

MACHINE (AIDED) TRANSLATION:
GENERALITIES AND GUIDES TO ACTION

David G. Hays

Machine translation is Golem astride the Tower of Babel. Golem the automaton is the symbol of man's horror of the thing that straddles the line between spirit and flesh. The crumbling tower symbolizes ethnocentricity and xenophobia. Combined, these irrational feelings can influence national policy and retard progress toward important goals. To move too fast is as much an error as not to move at all. The principles of the first section summarize my reaction to the contributions presented at the conference; the guides of the second section express my opinion about the making of decisions in a fairly broad area.

GENERALITIES

1. Almost everyone hates computers, including most computer scientists. In "Information Handling" (Current Trends in Linguistics, ed. T. A. Sebeok et al., volume 12, pp. 2719-2740), I noted that professors who give their students clever tricks for skimming technical articles refuse to permit their computer programs to use the same tricks; the computer must work the hard way, in accordance with general theories of the structure of information. A friend suggests that hatred of the machine must be responsible. Anyone who hates computers is likely to design cumbersome systems.

2. The more programmers there are, the lower their average skill. In the early days of computation, the few programmers were brilliant; as the number has increased, the number of brilliant programmers has gone up, but the number of adequate or inadequate programmers has gone up faster. The buyer of a system must ask which kind will make it.

3. The best in computing is vastly better than ever before, but almost everything is worse. Tasks that required senior professionals long hours ten years ago can now be accomplished by students in courses, because the software is more powerful. Yet systems that cost too much for each transaction are in general use, thwarting their customers' hopes, and the public is led to believe that inflexibility and intolerance are characteristic of machines.

4. Scientists care how a system works; engineers care only how well it works. The buyer of a system for use is with the engineer, but the buyer of development is with the scientist. The claim that a system works "as a human does" needs to be checked by psychologists; but the claim has nothing to do with operating effectiveness, and not much to do with developmental promise.

5. A computer system is like zuppa inglese. English soup is an Italian dessert, made in a large hemispherical bowl. Layers of cake, soaked in liqueurs, are separated with thin layers of jam and covered with a thick layer of whipped

cream. The layers of a system are hardware, software, application programs, data base formats, data base contents, and so on. Claims of universality, simplicity, and the like are often no more than the assertion that a layer of whipped cream can cover anything. Deep probes are necessary to evaluate such claims.

6. If everyone optimizes his own cost effectiveness, the system goes to pieces. The classic example is the war against German submarines in the Mediterranean. It was so successful that the Germans moved into the North Atlantic and nearly starved the British. Translation is not the end of the whole system; to raise internal costs can make the system at large much more effective if done right.

7. Brevity counts. The time of the reader has to be reckoned into the cost of the system; translations of key points can be more suitable than full translations. The machine may be more useful in finding passages than in translating them.

8. You cannot make a jumbo jet out of an elephant by pulling its ears. Martin Kay suggested that Hannibal was wiser to buy elephants to cross the Alps than he would have been if he had let a development contract for jet transport. Contrariwise, suitability as a chassis for the future jet is no criterion for selection of a first-stage machine; sooner or later it will be necessary to scrap the whole system and start over. What counts in the first installation is whether or not it works as installed, for however limited a purpose has been selected.

9. Almost everyone hates translators. They arouse our xenophobia by bringing the enemy into our camp. To give them help in their task, or credit for doing it, is loathsome.

10. Big ideas are easier to understand than little ones. Some examples of big ideas mentioned in the conference are words (as opposed to characters) as objects for optical recognition; syntactic patterns (as opposed to diagnostic contexts) in language processing; and scripts or frames (as opposed to grammatical and syntactic structures) as objects for computers to seek in texts. It might be easier to find that a news story is about a certain frame (detente), and that the source is Sadat. than to translate the whole; and the summary ("Sadat endorses detente") might be more helpful to the user than the translation would be.

GUIDES

1. A prima facie case has been made for gradual introduction of language-processing capacity into intelligence facilities.

2. System design and cost analysis remain the essential prerequisites to procurement.

3. The design should take into account as fully as possible the needs of users of translations.

4. No adequate reason for selecting a single system and excluding the rest has come to light thus far.

5. The main developmental track for a few years ahead is from character processing (editing systems) to word processing (dictionaries).

6. A plausible further development for the three to seven year prospect is automatic recognition of topic (for example, of requirements), and the matching of new text against old for partial identification of redundant, and therefore omittable, information.

7. The operational suitability of language-processing systems depends crucially on the smallest details of their design. As yet, only those of clearly superior knowledge, taste, and judgment can be entrusted with the work.

8. Several classes of systems are fundamentally different and cannot usefully be intermingled. Current commercial MT systems, which make no provision for editorial intervention between the earliest and latest stages of processing, are not suitable bases for machine-aided (editorial) systems; and the latter are not necessarily suitable bases for full-scale language-processing systems that may reach installability in as little as ten years if research and development are well supported.