THE TEXTUAL IDEA AND ITS RELATIONSHIP TO DISCOURSE STRUCTURE

Cosmin Danut BOCANIALA

Department of Computer Science, University "Dunarea de Jos" of Galati

Abstract: The textual idea could be characterized through a series of either detailing the current form of idea or continuing a previous form of it. Every element of the series has associated a span of the given text that carries it. If we are to relate different spans of text according to relationship between corresponding elements of the series, we will obtain a possible representation of discourse structure. In this paper we take a set of texts and represent their structure in the above manner. The purpose is to give researchers the possibility to asset our manner to represent discourse structure.

Keywords: discourse structure, textual idea, text summarization.

1. INTRODUCTION

The textual idea could be characterized through a series of either detailing the current form of idea or continuing a previous form of it. Every element of the series has associated a span of the given text that carries it. If we are to relate different spans of text according to relationship between corresponding elements of the series, we will obtain a possible representation of discourse structure. In this paper we take a set of texts and represent their structure in the above manner. The purpose is to give researchers the possibility to asset our manner to represent discourse structure.

First we show the way we split a text into its elementary units. Our inspiration sources were Rhetorical Structure Theory (RST) (Mann, Thompson, 1987) and action decomposition from Partial-Order Planning as shown in (Young, Pollack, Moore, 1994) and used in (Bocaniala, 1999). Then we take a set of texts, represent their structure in our manner and emphasize practical use. Finally we draw some conclusions on our manner to represent discourse structure.

2. THE ELEMENTARY UNITS OF A TEXT

For a given text we consider its clauses as the elementary units. We present below the facts that lead us to this choice. Our inspiration sources were Rhetorical Structure Theory (RST) (Mann, Thompson, 1987) and action decomposition from Partial-Order Planning as shown in (Young, Pollack, Moore, 1994) and used in (Bocaniala, 1999).

For Thompson and Mann (1987) the size of a elementary unit of a text is not a theoretical matter. They affirm in the same paper (Thompson, Mann, 1987) that elementary units are "roughly equivalent to clauses". That is, in their view a elementary unit is generally, but not always, a clause.

Does that fit to our approach? Is it suitable to the way we relate the components of the idea's flow to the spans of the given text? We consider that predicates are the textual elements that carry the constituents of the idea. Therefore, a elementary unit must be at least a clause in order to be able to express a component of textual idea.

This paper was recomended for publication by V. Mînzu

Young, Pollack and Moore (1994) perform partial-order planning putting a great emphasis on action[1] decomposition. Bocaniala (1999) uses their approach on text analysis. He creates an oriented graph where the nodes are the actions found in text. The actions are expressed through the predicates found in text.

How does action decomposition and its relationship to text predicates relate to our approach? Each action could be related to a span of text formed by a predicate and its determinants (nouns, complements, adverbs) – that is simply a clause! We believe that the actions found in text relates in a certain sense to text idea.

The conclusion of this paragraph is that we consider the clauses of a text as its elementary units because the clause's predicate and its determinants express a component of textual idea.

3. TEXT ANALYSIS

As we said in introduction, the textual idea could be characterized through a series of either detailing the current form of idea or continuing a previous form of it. According to previous paragraph, every element of the series has associated a clause of the given text that express it. If we are to relate different clauses of text according to relationship between corresponding elements of the series, we will obtain a possible representation of discourse structure.

We first propose a list of the possible situations that could occur; for each situation we offer a small sample text. Then, we introduce a graphical representation for our manner to understand discourse structure. Finally, we analyze two bigger sample texts and emphasize practical use. Before analyzing a text, we split it into a numbered list of clauses.

3.1. The list of possible situations

Let us consider three consecutive clauses c_1, c_2 and c_3 – c_2 follows c_1, c_3 follows c_2. We propose the following list of possible situations:

a) c_1 details c_2

That is, a clause that express a detail of the idea carried by another clause goes before it. A sample text for this situation is taken from the book "Tess of the d'Ubervilles" of Thomas Hardy, electronic edition, www.hti.umich.edu, page 32.

1.   When old Mr. Simon Stoke, latterly deceased, had made his fortune as an honest merchant (some said money-lender) in the North,

2.   he decided to settle as a county man in the South of England, ...

The 1st clause details the 2nd one by telling the time when Mr. Simon Stoke decided to settle as a county man in South of England.

b) c_2 details c_1

It is a common met situation. A clause gives details on the antecedent one. We have a sample text taken from the same source, "Tess of the d'Ubervilles" of Thomas Hardy, electronic edition, www.hti.umich.edu, page 9.

1.   This fertile and sheltered tract of country,

2.   in which the fields are never brown and the springs never dry, ...

The 2nd clause goes after the 1st giving details on the quality of region's land.

c) c_2 continues c_1

It is another common met situation. The idea from the 1st clause is continued by the idea from the 2nd clause. We also have a sample text taken from "Tess of the d'Ubervilles" of Thomas Hardy, electronic edition, www.hti.umich.edu, page 9.

1.   The forests have departed,

2.   but some old customs of their shades remain.

A similar situation is that when after c_1 follows a series of detailing clauses $c_2...c_{n-1}$, and c_n continues the idea of c_1. A sample text for this situation is taken from the same source, "Tess of the d'Ubervilles", page 34.

1.   Tess wished

2.   to abridge her visit as much as possible;

3.   but the young man was pressing, ...

The 2nd clause's idea details the idea of the 1st clause. The 3rd clause's idea continues the idea of the first one.

d) c_3 and c_2 detail c_1

The sample text is also taken from "Tess of the d'Ubervilles" of Thomas Hardy, electronic edition, www.hti.umich.edu, page 27.

1.   He ... sped on his way;

---

[1] An action represents here a mental representation of the meaning of a predicate considered together with its determinants (nouns, complements, adverbs).

2. while Tess stood

3. and waited.

The 2nd and 3rd clauses both detail the idea of the 1st clause. The two clauses explain what Tess did while the other person sped on his way.

Another sample text from page 32.

1. ... and in doing this he felt the necessity of recommencing with a name

2. that would not too readily identify him with the smart tradesman of the past,

3. and that would be less commonplace than the original bald stark words.

The 2nd and 3rd clause give the reader details on the characteristics the person wanted for his name.

We could add that there could be also a 4th, 5th, ... clauses that also detail the 1st one. Also, we could've had a series of clauses where all of them except last detail the last clause, i.e. c_1 and c_2 detail c_3.

3.2. A graphical representation

Let us consider the next phrase, taken from the same source as the previous sample texts, "Tess of the d'Ubervilles", page 32.

1. When old Mr. Simon Stoke, latterly deceased, had made his fortune as an honest merchant (some said money-lender) in the North,

2. he decided to settle as a county man in the South of England, out of hail of his business district;

3. and in doing this he felt the necessity of recommencing with a name

4. that would not too readily identify him with the smart tradesman of the past,

5. and that would be less commonplace than the original bald stark words.

The relationships established between text's clauses are next. The 1st clause specifies the time when Mr. Simon Stoke decided to settle in the South of England (2nd clause). The 3rd clause continues the idea from the 2nd clause. The 4th and the 5th clauses both offer details on the new name Mr. Stoke wanted.

We will graphically represent the discourse structure of our text using a graph. The nodes of the graph
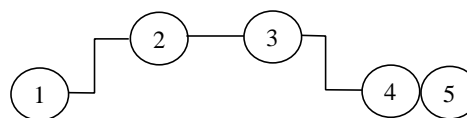

Fig.1. Graphical representation for our sample text.

stand for the text's clauses. A node's information is the number of the corresponding clause. The edges of the graph mark the relationships between clauses.

The graph is drawn using next rules:

−   if two clauses fall in case *a* or *b* (see section 3.1), the one that details the other is drawn lower;

−   if two clauses fall in case *c*, they are drawn at the same level;

If we have a series of nodes that fall in case *d*, we stick together the nodes of the graph that contains the numbers of detailing clauses. Then we follow the representation rule for case *a* or case *b*. We will have a single edge that unifies stick nodes with the node they detail.

The graph for our sample text is shown in Figure 1.

3.3. Two other sample texts analyzed

Let us consider next sample text taken from "Tess of the d'Ubervilles" of Thomas Hardy, electronic edition, www.hti.umich.edu, page 37.

1. The letter was in the third person,

2. and briefly informed Mrs. Durbeyfield

3. that her daughter's services would be useful to that lady in the management of her poultry-farm,

4. that a comfortable room would be provided for her

5. if she could come,

6. and that the wages would be on a liberal scale

7. if they liked her.

The 2nd clause's idea continues the idea of the first one. The 3rd, 4th and 6th clauses detail the 2nd clause – which were the informations in the letter addressed to Mrs. Durbeyfield. The 5th and the 7th clauses detail conditions for the ideas expressed by the 4th, respectively 6th clauses.

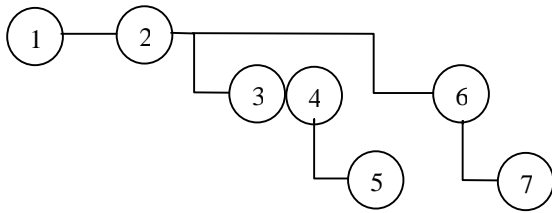The graphical representation for this sample text is shown in Figure 2.

Fig.2. Graphical representation for the first bigger sample text.

The second sample text we consider in this section is taken from the same book, "Tess of d'Ubervilles", page 47.

1. The descendants of these bygone owners felt it almost as a slight to their family

2. when the house

3. which had so much of their affection,

4. had cost so much of their forefathers' money,

5. and had been in their possession for several generations

6. before the d'Urbervilles came

7. and built here,

8. was indifferently turned into a fowl-house by Mrs. Stoke-d'Urberville

9. as soon as the property fell into hand according to law.

The 2nd clause details the 1st clause's idea – the time when the former owners of the house felt insulted. The 3rd, 4th and 5th clauses detail the value that the house had for the former owners – as shown by 2nd clause. The 6th and 7th clause detail the 5th clause idea. The 8th clause details the 2nd unit idea – it specifies what the house became. And, finally, the 9th
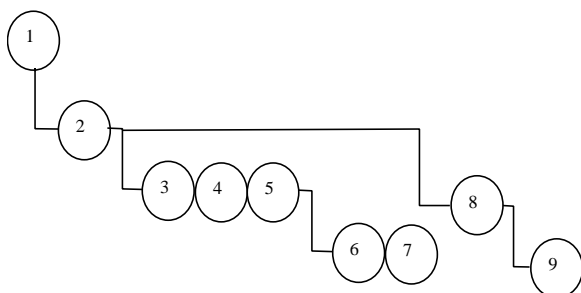

Fig.3. Graphical representation for the second bigger sample text.

clause details the time when the house became a fowl house – as shown in the 8th clause.

The graphical representation of this sample text is shown in Figure 3.

3.4. A practical use

We could use our manner to represent discourse structure in order to summarize texts.

If we take a closer look at Figure 2, we see that text's clauses are organized on three levels. The first level contains the 1st and the 2nd clauses, the second level contains the 3rd. 4th and 6th clauses, and the third level contains the 5th and the 7th clauses.

If we are to read the span of text formed by the first level we obtain a possible summary for our text.

1. The letter was in the third person,

2. and briefly informed Mrs. Durbeyfield ...

If we read the span of text formed by the first and second level we obtain a more detailed summary for our text.

1. The letter was in the third person,

2. and briefly informed Mrs. Durbeyfield

3. that her daughter's services would be useful to that lady in the management of her poultry-farm,

4. that a comfortable room would be provided for her

6. and that the wages would be on a liberal scale ...

If we bring together all three levels we will obtain the sample text itself.

Let us take a look at the Figure 3. The first level contains one single clause, the 1st. The second level contains one single clause too, the 2nd. The third level is richer and contains the 3rd. 4th, 5th and 8th clauses. The fourth and last level contains the 6th, 7th and 9th clauses.

If we read the span of text formed by the first level, we obtain a summary for our text.

1. The descendants of these bygone owners felt it almost as a slight to their family ...

It is noticeable that this is a good summary for our text. Even if we don't know what precisely – *it* word – insulted the former owners (of what?), the text could be understood: something made the former owners of some thing feel insulted.

When we put together the clauses from the first snd second level, we obtained another summary for our text.

1. The descendants of these bygone owners felt it almost as a slight to their family

2. when the house ...

This summary is bigger than the first one but unclear. The ambiguity is introduced by the 2nd clause, "when the house", which needs some more explanations.

A much more spectacular result is obtained when we add to the first two level the clauses from the third one.

1. The descendants of these bygone owners felt it almost as a slight to their family

2. when the house

3. which had so much of their affection,

4. had cost so much of their forefathers' money,

5. and had been in their possession for several generations

8. was indifferently turned into a fowl-house by Mrs. Stoke-d'Urberville ...

The span of text is clear and represents a good summary of our text. As a general observation, it seems that we cannot affirm that a summary will be better and/or clearer if we go on a lower level.

4. CONCLUSIONS

In this paper we presented a new manner to represent discourse structure. The starting point was the observation that the textual idea could be characterized through a series of either detailing the current form of idea or continuing a previous form of it.

First we showed that every element of the series has associated a clause from text that expresses it. Therefore, different clauses of text relate according to relationship between corresponding elements of the series. These induced relationships form a possible representation of discourse structure. We then took a set of texts and represent their structure in the above manner. We also emphasized the practical use of our manner to represent discourse structure in texts summarizing.

The purpose of this paper was to give researchers the possibility to asset our manner to represent discourse structure.

5. REFERENCES

Bocaniala, C.D. (1999). *Applications of planning in discourse analysis*, Graduation Paper, University "Al. I. Cuza", Iasi.

William C. Mann and Sandra A. Thompson (1987). *Rhetorical Structure Theory: A theory of text organization*, Technical Report RS-87-190, USC Information Sciences Institute, Marina Del Rey, CA.

William C. Mann and Sandra A. Thompson. (1987). *Antithesis: A study in clause combining and discourse structure*, pp. 2, Technical Report RS-87-171, USC Information Sciences Institute, Marina Del Rey, CA.

Young, R. M., Pollack, M. E., Moore, J.D. (1994). Decomposition and Causality in Partial Order Planning, In: *Proceedings of the Second International Conference on Artificial Intelligence and Planning Systems.*