An Architecture for Aggregation in Text Generation

James Shaw  
Dept. of Computer Science  
Columbia University  
New York, NY 10027, USA  
shaw@cs.columbia.edu

Kathleen McKeown  
Dept. of Computer Science  
Columbia University  
New York, NY 10027, USA  
kathy@cs.columbia.edu

3  Aggregation Operators

The following aggregation operators are applied to the propositions in the order described:

- **Semantic**: Ontological subsumption substitutes a more general concept from the domain ontology in place of the set of the children underneath the general concept (e.g., "drugs" vs. "nitroglycerin and levophed").
- **Referential**: The system selects an adequate description to refer to entities (e.g., "both arms" vs. "left arm and right arm").
- **Hypotactic**: Various modifying syntactic constructions, such as adjectives, PPs, present/past participle clauses and relative clauses, are used in place of a full sentence (e.g., "diabetic patient" vs. "the patient has diabetes").
- **Paratactic**: It takes advantage of parallel structures at the syntactic level to delete repetitive constituents and shorten expressions by using conjunction.

The sentence planner works at a semantic level and uses lexical information to guarantee that there is at least one way to express the combined propositions. The task of paraphrasing and exact word choice are carried out at the lexical chooser stage. This division allows the sentence planner to concentrate on operations which involve multiple propositions – detecting similarities and removing them. In contrast, the lexical chooser concentrates on paraphrasing and word choice, which occur after the sentence boundaries have been determined.

4  Conclusion

The aggregation operators use common grammatical constructions to produce concise sentences. We have built a prototype in our MAGIC system which generates coordinated multimedia presentation for care-givers in the Intensive Care Unit. Currently we are adding additional operators and scaling up the system.