Abstract Linguistic Resources for Text Planning

Marie W. Meteer
BBN Systems & Technologies Corporation
10 Moulton Street
Cambridge, Massachusetts 02138
MMETEER@BBN.COM

Abstract

In this paper, I define the notion of an abstract linguistic resource which reifies as a term for use by the text planner just those combinations of concrete linguistic resources (the words, morphological markings, syntactic structures, etc. that actually appear in a stream of text) that are expressible. I present a representational level, the Text Structure, which is defined in these abstract linguistic terms and which mediates and constrains the commitments of a text planner to ensure that the utterance being planned will be expressible in language.

1. Introduction

Natural language generation is the deliberate production of text to meet the communicative goals of some underlying application program. It consists of the following major activities: (1) determining what information is to be communicated, (2) imposing a suitable order on the elements of this information consistent with the constituent structure of language and expressing the relative salience and newness of the elements, and (3) determining what wording and syntactic constructions to use. The first two of these activities are generally considered "text planning" and its output is the "plan". The third activity is "realization" and generally handles all of the linguistic decision making.

While it is recognized that this division is problematic (Hovy, et al. 1988), nearly all generation systems today make this division. One of the chief accomplishments of my work has been to bridge the gap between these two activities through the introduction of a new representational level that simplifies both their responsibilities. It both provides the choices available to the text planner to allow it to take advantage of the expressiveness of natural language and, through that control of the choices, prevents it from composing an utterance that is not expressible in the language.

Most state of the art text planning systems follow a common design (see for example McKeown 1985, Derr & McKeown 1984, Paris 1987, or Hovy 1988). They start from a set of propositions, each typically verb-based and able to be realized independently as a simple sentence. Then they organize the propositions into a coherent discourse by combining them according to the dictates of predefined "schemas" representing plausible discourse relationships. Subsequent choices of concrete surface resources[1] are all local to the propositions and not sensitive to the schemas or other context, except for the discourse-level connectives used in combining the propositions into complex sentences and occasionally a shallow discourse history governing the use of pronouns.

In these approaches to natural language generation there is a gap between the plan, which is usually represented in the terms of the application program, and the resources used by the realization component to carry out that plan, which are the concrete words, syntax, morphemes, etc. The gap occurs because the text planner selects units from the application program and organizes them without simultaneously considering what linguistic resources are available for expressing them. These systems thus have no principled way of ensuring that their message is *expressible* in language.

Of course, they do successfully produce texts: they ensure their plans are expressible by accepting limits to their expressive competence, e.g. each atomic unit in the plan is required to be a proposition, and thus can always be realized as a clause. Each unit can be independently translated into language using the linguistic realization component since there are few restrictions on the connection between clauses. Clauses can be connected with coordinate or subordinate conjunctions (e.g. "and", "because") or simply made into separate sentences.

However, this kind of approach does not take advantage of the full expressive power of language, in which units can be much more tightly composed. In order to exercise the full expressiveness of language, text planning needs to address the *internal* composition of clauses and not just their organization into larger structures. Clauses in actual texts reflect a combination of multiple atomic units. Systems that ignore this and begin with units that are inevitably realized as kernel clauses (e.g. Mann & Moore 1981, Derr & McKeown 1984, Hovy 1988) have two major deficiencies: (1) they are presuming underlying programs have units of this size that may be simply selected for inclusion in the message and then realized intact, and (2) they are under-utilizing the power of natural language, which can use complex noun phrases, nominalizations, adverbial phrases, and other adjuncts to pack information from multiple units into one clause.

[1] The surface linguistic resources are all the syntactic structures, words, and grammatical features of the language available to the speaker.

Moreover, the process of composing multiple units into one clause is a much more complex problem than simply ordering propositions in a text plan. What compositions are possible depends on what linguistic resources are available to realize the units involved. For example, an Army battalion that has been assigned a defensive mission can be said to be "defending", but if we say that a battalion that has been assigned an offensive mission is "offending" we mean something very different. There is no comparable resource available to fit that textual niche and either a different, more complex resource must be used or the whole text changed (e.g. "attack". Furthermore, different types of resources have different constraints on their composition: one can "make an important decision", but one cannot "decide importantly".

In this paper, I address the problem of how we can constrain the actions of a text planner to ensure that it will never compose an utterance that cannot be realized and can still make use of the full expressive power of language. To do this, I introduce the notion of an *abstract linguistic resource* which groups together as the reified terms to be used by the planner just those combinations of concrete linguistic resources that are expressible. I have defined a level of representation in terms of these abstract resources, the Text Structure, which is used by the text planner to state its "plan". This intermediate level of representation bridges the "generation gap" between the representation of the world in the application program and the linguistic resources provided by the language.

The terms and expressions in the Text Structure are abstractions over the concrete resources of language. They provide the text planner with terms that define the possible combinations of features that express the semantic categories available in the language, such as event, process, or instance-of-a-kind. By providing the text planner with a set of semantic categories, rather than letting it freely choose from the individual linguistic features that define the categories, the planner is prevented from choosing a combination of features that is not realizable. These abstractions in Text Structure further constrain the composition of the utterance by defining what kinds of constituents can be extended and how the semantic categories can compose.

In this paper, I define the notion of an abstract linguistic resource for text planning by looking at (1) what are the concrete resources that the language provides and what are their abstract properties, and (2) which of those properties are appropriate to a text planner trying to compose an utterance from an application program. I then show a preliminary set of abstractions which is used in the text planner of the Spokesman generation system. This set was arrived at by both applying research results from linguistics and lexical semantics and empirically by connecting the generation system to an application program and producing text. The long range challenge of this work will be the continued development and refinement of this vocabulary of abstract linguistic terms to cover the full expressive power of natural language and determining the compositional constraints which will maintain the expressibility criteria.

2. Linguistic Resources

In this section, I address the question of what the linguistic resources are and what abstractions we can and should make over them. I begin by looking at the *concrete* resources, that is, those that actually appear in a stream of text. I then look at what various complexes of these resources express taken as a group. In Section 2.3, I look more generally at how work in linguistics can help develop a more complete vocabulary of abstractions.

2.1 The concrete resources of language

The concrete linguistic resources are all the syntactic structures, words, and grammatical features available to the speaker of the language. We can divide linguistic resources into two general classes:

- The lexical resources: These are what are often called the *open class* words (the nouns, verbs, and adjectives), and they carry most of the content.
- The grammatical resources: These include the *closed class* words, morphological markings, and phrase structure.

In what follows we ground the notion of concrete resources by looking closely at one fairly simple sentence:

Karen likes watching movies.

This sentence has lexical resources, such as "Karen" and "watch", and morphological resources, such as "-ing", the gerund marker on the verb "watch" which emphasizes the process aspect of the action, and "-s", the plural marker on the noun "movie". The phrase structure is also a concrete resource, which expresses how the constituents group together and certain kinds of relations between them; in this example, the phrase structure tells us that "movies" are what is watched, that "watching movies" is what is liked, and that "Karen" is the one that likes watching movies.

What is not there also expresses information. The fact that there is no determiner ("a" or "the") with "movies" indicates that it is not a particular set of movies being referred to (as in "the movies") but a general sample of movies. Note that it is not just the lack of the determiner that provides this information, but the features of the whole constituent the particular noun is in and the fact that it is plural: if the noun phrase were singular, then there would have to be a determiner before "movie" (*"Karen likes watching movie"). For other nouns in head position, the lack of a determiner can mean other things. For example, there is also no determiner in the first noun phrase in the sentence ("Karen"); however, in this case, since the head is a proper noun, it does refer to a unique individual. If a determiner is used with a proper noun, it has a more general meaning of "an entity with that name" (as in "All the Karens I had ever met had dark hair and then I met a Karen with red hair").

We will term this kind of composition, where the same resource means different things in different contexts, "non-linear" composition; this is in contrast to "linear" composition, where each resource contributes an identifiable part of the whole and what it contributes is not context dependent. The identification of which grammatical resources non-linearly co-occur and grouping those sets into single ábstract resources is a powerful method of constraining the text planner to keep its choices only those that are expressible in language, as we shall see in the next section where we develop abstraction resources for the sets of concrete resources that appear in the example.

2.2 Abstractions over concrete resources

Allowing a generation system to select concrete resources directly, as is done in virtually all other generation systems, makes available many more degrees of freedom than the language actually permits. As we saw in the previous section, some combinations of concrete resources occur in language, while others do not. Furthermore, we saw that the combination of the lexical resource in the head of a constituent and the grammatical resources in the constituent as a whole can combine non-linearly, so that the choice of the lexical and grammatical resources cannot be made independently of each other.

In this section, we look at how we can abstract over combinations of concrete resources by treating a particular set as a whole and naming it, rather than treating the resources as a set of independent features that happen to have appeared together. The vocabulary of abstractions we derive then becomes the terms in which the text planner makes its decisions. It is incapable of selecting a set of resources that is not expressible because it is not allowed to choose them independently.

For example, the two noun phrase constituents in our example ("Karen" and "movies"), express two different perspectives on the entities they refer to. "Karen" is expressed with the perspective NAMED-INDIVIDUAL and "movies" is expressed as a "SAMPLE-OF-A-KIND".[2] We can think of these perspectives as **semantic categories**; "semantic" because they represent something about the meaning of a constituent, not just its form, e.g. "Karen" is referring to a person as a unique individual with a name, in contrast to referring to her as an anonymous individual (e.g. "a woman"). A surface constituent can then be abstractly represented in the Text Structure for the purposes of the text planner as the combination of a lexical item and a semantic category.

Figure 1 shows the "abstract" resources for the two noun phrases of our example and the other constituents of our sentence: *Karen Likes watching movies.* [3]

The upward arrows begin at the surface constituent being abstracted over and point to the boxes showing abstract resources: the lexical item in italics and semantic category following it. This tree of boxes is an example of the Text Structure intermediate level of representation. We will return to how to develop a complete set of semantic categories in the next section.

In addition to abstracting over combinations of concrete resources by only representing the semantic type of a constituent, we can also represent the structural (syntactic) relations between the constituents. In Figure 1 the concrete relations of subject, direct object, etc. are represented abstractly as arguments and marked with a semantic relation.[4]

In this example, we have identified three kinds of information that are essential to an abstract representation of the concrete resources language provides:

- the constituency,
- the semantic category of the constituent, and
- the structural relations among the constituents.

In the next section we look at some of the motivations for these abstractions, and in Section 3, show how they can be used for text planning.

---

[2] "Sample" intended to mean "indefinite set"; the choice of names for categories is meant to be evocative of what they mean, while staying away from terms that have special meanings in other theories. Within Text Structure, these terms only need to be consistent. The names themselves do no work, except to help the observer understand the system.

[3] The notation of the phrase structure in the diagram is the notation used in the linguistic component Mumble-86. While it is slightly unconventional in that it explicitly represents the path that will be followed in traversing the structure, it is in other respects fairly standard in its terminology. I use it here since it is the notation I work with and because it lends a concreteness and reality to the diagrams since this is the structure the linguistic component will actually build when generating this sentence.

[4] The semantic relations shown here, "agent" and "patient", are capturing internally consistent relations. They are not attempting to carry the kind of weight and meaning as, say, the terms in theta-theory.

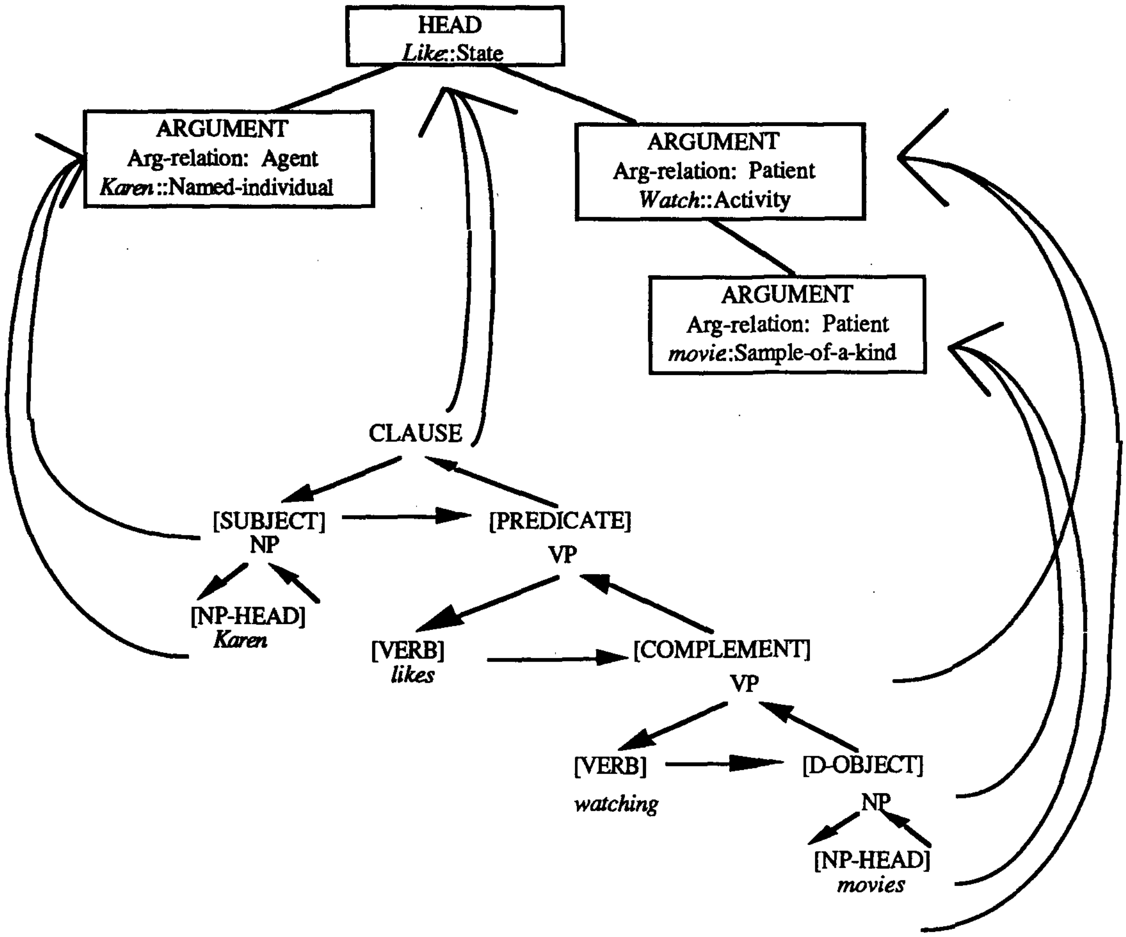


Figure 1 Abstractions over concrete resources

2.3 Developing a set of abstractions

The development of the full vocabulary of particular abstract resources is an ongoing process. The motivation for determining the abstractions comes from analysis of the language and what is expressible. A great deal of work has already been done in linguistics that can contribute to defining the vocabulary of abstractions. In this section, I look at the work of four linguists in particular who have influenced my development of the current set of semantic categories: Jackendoff, Talmy, Pustejovsky, and Grimshaw. While their work is very different in character, all explore regularities in language using a more semantic than syntactic vocabulary.

The notion of a semantic category used here was initially influenced by the work of Jackendoff (1983) who makes the following claim about the relationship between language structure and meaning:

> *Each contentful major syntactic constituent of a sentence maps into a conceptual constituent in the meaning of the sentence.*[5]

Included in his vocabulary of conceptual categories are Thing, Path, Action, Event, and Place.

However, while Jackendoff's categories are useful in that they span the entire language (since they are projections from the syntactic categories), they are not discriminatory enough to capture the constraints necessary to ensure expressibility. For example, two of the semantic categories in the example above, NAMED-INDIVIDUAL and SAMPLE-OF-A-KIND, are subsumed by the same category in Jackendoff's set, OBJECT. Similarly, his category EVENT has finer distinctions available in the actual resources: a finite verb (one which expresses tense) with its arguments expresses what I call an EVENT (*Peter decided to go to the beach*), whereas a nonfinite verb can express a generic ACTIVITY (*to decide to go to the beach*). Nominalizations make the event or activity into an OBJECT and different forms of nominalizations can pick out different aspects of the event, such as the PROCESS (*Deciding to go to the beach took Peter all morning*) or the RESULT (*The decision to go to the beach caused controversy*).

Figure 2 shows a partial hierarchy of semantic categories that reflects these distinctions.

---

[5] Jackendoff, 1983, p. 76.

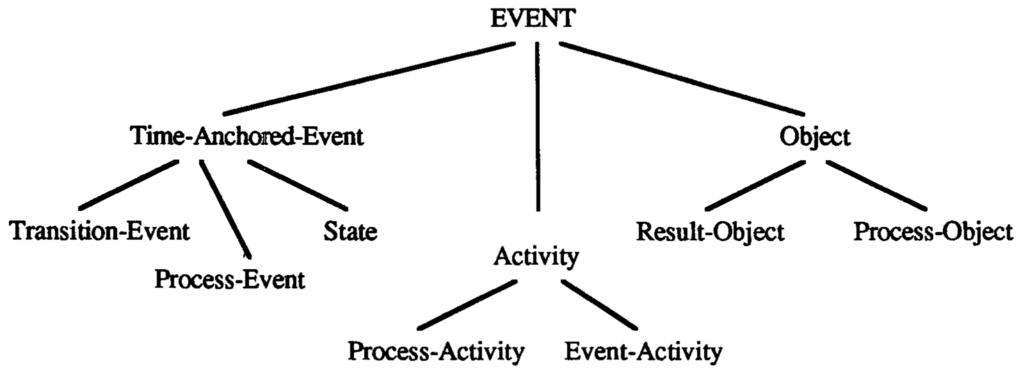


Figure 2 Partial hierarchy of semantic categories

In using these finer semantic categories in the planning vocabulary for generation, we are making a stronger claim than Jackendoff's, namely that these categories define what combinations of surface resources are possible in the language. For example, an ACTIVITY cannot have a tense marker, since by definition it is not grounded in time. The categories also serve to constrain how constituents are composed. For example, if we choose the EVENT perspective (e.g. *Michael decided to go to the beach*), we can add an adjunct of type MANNER to it (*Michael **quickly** decided to go to the beach*) but we cannot add an adjunct of type PROPERTY (**Michael **important(ly)** decided to go to the beach*)[6]. However, if we choose an OBJECT perspective (*Michael made **a decision***), the PROPERTY adjunct oan be added (*Michael made an **important** decision*). Both perspectives are available, and the text planner's choice must be consistent with the kinds of adjunctions it intends to make.

Research in lexical semantics has contributed a great deal to defining these finer grained semantic categories. Talmy's (1987) extensive cross language research resulted in a set of categories for the notions expressed grammatically in a language. Pustejovsky's (1989) Event Semantic Structure makes a three way distinction of event types (state, process, transition) which both captures the effects of nonlinear composition of resources and provides constraints on the composition of these types with other resources. Grimshaw's analysis of nominals (1988) contributed to the definition of object types which convey particular perspectives on events, such as result and process.

3. Using Abstract Resources for Text Planning

In order to plan complex utterances and ensure they are expressible in language, i.e. can be successfully realized as grammatical utterances, the text planning process must know (1) what realizations are available to an element, that is, what resources are available for it, (2) the constraints on the composition of the resources, and (3) what has been committed to so far in the utterance that may constrain the choice of resource. The first two points are addressed by the the use of abstract linguistic resources discussed in the previous section. The third is addressed by the ongoing Text Structure representation of the utterance being planned, which is also in abstract linguistic terms. In this section, I describe the Text Structure and how it mediates and constrains the text planning process.

3.1 Text Structure[7]

Text Structure is a tree in which each node represents a constituent in the utterance being planned. Figure 3 shows an example of the Text Structure representation for the utterance: *"Karen likes watching movies on Sundays"* .

Text Structure represents the following information:

Constituency: The nodes in the Text Structure tree reflect the constituency of the utterance. A constituent may range in size from a paragraph to a single word.

Structural relations among constituents: Each node is marked with its structural relation to its parent (the top label) and to its children (the bottom label on nodes with children). Structural relations indicate where the tree can be expanded: composite nodes may be incrementally extended whereas a head/argument structure is built in a single action by the planner, reflecting the atomicity of predicate/argument structure.

[6] Following the general convention in linguistics, we use a "*" to mark ungrammatical sentences, and a "?" to mark questionable ones.

[7] Note that I will not attempt a formal definition. I agree with the text linguist Beaugrande that "Formalism should not be undertaken too early. Unwieldy constructs borrowed from mathematics and logic are out of place in domains where the basic concepts are still highly approximative. Such constructs give a false sense of security of having explained what has in fact only been rewritten in a formal language." Beaugrande & Dressler, 1981, p.14.

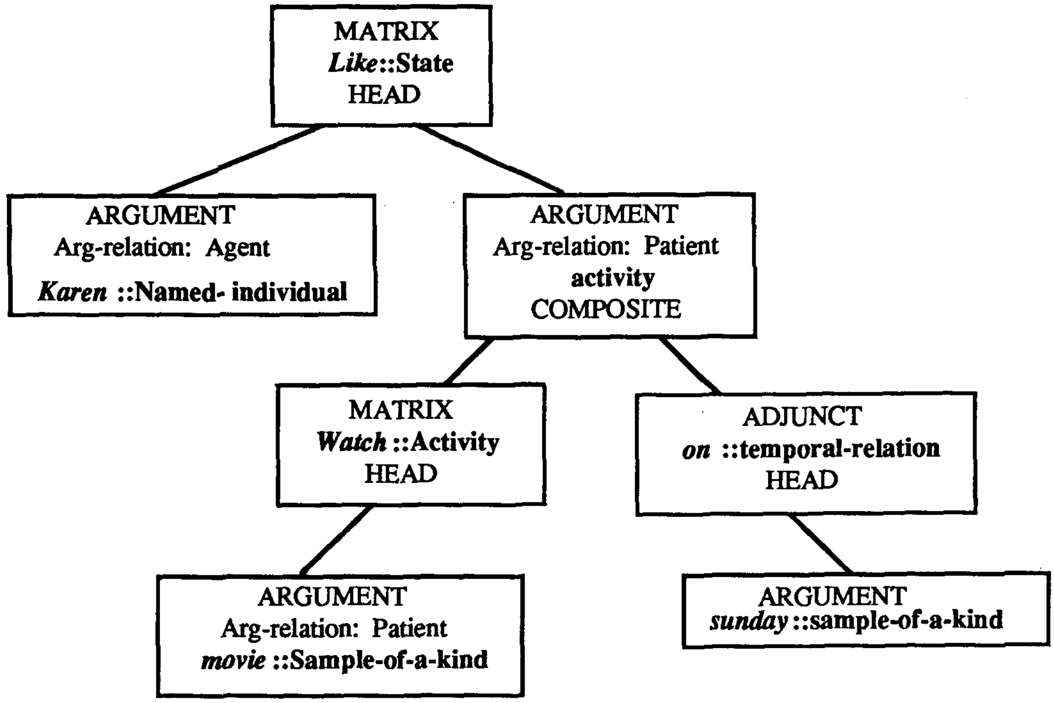


Figure 3 Text Structure for "Karen likes watching movies on Sundays"

Semantic category the constituent expresses: The labels in the center of the node (in bold) show the lexical head (when applicable, in italics) and the semantic category the constituent expresses.

3.2 Using the Text Structure for Text Planning

The abstract linguistic terms of our planning vocabulary can provide constraints on the composition of the message to ensure that it will continue to be expressible as we add more information. For example, the semantic category of a constituent can constrain the kind of information that can be composed into that constituent. Consider the earlier example contrasting "decide" and "make a decision", where in order to add an adjunct of type PROPERTY, the RESULT perspective of the EVENT must be explicit in the utterance, as shown in Figure 4.

In summary, the Text Structure can constrain the following types of decisions within the text planner:

- where additional information may be added (e.g. structure can only be added at leaves and nodes of type COMPOSITE; furthermore, in an incremental pipeline architecture such as this, information can only be added ahead of the point of speech)
- what functions and positions are available for the elements being added in (e.g. matrix or adjunct)
- what form the added element must be in (e.g. an object of type *property* can be added to a *thing* but not to an *event*)

The Text Structure representation is used in the text planner of my SPOKESMAN generation system (Meteer 1989). It serves as an intermediate representation between a variety of application programs and the linguistic realization component Mumble-86 (McDonald 1984, Meteer, et.al 1987). Portions of the outputs for three of these applications are shown below.

THE MAIN STREET SIMULATION PROGRAM (ABRETT, ET AL 1989)

Karen 10:49 AM: Karen is at International-conglomerate, which is at 1375 Main Street. Her skills are managing and cooking. Karen likes watching movies. She watched "The Lady Vanishes" on Sunday.

SEMI-AUTOMATED FORCES (SAF) PROJECT[8]

C/1 TB is to the east and its mission is to attack Objective GAMMA from ES646905 to ES758911 at 141423 Apr. A/1 TB is to the south. B/1 TB and HHC/2 are to the east.

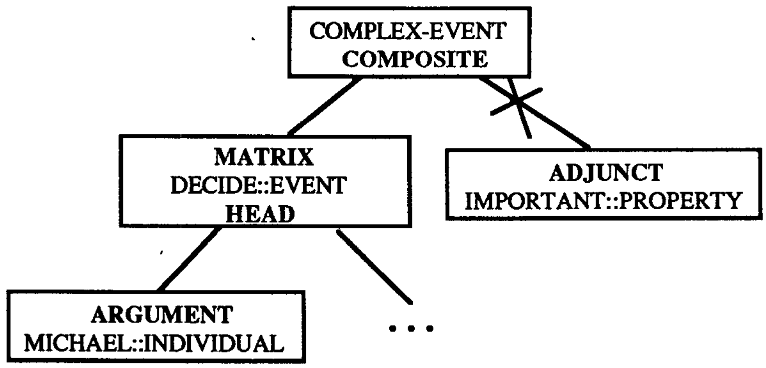
AIRLAND BATTLE MANAGEMENT PROJECT[9]

Conduct covering force operations along avenues B and C to defeat the lead regiments of the first tactical echelon in the CFA in assigned sector.

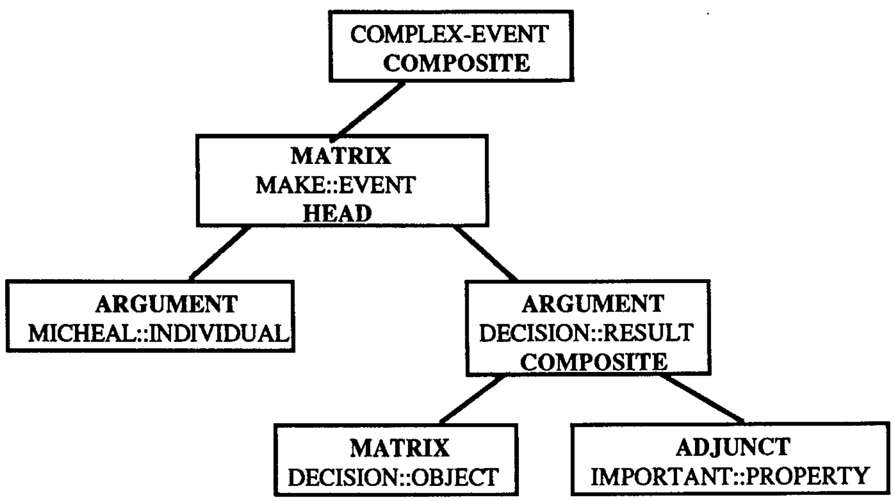
---

8 SAF is part of DARPA's SIMNET project, contract number MDA972-89-0600.

9 Sponsorship of ALBM is by the Defense Advanced Research Projects Agency, the US Army Ballistic Research Laboratory, and the US Army Center for Communications, contract number DAAA15-87-C-0006.



"Michael decided ..."



"Michael made an important decision"

Figure 4

4. Contrasting Approaches

The greatest difference between other approaches to NLG and ours is that they work directly in terms of concrete resources rather than introducing an abstract intermediate level as I have proposed here. Approaches fall into two classes: (1) those that use a two component architecture in which a text planner chooses and organizes the information to be expressed and passes it to a separate linguistic component that chooses the concrete resources to express the plan (e.g. McKeown 1985, Paris 1987, or Hovy 1988); and (2) those that use a single component which does the planning of the text directly in terms of the concrete resources (e.g. Nirenburg et al. 1989, Danlos 1987).

The limitation of the two component architecture is that the text planner is not working in linguistic terms, and so it cannot be sure that the plan it builds is expressible, i.e. can have a successful realization. Most such systems avoid this problem by limiting the expressiveness of the system overall. The planner begins with a set of propositions, each verb-based and able to be realized independently as a simple sentence. It then organizes the propositions into a coherent discourse by combining them according to predefined "schemas" representing plausible discourse relationships. Subsequent choices of linguistic resources are all local to the propositions and not sensitive to the schemas or other context, except for the discourse-level connectives used in combining the propositions and occasionally a discourse history governing the use of pronouns.

However, clauses in actual texts by people reflect a combination of multiple atomic units. Systems that ignore this and begin with units that are inevitably realized as kernel clauses under-utilize the expressive power of natural language, which can use complex noun phrases, nominalizations, adverbial phrases, and other adjuncts to pack information from multiple units into one clause.

The second approach, using single component architecture, recognizes the limitation of separating text planning from the choice of linguistic resources, and removes this division, letting the text planner manipulate concrete resources directly. However, this increase in complexity for the text planner has repercussions for the complexity of the architecture overall. For example, Nirenburg uses a blackboard architecture that must backtrack when the text planner has chosen incompatible concrete resources.

5. Conclusion

I have argued that an intermediate level of representation is needed within the text planner in which to compose the utterance and that this representation should be in abstract linguistic terms. Making the vocabulary in which the text planner makes its decisions be an abstraction over the concrete resources of the language simplifies the decision making in the composition process, since the text planner need not deal with the particular grammatical details of the language. Furthermore, since the abstract vocabulary captures all and only those combinations of resources that occur in the language and since its terms constrain the composition with other terms, the representation serves to ensure that the decisions that the text planner makes when composing the utterance will not have to be retracted, that is, that the utterance the text planner composes will be *expressible* in language.

I have shown how a preliminary planning vocabulary can be developed by approaching the problem from two sides: (1) using research in linguistics and text analysis to determine a set of abstractions over concrete linguistic resources and (2) using these terms in a text planner generating text from a real application to empirically test the usefulness of this set for generating. The long rang challenge of this work will be continuing this bidirectional development and testing process to define an intermediate representation that both covers the expressiveness of natural language and ensures the expressibility of the generator's text plan.

References

Beaugrande Robert de, & Wolfgang Dressler (1981) *Introduction to Text Linguistics.* Longman. London, England.

Abrett, Glen, Mark Burstein, & Stephen Deutsch (1989) Tarl: Tactical Action Representation Language, an environment for building goal directed knowledge based simulation. BBN Technical Report No. 7062. June 1989.

Danlos, Laurence (1987) *The Linguistic Basis of Text Generation,* Cambridge University Press, Cambridge, England.

Derr & McKeown (1984) "Using Focus to Generate Complex and Simple Sentences", *Proceedings of Coling-84, Stanford University,* July 2-6 1984. p.319-326.

Grimshaw, Jane (1988) "On the Representation of Two Kinds of Noun" Presented at Theoretical Issues in Computation and Lexical Semantics Workshop, Brandeis University, April 1988.

Hovy, Eduard (1988) "Planning Coherent Multisentential Paragraphs" In *Proceedings of the 26th Annual Meeting of the Association for Computational Linguistics,* Buffalo, New York, June 7-10, 1988, p. 163-169.

Jackendoff, Ray (1983) *Semantics and Cognition,* MIT Press, Cambridge, Massachusetts.

McDonald, David D. (1984) "Description Directed Control", *Computers and Mathematics 9(1)* Reprinted in Grosz, et al. (eds.) *Readings in Natural Language Processing,* Morgan Kaufman Publishers, California, 1986, pp.519-538.

McDonald, David D. & Marie Meteer "Adapting Tree Adjoining Grammar to Generation", submitted to 5th International Workshop on Natural Language Generation.

McKeown, Kathleen (1985) *Text Generation,* Cambridge University Press, Cambridge, England.

Meteer, Marie W. (1990) *The Generation Gap: The problem of expressibility in text planning.* Ph.D. thesis, Computer and Information Sciences Department, University of Massachusetts, Amherst, Massachusetts. February 1990.

Meteer, Marie W. (1989) *The SPOKESMAN Natural Language Generation System,* BBN Technical Report 7090.

Meteer, Marie W., David D. McDonald, Scott Anderson, David Forster, Linda Gay, Alison Huettner, Penelope Sibun (1987) *Mumble-86: Design and Implementation,* UMass Technical Report 87-87, 173 pgs.

Nirenburg, Sergei, Victor Lessor, & Eric Nyberg (1989) "Controlling a Language Generation Planner", *Proceedings of IJCAI-89,* Detroit, Michigan.

Paris, Cecile L. (1987) *The Use of Explicit User Models in Text Generation: Tailoring to a User's Level of Expertise,* PhD Thesis, Columbia University, Department of Computer Science.

Pustejovsky, James (1989) "The Generative Lexicon", submitted to *Computational Linguistics.*

Talmy, Leonard (1987) "The Relation of Grammar to Cognition" (ed) B. Rudzka-Ostyn, *Topics in Cognitive Linguistics,* John Benjamins.