TOWARDS AN ELECTRONIC GREEK HISTORICAL LEXICON *

The writer sets out for the first time the plans for the progression of the *TLG* project in the near future. In a first stage, the production of a «Greek Electronic Lexicon» is intended. This job will involve the lemmatization of all the forms included in the *TLG* Index and the morphological categorization of words. Consequently, it will become possible the grouping of words according to derivation and composition. Automatized access to the texts combining information on their date, authorship, genre, and morphological characteristics will give rise to new, unforeseen research subjects and perspectives. The semantic analysis of words is postponed for a later stage in the project.

A *Thesaurus Linguae Graecae* along the lines of the *Thesaurus Linguae Latinae* is simply unfeasible. This is old news, but Hellenists continue to feel the need for a comprehensive historical lexicon to ancient Greek. We have, of course, nothing to complain of. We can agree upon our admiration and appreciation of the work embodied in Liddell-Scott-Jones; and the *Diccionario Griego-Español* will build upon, and improve upon, LSJ. Lexica such as LSJ and the *Diccionario* are obviously essential for all our daily work; they are among the basic tools which sit at the elbow of everyone who studies the ancient Greek language, literature, or culture. But as the *TLG* data bank nears completion, and along with it the other Greek data banks (that is, those of the Greek papyri and inscriptions), one naturally wonders whether it may be possible to create a yet fuller description of the Greek language.

When one speaks of assembling a comprehensive historical lexicon of Greek, the difficulties —and there are many— are plain enough. The 1905 pronouncement of Diels is much quoted, and much of it is still true, but it is worth our attention also for what it teaches us about how much has changed.

---

* Paper read at the International Conference *Thesaurus Linguae Graecae et Thesaurus Linguae Latinae: New Directions in Greek and Latin Lexicography*, Irvine 17-18 december 1993.

> Even if we were to assume that [the Greek texts from Homer to Nonnus] had all been worked over, slipped, or excerpted by a gigantic staff of scholars, and that a great house had preserved and stored the thousands of boxes, whence would come the time, money, and power to sift these millions of slips and to bring Νοῦς into this Chaos? Since the proportion of Latin to Greek Literature is about 1 : 10 [1], the office work of the Greek Thesaurus would occupy at least 100 scholars. At their head there would have to be a general editor, who, however, would be more of a general than an editor. And if this editorial cohort were really to perform its task punctually, and if the Association of Academies were to raise the [money] necessary for the completion of (say) 120 volumes; and if scholars were to become so opulent that they could afford to purchase the *Thesaurus Graecus...* — how could one read and use such a monstrosity?

Now clearly we no longer need to be so very pessimistic about the enormous task of comprehensive slipping and the imposing logistical problems of setting up a «great house» with «thousands of boxes». There is still work to be done, but the electronic *Thesaurus Linguae Graecae* has largely taken care of that — and offers of course a wealth of immediate benefits as well. And here again, we should not forget the important work of the parallel projects at Duke, Cornell, Ohio State, and elsewhere (all under the aegis of the Packard Humanities Institute) which are performing the data entry of Greek papyri and inscriptions. Yet many of the problems of bringing «Νοῦς into this Chaos» remain, and there is to my knowledge no plan and little hope for a *TLG* which will function in some sense as a Greek counterpart to *TLL*. It is not hard to see why this is so. Given traditional techniques and goals, there is simply too much in the way of data and too little in the way of resources to attack the problem successfully.

The question then arises whether technology can help further in the goal of creating a comprehensive historical lexicon. The answer is, of course, «yes», and to be sure we can see profitable use of technology in ongoing lexicographical research. The *TLL* for example now uses word processing programs to improve the efficiency of redaction and printing, and it uses data banks available on CDROM to check and augment the materials in its *Zettelarchiv*. One can imagine ways in which a lexicon project could more thoroughly integrate computer resources into its day-by-day work. Simply not having to retype and proofread the passages one quotes could, for instance, offer considerable advantages of efficiency. And indeed in the companion paper to this one Professor Adrados describes in some detail how the *DGE* is applying tech-

---

[1] A ratio of perhaps 1 : 6 appears closer to the truth.

nology to speed and improve its lexicographical research. But for the goal of a comprehensive Greek historical lexicon, thinking along these lines will not get us very far, since the nature of the problem, as we have already seen, is not incremental but fundamental: in the construction of a c o m p r e h e n s i v e Greek lexicon, the question is not one of e f f i c i e n c y, but one of f e a s i b i l i t y.

We need then to ponder more radical steps, ones which may take us far away from the methods and format of a printed *Thesaurus*. The goals will need in some way to be reformulated for the project to become feasible. We must ask not whether technology can help in the production of a Greek *Thesaurus*, but whether technology can offer us a different strategy altogether for attacking the problem. Or to put the question more specifically: In what ways might an e l e c t r o n i c lexicon be fundamentally different from one produced and distributed through the medium of print, and how might these differences be exploited to help in the goal of understanding the historical development of Greek words?

I will begin with two observations which though obvious are potentially very powerful. The first is simply that an electronic lexicon need not be worked on in alphabetical sequence. Early in the history of *TLL* Leo urged that the same scholar work on both the base word and all of the prefixed forms: thus *colere*, but also *excolere*, *incolere*, *percolere*, etc., would be studied together to understand better the similarities and differences among these related words. For *TLL*, the logistics of production and pressures of publication made this strategy not possible, but no such constraint need apply to a lexicon created and distributed in electronic form.

My second simple observation is that Greek has a very strong tendency to compound formations. Prefixes and suffixes of many kinds are productive in ancient Greek. The consequence of this productivity is a large vocabulary, much larger than Latin. From the point of view of creating a comprehensive lexicon, this large vocabulary is as much of a hindrance as the size and variety of the corpus. But the size of the vocabulary is, to a certain extent, a phantasm. A reader of Greek will rely not so much on knowledge of every individual word as on knowledge of the base words, prefixes, and suffixes, and on knowledge of how these elements interact. The reader, regardless of the level of expertise, will commonly deduce the part of speech, meaning, and even the usage of an unfamiliar word w i t h o u t   t h e   n e e d   o f   c o n s u l t i n g   a   l e - x i c o n. Once one knows the word γάμος, for example, a reader with some knowledge of the Greek suffixes and prefixes will hardly need a

lexicon to guess the meaning of γαμέω and γαμέομαι, γαμέτης, γαμικός, νεόγαμος, σύγγαμος, and a host of related words. But for our purpose, there is another, perhaps more important consequence of the close relationships among Greek words. What I mean is simply this: For the scholar researching the history of the word and its usage, looking at γάμος in isolation will u s u a l l y be inadequate. That is, a scholar will often want to study at least some of the related formations. How ideal then is a Greek lexicon organized primarily by separate lemmata? The different set of habits and needs among Greek readers and researchers raises, I think, some interesting and fundamental questions about lexicographical principles of organization[2].

What if then we do not begin with alpha? And what if we do not think of individual words as isolated items but as the members of various word groups? What advantages does this offer us? What I would like to propose is a sort of «bare bones» Greek historical lexicon, not anything to compare to *TLL*, but rather a skeleton or a framework for future efforts, a framework which however will itself be useable and useful. The advantages I claim for this «electronic lexicon» are three: (1) I believe it adapts more congenially to the actual habits and needs of Greek scholars and students than traditional printed lexica; (2) it will gracefully accommodate whatever future efforts are made to expand the lexicon, including easy incorporation of publications pre-existing in electronic form; and (3) the source materials needed to create this «electronic lexicon» will be useful in other applications, such as in pedagogy and a wide range of research.

The basic idea is straightforward. Take as an example the noun μανία. Let us imagine that you ask our «electronic lexicon» for this word. You are presented then with part of speech, and with forms and gender (as appropriate). You are also presented, on demand, all the extant passages for this word in a chronological arrangement, followed, again on demand, by a list of relevant bibliography. So far, it is more or less as though you have been presented with an electronic form of the word slips organized in the *TLL Zettelarchiv*. But now imagine that you would like to see more about this word, which in Greek will often mean wanting to know more about the related words. Ask our «electro-

---

[2] This point has long been realized by Greek lexicographers. The *Thesaurus* of Stephanus was organized by word groups, but this was criticized by users as bothersome — and thus the pirated reprint by Scapula, which was organized entirely alphabetically, became the more widespread edition. Similarly, Passow originally planned to organize his dictionary by etymological word groups, but was dissuaded by the vigorous criticism which this plan evoked.

nic lexicon» for a conspectus of related formations. Look at the adjecti-
val derivatives, and you find μανικός; look into the derivatives of that,
and you find the adverb μανικῶς. These words may be segregated, in
the traditional way, under their own lemmata, but, if you like, you can
request a combined chronological listing of passages for
all three related words. That is, we can view all the extant pas-
sages for all three words combined into a single chronologically-ordered
listing. Cast the net a bit further, and you can include other derivative
words, such as ἀμανικός, ἀμανία, μανιώδης. Or you can look further
into the word group and include μαίνομαι, and then include the deriva-
tives to this word by prefix and suffix, such as ἀπομαίνομαι, ἐκμαίνομαι,
ἐπιμαίνομαι, μαινάς, μαινόλης, and so forth. Again, all of these can be
looked at as individual entries, or they can be collected together in
groups or as a whole into a single chronological listing of the passages
in which they occur. The chronological listing can itself be restricted in
various ways, as for example to include only verse, say, or only epic, or
only occurrences before the third century B. C.

    With the words organized in this way —what in the technical docu-
mentation I call «morphological categorization»— it becomes possible
for the student to see how the various words interrelate and for the
scholar to see how the group of words, not just a single member,
developed over time. Best of all, the researcher can himself define what
portion of the group interests him, from a single word to a cluster of
closely related derivatives to the group as a whole. The researcher inter-
ested in μανικός, for example, may want to know not only that the
word first securely appears in the generation of Isocrates, Plato, and
Xenophon, but also that related words such μανία, μαινάς, and μαίνομαι
are found in much earlier authors such as Ibycus and Homer and the
Hymns. The researcher interested in the instances of ἐρωμανία in the
Greek Anthology will want to know not just that the word occurs only
there and in Eustathius, but also that ἐρωτομανία and ἐρωτομανής are
the more widespread compounded formations. In a similar way one can
reasonably argue that just as important to the history of the simplex
μανία are its many compounds, words like ἀνδρομανία, αὐτομανία,
γυναικομανία, δοξομανία, εἰδωλομανία, θεατρομανία, θηλυμανία, ἱερομα-
νία, ἱππομανία, λιθομανία. With the sort of electronic lexicon I have des-
cribed, it would be possible to look at all of these at one and the same
time, grouped together and then, if you like, restricted by work, author,
genre, or chronological period.

    But one need not always begin from the point of view of the word
or the word groups. Let us imagine that an interest in, say, μανικῶς

extends to more general questions about the suffix -ικῶς. Just as μανικῶς, in our hypothetical lexicon, has been classed as an adverb ending in -ῶς, so it has also been put down as a derivative of μανικός, which is itself classed as an adjective ending in -ικός. Thus it would be possible for the researcher to define an author, genre, or time period of interest and then to retrieve all adverbs with the -ικῶς suffix, or all adverbs in -ικῶς and all adjectives in -ικός. One could then study the relationship between adjective and adverb, between adjective/adverb and noun; the words could be organized into groups and these could be presented as a combined chronological listing of passages; one could look at charts of distribution and frequency over time; and so forth. Tools of this sort would make it possible for the scholar to answer easily very difficult questions, such as whether or not the appearance of μανικός and μανικῶς, in the first half of the fourth century reflects a more general trend towards productivity in the -ικός suffix in that period.

Now the reader may want to reply at once that the answer to this question, at least, is already known. C. W. Peppler in a 1910 article already discussed the rise in productivity in the -ικός suffix, and his work forms the base for that part of the discussion in Buck-Petersen's standard reference[3]. And yet this is a good case in point. Peppler argued (and here I quote the summary in Buck-Petersen) that «adjectives [in -ικός] were in the service of the 'new culture', and symbolize the expansion of intellectual interest; they were the tools of the sophists and of abstract thought. They made their way into the language of the polite society of Athens, affected as it was by the sophists, but the language of the common people was relatively free from them». The figures which Peppler presents to support his case are at first glance impressive. But on second consideration, one may harbor doubts. He notes for instance that in the older Aeschylus and in the conservative Sophocles, respectively 12 and 8 adjectives in -ικός are found, whereas in Euripides there are 24. The «new-fangled» Euripides uses these words, then, two to three times as often. But since the corpus of Euripides is three times that of Aeschylus or Sophocles, this is hardly solid evidence for a link between adjectives in -ικός and the encroachment of the «new culture». Similarly, the «conservative» Isaeus uses 7 -ικός adjectives, while the «liberal» Isocrates uses 55; but not only is the corpus of Isocrates 4 times as large but it is also much more varied in subject

---

[3] Charles W. Peppler, «The Termination -κός, as Used by Aristophanes for Comic Effect», *AJPh* 31, 1910, pp. 428-444; Carl Darling Buck and Walter Petersen, *A Reverse Index of Greek Nouns and Adjectives*, Chicago 1945, p. 636 ff.

matter; indeed the extant writings of Isaeus are all forensic, rather spe-
cialized speeches dealing with questions of inheritance, whereas the
writings of Isocrates comprise politics, ethics, theory of education, and,
of course, many other topics. Now Peppler's main lines of argument
may not be too far off the mark — the parodies he points to in Aristo-
phanes are still strong evidence, after all, and we are all aware of the
extensive use of such words by Plato and Aristotle. But the history of
productivity for this suffix clearly needs more detailed examination. An
«electronic lexicon» organized along principles of morphological cate-
gorization will obviously facilitate much more precise investigations of
this nature.

I have already remarked that an «electronic lexicon» of this sort will
be useful in a wide range of applications, and our example of the deve-
lopment of words in -ικος and its relationship to the intellectual history
of Athens should begin to give an idea of the sorts of research which
may become possible. But the construction of this «electronic lexicon»
will carry with it benefits aside from those associated with the lexicon
itself, and I would like to detour for a moment to consider these bene-
fits.

First of all, there would be immediate and practical benefits for the
TLG data bank as it stands. In the course of the lemmatization and
morphological categorization of the words, we should effectively elimi-
nate most of the errors which (despite the intensive efforts at data cor-
rection) inevitably remain in the *TLG* data bank. Another practical
benefit would be the identification of all proper names in the *TLG* data
bank, that is, the differentiation of proper names from words capita-
lized for other reasons (as at the start of sentences). This should make
it possible to produce automatically a complete Onomastikon to the
literary texts.

A more far-reaching benefit of the construction of this «electronic
lexicon» would be the availability of the parsing and lemmatization of
all the Greek word forms. This availability should have a great impact
on what can be done with software, both in the realm of research and
in the realm of pedagogy. On the research side, David Packard many
years ago demonstrated a search program which allowed scholars to re-
trieve examples where, for instance, a noun in the accusative precedes a
noun in the nominative, or where a dative follows a particular prepo-
sition. On the pedagogical side, the Perseus project has developed a
parser which in many cases can generate on-line reading aids. But ef-
forts such as these have been hampered, and in the case of research seri-
ously so, by the incompleteness of the data on which the software

depends. A complete parsing and lemmatization of the Greek words should allow the development of far more sophisticated and accurate software tools for the analysis of Greek texts.

But what about the lexicon itself? As a historical lexicon, this «framework» is indeed «bare bones», for no matter how usefully the material is organized, combined, interrelated, and selected, you would not have available the wealth of semantic and syntactical a n a l y s i s found in *TLL*. In fact, in its first incarnation, the «electronic lexicon» would have very little semantic information at all. To root out the nuances of meaning and the habits of usage, you would, like the *TLL* lexicographer, have to read through the passages. Given the lack of semantic information, we might better use the general term p h i l o - l o g i c a l t o o l than the specific term l e x i c o n. But I have chosen the word l e x i c o n because, as I said at the outset, this preliminary concept is intended as a suitable framework on which to hang future research. And what would be the nature of this future work? Some possibilities are, first of all, full-scale lexicographical articles for specific words or word groups; but also articles on the functions of the various prefixes and suffixes; discussions of problems of dialect and spelling and ety-mology; cross links to specialized lexicographical discussions already existing in electronic form; cross links to pictorial data banks of archaeological artefacts; and so forth. The great advantage of the stra-tegy I have outlined is that it will put together many of the lexicogra-phical and general philological tools which would be desirable for the construction of a full-scale historical lexicon, but in such a way that'the tools themselves would be available to everyone.

The construction of this «framework», this «electronic lexicon» which I have described would be thus but the first step in what I hope would be a long-term effort to generate a comprehensive Greek histori-cal lexicon. But if that is our hope, we must try to anticipate how that might realistically come about. In this day and age it will simply not be possible to assemble a large group of scholars to work over a period of many decades in the manner of *TLL*. I think we all agree on this point. But here I would like to make one final obvious point about techno-logy: namely, that i t f r e e s u s f r o m t h e n e c e s s i t y o f a l l w o r k i n g i n t h e s a m e p l a c e. If we are to make progress on the construction of a comprehensive Greek historical lexicon, logistics will demand that progress come at the hands of many people in many places all over the world. Technology makes that feasible. But before we can begin to think about organizing international teams to take on the lexicographical work —or to think about whether that is truly de-

sirable— we must first assemble our equivalent to the *TLL Zettelar-chiv*; and naturally it will be critical that this be available not in a centralized spot but wherever the scholar may be. Technology also makes that feasible.

The possibility I am putting forward can be viewed, then, as an attempt at further progress towards that necessary first stage, that is, the creation of our equivalent to the word slips in the *TLL Archiv*. We have already, as it were, fashioned the slips. But now we must begin to organize them in some way, we must begin to make «Νοῦς from this Chaos», in Diels's phrase. Since these «slips» are nothing but the discrete elements of the electronic texts residing in the *TLG* data bank, it seems only logical to try to exploit their electronic format. I believe that the «morphological categorization» which I propose does that. The proposal satisfies several conditions which are, in my view, essential. (1) The task, if quite difficult in certain respects, is for the most part straightforward, and one which could be completed by a team at *TLG* in a reasonable amount of time. To fashion a stable framework of some kind is a critical prerequisite to the extensive use of collaborators. (2) The organizational framework should in and of itself carry immediate and foreseeable benefits, and not depend entirely on future work for its justification. (3) The framework must be one which is open-ended, and which allows the easy incorporation of other types of electronic data.

This proposal would be then in any case but a first step, a framework or, to use a different metaphor, a n u c l e u s for future work of many kinds, to be pursued, we would hope, in many places. A proposal of this type would offer, I think, some exciting possibilities for a fuller description of the Greek language, which is, after all, the goal of the lexicographer.

WILLIAM A. JOHNSON