Introduction

It has long been recognized that one of the main prerequisites for a successful computational investigation of discourse is the development of a comprehensive theory which would take into account a sufficiently broad range of perspectives on discourse (e.g., linguistic, cognitive, planning and problem solving) and put it into an integrated computational architecture, thus providing the infrastructure for describing actual discourse-related phenomena.

In building such a framework, one of the basic tasks is developing a (preferably, open-ended, extensible) knowledge representation scheme which would allow the representation of a wide variety of domain and discourse meanings. An important property of such a scheme should be its capacity to express the complex interrelations among all the types of knowledge relevant for treatment of discourse. At present, we believe that the minimum of this knowledge includes knowledge about the domain (world), about the goals and plans of speakers/hearers, and about the structure of a discourse.

The next task in building a discourse treatment framework is the delineation of the regular ways of realizing certain discourse meanings and relations in a particular language.

Following that, one needs to create a set of heuristic rules for mapping between meanings and their realizations (because, in principle, the knowledge acquired at the previous step can lead to a situation where one meaning can be realized by several realization means, and a realization means can realize more than one discourse meaning). We believe that one of the most important sources for such heuristics is the capability of reasoning about the goals and plans of the speakers [1, 2].
The program we sketch above is vast. In this paper we discuss only a text meaning representation scheme which we developed for use in a natural language generation environment, DIOGENES [3].

1 Motivation for Text Meaning Representation

In a language communication situation the consumer (hearer, reader) understands the text communicated by the producer (speaker, writer) by

1) understanding the meaning of the natural language utterances comprising the text;
2) uncovering rhetorical relations among these utterances and their components;
3) recreating intentions behind the speaker's discourse (in other words, those of speaker's plans and goals which are relevant to discourse) and
4) detecting the attitudes that the speaker holds toward the content of the discourse.

Goals, plans and actions that relate to dialog, text or generally, discourse distinguish themselves from all others by the fact that they belong to the context of a multiagent situation. In a dialog situation, this is obvious and explicit. But in fact any true act of communication, any message that carries meaning, does so on the basic assumption that the producer of the message knows and takes into account the existence of a consumer. The production of a message is an action which we perceive as a step in a plan to achieve one of the producer's goals. These plans take into account the knowledge the producer has (or assumes he has) about the target audience. A theory of discourse goals must, therefore, follow the prior introduction of a model of a participant in a language communication situation.

Based on what we feel is relevant to provide a fine-grained theory of discourse (both producer- and consumer-oriented) we will take into account the following components in an agent's model:

(i) Knowledge about the world, which we find useful to subdivide into
   * concept memory containing knowledge about types of things in the world
   * episodic memory containing knowledge about tokens of things in the world
   * Attitudes to contents of the concept and episodic memory
   * Models of other agents, with all their components

(ii) An agenda of active goal and plan instances (the intentional plane of an agent)

Speaker goals will relate to these different components. Thus a speaker may want to achieve the following types of goals:

(i) Modify the hearer's concept memory, for example by giving a definition.
(ii) Modify the hearer's episodic memory, for example by stating a fact, describing an object or relating an event.
(iii) Modify the hearer's model of the speaker, for example by expressing his attitude towards some fact, as in Unfortunately, Peter will come too.
(iv) Modify the hearer's attitudes to facts of the world.
(v) Modify the hearer's agenda, for example by threatening, giving an order or asking a question.

The speaker's processing during generation can be sketched as follows. Given an input stimulus, the speaker will activate a goal, choose a rhetorical plan to realize that goal, and, with the help of his knowledge about
a) the world;
b) the hearer;
c) language (at both the sentence and discourse levels) and
d) the relevant pragmatic constraints, he will generate a discourse.

2 Knowledge Representation for Text Meaning

A natural language clause is the combination of a single predicate with its case roles and their modifiers. A natural language sentence can contain more than one clause. A natural language text is a sequence of natural language clauses grouped into sentences, paragraphs, etc. In our opinion, natural language text understanding consists in representing the semantic and pragmatic (discourse, attitudinal, speech act) information from each natural language clause, augmented by the representation of domain-related and text-related connections among natural language clauses or sets thereof.

In fact, the final result of the process of text understanding includes not only the information overtly present in the source text. It also includes the results of reasoning by the consumer, with the purpose of
a) filling in elements of representation not directly assignable from the source text and
b) reconstructing the agenda of rhetorical goals and plans of the producer active at the time of text production and connecting its elements to chunks of meaning representation.

In what follows we define the format of our text representation language.

The representation of the meaning of a text is a pair

\[ \text{TAM} = \langle C, R \rangle \]

where \( C = \{\text{clause}_1, \text{clause}_2, ..., \text{clause}_n\} \) is a set of representations of the meaning of natural language clauses and \( R = \{\text{relation}_1, \text{relation}_2, ..., \text{relation}_k\} \) is a set of connections among elements of the representation (clause components, clauses, sets of clauses).

2.1 TAMERLAN Clauses

The representation of a clause is a triple

\[ \text{clause}_i = \langle S_i, P_i, \{A_{i,j}\} \rangle \]
S_i stands for the semantic content of the natural language clause represented by clause_i. The representation language we use for representing semantic content of natural language clauses is a modified version of the interlingua knowledge representation language used in the KBMT-89 system [4].

P_i is a pointer to the action or plan in the author (speaker) agenda that is realized by the uttering of the clause represented by clause_i. We therefore overtly refer to the producer’s goals in each TAMERLAN clause. The agendas contain structures of goal and plan instances. Since we are interested in discourse situation, we describe only the goals and plans that a) presuppose the situation with at least two cognitive agents and b) relate to rhetorical realizations of goals. These pointers are not directly realized, but serve as background knowledge during the realization process of other elements of input. Reasoning about speaker goals and plans is a widely accepted technique in natural language generation [5, 6, 7]. The plans to which these pointers point are realized in the text only when it is decided to produce a direct speech act, in which case the realization usually involves generating a separate target language clause.

To illustrate our use of the pointers, suppose the speaker wants to promise to return to his current location at 10 o’clock. Depending on the context and other parameters, he may decide to produce I will return at 10 or I promise to return at 10. In the latter case the decision is made to realize overtly the meaning of the speech act (using a performative verb). The mechanism for this is as follows: traversing the pointer P_m the speaker gets to the relevant point in the agenda, which is the (primitive) plan PROMISE (or THREAT, as the case may be). Since the realization rules for speech acts prescribe their realization as first-person-singular clauses with the lexical realizations of the names of appropriate speech plans (acts), the natural language clause I promise X gets produced, and eventually X is expanded into the subordinate natural language clause to return at 10. The central point is that the former natural language clause I promise is the realization of a pointer in the input, not of an entire text representation clause.

In an interpretation environment, conversely, there will be some full-fledged natural language clauses whose real meaning is purely pragmatic, representable, for instance, as a pointer P or as an attitude, A_i.j.

{A_i.j} is a set of author (speaker) attitudes towards S_i or parts thereof. The attitudes are represented as quadruples of the following form:

A_i.j = < type_i.j, value_i.j, attributed-to_i.j, scope_i.j >

The possible types of attitudes include:
epistemic (with values taken from the (0,1) interval to account for expressions like perhaps; the endpoints of the interval intuitively correspond to the values of impossible and necessary);

* deontic (with values taken from the same scale, with the endpoints interpreted as roughly absence of obligation as in I needn’t go and absolute obligation as in I must go);

* expectation (ranging from total surprise to total expectation)

* evaluative (with values taken from a similar scale, with the endpoints interpreted as, roughly, the worst, the best, the midpoint as neutral. It goes from a very negative attitude to a very positive attitude. Depending on the scope, realization will greatly vary);

The organization of the above types is similar - their value ranges are all one type of scale. The differences among them are semantic. The above classification is an enhancement of Reichman’s treatment of "context spaces" [8, p. 56]. We use the terminology (if not exactly the spirit) of her distinction among the epistemic, evaluative and deontic issue-type-context spaces. Context space is Reichman’s term for a discourse segment. The issue context space corresponds to our attitude component, while the non-issue context space provides a shallow taxonomy for discourse segment types (Reichman gives comment, narrative support, and non-narrative support as the non-issue type values).

We envisage adding other types of attitudes. It is clear that a more complex and complete taxonomy of attitudes will have to be developed. In particular, the evaluative attitude is a gross simplification of what should become a complex net of categories enabling us to represent the meaning of evaluative verbs, adjectives and adverbs featuring scalar behavior. When trying to represent the meaning of sentences containing adjectives such as interesting, important, or boring; one realizes that to use a single evaluative scale ranging from good to bad to represent all three adjectives is inadequate, especially in view of the problem of lexical selection during generation. (Using the value of the evaluative attitude scope, however, provides a partial answer to this problem (see below)). Our solution nevertheless suffices as a first approach to capture argumentative orientations, which, as discussed by Anscombe and Ducrot in [9], play an important role in distinguishing, for example, but from and.

The attributed-to component of the attitude simply binds the attitude to a particular cognitive agent (which may be the speaker of the utterance or some other known or unknown agent), who endorses the responsibility of the content of the utterance. This is important for understanding reported speech, and more generally the polyphony phenomena, in the sense of Ducrot [10]. Ducrot’s theory of polyphony, an approach to extended reported speech treatment, provides a framework for dealing with the interpretation of a number of semantic and pragmatic phenomena, e.g. the difference in meaning and use between since and because, certain particularities of negative sentences, etc.
The scope of the attitude representation pinpoints the entity to which this attitude is expressed. The values of the scope can be an S_i, a part of S_i or another attitude value, with its scope. In understanding the text the reader (hearer) notes the attitudes of the author (speaker) to the content. The attitudes can be expressed toward events, see (1), objects (2), properties (3) or other attitudes (4).

(1) The train, unfortunately, left at 5 p.m.
(2) This book is interesting.
(3) The meeting was reprehensibly short.
(4) Unfortunately, I ought to leave.

Knowing what the scope of an evaluative attitude is guides lexical selection during generation. For instance, if the scope is an event, adverbs like fortunately and unfortunately will be used. If the scope is the physical appearance of a person, the attitude will be realized as attractive or ugly. If the scope is a small amount, a little or little will be used.

2.2 TAME.RLAN Relations

R is a set of relations, each of which is a 4-tuple:

\[ \text{relation}_i = (< \text{rel-type}_i, \text{from}_i, \text{to}_i, \text{A}_i>) \]

where \( \text{rel-type}_i \) is taken from a predefined set of relation types organized as a hierarchy (for the complete hierarchy, see [14]). A number of studies have been devoted to enumerating and classifying relations [11, 8, 12, 13 to name but a few]. We are developing our own taxonomy, taking into account semantic and pragmatic criteria. The top level of our taxonomy comprises three types. Type I, with subtypes cause, conjunction, alternation, co-reference, time and space, regroups domain relations which connect facts, events or objects of the world. Type II, which includes subtypes particular, reformulation, elaboration, and conclusion, are text relations: they reflect the discourse plans and goals of the speaker. Type III relations connect domain-related text components to speech acts or speech act components such as the time of utterance. We have isolated two subtypes so far: (i) the temporal relation, which relates the speech act time of an utterance to the time of the action or event expressed by the utterance, and which is realized linguistically through tense and/or temporal adjuncts; (ii) the causal relation, which is illustrated in (5) and (6). In both examples, what is justified is not a fact of the world, but a speech act (an assertion in (5) and a question in (6)).

(5) Your mother didn't come back last night. Because the mail is still in the mail-box.
(6) Can you give me the time? Because I have a meeting at three.

From stands for the domain of the relation and to, for its range. Both from and to can have as value a single TAME.RLAN clause, a set of such clauses or a component of a TAME.RLAN clause.
The \( A_i \) slot represents the producer attitude to relation \( _i \). Very often, the producer will want not only to express that two facts or events are connected by a relation like cause or contrast. He may also, depending on his intentions and/or perception of the structure of the world, adopt different points of view, have marked attitudes towards the relation itself. For example, if he wants to express a relation of contrast between two facts, he can realize it either as an adversative (\( \text{but, nevertheless, whereas, etc.} \)) or as a concessive (\( \text{although, however, etc.} \)). Through a concessive or an adversative, the speaker adds meaning (lets the hearer know about his intentions and opinions) to a simple, attitudinally neutral \( A \) is different from \( B \). To mention but one other example illustrating the pertinence of adding an attitude slot to the relations, compare the connectives \( \text{finally or secondly} \) which are attitudinally neutral with \( \text{last but not least, which is not.} \)

2.3 A Sample Text Representation

To illustrate, let's take sentence (1) as the input text. The expected representation will be as follows. (For readability, we use a simplified version of our meaning representation language).

\[
\text{TAM} \_1 = < \text{C} \_1, \text{R} \_1 > \\
\text{C} \_1 = \{ \text{clause} \_1 \} \\
\text{R} \_1 = \{ \text{relation} \_1 \}
\]

\[
\text{clause} \_1 = < \text{S} \_1, \text{P} \_1, \{ \text{A} \_1, \text{A} \_2 \} > \\
\text{S} \_1 = (\text{head} = \text{move} \\
\text{agent} = \text{engineer} \\
\text{theme} = \text{train} \\
\text{source} = \text{station} \\
\text{destination} = \text{unknown} \\
\text{aspect} = \text{inchoative} \\
\text{time} = \text{time} \_1 \\
\text{(value} = 5 \text{p.m.}) \\
\text{S} \_1 \text{ represents} \text{The train leaves at 5 p.m.} \\
\text{P} \_1 = (\text{plan} \_1 \\
\text{type} = \text{inform-hearer} \\
\text{direct} = \text{yes} \\
\text{time} = \text{time} \_2) \\
\text{P} \_1 \text{ represents the inform speech act. Time} \_2 \text{ is the speech act time, the time of utterance.} \\
\text{A} \_1, \_2 = (\text{type} = \text{evaluative} \\
\text{value} = 0.2)
A_I,1 represents the meaning of unfortunately.

\[ A_{I,2} = \text{(type = epistemic, value = 1, scope = S_1, attributed-to = speaker)} \]

A_I,2 represents modality.

\[ R_1 = \{\text{relation}_1\} \]
\[ \text{relation}_1 = \text{(type = TypeIII.temporal.after from = time_2 to = time_1 A_1 = neutral)} \]

Time_2 (the utterance time) follows time_1 (the time of the event S_1). Relation_1 is realized in the sentence through the past tense of left.

3 Summary

This paper describes the results of an initial investigation into discourse and pragmatic phenomena relevant to natural language generation. We have formulated the paradigmatic environment for the study of discourse phenomena by detailing a model of verbal communication among intelligent agents, based on the information processing metaphor and developed a text meaning representation language.

In the immediate future our research will address the following tasks:

(i) acquiring a set of domain- and discourse-related goals and plans to 'populate' the model of the speaker; this task will also involve acquiring knowledge about world and discourse objects which appear as roles in these goals and plans;

(ii) extending and improving the taxonomy of discourse relation types;

(iii) describing and acquiring knowledge about the various scales that serve as the basis for representing meanings of attitudes.

4 Literature


