

Constructing an Annotated Story Corpus: Some Observations and Issues

Oi Yee Kwong

Department of Chinese, Translation and Linguistics
City University of Hong Kong
Tat Chee Avenue, Kowloon, Hong Kong
E-mail: Olivia.Kwong@cityu.edu.hk

Abstract

This paper discusses our ongoing work on constructing an annotated corpus of children's stories for further studies on the linguistic, computational, and cognitive aspects of story structure and understanding. Given its semantic nature and the need for extensive common sense and world knowledge, story understanding has been a notoriously difficult topic in natural language processing. In particular, the notion of story structure for maintaining coherence has received much attention, while its strong version in the form of story grammar has triggered much debate. The relation between discourse coherence and the interestingness, or the point, of a story has not been satisfactorily settled. Introspective analysis on story comprehension has led to some important observations, based on which we propose a preliminary annotation scheme covering the structural, functional, and emotional aspects connecting discourse segments in stories. The annotation process will shed light on how story structure interacts with story point via various linguistic devices, and the annotated corpus is expected to be a useful resource for computational discourse processing, especially for studying various issues regarding the interface between coherence and interestingness of stories.

1. Introduction

Story understanding has been a notoriously difficult topic in natural language processing. In addition to basic lexical, syntactic, semantic, and discourse processing, the task must also draw on extensive common sense and world knowledge.

As far as discourse processing is concerned, the role of story structure has received much attention in past studies. Nevertheless, stories often exhibit something more than just a coherent structure. As a special kind of discourse, stories are often characteristic for their "interestingness". They have a "point". The interface between story structure and story point, however, has not been sufficiently addressed in the literature. We therefore aim at filling this gap by first constructing a story corpus annotated with various kinds of structural and semantic information deemed important for their comprehension. The annotation process will enhance our understanding on how story structure interacts with story point via various linguistic devices, and the annotated corpus is expected to be a useful resource for computational discourse processing and related studies.

This paper discusses our ongoing work on the construction of such a corpus. To start with, we will focus on one special type of story, namely fables. Fables are chosen for their surface simplicity and deep complexity, as will be elaborated in Section 3. Based on the observations from introspective analysis, we propose a preliminary annotation scheme covering the structural, functional, and emotional aspects connecting discourse segments in a coherent story. Pilot annotation based on part of this scheme is underway, and we will tentatively refer to the resulting corpus as SPAS (Structure and Point Annotation of Stories).

The background and motivation for constructing SPAS will be further explained in Section 2. The observations

from introspective analysis on story comprehension will be discussed in Section 3. Section 4 outlines a preliminary annotation scheme and reports on our pilot annotation based on the proposed scheme. Section 5 discusses some potential issues to be investigated with the resulting corpus, followed by future directions and conclusion in Section 6.

2. Background and Motivation

2.1 Discourse Structure

The seminal paper by Grosz and Sidner (1986) on attention, intentions, and the structure of discourse, gives one of the most comprehensive computational theories of discourse structure to date, which has been shown to satisfactorily account for interruptions in dialogues in particular. The structure of any discourse is composed of three interacting components: a structure of the actual sequence of utterances, a structure of intentions, and an attentional state. Their combination thus gives a comprehensive account of discourse structure and the interpretation of certain discourse relations and referring expressions. The linguistic structure consists of the discourse segments and some embedding relationship that can hold between them. The intentional structure accounts for the overall purpose of the discourse known as the discourse purpose, and the purposes of its component segments known as discourse segment purpose. The attentional state is an abstraction of the participants' focus of attention as their discourse unfolds, dynamically recording the objects, properties, and relations that are salient at each point in the discourse, modelled by a set of focus spaces. The intentional structure is apparently the most difficult to identify as it might not be readily indicated by surface linguistic devices, and is at the same time closely related to discourse participants' beliefs and shared knowledge.

Another classic computational theory for discourse is the more functionally oriented Rhetorical Structure Theory (RST) proposed by Mann and Thompson (1987). RST aims at giving a descriptive account of discourse relations holding between adjacent text spans, thus indicating the coherence and structure exhibited among natural text. A text is divided into units, essentially clauses, based on some theory-neutral classification, which are hierarchically structured and functionally organized with respect to a set of discourse relations like EVIDENCE, ELABORATION, CONCESSION, etc. RST accounts for mostly the intentional structure in Grosz and Sidner's theory. Although the discourse relations in RST might have mixed different levels of analysis (e.g. Moore and Pollack, 1992), their identification relies more explicitly on cue phrases in the surface text.

2.2 Story Structure and Story Grammar

Knowledge of discourse structure is critical to discourse understanding in computational linguistics, especially for the analysis and processing of various discourse phenomena like temporal relation extraction, anaphora and co-reference resolution, subjectivity identification, etc. Researchers have thus attempted to trace the patterns responsible for coherence and develop "text grammars" in various forms, for discourse in general (e.g. Halliday and Hasan, 1976; Hobbs, 1982; Scha and Polanyi, 1988) and for specific genres such as stories (e.g. Rumelhart, 1975), news articles (e.g. van Dijk, 1985), conversations (e.g. Reichman, 1985), and argumentative discourse (e.g. Cohen, 1987).

Stories have their own typical structure which distinguishes them from other genres. Stories are created and retold, during which the details may not be repeated verbatim, but people apparently have a common understanding of the elements required to tell a sensible story. The proper organisation of such elements is essential to story understanding and recall. While the concept of story structure is mostly accepted (e.g. since Propp's (1968) analysis of Russian folktales), and its psychological validity has been demonstrated (e.g. Mandler and Johnson, 1977; Haberlandt, 1980), the appropriate formalism to represent such structure remains controversial. Rumelhart (1975) put forth a story schema in the form of phrase structure grammar rules, and it soon invited vigorous debates. Criticisms include the "constituents" in the grammar rules are constrained to appear in sequential order but that is hardly true in real stories, and they fail to isolate content from structure. Moreover, Rumelhart's story grammar was intended for written monologues. For stories involving more than one protagonist and conversations between protagonists, as well as with direct speech from individual protagonists, it is often difficult to assign any of the constituent in story grammar to a text segment in the story.

2.3 Story Point

Amongst others, Wilensky (1982) argued most strongly that the surface ordering of the propositions does not

account for a good story, but the "point" of a story matters more than anything else. Work on story understanding along this line has thus focused on knowledge structures like scripts and frames (e.g. Cullingford, 1986; DeJong, 1982), and goals and plans (e.g. Schank and Abelson, 1977; Lehnert, 1979; Wilensky, 1983), capturing typical events, course of actions, and memory structures expected in various scenarios. Current research in artificial intelligence on story telling also mostly focuses on comprehensive world knowledge databases and planning strategies (e.g. Oinonen et al., 2006).

2.4 Our Work: Structure-Point Interface

In view of the conflict between story grammar and story point, which has remained largely unresolved, we attempt to construct an annotated corpus of children's stories. Through the structural and semantic annotations capturing discourse coherence and plot development, we can revisit the notions of story grammar and story point for possible reconciliation and see how they interact in a story via various surface linguistic devices. While the structural components can follow Rumelhart's constituents in story grammar, and discourse relations can follow RST, it is less obvious as to how the "point" of a story could be annotated. The following introspective analysis on story comprehension illustrates the subjectivity of story understanding, and thus the difficulty for annotating story point reliably. Nevertheless, the analysis also suggests some textual clues which might somehow relate to the story point, thus making at least a partial annotation of it feasible.

3. Subjectivity of Story Understanding

Story processing and comprehension has been studied by psycholinguists with respect to memory structure (e.g. Mandler and Johnson, 1977). Hence the unfolding of episodes in a story is closely associated with the units for remembrance of the various parts of a story.

Story retelling is another way to probe the relation between the structure of stories and memory. A story can be retold in numerous ways, suggesting there are core and optional elements. The ability to retell the core elements is an important sign of story comprehension. However, retelling could also be a result of rote learning.

A better or more convincing way to show understanding, especially by machine, is to answer questions related to the story. This can involve the recall of the events explicitly mentioned in the story, or more importantly questions requiring more sophisticated inferencing (e.g. Charniak, 1972).

However, story understanding is by and large a subjective experience. People's interpretation of stories might not always agree with each other. In the following we will demonstrate the subjectivity of the task, based on some introspective analysis.

We have chosen to start with stories from Aesop's Fables. Fables form a special kind of stories. They are simple on the surface, often consisting of only a few sentences in the original version. For specific readers, such as children,

the fables might be retold in a longer way with more elaboration and details. Despite the surface simplicity, fables are in fact very difficult to process as they often have an underlying message, or the moral, to convey. Whether a reader can readily draw an association between the story and the moral depends on how well they understand the story, which often requires sophisticated processing, posing a challenge for both humans and machines. Thus what can be learned from the analysis of fables could be very useful and applicable to understanding other simpler stories.

3.1 Example: The Hare and the Tortoise

Let us start with a well-known fable “the Hare and the Tortoise”. Three annotators were asked to read one version of the story, and for each sentence, indicate the keywords which bear some relation with the moral. They were also asked to state the information provided by and inferred from such keywords, and how they think such information relates to the moral. Table 1 shows the results from three annotators (indicated as 1 to 3) on the first, second and sixth sentence of the story.

It is quite clear from Table 1 that the semantic interpretations involved are considerably subjective and are very difficult to be measured in terms of inter-annotator agreement. For the keywords, for which the unit of interpretation is more restricted, we could still see how many words are found in common from the various annotators. For the main information, although the expressions might be different, the annotators tend to agree with one another most of the time, e.g. “... tortoise was not bothered” and “... tortoise doesn’t care” for Sentence 1, and “The tortoise thought she could win a race” and “The tortoise thinks she can win” for Sentence 2. For the inferred information and the relation with the moral, however, it would be far too difficult to control on the annotators where world knowledge is involved. This type of human judgement information is nevertheless important for informing the development of higher level language understanding systems.

So where is the “point” of the story? The readers expect the hare to win all along until an unexpected result comes up toward the end of story. How can an annotator mark this up in the text? Are there any more objective and consistent way to annotate this?

Keywords	Main Information	Inferred Information	Relation with Moral
<i>Sentence 1: A hare one day ridiculed the short feet and slow pace of the Tortoise, who replied, laughing:</i>			
1. ridiculed, laughing	Hare teased tortoise of his physical shortcoming but tortoise was not bothered.	Hare is arrogant and makes fun of others' shortcomings, but tortoise does not feel bothered by the tease.	Shows that tortoise has shortcomings but was not bothered.
2. ridiculed	A hare ridiculed the tortoise.	The hare thought he is stronger than the tortoise.	Even though the hare was stronger than the tortoise, it was not a must for him to win the race.
3. ridicule, short, slow, laughing	The hare thinks that the tortoise is inferior, but the tortoise doesn't care.	The hare runs more quickly than the tortoise.	The hare thinks she can win in any case, but ...
<i>Sentence 2: "Though you be swift as the wind, I will beat you in a race."</i>			
1. swift, race	There will be a race between hare and tortoise and the latter knows that the former runs fast.	Despite the shortcomings it has, tortoise is confident about winning the race, it is optimistic.	Tortoise is confident, not let down by the fact of being disadvantaged.
2. swift, beat	The tortoise thought she could win a race.	The hare is challenged by the Tortoise.	Nil
3. swift, beat, race	The tortoise thinks she can win.	The tortoise guess (implied by the use of subjunctive) the hare can run quickly, but she has some way to beat her down.	The tortoise does not admit she's inferior.
<i>Sentence 6: The Hare, lying down by the wayside, fell fast asleep.</i>			
1. fast asleep	Hare stopped during the race.	Hare is not hard-working and does not focus on the goal.	People can only reach their goals with hard work and focus on the goal.
2. lying down, fell fast asleep	The hare lying down.	The hare had stopped during the race.	Even though the hare ran faster than the tortoise, he stopped and looked down upon his competitor, so he could not win the race.
3. lying down, asleep	The hare stopped.	The hare does not keep her goal and thinks she can still win even if she resumes after taking a rest.	Arrogance is not a good thing.

Table 1 Preliminary Story Analysis by Three Annotators

3.2 Example: The Fox and the Grapes

The interestingness of a story often arises from a twist in the story. In many fables, this is often manifested through a contrast between the discourse segments. Take the following example of “The Fox and the Grapes”:

A famished fox crept into a vineyard where ripe, luscious grapes were draped high upon arbors in a most tempting display. In his effort to win a juicy prize, the fox jumped and sprang many times but failed in all its attempts. When he finally had to admit defeat, he retreated and muttered to himself, “Well, what does it matter anyway? The grapes are sour!”

One can, in a relatively straightforward way, say the first sentence lays the BACKGROUND for the second sentence, in RST’s functional terms; and the first sentence could alternatively be a SETTING consisting of an EVENT, following Rumelhart’s story constituents. The “point”, on the other hand, comes from a contrast between the fox’s bitter remark on the grapes (i.e. they are sour), while the grapes have in fact been described as ripe and luscious in the beginning of the story. The “point” of a story therefore cannot be understood from a single sentence or discourse segment, but it is a result of the interaction of various segments as the story unfolds.

If the keywords are more reliable units for ensuring annotation consistency, as seen from Table 1, we should take a closer look at how the keywords deemed important by the annotators bear on the story point. They are certainly content words, but in addition, they often carry certain sentiment or polarity, and perhaps value judgement. The contrast between discourse segments which constitutes the twist in the story can often be

partially traceable through the lexical items used. For instance, the opposite or contrasting semantic relation between adjectives like luscious and sour is an important clue. The structure and semantics of a discourse are, after all, based on the rhetorical style and lexical choices. If one can identify these useful clues from the text, one could at least partially relate the “point” of a story to its coherent structure by means of certain surface linguistic devices.

4. A Preliminary Annotation Scheme

Based on the above crude but useful reflection, we can at least learn the following for the annotation of stories:

- The structure of a story can be decomposed into individual discourse units such as clauses, sentences, or larger discourse segments. Each of them can be labelled according to some structural categories, and the relation between segments can be discretely labelled.
- The “point” of a story, however, cannot be based on the labelling of individual segments alone because the “point” actually emerges from the interaction of the segments.
- Nevertheless, certain surface linguistic means are responsible for indicating the “point”. For instance, a contrast in polarity of the lexical items in the text could be a clue to the twist of the story. This would have to work together with other linguistic devices like discourse connectives and action verbs, and rhetorical style like direct speech and personification, amongst others. Annotating such linguistic devices could therefore at least partially represent the “point” of a story, as they often provide the means by which story structure and story point interact.

Level of Annotation	Description	Examples
Structural	The constituents in Rumelhart’s story grammar are assigned to discourse segments in a story, e.g. STATE, EVENT, REACTION, GOAL, ATTEMPT, OUTCOME, etc. Currently we only treat them as tags, ignoring their combination according to the grammar rules.	<GOAL>In his effort to win a juicy prize, </GOAL> <ATTEMPT>the fox jumped and sprang many times</ATTEMPT> <OUTCOME>but failed in all its attempts.</OUTCOME>
Functional	The functional relations between discourse segments according to the RST relations are annotated, e.g. ELABORATION, CIRCUMSTANCE, BACKGROUND, CONTRAST, etc., insofar as the subjective judgement from the annotator can justify.	“When he finally had to admit defeat” CIRCUMSTANCE “he retreated and muttered to himself...”
Emotional	The emotional level of annotation marks up the sentiment-bearing words which relate to the development of the story leading to the “point”. These keywords are marked for their positive or negative polarity and the target.	Famished: polarity = negative target = fox Luscious: polarity = positive target = grapes

Table 2 Preliminary Annotation Scheme

4.1 Levels of Annotation

We therefore propose a preliminary annotation scheme for SPAS, which consists of three levels of annotation:

- Structural – referring to the “structure” of stories in the tradition of Rumelhart’s story grammar, and the labels will follow the constituents such as SETTING, EPISODE, EVENT, GOAL, etc.
- Functional – referring to the functional relation of individual discourse segments toward one another, in the tradition of Mann and Thompson’s RST, so one segment could be the CONDITION, CIRCUMSTANCE, ELABORATION, etc. with respect to another segment.
- Emotional – referring to the polarity and value judgements expressed in the story through individual lexical items, which can serve as a clue for the “point” of a story.

Table 2 gives a brief description and examples of these levels of annotation. The examples are presented here for explanatory purpose, and the representation may differ from the actual format used in the annotated corpus.

4.2 Progress of Pilot Annotation

Our preliminary corpus is based on a set of 100 stories in English from Aesop’s Fables from printed and electronic sources. At least two different published versions of each story were collected. The stories were first annotated with textual information including paragraphs, sentences, and the moral; as well as basic linguistic information like POS and pronouns and their antecedents. For the time being, we have been focusing on the structural and emotional levels of annotation. The functional level is apparently more difficult and is expected to lead to a lot of different interpretations and judgements, and is therefore left for a later stage.

5. Potential Investigation with the Corpus

The annotated corpus is expected to provide a useful resource for further research on story structure and story understanding, and most importantly on the interface between story structure and story point. From the perspective of natural language processing, the following are a few possible areas of investigation.

5.1 Core and Optional Elements in a Story

Stories can be retold in numerous ways and in various lengths. The difference in length could be a result of the complexity of the story structure. For instance, the number of Goal-Attempt-Outcome cycles in “The Fox and the Grapes” could be varied (e.g. whether the fox jumps to the grapes two or three times); and the description of the setting could have different amount of details (e.g. simply saying the fox saw the grapes, or verbosely describing the environment of the grape yard). The annotation could allow us to analyse for the core and optional elements in stories, and to align different versions of the same story structure-wise.

5.2 Interestingness vs Coherence

Some stories, especially longer versions, often include extra non-essential, if not irrelevant, details. For example, whether the Fox is asked to be the judge in the Hare-Tortoise race does not make a difference to the ending of the story nor the moral to be conveyed. Such non-essential details are only for the interestingness of a story but not necessarily the coherence of the story. The annotation could allow us to analyse for the relevant and irrelevant details, and to devise ways to tell them apart in story understanding.

5.3 Interface between Structure and Point

The moral of a story can be explicitly marked by means of a simple statement at the end, or a quote from one of the protagonists, or only implicitly inferred from the text. If the moral is explicitly present as part of the story text, it can be easily annotated. However, if it is absent from the surface text, the annotation should indicate the implicit moral, giving an account for how it can be inferred from the structure of the story. As mentioned earlier, the “point” in many fables often arises from some unexpected events. For example, no one would expect the slow Tortoise would beat the fast Hare and win the race, or the mouse could somehow rescue the lion. Such twists in stories could partially be reflected from the surface text, e.g. lexically through the use of opposites or words with reverse polarity. The annotated corpus could allow us to have more in-depth investigation into the interface between story structure and story point.

5.4 Interaction with Linguistic Devices

Surface linguistic cues are often exploited in computational discourse processing. Cue phrases, discourse connectives, adverbials, pronouns and reference expressions, lexical chains, psychological and perceptual verbs, focus entities, subjective elements, etc., are important indicators of discourse structure and coherence. Apart from indicating the twist with opposites or words with reverse polarity, we also observed some other common linguistic features in stories. For example, some versions, especially those written for younger children, tend to dramatize a story by introducing considerable direct speech of individual protagonists and conversations among them. Personification (of animal characters) is also often found in stories. In addition to dialogues between characters, personification is achieved through actions performed by them (e.g. the Hare mocks the Tortoise) and attributes assigned to them (e.g. the greedy Fox). The way the characters are addressed in stories (e.g. Mr. Fox indicating the Fox is respected, big brother Hare (in some Chinese versions) indicating the Hare is superior, etc.) may also suggest important connection with the story point. Expanding our annotation to cover the effect of direct speech, action words like mocking and boasting, attribute words like cunning, proud, and slow, modes of address, etc., could provide important and useful cues for story structure, especially in association with plot progression, goal conflict, and the twist of the story.

6. Future Work and Conclusion

While the pilot annotation is underway, we have planned for the following future directions in the next phase:

- Design a more detailed annotation scheme with more precise and well-defined guidelines to facilitate subsequent large-scale annotation of fables and other kinds of stories.
- Further (re-)consider and refine the details of the various levels of annotation to render subjective judgement in a more measurable way.
- Revisit the notion of story grammar and structural rules to capture the flexible combination of various core and optional story constituents.
- Investigate and analyse the importance and functions of various surface linguistic devices which serve to interface between story structure and story point.

Hence in this paper we have reported on our ongoing work on the construction of a corpus of children's stories, starting with fables. For our annotation, we intend to include not only discourse structure, or more specifically story structure, but also clues for the story point. We propose a preliminary annotation scheme covering the structural, functional, and emotional aspects of story texts. The annotated corpus is expected to facilitate further research in computational discourse processing, in particular story analysis and story understanding, especially with respect to the interface between story structure and story point, and how they interact by means of various surface linguistic devices including cue words, discourse connectives, sentiment bearing words, etc.

7. Acknowledgements

The work described in this paper was fully supported by a grant from City University of Hong Kong (Project No. 7200115).

8. References

- Charniak, E. (1972) *Toward a Model of Children's Story Comprehension*. PhD Thesis, MIT.
- Cohen, R. (1987) Analyzing the Structure of Argumentative Discourse. *Computational Linguistics* 13(1-2): 11-24.
- Cullingford, R. (1986) SAM. In B.J. Grosz, K. Sparck Jones and B.L. Webber (Eds.), *Readings in Natural Language Processing*. Morgan-Kaufman, Los Altos.
- DeJong, G. (1982) An Overview of the FRUMP System. In W.G. Lehnert and M.H. Ringle (Eds.), *Strategies for Natural Language Processing*. Hillsdale, NJ: Lawrence Erlbaum Associate, Publisher.
- Grosz, B.J. and Sidner, C.L. (1986) Attention, Intention, and the Structure of Discourse. *Computational Linguistics*, 12(3):175-204.
- Haberlandt, K. (1980) Story grammar and reading time of story constituents. *Poetics*, 9:99-118.
- Halliday, M. and Hasan, R. (1976) *Cohesion in English*. Longman Group.
- Hobbs, J. (1982) Towards an understanding of coherence in discourse. In W. Lehnert and M. Ringle (Eds.), *Strategies for Natural Language Processing*. Hillsdale, NJ: Lawrence Erlbaum Assoc.
- Lehnert, W.G. (1979) The Role of Scripts in Understanding. In D. Metzger (Ed.), *Frame Conceptions and Text Understanding*. Berlin: de Gruyter.
- Mandler, J.M. and Johnson, N.S. (1977) Remembrance of things parsed: story structure and recall. *Cognitive Psychology*, 9:111-151.
- Mann, W.C. and Thompson, S.A. (1987) *Rhetorical Structure Theory: A Theory of Text Organization*. ISI Research Report ISI/RS-87-190, University of Southern California.
- Moore, J. and Pollack, M. (1992) A problem for RST: The need for multi-level discourse analysis. *Computational Linguistics*, 18(4):537-544.
- Oinonen, K.M., Theune, M., Nijholt, A. and Uijlings, J.R.R. (2006) Designing a story database for use in automatic story generation. In R. Harper, M. Rauterberg and M. Combetto (Eds.), *Entertainment Computing – ICEC 2006*. LNCS Vol.4161, pp.298-301, Springer Verlag, Berlin.
- Propp, V. (1968) *Morphology of the Folktale* (translated by Laurence Scott). University of Texas, Austin.
- Reichman, R. (1985) *Getting Computers to Talk Like You and Me: Discourse Context, Focus, and Semantics (An ATN Model)*. The MIT Press.
- Rumelhart, D. (1975) Notes on a schema for stories. In D. Bobrow and A. Collins (Eds.), *Representation and Understanding: Studies in Cognitive Science*. New York: Academic Press.
- Scha, R. and Polanyi, L. (1988) An augmented context-free grammar for discourse. In *Proceedings of the 12th International Conference on Computational Linguistics (COLING'88)*, Budapest, Hungary, pp.573-577.
- Schank, R. and Abelson, R. (1977) *Scripts, Plans, Goals, and Understanding*. Hillsdale, NJ: Lawrence Erlbaum Assoc.
- van Dijk, T. (1985) Structures of News in the Press. In T. van Dijk (Ed.), *Discourse and Communication: New Approaches to the Analysis of Mass Media Discourse and Communication*. De Gruyter.
- Wilensky, R. (1982) Story Grammars Revisited. *Journal of Pragmatics*, 6:423-432.
- Wilensky, R. (1983) *Planning and Understanding: A Computational Approach to Human Reasoning*. Massachusetts: Addison-Wesley Publishing Company.