

Letters to the Editor

Adaptive Parsing

I am concerned by a comment made in Julia Johnson's review of my book *Adaptive Parsing* (Kluwer Academic Publishers, 1992) in *Computational Linguistics* 18(3) (September 1992). In the review, Dr. Johnson poses a number of thought-provoking questions that underscore open issues in this research, and seems to have been, overall, a thoughtful and attentive reader. Toward the end of the review, however, she states, "The performance improvements realized with adaptive parsing over a particular kernel grammar without adaptation were not strong." This statement does not agree with the results in the book. As shown in the utility analysis on pages 194–200 and 207–210, performance of the system using the kernel grammar without adaptation gave an acceptance range from 7% to 24% of utterances; with adaptation, acceptance increased to 81% to 91%. I find it difficult to interpret this data as anything but a very strong performance improvement.

Since the perceived usefulness of adaptation rests in great part on the performance improvements it affords over a static sublanguage, I am grateful for the opportunity to point out and correct this misperception.

Jill Fain Lehman
Department of Computer Science
Carnegie Mellon University
Pittsburgh PA 15213-3890

I wish to respond to Jill Fain Lehman's letter concerning my review of her book *Adaptive Parsing*. I interpret the data cited by Dr. Lehman as showing only a weak performance improvement.

Lehman examines the relationship between the number of accepted inputs and the number of new constructions in the grammar. In on-line experiments with the adaptive interface CHAMP, a small increase in search time was required to produce a proportionately large increase in the number of accepted inputs. But the more adaptive the *user* (as opposed to the system), the greater the number of accepted inputs relative to the increased size of the grammar.

My interpretation of Lehman's performance data is consistent with the observations of the experiments:

User 10's natural grammar was quite different from the kernel, resulting in both a high rejection rate and high number of new constructions on the first day. Of the ten users studied, however, she was by far the most adaptive and rapidly settled into a comfortable mix of kernel forms and derived forms. In fact, at the end of her third session, User 10 commented that she found it more natural to use a form she believed the system would understand than to use an alternative form

that she preferred but that she believed the system could not understand. (pp. 204–205)

The statement that User 10 was the most adaptive among the users confirms the assumption that the more adaptive the user, the greater the number of accepted inputs relative to the increased size of the grammar. Lehman states that User 10 gains more than any other user in terms of increased acceptance relative to the need for additions to the kernel grammar: “almost eight times as many accepted sentences with virtually no increase in search” (p. 208).

Lehman is assuming that a greater level of acceptance of utterances indicates a greater performance in terms of the system’s adaptation to the user’s idiosyncratic grammar. Learning (by the system) occurs if the system can accept an increasing number of utterances with a decreasing growth in the grammar. For Lehman, grammar growth is measured by the number of nondeviant sentences accepted and cost of parsing by the average number of states to accept. However, I take the average number of states to accept to be a better measure of the growth of the grammar, a measure that is more independent of the sentences that the user happens to type. The intuition is that with a larger grammar there would tend to be more search required.

Study of the data that Lehman gives in the tables of her Chapter 9 leads me to conclude that CHAMP does some learning, but the user also does some learning, as Lehman argues. However, I think her data interpreted as I suggest could be used to build a more convincing argument (than the one provided in the book) that the system adapts to the user’s idiosyncratic grammar and that the grammar of the user and the grammar of the system tend to converge.

Thus, it would seem that what is required to establish adaptivity is not a simple measure of acceptance but rather a measure of *intelligent* acceptance. Can the system discriminate the utterances that should be used as a basis for modifying the kernel grammar?

Julia Johnson
 Department of Computer Science
 University of Regina
 Regina, Saskatchewan
 Canada S4S 0A2

In her reply, Dr. Johnson argues that the performance improvement in CHAMP should be attributed to human rather than machine adaptation.

Is it possible that the performance improvement is totally attributable to the user? It seems highly unlikely. First, let’s be clear about what we mean by user adaptation. There is no question that adaptive parsing relies on the adaptability of the user. It is, in some sense, the whole point: that there is a natural, easy type of adaptability in the user—the user’s tendency to re-use what worked in the past—that has heretofore been ignored and that can be used to constrain the linguistic coverage required of an interface. Note, however, that this is a different kind of adaptability than that required by a static sublanguage, i.e., the adaptability required of a user to find the specific restricted subset of English chosen by the interface designer. The latter type of adaptability is hard. See, for example, the behavior of User 8, who had great difficulty performing the task with a nonadaptive version of the kernel (pp. 40–44) and the more general arguments of Watt (1968), Burton and Brown (1979), and Tennant (1981).

Second, let's look at the expected ratio of acceptances to learned components under the assumption that the users are doing all of the productive adaptation. The data show that the acceptance level is attributable primarily to learned forms (the kernel parses only 7–24%). If we assume that the system's adaptation is not responsible for the increase in acceptances, then the system must be learning components that don't reflect the user's natural usage. Thus, the scenario dictates, the number of acceptances can increase while the number of new components decreases only if the user is guessing these bad rules and conforming *her* behavior to *them*. Why does this seem unlikely? Because users aren't very good at inducing a system's grammar (see above). Because only one protocol (User 10) gave any indication that a user was attempting to do so and her data clearly show that she wasn't particularly good at it (i.e., CHAMP did ample adaptation as well). Because these were positive experiences for the user (her sentence was accepted), so it is difficult to believe she would respond by trying to induce an *incorrect* rule to attribute to the system, then adapt her behavior accordingly. And, finally, because there is a simpler explanation: that the user responds to the positive experience of having her utterance accepted by using her form again in the future, and that the system learns grammatical components that capture the conditions on her usage well enough to converge fairly rapidly on her self-bounded grammar.

That the system can accept an increasing number of utterances with a decreasing amount of adaptation follows from building general, useful rules. Of course, by augmenting the grammar appropriately, the system moves the processing burden off those portions that do adaptation and generalization and onto those portions that do simple, efficient parsing. Thus, the system continues to act in a responsive, perceptive, intelligent manner, but it does so by relying on its parsing components.

Dr. Johnson's remarks about the growth of the grammar suggest that she has misinterpreted some of the tables in Chapter 9. I measure the growth in the grammar directly by the number of new components added in each session (shown, for example, in Figure 9-15), and not, as she implies, just by the number of sentences accepted. The average number of states to accept (e.g., Figure 9-21) is indeed a measure of work done to accept a sentence by the grammar in each stage of its development. The growth of the grammar cannot be inferred from this.

"Can the system discriminate the utterances that should be used as a basis for modifying the kernel grammar?" I continue to maintain that that's exactly what it does.

Jill Fain Lehman
Department of Computer Science
Carnegie Mellon University
Pittsburgh PA 15213-3890

References

- Burton, R. R., and Brown, J. S. (1979). "Toward a natural language capability for computer-assisted instruction." In *Procedures for Instructional Systems Development*, edited by H. O'Neil. Academic Press.
- Tennant, H. (1981). *Evaluation of natural language processors*. Doctoral dissertation, University of Illinois at Urbana-Champaign.
- Watt, W. C. (1968). "Habitability." *American Documentation* [now *Journal of the American Society for Information Science*], 19(3), 338–351.