

A SAGA CONCORDANCE PROJECT
L. Michael Bell, U. S. A.

1

SUMMARY

Recognizing the potential usefulness of research and teaching aids produced by the computer, a small band of scholars is experimenting with a coordinated effort to prepare Old Norse texts for computer processing. To date, they have produced experimental concordances to at least six fornaldarsögur and to a number of the sagas in Heimskringla. Samples of some of these concordances, and other research tools, will be displayed at the conference.

This paper is a progress report on the project; an announcement of future plans; a discussion of some products of the project; a discussion of some computer-aided research results; a proposal for more cooperation on such projects; suggestions for projects of a similar nature; and a solicitation of recommendations from the scholarly community for future projects and directions.

As of May 1, 1976, the project--tentatively entitled "Computer Research on Early Scandinavian Texts," or CREST--had put between 230,000 and 250,000 words of O.N. prose texts in machine-readable form (IBM punched cards). These texts are now ready to be turned into various sorts of "first-generation" teaching and research aids, such as frequency lists and reverse alphabetizations.

As of the same date, concordances have been produced to some sagas in Heimskringla (Ynglinga, Haralds hárfagra, Hálfðanar svarta, Hákonar góða, Haralds grafeldar) and some fornaldarsögur (Áns bogsveigis, Bösa, Hálfs, Yngvars víðforla, Illuga Griðarfóstra, and Sorla þáttr). It is expected that several more concordances will be produced between May and July, and as many of these as possible will be brought to the conference.

The project's participants plan to publish a "pilot" concordance to five fornaldarsögur via microfilm, with book-form copies available from microfilm. This publication should be available by early 1977. It is hoped that this experimental concordance will serve not only as a useful research tool but as a specimen or prototype, useful in improving concordance design. Such a result, however, will require circulation of the concordance throughout the scholarly community, and informing project members of its good and bad features, as tested in research use.

The project feels that it has shown the feasibility of producing interim results quite quickly (the fornaldarsaga project began in May 1975). Concordances in particular are easily produced by computers, and they are useful in tracking down the widely scattered and unobtrusive lexical clues to certain literary tendencies. This paper will present examples of the above: An apparent pattern of scene structure in a subplot of Egla (Table 6), and a similar pattern which has echoes in other sagas and even in Eddic verse (Table 7). A third example will appear in the next issue of Arkiv för nordisk Filologi, the author's article on "Oral Allusion in Egils saga."

Because the computer can liberate the scholar from previous limitations on time, energy, money, and similar resources, the project's members feel that they are not being precipitous in proceeding with their pilot project before making a careful survey of the needs of the field, and without being overly

L. Michael Bell
U.S.A.

A SAGA CONCORDANCE PROJECT
L. Michael Bell, U. S. A.

2

concerned about the nature of the text used for computer processing. One has to start somewhere; and a second desideratum is to avoid the tendency of scholarly groups to plan and plan forever, over-structure their projects, and not produce any solid results until those results are obsolescent. At the very least, later projects can learn valuable lessons from this project's trials and errors.

Furthermore, decisions about the usefulness and direction of such projects and their results, both with regard to the present and the future, are best made not in a vacuum but with some interim results in hand. It is in this spirit that we want to suggest some future directions and solicit many more.

For instance, it seems possible that computers could render some aid to the dictionary project in Copenhagen. (The Old English Dictionary Project in Toronto is relying heavily on the computer for various crude, quantitative tasks of tabulation.) Second, there are various ways in which the computer can help the scholar preparing additions, collating texts, or searching out various kinds of patterns--lexical, morphological, idiomatic, thematic, structural, and possibly even phonological. Authorship studies are greatly facilitated by even the crudest of computer-produced research aids.

Teaching aids are not the least of the possibilities opened up by computer processing. Idiom lists, grouped-frequency word-lists, morphological analyses, and finding lists of verb-forms are among the types of teaching aids which could easily be produced by the computer. Longer-range possibilities include automatic syntax analysis. As long as the research is properly designed, the computer presents us with the possibility of making a rigorously empirical normative picture of O.N. on almost any desired level. The genius, or the experienced teacher, relies on a "feeling for" a particular language or literature. This "feeling" is, in a sense, a glimpse of the complete picture which computation can help us make. And computation could have no more legitimate purpose than to make that picture available to the teacher or student with less keen intuition or less experience.

The members of the project would like to propose a cooperative effort to harness the computer for the benefit of our field. We first need to know the field's perception of its needs, in the form of nominations for neglected areas of investigation, of procedural and organizational suggestions, and the like. Despite all the acknowledged pitfalls in computer research in the humanities, we feel that we have shown the feasibility of achieving interim results quite quickly. Now, a wider symposium is needed--comments, criticisms, and suggestions from all around the field--to assure that, when research tools are made for the major genres of Old Norse, the job is done right.

L. Michael Bell

U. S. A.

I. Introduction

Eight years ago, after completing a research project which involved searching through about two and a half million words of Old Norse, Professor Peter Hallberg expressed this vision of the future:

"...datamaskinstekniken numera har gett oss förut oanade möjligheter att penetrera och kartlägga en text. Sannolikt skulle ingenting annat innebära en lika stark stimulans för forskningen på området, som om man nu kunde komma överens om ett centrum för databehandling av sagatexter....Där skulle snabbt kunna utarbetas konkordanser över enskilda texter. Ett studium av vokabulär och fraseologi på så bred basis, att det förut tätt sig praktisk ogenomförbart, skulle med hjälp av datamaskin kunna bli en snabb affär, en helt alldaglig uppgift. Inte minst skulle en forskare så gott som omgående kunna få sina intryck och hypoteser på olika punkter verifierade eller falsifierade. Han skulle i ett tidigt skede kunne hejdas inför många blindgator och bli besparad mycket onödigt arbete. Hans krafter skulle frigöras för planläggning och lösning av metodiska problem, medan maskinen övertog rutinarbetet....

"Ett centrum för databehandling av sagatexter skulle ge vår forskning möjlighet att formulera och lösa uppgifter, som hittills har tätt sig utopiska-- i den mån vi ens har varnat dem."¹

One year ago, a small and loosely-knit band of Old Norse scholars decided to experiment with a cooperative approach to preparing ON texts for computer processing. There had been previous sporadic rendering of ON into "computerized" form, but it had always been disorganized, and appeared to be behind some other medieval Germanic languages in both quantitative scope and qualitative results.

Today we would like to give you a First Annual Report on that project-- on its genesis, progress, and direction; and we would like to suggest and solicit both directions and projects for the future. We hope that this will be one pätr in a long saga with a pleasant ending.

Although we intend a bit of hvøt, this will not be another talk on "what computers can do for you." Exhortations of that sort have acquired a rather sad and sour undertone in recent years, after the euphoria of a decade or so ago. "Humanistic computation" is one field that has been retarded by its very promise. One reason for this phenomenon is the nearly ungraspable magnitude of the possibilities which computers hold out to us as humanists. A second is that "the basic questions of literary scholarship...cannot be quantified,"² with consequent difficulties in working out "the relationship between exact quantity statements and fundamental questions."³ Professor Louis Milic, one of the most sensitive observers of the whole computer-humanist scene, has pointed out:

"The real [literary] questions tend to take much vaguer forms [than do scientific questions], partly because we don't know what kind of answer we are looking for or whether an answer is called for at all....This is not to say that literary problems cannot be formulated which can elicit precise numerical answers... Where does Keats use exclamation points? What is Macaulay's adjective-noun ratio? ...[Humanists] have neglected to be very accurate about the peripheral questions which often furnish the material for attempted answers to the fundamental questions...the literature is in consequence muddled with inaccurate and imprecise

L. MICHAEL BELL
U.S.A.

quantity statements for which exact quantities could have been easily substituted." [Milic then cites the example of a type of Swift scholar who "would mention his feeling or impression that in Swift's work verbs were quite frequent or adjectives very scarce but without...taking the trouble of counting any."] ⁴

But at the other extreme, as Milic has also observed, "...comfort in exactness represents a dangerous lure to the humanistic scholar....More dangerous... is the possibility that matters basically qualitative will be distorted into a quantitative form....The illicit mis-shaping of literary problems into quantitative problems seems bound to damage the study of literature." ⁵

His warning, uttered in 1964, has been amply justified in subsequent years; but the more hopeful of his conclusions is also still true: "There is now no reason why we should not have up-to-date concordances of all the poets, indices verborum of the major prose writers, bibliographies and indexes of great variety and completeness, as well as variorum texts of practically everything....These tools will make the fundamental work of the literary scholar easier and less likely to be interrupted for the lack of a simple piece of information." ⁶ The alternative, of course, is for a scholar to re-read his entire text, author, or corpus every time he wants to test the consistency of a pattern he is interested in: syntax, metaphor, concept, image, and so on.

A third reason for this lack of progress is a manifest tendency to go off chasing will-o'-the-wisps like automatic syntax-analysis, machine translation, computer-generated stemmas, and the like; and to neglect the solid but unglamorous tabulation work that the computer can do. Humanists who believe that the computer can do their thinking for them have, in recent years, swelled the body of evidence for the programmer's proverb: "Garbage In, Garbage Out." Translation: Badly-designed research, when entrusted to the computer, results in mistakes more preposterous than generations of scholars could make by hand. There comes to mind the Church of Scotland minister who, in 1965, used a computer to show that six different authors had written the New Testament epistles commonly attributed to St. Paul. He wrote up his results in two long articles in The Observer (London); a fellow cleric from Massachusetts applied the Scottish minister's procedure to those articles and showed, with the same degree of certitude, that they had been written by several different people.

Gaffes such as these have, of course, been instructive and cautionary for those who want to harness the computer to their own specialties. Perhaps the best metaphorical caution would sound something like this: If the phrase "computer research" evokes a picture of a mechanical brain which, when its button is pushed, says, "Yes, Snorri wrote Njála"--the image could hardly be further from reality. If it evokes a picture of a mechanical slave going through index cards with superhuman speed, reorganizing them according to very elementary instructions, then the picture is quite close to reality. And perhaps the best place to begin the main body of this paper is to take it for granted that, despite all the evident pitfalls, the computer has much to offer us; that it is simply a tool which can aid us as traditional humanist thinkers. That granted, we can get down to specifics, and entertain philosophical and procedural objections as they arise.

II. Plans and Progress

The first ten scholars listed in Table 2 ("Fornaldarsaga Project - Progress Chart") agreed in May 1975 to cooperate in the production of an experimental concordance to a sample of Old Norse prose, more or less randomly chosen. We chose the fornaldarsögur because their secondary importance would make the consequences of error less drastic. Sagas in the 8,000-9,000-word range were chosen, partly for the same reason. I incorrectly chose as our text the 1943-44 Guðni Jónsson edition published by Forni (3 vols.); more on this point below. Our small group calls itself CREST, for Computer Research on Early Scandinavian Texts.

Each member "keypunches" one saga onto IBM punched cards, with duplication of the keypunching, so that the punched cards can be automatically collated and most, if not all, keyboard errors isolated and corrected. The cards are collated and corrected at the University of Colorado, and then concorded; each keypuncher receives a "rough" concordance to his saga. (Samples of the collation-program output and the rough concordances are on display at this conference.)

As of 1 May 1976, we have finished keypunching 39,000 of the approximately 41,500 words in the six "pilot" sagas. Most keypunchers have received rough concordances to their sagas. Some additional samples of research aids have been produced from these keypunched texts, and were presented and discussed in a special session of the Society for the Advancement of Scandinavian Study, meeting in Austin, Texas, in April 1976. These aids, drawn from Yngvars saga víð-förla, included "reverse alphabetizations" (see Table 8); a list of word-forms in the saga in order of frequency; and a similar list, with frequencies, but in alphabetical order.

Some members of the project have gone on to keypunch a further 39,000 words of other fornaldarsögur (Table 2); for the longer and more important of these sagas, such as Völsunga and Hervarar, we have switched to the four-volume Íslendingasagnautgáfan edition done by Guðni Jónsson; this text is identical to the Forni edition except for pagination and length of lines. The problem of converting the latter text to the former pagination and line-numbering will be dealt with at the University of Colorado.

Early on in our pilot project, it occurred to us that we might be able to produce at least a rough concordance to Heimskringla, or a significant part of it, in time for this conference. Accordingly, six of the original members, along with a later recruit, set about getting Heimskringla keypunched while still in the midst of the pilot project. The same procedures of double keypunching were followed. The keypunch codes were changed slightly to reflect the change from the semi-popular normalization of the fornaldarsaga edition to the more rigorous Fornrit norm; and provision has been made to incorporate footnotes.

The organization and progress of the Heimskringla project can be followed in Table 1. Of the 229,000 words estimated by P. Hallberg to be in Heimskringla, I have as of 1 May received either cards, or formal notification of completion, representing 150,000 words, give or take a few thousand. Rough concordances to the first three sagas following the Prologus have been made, and we expect to

L. Michael Bell

U. S. A.

make several more by July and bring as many as possible to this conference.

In effect, then, we have already keypunched the equivalent of Heimskringla-- an absolute minimum of 228,000 words. The texts represented by that figure do not overlap ideally, so that the entire volume is not ready for collating and turning into research tools; but certainly a quick interim result has been achieved. The Heimskringla branch of the project began only last Fall.

Our plan for disseminating the results of our work is to publish a prototype concordance to six fornaldarsögur by a method that has been pioneered by Dr. Michael J. Preston, coordinator of the Center for Computer Research in the Humanities at the University of Colorado, and technical director of our project. This method is simply to publish the concordances the way most Ph.D. dissertations are "published" in the U. S.-- through Xerox/University Microfilms Corporation, the publisher of Dissertation Abstracts. Esthetically acceptable printing is achieved by printing-out not with a standard computer print-chain, but through a computer-driven IBM Selectric typewriter, essentially similar to the self-correcting, "programmable" MT/ST model used in many offices in the U.S. and abroad. These typewriters can be equipped with interchangeable "type balls" offering a great diversity of characters and type faces. In some cases, a typographically sufficient ball is available in IBM stock. Such is the case with respect to semi-popular editions such as the Íslendingasagnautgáfan, which reduce ö, ó, and œ to o and æ; a concordance to those editions can be produced with the IBM Icelandic type ball whose print you are now reading. In other cases, a type ball may have to be modified by creating and adding special characters, usually at a cost of a few hundred dollars.

When a carbon ribbon is used, this method produces quite acceptable books, sold by Xerox for \$15.00 (softbound) to \$18.00 (hardbound), or in microfilm for about \$6.00. Dr. Preston has published several concordances by this method, as well as two by conventional publication. Not only does the microfilm-to-book method produce research tools at a reasonable price, but it avoids the enormous subventions required by conventional publication, which in turn necessitate high book prices. For Preston's recent Concordance to the Middle English Shorter Poem (Leeds: Maney, 1975), the subvention was \$3,000; and the 2,500-page concordance costs \$87.00.

Furthermore, the microfilm-to-book method introduces a new flexibility to the design and production of concordances. For example, if it should turn out that our prototype concordance is worthless because of bad design--poor format, insufficient context, too few citations, wrong treatment of compound words, etc.--the economic consequences of discarding it and starting afresh will certainly be smaller than if hundreds of thousands of dollars and man-hours had gone into its production and purchase. To adapt one of P. Hallberg's observations: We now have at least the potential freedom to concentrate on how concordances are best designed, rather than on the brute work of making them.

Without going into much technical detail,⁷ I should like to mention a few features of the proposed format of our fornaldarsögur concordance. We plan to split it into two sections, reserving one section for very-high-frequency (VHF) words such as ok, er, þat, and til. Previous experience⁸ suggests that such words,

L. MICHAEL BELL

U. S. A.

though they may represent fewer than 50 separate forms, will account for fully half of any given ON text. The format for this "VHF" section will resemble that in Table 5, with citations not in text order but grouped by similar following or "right-hand" contexts. The reasoning behind this decision is that such words, "function" rather than "content" words, are useful primarily in more purely linguistic investigations, where text order is not of primary interest. Furthermore, a concordance would hardly be of maximum value if it offered its user only the intimidating prospect of analyzing, by eye alone, the 3,298 occurrences of ok in Egla. The alphabetization of right-hand context is a crude but useful preliminary processing to aid those types of research involving VHF words which we presume (perhaps erroneously) will be primary.⁹

The main section of the concordance will resemble Table 4: In order to achieve more context for each word (since such words are more often analyzed for contextual meaning), the lines will follow the long axis of the page. We plan to include several appendices--ranking-frequency lists, various statistics, perhaps reverse alphabetizations--without going into elaborate predictions as to how many of them will be useful. (This matter will be taken up again in a moment.)

It is hoped that this concordance will circulate among scholars as a prototype, and that enough "field testing" will result in the CREST project being informed about the concordance's good and bad features. The ideal result of such interaction, of course, will be ever more useful concordance designs.

At this point, it might be asked why we embarked so precipitously on such a project, without carefully surveying the needs of our field, and without putting our work on a scrupulous and secure epistemological basis by going back to the manuscripts--or at least to the available manuscript editions. The basic answer is that one has to begin somewhere--even with a popular edition, if it is the most comprehensive available. We did not want to over-structure our project, and we wanted to avoid the scholarly committee's tendency to plan and plan forever, and not to achieve solid results until they are obsolescent. Nor did we want to waste time by planning in a vacuum--especially since the humanist typically cannot know, at the beginning of his research, what specific questions he will want to ask the computer. Accordingly, we decided to strive for what some scientists call a "quick and dirty" result, and to base future planning on interim results which would be concrete to a greater degree. At the very least, both our project and later ones can learn from our trial and error.

The disadvantage to our being a fairly cohesive, quickly-interacting small group--though it may not be obvious at first sight--is that we are not a majority of the scholarly field to which we belong and which we aim to serve. That is why we are here today, announcing our first results and asking for "feedback" from you, both now and in the future.

As I will mention below, the first area in which we need this "feedback" is in your views of the usefulness of our prototype concordance, which should be published by early 1977 at the latest.

L. Michael Bell

U. S. A.

though they may represent fewer than 50 separate forms, will account for fully half of any given ON text. The format for this "VHF" section will resemble that in Table 5, with citations not in text order but grouped by similar following or "right-hand" contexts. The reasoning behind this decision is that such words, "function" rather than "content" words, are useful primarily in more purely linguistic investigations, where text order is not of primary interest. Furthermore, a concordance would hardly be of maximum value if it offered its user only the intimidating prospect of analyzing, by eye alone, the 3,298 occurrences of ok in Egla. The alphabetization of right-hand context is a crude but useful preliminary processing to aid those types of research involving VHF words which we presume (perhaps erroneously) will be primary.⁹

The main section of the concordance will resemble Table 4: In order to achieve more context for each word (since such words are more often analyzed for contextual meaning), the lines will follow the long axis of the page. We plan to include several appendices--ranking-frequency lists, various statistics, perhaps reverse alphabetizations--without going into elaborate predictions as to how many of them will be useful. (This matter will be taken up again in a moment.)

It is hoped that this concordance will circulate among scholars as a prototype, and that enough "field testing" will result in the CREST project being informed about the concordance's good and bad features. The ideal result of such interaction, of course, will be ever more useful concordance designs.

At this point, it might be asked why we embarked so precipitously on such a project, without carefully surveying the needs of our field, and without putting our work on a scrupulous and secure epistemological basis by going back to the manuscripts--or at least to the available manuscript editions. The basic answer is that one has to begin somewhere--even with a popular edition, if it is the most comprehensive available. We did not want to over-structure our project, and we wanted to avoid the scholarly committee's tendency to plan and plan forever, and not achieve solid results until they are obsolescent. Nor did we want to waste time by planning in a vacuum--especially since the humanist typically cannot know, at the beginning of his research, what specific questions he will want to ask the computer. Accordingly, we decided to strive for what some scientists call a "quick and dirty" result, and to base future planning on interim results which would be concrete to a greater degree. At the very least, both our project and later ones can learn from our trial and error.

The disadvantage to our being a fairly cohesive, quickly-interacting small group--though it may not be obvious at first sight--is that we are not a majority of the scholarly field to which we belong and which we aim to serve. That is why we are here today, announcing our first results and asking for "feedback" from you, both now and in the future.

As I will mention below, the first area in which we need this "feedback" is in your views of the usefulness of our prototype concordance, which should be published by early 1977 at the latest.

L. Michael Bell

U. S. A.

III. Some Results of Concordance-Based Research

Before going into more detail about the future, can any advances be claimed that have already resulted from computer-aided research? My answer may be suspect, because I have read of no other published results of such research. But I can at least give you the wherewithal to judge my claimed results for yourselves. I will draw a rapid sketch of some results of my 1966 concordance to Egla, and dwell on two or three of them in a bit of detail--much of which detail is found in Tables 6 and 7.

In skimming through a concordance, one finds patterns leaping to the eye. In tracking these patterns down by verifying them in broader context, I turned up the following observations, large and small, about Egla: (1) In character description (mannlýsing), precocity is a quality only of sympathetic characters, never of villains or adversaries. (2) That the motif of "ominous silence" is almost a royal prerogative, in that a separate lexicon is used to describe the ominous silence of non-kings. (3) That certain scene-types are cut to the same formulaic pattern, right down to the lexical level. (4) That certain formulaic patterns are used to build key scenes, which in turn are the structural units of plot and subplot. (5) That certain phrases, innocuous on the surface, are apparently the cues for specific turns of plot, and may endocentrically be echoes of heroic poetry. (6) That the commonest lexical tags of battle-scenes are not words for "battle," but phrases of the type fell mart, en sumir flyðu; and that the words for "flight" are manipulated so as to show the author's ethical slant.¹⁰

A seventh finding will appear in the next issue of Arkiv för nordisk Filologi, an article entitled "Oral Allusion in Egils saga." It examines the oral-allusion phrase-types which have been treated (rather subjectively) by K. Lies-tøl and W. Baetke, and carefully analyzed by T. Andersson. My study shows, I believe, that the scope of such inquiries must be broadened to include related phrase-types such as varð ekki til tíðenda í þeira ferð, in order to understand the full rhetorical range of the more "canonical" oral-allusion phrase-types. This analysis required constant tabulation and cross-checking, for which a concordance was indispensable. It also turned up some interesting minor patterns of restrictedness in the lexicon--for instance, that geta is used almost exclusively in oral allusion, and that mæla is used only if the allusion is a proverb.

As a more detailed example we may choose item (4) above, which is illustrated in Table 6. If one examines phrases of the type "þú munt ráða" in Egla, one finds that a high proportion of them occur at turning-points in the life of Þórólf Skallagrímsson, Egill's brother. If one then examines all the turning-points in the subplot delineating his life, one notices that this subplot is a skillfully constructed chain of events, all of which (1) proceed from his friendship with Björn Brynjólfsson or (2) are precipitated by the complications in other men's friendships, such as Egill-Arinbjörn. Four pivotal scenes are constructed as father-son confrontations with striking lexical and thematic parallels or echoes--including phrases which occur only in these scenes. This scene-type is fore-shadowed by a scene set in the previous generation. Furthermore, all of the cruxes of Þórólf's life which are not constructed in this way

L. MICHAEL BELL

U.S.A.

can be shown to have been carefully fore-shadowed in some other way. The whole is, among other things, a demonstration of plot-symmetry which would make Chaucer envious.

A similar pattern, item (5) above, is outlined in Table 7. Six of the seven occurrences of weak mikli, -a occur in variants of "sá, inn mikli (maðr)." This phrase in turn occurs only in scenes where the hero's life is being overtly and verbally threatened. Narrowing our focus still further, we find that the phrase is characteristically spoken by one who will soon order his (or her) men to chase and kill the man whom the phrase describes. These scenes are further knit together by lexical echoes which are certainly not bound to the semantic context, though they may be bound to the traditional thematic context. Fragments of the same pattern may be found in Njála; and independent investigation indicates that sa + inn by itself has vestigial mythic resonance.

These unexpected but articulated and consistent patterns must bespeak unexamined levels of artistry in at least this one saga--with a strong presumption that other, unsuspected types of stylization remain to be discovered in classical ON prose. The patterns are clearly grounded in observable lexical patterns (though there is room to over- or underestimate their consistency). Thus do concordances throw into relief patterns and tendencies of which the saga-writers themselves may have been unaware. Such research aids cut tunnels through mountains which once blocked our broader view, or give us X-rays of the constituent structure of these mountains as wholes. The very least that they do is hint at a hitherto unnoticed subtlety in the art of the saga.

IV. Proposals

We suspect that there may be more such surprises in store for us if we enlist the computer in our research. Therefore, from the perspective of a group with some nuts-and-bolts experience in preparing sagas for the computer, and in using computer-produced concordances, we now want to make some suggestions for the future and solicit many more. The suggestions we make here will necessarily be random and diffuse, since it would require a simultaneously microcosmic and macrocosmic view of our field to specify all the needed projects, and to propose specific tools or techniques for each one. Accordingly, the following proposals are highly idiosyncratic.

First, it seems likely that we could render some aid to the dictionary project in Copenhagen. As precedent we might cite the current Old English Dictionary project in Toronto, which is entrusting many strictly quantitative tasks to the computer.

Second, the computer could aid scholars now preparing editions. A scholar working with one of the numerous unpublished romances, for instance, would have his decision-making greatly eased by a compilation of all the examples of a given form in their contexts--not even necessarily drawn from the unpublished mss. themselves, but just from published romances. Such a compilation tends to help define the gross distinctive characteristics of a given corpus. Once the keypunching is done, rough concordances are easily produced; and the effort would hardly have been wasted, even if only one scholar ever used the rough concordance.

L. Michael Bell

U. S. A.

Third, on the literary level: Do the thematic and lexical patterns which I think I found in Egla hold good for other sagas? Are there discriminators of saga or genre style to be found in such patterns? No doubt there are even more interesting types of literary investigations to be suggested.

This brings us to questions of authorship. My personal view is that such determinations, when purely statistical, are sterile and possibly misleading. The most exciting prospect which they raise is that of pedagogic by-products. Before a computer could confidently assert that Snorri wrote Egla, it would have to know so much about norms of syntax and style, and about relative deviations from them, that its by-products would make exciting teaching tools.

Such by-products could include things like the Grouped-Frequency Word-List familiar to many beginning students in Anglo-Saxon in the U.S.; complete lists of attested verb-forms, of the sort available to students of Classical Greek; lists of idioms; or any of the numerous aids used by students and teachers of the more "popular" languages, living and dead. As teachers of Old Norse, haven't most of us wished that we could instantly produce many examples of troublesome but typical constructions? As it is, we must generalize from mostly amorphous experience, from our "feeling for" a language or literature. In the case of a great teacher, this "feeling" deserves more respect than any product of a computer. But not everyone, least of all the inexperienced teacher, has the "feeling"--it is, in a sense, an intuitive glimpse of the complete picture that computation can give us. And computation could have no more legitimate humanistic purpose than to make that picture available to a teacher or student with less keen intuition, or less experience.

It might be worth mentioning here that one such experiment has already borne pedagogic fruit. When Kenneth Chapman was preparing his Graded Readings and Exercises in Old Icelandic, he went through a hand-made word-count to 400,000 words of ON, done under Einar Haugen's direction in 1939-42, to find words and phrases whose high frequency assured them of being characteristic of ON, and so important for introductory classes. We would like to see the same thing happen on a larger scale, and done by machines.

But first of all, we need to know the field's perception of its needs. We need nominations for neglected areas of research, suggestions for research aids and procedures, and the like. Is manuscript collation a pressing need? Should we be producing reverse indices (Table 8) for normative morphological studies? Should we be devising procedures for encoding the manuscripts themselves, as John Weinstock has begun to do at the University of Texas? Should we be analyzing skaldic verse? Should we be thinking about typological studies of runic inscriptions? -- It should be mentioned here that the University of Colorado has contracted to purchase some computer equipment exclusively for humanistic applications, and that we want to entertain suggestions for computer-oriented projects in Old Norse which could be carried out on this equipment (at present, graphics screens and typewriter terminals).

Finally, I would like to urge that our prototype concordance be circulated widely within the field, and suggestions concerning it be sent to us. (To avoid sounding like a salesman, I might propose that some philanthropic foundation

provide the wherewithal to distribute free copies to large numbers of active scholars in the field.) We need to know your opinion of this concordance and its design, as tested in research use. Is it inclusive enough? Too inclusive? Legible? Easy to work with? What would be your suggestions for improving concordances of this type? What appendices would you like to see in the ideal concordance? (For an example of some of the possibilities in concordance and appendix design, I heartily recommend Einar Bjorvand's A Concordance to Spenser's "Fowre Hymnes" [Oslo: Universitetsforlaget, 1973], though it illustrates more of the conventions of a poetic concordance.)

For reasons already given, we will refrain from suggesting a symposium to demarcate the most pressing needs of our field. But we want to make sure that, when concordances are made to the major genres, they are done right.

Notes

1. Peter Hallberg, Stilsignalement och författarskap i norrön sagalitteratur (Göteborg, 1968), p. 170.
2. Louis T. Milic, "Some Risks of Technological Overindulgence for the Humanities," in Literary Data Processing Conference Proceedings, ed. by Bessinger, Parrish, and Arader (White Plains, N.Y.: IBM Corporation, 1964), p. 61.
3. Milic, p. 60.
4. Milic, pp. 58-60, passim.
5. Milic, p. 62.
6. Milic, p. 61.
7. It should be understood here that broad criteria of usefulness--what to do with compound words, whether to lemmatize or disambiguate--require some highly technical decisions and processes at the computer end of such an operation; but that technical considerations need not, in general, deter the humanistic scholar from making his plans for a concordance or similar project.
8. This experience was gathered during the author's year as IBM Fellow at the Harvard Computing Center in 1965-66, during which year a concordance to Egils saga was produced under his direction. Of the 65,517 words which the concordance program found in Egla, 33,978 were comprised by just 104 forms which occurred a minimum of about 100 times each. For details consult the author's "A Computer Concordance to Egils saga Skalla-Grimssonar," Studies for Einar Haugen (The Hague: Mouton, 1972), 58-68.
9. Here we should mention a possibility which can serve as an "escape valve" from potential error in our assumption: If there is a scholar who desperately needs a concordance of all the occurrences of ok in text order, it is a simple matter to make, for him alone, a rough concordance to ok, or to all the conjunctions in the text in question. Our strong presumption is that there are unlikely to be large numbers of scholars desirous of having such compilations--still fewer desirous of going through them more than once.

SAGA CONCORDANCE PROJECT
L. Michael Bell, U. S. A.

12

10. Complete presentation of all these patterns is made in the author's doctoral dissertation, "Lexical Patterning in Egils saga Skalla-Grímssonar: A Computer-Aided Approach," Harvard University Ph.D. Diss., 1973.

L. MICHAEL BELL
U.S.A.

SAGA CONCORDANCE PROJECT
L. Michael Bell, U.S.A.

Table 1: Division of Labor in the Heimskringla project -
Sections averaging 120 pp. each (Íslensk Fornrit edn.)

SECTION	PP.	CONTENTS
1. First 4 sagas in Vol. I	142	Prologus, Ynglinga, Hálfðanar svartá, Haralds hárfagra
2. Ólafs Tryggvasonar	147	
3. Ólafs helga (1/3)	136	(pp. 3-138)
4. Ólafs helga (1/3)	135	(pp. 139-274)
5. Ólafs helga (1/3)	141	(pp. 275-415)
6. Remnants of Vols. I-II	105	Hákonar góða, Haralds gráfeldar, Úr Ólafs sögu sérstöku
7. Haralds Sigurðarsonar	134	
8. Remainder of first 5 in Vol. III	136	Magnúss góða, Ólafs kyrra, Magnús berfoætts, Magnússona
9. Last 4 in Vol. III	136	Magnúss blinda, Haraldssona, Hákonar herðibreiðs, Magnúss Erlingssonar

SECTION	FIRST KEYPUNCHER	SECOND KEYPUNCHER
1.	✓ Lindow (Berkeley)	Bell (Colorado)
2.	Weinstock (Texas)	Mitchell (Minnesota)
3.	✓ Weinstock	✓ West (Seattle)
4.	✓ Weinstock	✓ West
5.	✓ Weinstock	✓ West
6.		✓ Lindow
7.		Lindow
8.	✓ Conroy (Washington)	Lindow
9.	✓ Conroy	Jorgensen (Riverside)

✓ = keypunching completed (as of 1 May 1976)

L. Michael Bell

USA

Table 2

FORNALDARSAGA PROJECT - Progress Chart (X indicates completion of a step)

Saga	Keypunchers		Collate & "Rough"	In Charge
	1	2		
Ans s. bogsveigis	X	X	Correct	Concordance
Bósa s. ok Herrauðs	X	X	X	X
Hálfdanar s. Brönufóstra	X	X	X	X
Hálfs s. ok Hálfssrekka	X	X	X	X
Yngvars s. víðförla	X	X	X	X
Illuga s. Griðarfóstra	X	X	X	X
Sörla þátr	X	X	X	X
Hrómundar s. Gripssonar	X	X	X	X
Völsunga s.	X	X	X	X
Orvar-Odds s.	X	X	X	X
Gríms s. Loðinkinna	X	X	X	X
Þátr af Ragnars sonum	X	X	X	X
Hervarar s. ok Heiðreks	X	X	X	X

Notes: (1) If a saga appears on the list at all, keypunching has begun.

(2) The projected final step is a concordance to 5 or 6 sagas combined, to be produced on a computer-driven typewriter terminal and published in microfilm (with book versions available from microfilm).

SAGA CONCORDANCE PROJECT
L. Michael Bell, U.S.A.

Table 3: Old Norse and Faroese Texts in Machine-Readable Form

All texts listed herein have been completely keypunched at least once, and many have been keypunched in duplicate. Many have also been concorded. Texts are listed in approximate chronological order of keypunching.

<u>Text</u>	<u>Principal Investigator(s)</u>
Egils saga	L. Michael Bell (Colorado)
Egils saga	Pardee Lowe, Jr. (Falls Church, Va.)
Sjúrðarkvæði	L. Michael Bell
Dialogues of Gregory	John Weinstock (Texas)
Grágás	H. Fix, H. Beck, et. al. (Saarbrücken)
Fóstbrœðra s., Hrafn- kels s., Egils s., Njáls s., chs. 1-50	Ralph A. West (Seattle, Wash.)
Bartholmeus s. postola	L. Michael Bell
Elder Edda (partial)	Joseph Harris (Stanford) & L. Michael Bell
FORNALDARSQUR	(see Table 2)
HEIMSKRINGLA	(see Table 1)

L. MICHAEL BELL
U.S.A.

Table 4. Proposed Concordance Format for Low- to Medium-Frequency Words, illustrating "KWIC" Format (headwords centered) with entries in text order. Sample is from concordance to Haralds saga hárfagra; numbers at left refer to page and line in Íslensk Fornrit edition.

HAUG (2)	þeir höfðu verit at þrjú sumur at gera haug eim. Sá haug var hlaðinn með
099 21	r hauginn. Hrollaugr konungr fór upp á haug þann, er konungar váru vanir at
100 01	
HAUGINN (2)	
099 26	Eptir þat gekk Herlaugr konungr í hauginn með tólfta mann. Síðan lét hann
099 27	a mann. Síðan lét hann kasta aptr hauginn. Hrollaugr konungr fór upp á ha
HAUGINUM (3)	
147 17	onungs, sá er lá yfir legi hans í hauginum, ok er steinninn hálfv fjoðrtá
147 19	nær tveggja álna breiðr. Í miðjum hauginum var leg Haralds konungs. Þar v
147 23	ar í kirkugarðinum, er þá váru í hauginum ok nú var frá sagt. Svá segja
HAUGR (3)	
099 21	þrjú sumur at gera haug einn. Sá haug var hlaðinn með grjóti ok lími ok
147 15	jálfan kirkjugarðinn í útnorðr er haugr Haralds konungs ins hárfagra. Fyr
149 08	igróðr fellu þar báðir, ok er þar haugr hvárs tveggja þeira á brekkurni,
HAUGRINN (1)	
099 22	jóti ok lími ok viðum górr. En er haugrinn var algórr, þá spurðu þeir br
HAUGSINS (1)	
099 25	þar lét Herlaugr konungr aka til haugsins vist mikla ok drykk. Eptir þat
HAUST (2)	
128 21	settisk hann í löndin. En bæði á haust ok um vetr ok um vár fóru víkígar
131 03	ör eyjunum ok kom aptr þegar sama haust ok kom þá óvart Hálfðani. Þeir hi
HAUSTDEGI (1)	
106 15	n skyldi eigi fara suðr um Stað á haustdegi. Þá setti Haraldr konungr Rög
HAUSTIT (5)	
104 05	menn ok treysti sér fólkitt, en of haustit bjósk hann at fara norðr til þr
106 14	konungr yfir. Þetta var síðla of haustit, ok gera menn þat ráð með Harald
109 19	n svá um sumarit of þat fylki. Um haustit fór hann upp á Raumaríki, ok fó
129 10	til handa honum. Sigldi Einarr um haustit vestr um haf. En er hann kom ti
134 09	ng. Fór konungr þá austr eptir um haustit. Var þat lengi síðan í Orkneyju
HÁBRÓK (1)	
144 19	glands ok fekk til stýrimann Hauk hábrók. Hann var kappi mikill ok inn kæ
HÁLEGG (1)	
131 13	r þannug til ok fundu þar Hálfðan hálegg ok tóku hann höndum. Einarr jarl
HÁLEGGR (4)	
126 19	, einn var Sigurðr hrísi, Hálfðan hálegg, Guðrórðr ljómi, Rognvaldr rétti
130 14	. Þá fóru til á einu vári Hálfðan hálegg ok Guðrórðr ljómi með mikla svei
131 01	gnvaldr jarl, faðir hans. Hálfðan hálegg kom vestr til Orkneyja ok mjök
138 07	n hvíti fell á Eistlandi, Hálfðan hálegg fell í Orkneyjum. Þeim Þorgisli

L. MICHAEL BELL
U.S.A.

Table 5: Proposed Concordance Format for High-Frequency Words, Illustrating "Extended KWIC" Format: entries arranged not in text order, but alphabetized by 20 characters following headword. This sample represents 36 of the 203 occurrences of the form er in a rough concordance to Haralds saga hárfagra.

ER (203)

100	24	náðu hirðvist með Haraldi konungi,	er	afreksmenn eru at afli ok hrey
105	07	tyrk, ok má auðna ráða sigri. Hinn	er	annarr kostr, ok er þat þó eng
143	10	il at ausa vatni eða gefa nafn. En	er	at þeiri stefnu kom, er þóru v
098	03	ðu griða, ok þat fengu allir, þeir	er	á konungs fund kómu ok gerðusk
120	09	mitt landit herjuðu víkingar, þeir	er	á vetrum váru fyrir vestan haf
111	22	anda. Riðr Haraldr konungr þannug,	er	áðr hafði riðit Eiríkr konungr
141	25	álfðan skyldi halda ríki öllu, því	er	áðr hafði hann haft, skyldi ha
107	07	æri. Tók Rognvaldr jarl skip þau,	er	átt hafði Vémundr konungr, ok
107	21	mir þreyta þetta með sér þar til,	er	báðir samna her. Þeir hittusk
098	15	nn setti jarl í hverju fylki, þann	er	dóema skyldi lög ok landsrétt
105	08	er þat þó engi kostr þeim mönnum,	er	eigi eru ótignari en Haraldr
120	16	tland ok drap þar alla víkinga, þá	er	eigi flyðu undan. Síðan siglir
096	08	ess at taka til manns þann konung,	er	eigi hefir meira ríki en nokku
097	10	hann, "er mér þykkir nú undarligt,	er	ek hefi eigi fyrr hugleitt."
110	21	vinattu vert fyrir góðvilja minn,	er	ek hefi lýst fyrir þer í heimb
135	23	at þar er komit. Hon segir, at þar	er	ekki komit. Finnum þykkir þat
096	10	þykki mér undarligt," segir hon,	er	engi er sá konungr, er svá vil
098	01	a menn alla ok brenna byggðina. En	er	fólkit varð þessa víst, þá fly
096	16	aldr er konungr svá ríkr, at henni	er	fullræði í. En þó at hon svari
097	04	i orð meyjarinnar ok telja, at hon	er	furðu djörf ok óvitr, ok segja
120	19	r. Hann drap þar marga víkinga, þá	er	fyrir liði reðu áðr. Hann átti
144	26	segir, at sá skal síðast út ganga,	er	fyrstr gengr inn, ok allir sta
118	09	ða. Margir váru þeir ok ríkismenn,	er	gengu til handa Haraldi konung
142	20	settisk hann optliga at stórbúum,	er	hann átti á Hórdalandi á Alrek
147	07	andi eða á Hórdalandi at stórbúum,	er	hann átti. Eiríkr ok Gunnhildr
107	08	undr konungr, ok allt lausafé, þat	er	hann fekk. Berðlu-Kári fór nor
107	20	ok kvezk því mundi halda, til þess	er	hann fyndi Harald konung. Jarl
110	26	n greiddi hann fram stórar gjafir,	er	hann gaf konungi. Síðan gekk Á
147	05	ldr konungr lifði þrjá vetr, síðan	er	hann hafði Eiríki gefit einval
138	06	eptir þat tók Ólafr við því ríki,	er	hann hafði haft. Hálfðan hvíti
094	09	ngu margir höfðingjar á ríkit, þat	er	hann hafði leift. Var inn fyrs
122	13	ástr, ok fekk hann þar af bana, ok	er	hann heygör á Ekkjalsbakka. Þá
109	16	hann þá sakar á hendr þeim böndum,	er	hann kenndi landráð við sik. B
135	02	ar orrostu mikla ok hafði sigr. Þá	er	hann kom aptr á Finnmark, þá f
097	20	k þaðan norðr um Döfrafjall, ok þá	er	hann kom ofan í byggðina, þá l
123	11	ði mjök í Austrvegu. Á einu sumri,	er	hann kom ór víkingu austan í V

SAGA CONCORDANCE PROJECT

L. Michael Bell, U.S.A.

Table 6

THE "ÞÓRÓLF II" PATTERN: An apparent pattern in scene-structure, detected with the aid of a computer-produced concordance to Egla. Illustrates a type of literary analysis aided by concordances, since it requires frequent cross-checking of words in context.

Scenes 2-5 and the fragment on pp. 139-40 represent five of the eight cruces in the subplot narrating Þórólf Skallagrímsson's life; the other three are not patterned, but each is carefully foreshadowed in its own way. Scene 1 is a proleptic occurrence, interpretable as a foreshadowing of the main pattern.

The hyphenated pairs 2-5 represent friendships or confrontations arising from friendships. Abbreviations: K = Kveldúlf, Þ = Þórólf Kveldúlfsson, Þ2 = Þórólf Skallagrímsson, B = Björn Brynjólfsson, Er = Eirík, SG = Skallagrím, E = Egill, A = Arinbjörn. Themes are given Roman numerals, sub-themes letters, and phrase-types Arabic numerals 1-4. Other Arabic numerals refer to page numbers in the Íslenzk Fornrit edition.

	Scene 1	2	3	4	5	Echo	Other occur- rences?
	K-Þ	Þ2-B	Þ2-Er	Þ2-SG	Eg-A		
I. Friendship resulting in filial intractability	14	88	92	96	105		?
1. títt...fylgjusamr		88			105		No
2. "eitt skal ganga yfir okkr báða"			92		114, 122	184	No
II. Grudging Consent							
A. (Paternal) reluctance vs. (filial) insistence	14		92	96	122	139	?
3. "þú munt ráða"	14	88		96	122	140	Common idiom
4. "ver til hans svá vel, sem þú vilt"		88	93				No
B. Prophetic warning	14		93	96	114	140	?

Note On pp. 93 and 140, the "Prophetic warning" theme uses the phrase munu...iðrask, which also occurs nowhere else in the saga.

This pattern shows a striking degree of articulated structure--probably conscious design on the part of an author--in the construction of crucial scenes. The whole subplot of Þórólf Skallagrímsson's life displays a remarkable symmetry; the sequence of Scenes 2-5 in particular represents a chain of causation which gives evidence of extremely careful, skillful, and subtle artistry.

L. Michael Bell

U. S. A.

Table 7

THE "SÁ INN MIKLI" PATTERN: Evidence for scene-structure patterning centered around the use of weak mikli, -a in Egla.

The scene-type below is the aftermath to a confrontation involving Egill or Skallagrím, except p. 203, which is single combat. The degree to which the lexical and thematic elements of this pattern are restricted to this scene-type is illustrated in the last column, and in the notes below. In the basic form of this pattern, the speaker of the "sá inn mikli" phrase-type soon issues instructions to chase and kill the person whom the phrase describes.

Abbreviations: SG = Skallagrím, H = Harald, Eg = Egill, Er = Eirík,
Lj = Ljótr inn bleiki, Av = Arnviðr jarl

<u>Phrase-type or motif</u>	Scene 1	2	3	4	5	Other occurrences?
	SG-H	Eg-Er	Eg-Er	Eg-Lj	Eg-Av	
sá inn mikli	63, 65	111	157	203	233	No
mega ætla/marka	65		153		233	m.marka, 299
"chase" } "kill" } Instructions	65	111	(160)		233	?
	65	111	160	(203)	233	?
komask í fœri	65		160			koma í fœri, 60
allfúss at berjask			158	203		No
Advice to Depart	65		160			

Notes

1. Six of the seven occurrences of weak mikli, -a are accounted for here. It can further be shown that the pattern above does not reflect any properties of the word mikill, the construct "sá + Article + mikill," weak forms of mikill (meiri, mesti, etc.), or anything similar. All 376 occurrences of forms of mikill in the saga were examined in context.

2. Scene 3 is extremely/diffuse from the viewpoint of a basic or "standard" form of the pattern (as represented in Scenes 1 and 5). It is full of "exceptions," but the number of elements which appear is striking.

3. Fragments of the same pattern appear in Njála, ch. 119, esp. pp. 298-302 (Fornrit), applied to Skarpheðinn. Variants of the phrase sá inn mikli, as a focusing device on a hero, can be found in other sagas and in Beowulf.

4. Independent investigation (Ulrike Sprenger, BGDSL 87:74-92) finds evidence that sá inn by itself has vestigial mythic resonance in Old Norse.

L. Michael Bell

U.S.A.

SAGA CONCORDANCE PROJECT
 L. Michael Bell, U.S.A.

Table 8

A sample of a "Reverse Alphabetization" of the words in an Old Norse text; sample taken from Yngvars saga víðförla. Such an index is potentially useful in morphological studies, among other areas. Only the first 41 of 9,200 words are listed. (Note: This version of the index, which is only experimental, puts all long vowels at the end of the alphabet; thus fimmtánda follows frænda.)

eirtrumba	grafa
halda	gefa
gjalda	knefa
vilda	.
fjǫlda	.
gapanda	.
standa	.
vanda	.
enda	.
tíðenda	
óhægenda	
senda	
samkunda	
frænda	
fimmtánda	
skaða	
eða	
vígða	
leiða	
reiða	
garða	
harða	
jarða	
verða	
ǫndverða	
heyrða	
dauða	
eyða	
græða	
ráða	
forráða	
bíða	
ríða	
skríða	
víða	
bjóða	
skrúða	
hafa	