INTRODUCTION

to

NATURAL LANGUAGE PROCESSING

CRITICAL DISCOURSE ANALYSIS

PROJECT REPORT

Turhan Doğan 121058-2
1. INTRODUCTION

Critical Discourse Analyze (CDA) is a critical approach to discourse analyze. Adding a critical dimension to discourse analyze is looking for hidden meaning of discourse. CDA, which is highly context-sensitive, is a social approach for discourse analyze. Although it include visual analyze of event, it can be carried out for only written texts. CDA reveals point of view of the author for the events and group of people. The method of CDA in addition to revealing social inequalities in a text it also gives some straight forward methods that conveys authors point of view about the social group in the text.

CDA could be done by using various methods. CDA can be divided into three parts, which are analyzing the text as a whole, sentence-by-sentence approach, word and phrase level analyze. This paper introduces these parts, excluding words and phrase level because of its highly sophisticated world knowledge, in details and demonstrates techniques, theory and algorithms to methods for carrying these two main parts. Use of looking text as a whole, use of sentence-by-sentence approaches, techniques applied in project and implementation issues are introduced in following part in order.

2. TEXT AS WHOLE

The type of text and what it is actually about is important in the context of CDA. The genre and frame of text is two important characteristics of a text for CDA. In the context of some type of text use of some words makes some distinction for the possible approaches of the author. While the author is expected to use a neutral, objective language, he or she will insert an occasional loaded word to slant the report. Genre knowledge, which is a general characteristic of the text to be analyzed, enables the analyst to detect and interpret such deviations critically [1].

Genre of text is one of the aspects of text that in its whole character. Genres of text inform us what type of foregrounding and back grounding is applied in the text. For example, for a text in news report type it is expected that important (foreground) sentenced comes before less important (back grounded) sentences. Topic of these utterances gives us point of view of author in the text. Genre of the text is supposed to be supplied as information by user in this project.

Frame of text is another characteristic of text which is actually overall attribute of the text. Framing refers to how the content of a text is presented, what sort of perspective (angle, slant) the writer is taking [1]. Frames can be in another frame. For example, a news report that is framed as story and in this frame a good guy vs. bad guy frame, in which a group of people given a favorable treatment, can be presented. Some choice of word in some genre context is also critical. Calling “demonstrators” or “protestors” are choices that convey the point of view of author. For example a paper that uses these words to mention examples of derived words can not be viewed as sided.
Background of a text can be put in the text as omission which is leaving certain things out of text. The omission of concept in a text occurs so that there is no way alternative for the thing to be explained. Detecting such an omission requires highly sophisticated world knowledge that includes also these alternatives. Text level omission is not dealt with in this project.

Presupposition is another text characteristic of text that embedded in the frame of text. One of the concepts in a text can be supposed more important than other in the text. This type of presupposition is not concerned in project implementation.

**3. SENTENCE BY SENTENCE**

Analyzing sentences is individually is second part in CDA. Sentences can be analyzed in terms of topic, agent patient relation and nominalization. In addition sentence level presupposition and insinuation can occur, however these kind of manipulation requires again a sophisticated level of world knowledge. These kinds of manipulations are again out of concern of this paper.

Topicalization of sentence is a form of foregrounding in sentence level. In other words it is what the sentence is actually about. Subject in the sentence can be though as what the sentence is about. Choice of topic in a sentence is a writers’ choice to affect readers’ attention. The overall sentences convey what the article is about. For example, in a text having 20 sentences about supporter of a law and 5 sentences about opponents of the law in parliament it can be concluded the topic of the sentence is about supporters. Topicalization of a text includes problems like anaphora resolution in computational approach. Subjects that are like they, those should be resolved for an accurate topicalization of text. Ignoring these sentences can be another approach.

Agent patient relation in text especially for some words it is important to pay attention. For some verbs passive usage without agent is severe affect on reader. Words like “arrest”, “shoot”, “kick”, “kill” requires agent and patient, otherwise it can be thought that the usage is some other purpose. For example, “the officers arrest demonstrators” or “the demonstrators are arrested” has different discursive affect on receiver. The latter give the receiver as if it is demonstrators’ responsibility of being arrested.

Nominalizations in sentences are another way of hiding agency in a sentence. Agent critical words as illustrated before and nouns associate these words and agentless use of them should be detected in sentences for CDA.
3. ANALYSE TECHNIQUES

The techniques used in project are straightforward applications of CDA. The topic of text determined in accordance to topics of sentences and the places of these sentences in the text. For example, the first sentence in a news report is supposed to have higher degree of significance and the latter sentence is the less significant. The topics of the individual sentences are supposed to be the head part of the subject noun phrase. Each sentence has a degree of importance so topics of each sentence do. The overall importance is found and by addition each significance degrees of sentences. The rate of importance of each group of people to whole importance additions gives each rate that associate the group of people. The higher rate gives the topic of the whole text. Then we can conclude that the text having highest importance rate as topic X is about topic X.

The agent patient relations are deduced from parsed and anaphorically resolved sentences of the text. As the sentences are parsed the head of the subject noun phrase is thought as the group of people that is to be referred. There are some verbs that are kept in lexicon as agent critical verbs. The agentless sentences with these verbs should be given as output to the user. These sentences should be investigated in terms of its possible agents. The author of the text should be graded as sided in accordance to verb and agent relation. On the other hand, the sentences having agent critical verbs with an agent should be taken under consideration. Such these sentences are supposed to be grade for author as he is supporting the particular side in accordance to agent patient relation. For instance;

Officers arrested demonstrators. Demonstrators are favored so author should be thought as sided for demonstrators
Demonstrators are arrested by officers. This sentence should be graded close to neutral value. A little favor for officers can given
Demonstrators are arrested. Officers are favored so author should be thought as sided for officers.

The weight of these grades should be considered in the context of genre of text and place of the sentence in the text. For example, in a news report a grade in the middle of the report should affect the overall grade in half weight. The overall grade should be given to the user as an estimate value. A deterministic approach would be a naïve positivist approach for a critical approach

Hidden agent in the nominalized noun phrase should be found and graded in accordance to previous manner. A noun related to any of the agent critical verbs should be graded accordingly. For example, the nouns “arrestment” is related to verb “arrest” is obvious. However the noun massacre is not directly related but implied. Such these relations can be linked to a verb like “kill”.

The presupposition and omission conveyed by social context of text is discarded in this report. Such these knowledge requires a sophisticated level of world knowledge. For example a politician claiming that US government should apply lower tax rate with
presupposition that US government apply high rates although it is the one of the least amount of tax rate in US when compared to other countries. These kind of information is not supposed to be evaluated, although for a relevant critical analyze it should be evaluated.

5. IMPLEMENTATION ISSUES

In the implementation phase of the project there are several assumptions that are made to simplifying the implementation part. This part introduces these assumptions for the implementation part.

Firstly there is no level of parsing in the implementation part. Therefore the head parts of noun phrases are found with some more additional assumption in the implementation. It is assumed that there are no relative clauses in noun phrase. Under this assumption it is also assumed that the head of subject noun phrase is the word just before auxiliary verb or the normal verb. In addition, the topic of the sentence is assumed to be the head part of the subject noun phrase or to whom it refers to. Anaphora resolution is not dealt with in any level. There is no resolution and if it is needed the sentence is passed without evaluation.

For this release, it is only assumed that the text is in news report genre. Each genre should have a different grade importance in accordance to its frame. In this implementation it is assumed that the significance of sentences are degraded as the text goes on as it is expected in news report texts.

6. CONCLUSION

Critical discourse Analysis can be used for conveying power relationship and side of the author in a text. Although some techniques requires sophisticated level of world and social world knowledge, some CDA techniques can be used and easily applicable for computer applications.

REFERENCES