COMPUTATION IN DEPARTMENTS OF LINGUISTICS

RICHARD FRITZSON

Department of Linguistics
State University of New York at Buffalo
Buffalo, New York 14261

That computers and linguists meet, for the most part, only in the still somewhat exotic field of computational linguistics is a sad statement about the state of ordinary linguistic research The time when computers were to be considered only the tool of the natural scientist or the statistically minded social scientist is long past, 'word processing technology' is now the specialty of a growing number of computer companies Not only can this technology be of great value in reducing the clerical burden of the linguist and linguistics student, but, linguists, as specialists who have been studying and manipulating language for years, are in a position to be contributing to this field In fact, in many areas of linguistic research the analysis of particular languages, the search for linguistic universals, the analysis of discourse and text, computer technology can be of help to the linguist, and, in many subfields of computer science automated language processing, the design of human/machine interfaces, the structuring of data bases, linguistics has much to offer the computer scientist, yet up until now, relatively few such cross contributions have been made Computer scientists have been slow to discover the value of linguistics to their work, the time has come for linguists to take the initiative and to train themselves (and their students) to make use of and contribute to the field of computer science.

Specialized training in the use of the computer within a particular discipline is not new Students in many social sciences now find themselves facing increasing pressure and mandatory requirements to take computer training within their department, linguistics is, in fact, unusual in not having such requirements or even oppurtunities At a time when graduating linguistics students are facing a shrinking job market, the oppurtunity to be trained in a 'commercially useful application of linguistics ought to be attractive to many students

Today, in most universities, computing is available to linguistics

departments only through the use of a large, central university computer which is expected to be of service to all university departments. But, as computer costs continue to fall, and, as large computing centers continue to be unresponsive to the needs of their new users, it will not be uncommon to find more and more departments purchasing their own computing facilities and buying or developing their own software This is already happening today, both by externally funded individual researchers and by entire departments in need of specialized computing facilities What kinds of computing equipment are available for a linguistics department trying to equip itself today?

My answer is structured, to some extent, by the organization of language It is widely understood, even by non-stratificational linguists, that the faculty of language is based on a stack of structured systems, each one building a large number of units above from a smaller number below, i.e a handful of phonetic features combine to form less than fifty phonemic segments which combine to form thousands of morphemes,tens or hundreds of thousands of words, an infinite number of sentences and texts expressing countless ideas and concepts It will not be surprising to find that as one climbs this stack, from phonology upward, the amount of computing power needed to perform useful tasks and research increases in proportion to the increasing number of units and the complexity of their structuring I will concern myself, mostly, with the possibilities available for the study of the lower levels This is because the type of linguistic work being done in the study of the semantic and cognitive levels is still primarily research and the people involved are more likely to already know their needs and options as far as computing goes Also, since the cost of computing in these areas is somewhat higher, it is less likely that departments will be doing their own purchasing for these purposes

HARDWARE FOR THE PHONOLOGIST

The student of phonology, morphology and linguistic field analysis is concerned primarily with the manipulation of linguistic text, expressed as a series of phonemic symbols or blocks of phonetic features The task is to identify identical or similar substrings, correlate their appearance with a particular meaning, and segment the text into these identified substrings As new substrings are identified, the text is often rewritten with a new organization based on new understandings, so as to improve the chances of finding new substrings, field workers often use index cards for this purpose Problem

after problem is solved in this way, with a not insignificant amount of time being spent in the reorganizing and recopying stages It is a tedious business because it is very mechanical In fact, efficient computer algorithms for doing much of the job already exist and have been implemented on nearly all computers in the form of text editors. The task is relatively simple and even the smallest computer available can do an adequate job

A linguistics department interested in providing its students with training in the use of computers for this kind of work (and they will become standard tools for the purpose very soon) would do well to purchase as many (one or more) identical, small (hobbyist size) computers as it can afford For educational purposes, the very smallest microcomputers, equipped with modest mass storage devices, such as tape cassettes or floppy discs, are just fine Assignments in classes can be distributed on departmentally owned or student owned tapes or discs (less that $10 each) These can be automatically duplicated just as assignments are now mimeographed, they are reuseable and usually contain enough room to store several assignments, including the partial results from day to da and final solutions For larger, research sized projects, involving alot of text, or more complicated analyses, such as automated analysis of phonological tactics, the fastest microcomputers, with larger mass storage devices, might be more appropriate

(Imlicit in the discussion of these types of machines is the fact that student use of them is via an interactive terminal Microcomputers are not typically operated in 'batch mode', and no benefit could be derived from doing linguistic analysis in any but an interactive mode of operation)

While microcomputers and associated memories are relatively inexpensive, linguists have a genuine need for sophisticated input and output devices which are somewhat more expensive Standard computer terminals generally provide all and only the characters available on a typewriter keyboard, some provide only upper case letters What is needed is a terminal with the same capabilities as the selectric style typewriter one with changeable type fonts, including the standard phonetic symbol alphabet, with diacritics CRT terminals (cathode ray tube terminals) can provide this type of operation more cheaply, more reliably, and more flexibly than printing terminals (there is no need to stop and change type fonts) CRT terminals which support user designed type fonts are available, and in fact, may be the only ones on which the standard phonetic alphabet can be currently supplied These terminals are somewhat expensive

(several thousand dollars each), but since they are very flexible and often support some degree of computer graphics display as well as having the potential to display texts written in any language, they are valuable educational tools

If all or most of the terminals in a department are CRT type terminals, it will be necessary to provide some means of producing 'hard copy' output on paper While most interactions with a computer can take place on a screen, some record of the results of a session will be needed for study and evaluation Printers which can handle the flexible type fonts needed by linguists are available They are fast they operate in the same way that copying machines work and simply transfer the contents of the CRT screen to the paper (including graphic materials) They are expensive However, a small department might well find that only one of these printers is necessary to meet their needs, the results of work done on any of the small microcomputers could be moved (either over communication lines or carried on a disc or tape) to the printer with little or no delay

HARDWARE FOR THE GRAMMARIAN

Syntax is, perhaps, the most widely studied subject in linguistics today Given that this is so, there is a real need for linguists, both professional and student, to understand the extreme difficulty of the task of writing a grammar for a language That attempts are made to do this without the aid of a computer is perhaps all the evidence one needs to see that the difficulties are not well understood. A formal grammar, particularly one written in the notations commonly used today, is very much like a computer program It is a list of instructions for generating a list of strings, a computer program is a list of instructions for performing some process (which might be generating a list of strings) Both need to be precise, both are very complex, both suffer from the fact that a change in one part of the ordered list may cause an unanticipated change in the effect of another part It would be very surprising to find that linguists were better at producing untested, yet correct, formal characterizations of complex processes than computer programmers I expect that testing a newly written grammar will be as enlightening an experience for a linguistics student as debugging a new complex program is for a computer science student.

Furthermore, just as the computer is of use in studying phonology and

morphology, it can also offer data organization services to aid in the study of syntax Automated tactic analysis of syntax is still a research project, the software necessary for it is not likely to be produced by a software house. But the research is probably best performed in a linguistics department

Having established a need, we must now recall a warning made earlier Useful contributions to the study of syntax by computers requires more computing power than is needed for similar contributions to the study of phonology and morphology While the need for sophisticated type fonts and input/output devices is lower (not necessarily a good eductional syntax program would permit the manipulation of syntactic trees on a graphics screen), there is a real need for faster processors and increased memory capacity To purchase the necessary computing power, a department would have to step up from the hobbyist microcomputer size machines to the scientific research minicomputer (e g the middle range PDP-11 series) These machines cost an order of magnitude more than the microcomputer and yet, when the subject is syntax, will probably only serve a few students at a time

An alternative, available to some departments, is to use the university's central computing facility. Money could be spent on the best available terminals and the needed communications equipment Grammar testers have been written by university researchers for typical university size computers (Friedman 1971, for transformational grammars, Kehler 1976, for ATN grammars) and are available at little or no cost

As I mentioned in the beginning, the use of the computer in the study of semantics and cognition is still very much a research topic and little, if any of the work being done currently can be performed on small computers I will not describe the requirements of such work since they vary widely depending on the nature of the work.

SOFTWARE FOR THE LINGUIST

What is missing from the computing facilities described so far is software, programs which are of use in solving linguistics problems The small computers are sold with a minimum of very traditional computer software, none of it of any use to the nonprogramming linguist In fact, at no level of computing power is there currently available commercial software which is of use to nonprogramming linguists For large computers, as mentioned above,

some of the results of university research work is available for some purposes However, for the types of machines that departments are likely to purchase, there is essentially nothing

This problem can be overcome in two ways The standard method is for a department to hire a student programmer to design and write the needed software This has several advantages it is relatively cheap (especially when university assistantships are available for the purpose), it is personal - the student can be instructed to write exactly the kind of program that is needed The disadvantages of this method are in the quality and durability of the systems produced in this way Student programmers are, in fact, students learning to program Often their work is lacking in the 'ease-of-use' or 'human engineering' features found in well written, commercially produced programs, and, it is just these features which are very important to users not familiar with or comfortable with computers Furthermore, programs produced by student programmers are not well known for their reliability, maintenance of them is difficult and usually restricted to the period of time that the original programmer is still available Again, to the user unfamiliar with computers, reliability is a very important feature It is very discouraging to try to do anything with semi-operational programs

An alternative is to create sufficient demand for this type of educational software so that a commercial software house or a well funded university programming group would consider the investment of its time and money profitable With linguists and linguistic educators providing input at the design level, very useful and reasonably priced software could be produced in this way The catch, however, lies in generating sufficient demand

A final comment about one other potential use of computers within a linguistics department The search for language universals (cross linguistic research) requires very large collections of information A collection of partial and complete grammars along with sample texts for a large representative sample of human languages is a formidable amount of information The kinds of questions posed by linguists using this information do not require immediate interactive response In fact, they traditionally require weeks or months of library research for answers It is therefore not unreasonable to consider the storage of this information on a small, even hobbyist size, computer equipped with large mass storage devices The task is a difficult one, but

of potential value to both linguists and computer scientists

Linguists need easier access to this information A computerized database, structured according to the needs of linguists, would be a very valuable tool which could be distributed to any department willing to make the necessary investment in hardware The database is large, but unlike many other large databases, it is one about whose structure a great deal is known Computer scientists are still looking for ways to effectively and efficiently organize databases, and linguists, with their intimate knowledge of the structure of language, have an opportunity here to provide an example of how to use the structure of a body of information in storing it on a computer effectively It is a task which requires the expert knowledge of several linguistic disciplines and it is a research project ideally suited to a department of linguistics