Technical Correspondence

Questioning the Need for Parsing Ill-formed Inputs

Parsing of ill-formed input may be an important issue in the design of robust natural language systems, as pointed out by Eastman and McLean 1981. The unpredictable range of user input in natural language requires a mechanism to deal with unacceptable syntactic constructs. However, human behavior is complicated, and the capacity to adapt our mode of expression to the varying contexts in which we communicate should not be overlooked. Because of our adaptability, particularly in patterns of speech, the ability to parse ill-formed inputs may not be a critical requirement for all natural language systems.

For example, we conducted an experiment that simulated a voice-driven office automation system, where subjects were provided with a tutorial and asked to solve simple problems using voice commands.[1] The purpose of the experiment was to test the usefulness of our proposed system and to gather data for the designers of the natural language processor we planned to implement (Biermann et al. 1983).

The subjects in the experiment were seated in front of a terminal display with a microphone and were instructed to speak their commands using discrete speech (about one word per second) or slow connected speech. The monitor sat in a separate room, listed to their requests on earphones, and typed the corresponding commands to the system. In this way, subjects experienced a system that responded to spoken English sentences. If a command was not judged to be recognizable by our voice recognition system or parsable by our implemented grammar, the subject received an error message and was forced to rephrase the request.

The surprising result of this experiment was that the subjects consistently spoke well-formed sentences when forced into the regimen of slow methodical speech and had very few inputs rejected on the basis of improper syntax. We had expected to find frequent occurrences of the three classes of ill-formed input described by McLean 1981, yet none of our subjects exhibited such errors to a significant degree.

Out of more than 1600 sentences spoken by our 15 subjects, only 10 utterances were rejected as ungrammatical inputs. Five of the ill-formed sentences contained incomplete noun phrases not processed by our system (e.g., "Combine the first paragraph with the second [paragraph]."), and the remainder involved skipped articles (e.g., "Remove period.", rather than "Remove *the* period."). We also found that relatively simple syntax was used, with a total of only six sentences containing conjunctions and with a notable absence of relative clauses.

The subjects were required to begin each utterance with an imperative verb and to use words from a list of recognized vocabulary (in this case, some 50 words from the office domain). These constraints may have contributed to the lack of syntactic complexity we noted, but it is still remarkable that during an hour-long session, under the pressures of problem-solving, subjects routinely managed to form grammatically acceptable inputs.

As a consequence of these observations, we have practically discontinued our efforts to parse ill-formed sentences. For pragmatic reasons, our work in that area has become a low priority in the overall development of a voice-driven natural language system.

Linda Fineman
Department of Computer Science
Duke University
Durham, NC 27706

References

Biermann, A., Rodman, R., Ballard, B., Betancourt, T., Bilbro, G., Deas, H., Fineman, L., Fink, P., Gilbert, K., Gregory, D., and Heidlage, F. 1983 Interactive natural language problem solving: a pragmatic approach. Paper presented at the Conference on Applied Natural Language Processing, Santa Monica, CA (February).

Eastman, C.M. and McLean, D.S. 1981 On the need for parsing ill-formed input. *AJCL* 7 4 (October-December), Technical Correspondence.

McLean, D.S. 1981 METASZK: a natural language front end to System 2000. M.S. thesis. Department of Mathematics and Computer Science, Florida State University, Tallahassee, FL (March).