, Exploring text recombination for automatic narrative level detection, in: Proceedings of the Thirteenth Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2022, pp. 3346–3353.

[8] V. I. Propp, Morphology of the folktale, University of Texas Press, 1968.

[9] M. d'Aulnoy, Contes des fées, L. Duprat-Duverger, 1866.

[10] E. D. Williams, The Fairy Tales by Madamme d'Aulnoy, Ph.D. thesis, Rice University, Houston, Texas, 1982.