

NEW TECHNOLOGIES FOR READING: THE LEXICON AND THE DIGITAL LIBRARY*

Introduction

Books—codices, really—have begun to share the stage with electronic documents, a coexistence that is likely to prove as enduring as that of manuscript and print. While electronic documents are still quite new, we now have enough experience to speak with something other than anticipation—whether fretful or eager—about how print and electronic media complement one another. This discussion will focus on one particular reference work: the entire text of the Liddell-Scott-Jones *Greek-English Lexicon* (ninth edition, 1940; cited hereafter as LSJ 9) has been available on the World Wide Web since the fall of 1995. Between July 9, 1996, and August 20, 1997, the electronic lexicon was accessed 336,649 times, with usage growing steadily (as I write in August 1997, the lexicon is used more than 1,000 times each day).

LSJ 9 has, of course, always pushed the limits of the single-volume codex: Its sheer bulk—LSJ 9 contains around 37 million characters of information, comparable to almost forty volumes of standard size or to all the literary texts that survive up through the death of Alexander—makes the printed volume ungainly at best. LSJ 9 is one of the few books that is arguably more useful in electronic than in print form: the book is too big, the print is too small, and the content is not narrative—readers generally read very short passages at any one time.

The codex remains a handy medium with which to read a text of Homer or an article from *Classical World*—or, at least, most articles. This article, in its print form, is an exception. The printed text cannot include dynamic links to the searches or other functions that it will describe. In the electronic version of this document, active links can replace these screen shots: readers can not only call up and reproduce the functions described here (the “reproducible results” that our scientific colleagues talk about so much); they can also pose new queries and explore the limits of these techniques. Those skeptical of the new medium should have little trouble finding errors in the data entry or rough edges in the software. But most such errors, when reported, can be fixed and will vanish like bad dreams (our electronic Liddell-Scott-Jones *Lexicon*, for example, rebuilds itself each night and incorporates any changes made the previous day).

Fluidity, of course, poses problems of its own (to which I will return later). Electronic access to Greek texts, morphological analyses, and lexica have made Greek texts accessible to intelligent readers who would otherwise have turned to some other activity: to take

* A draft version of this article is available on the Perseus web site at <http://www.perseus.tufts.edu/~gcrane/lcj.cw.html>. The Electronic LSJ and the various searches cited here are also available at <http://www.perseus.tufts.edu/>.

only three of thousands of examples, a computer scientist at the University of Tokyo, a high school student in Kansas, and an Australian naval officer, who had no practical access to the conventional tools which we professional scholars take for granted, were able to use the electronic LSJ 9 to read Greek. Readers from homes, from private companies, military bases, government agencies, dorm rooms, and faculty studies have used the Internet to study Greek texts to which they would often not otherwise have had access. Those of us who read texts quickly turn happily to our cleanly printed, if dog-eared codices, but, even for professional scholars, the case is much less clear when we study one twenty-line passage by examining several dozen others. For those who read Greek and who are not professional classicists—in other words, for all of those who, by paying tuition, donating to projects, or even buying our books, justify the existence of those of us who are professional classicists—new technologies of reading can make the difference between continuing Greek and giving up. In many universities, keeping eight students instead of six (or sixteen students instead of thirteen) can mean the difference between holding and canceling a class and, ultimately, between maintaining classics as a separate department.

We must be careful to distinguish in our minds between tools that will help us do our research more effectively and those which not only help us but which can increase the number of those who read Greek (or Latin or Akkadian or even math or any subject). I can think of few, if any projects, that need be relegated to the first category. When we first contemplated putting the ninth edition of the Liddell-Scott-Jones *Greek-English Lexicon* on-line more than ten years ago, we certainly had professional scholars in mind. To our surprise and delight, we came to believe that a tool intended for classicists could help expose Greek texts to a wider audience. (The papyrological database which Roger Bagnall outlines in this issue of *Classical World* provides an excellent example of this process, for the electronic medium promises to allow more people to make even more creative use of this crucial category of text.) Many of the books that we now hold in our hands could, in an electronic form, be reconceived in such a way as to serve both the general reader and the specialist more effectively: our initial experiences with the on-line LSJ 9 provide only one of what we hope will be many standard reference works that begin to reach new, far more diverse audiences.

A tremendous amount of work needs to be done, and our efforts have barely scratched the surface, but new technologies of reading, judiciously employed, are a necessary, though clearly not sufficient, condition if we are to expand the study of ancient Greek in the modern world. Nor does Greek alone stand to profit. The same techniques can obviously be applied to Latin, while the less-studied languages of the ancient world—Sumerian, Akkadian, Hittite, and Egyptian, to name several—may finally be able to play a role in general scholarship of antiquity more commensurate with their immense importance.

Even the modern languages will change: Mandarin, with its billion or so speakers, may not face the same problems as Sumerian, but tens of thousands of students in U.S. universities give up Chinese long before they have been able to read a Chinese newspaper by themselves—a skill that often does not develop now until the fourth or fifth year of study.

To what extent our work contributes to this larger goal remains to be seen: even if our work should appear promising, the worst dead ends stand at the end of paths that originally seemed most exciting and subsequently attracted the most (wasted) effort. Likewise, it may be too early to gauge the success of the particular decisions that we made as we matched what we wanted against what was possible. As classicists we all recognize the extent to which the past imparts momentum to the present. Before considering our work with the lexicon and its relation to Greek texts, we therefore need to sketch briefly the background of our decisions.

Background: Perseus and Greek Lexica

The Perseus Project is probably best known for the many color images of Greek art and archaeology which it has assembled. The initial Perseus CD-ROM (*Perseus 1.0*, Yale 1992), already delivered 700 drawings and 7,000 color images on a single disk, making it arguably the best-illustrated “single volume” publication in the history of classics. *Perseus 2.0* (Yale 1996), contained not only the same 600 drawings but 1,000 entirely new maps and 24,000 color images, each of which has, on the average, twice as much detail as its counterpart in the first edition.

But if the imagery has attracted the most attention and indeed absorbed the majority of our resources, the representation and manipulation of on-line text predated our work with images and remains technically the most developed aspect of our project.¹ Thanks to Elli Mylonas, who worked on the project for six years, we were committed from the start to working with “Standard Generalized Markup Language” (SGML), the most powerful standard for encoding aspects of an electronic text. By collaborating with the NEH- and EC-

¹ This statement needs careful qualification. The atlas in *Perseus 1.0* was a relatively simple tool, which allows users to plot longitude and latitude coordinates on any map. The major contribution of the *Perseus 1.0* atlas was its ability to plot sites on any map whatever the scale, thus keeping separate maps and the data needed to plot objects. Nevertheless, the capabilities that we could publish in *Perseus 1.0* did not reflect the sophistication of the work that was done. The limitations of our delivery environment in 1992 (Apple Computer's *Hypercard 1.2*) prevented us from including in Perseus the color maps or satellite images which had been created for the project. *Perseus 2.0* contains ca. 1,000 maps drawn not only from satellite images but also from the Digital Chart of the World (the best general source for mapping information currently available). Fritz Hemans inaugurated this dimension of our work at the Center for Remote Sensing at Boston University. Nick Cahill and Neel Smith, now of the University of Wisconsin at Madison and the College of the Holy Cross respectively, carried on the work.

funded Text Encoding Initiative (TEI), our experiences in developing texts for Perseus contributed to the emerging TEI standards for electronic texts even as we benefited from the varying experiences of our collaborators in the TEI.

My own interest in Perseus grew out of work that I began in 1982 as a graduate student entrusted with developing software for the first public versions of the Thesaurus Linguae Graecae (TLG), the monumental database of ancient Greek texts developed at the University of California at Irvine. Those of us who began planning Perseus during the 1985–1986 academic year saw our work as complementary to the TLG. Because the TLG was engaged in creating an exhaustive database of Greek texts, we felt that we could concentrate on collecting a broader range of materials for a smaller chronological time period. We set out to build what would now be called a digital library: a heterogeneous collection of electronic materials about the ancient Greek world. Our emphasis was on breadth and flexibility of coverage. We wanted to represent within this new medium as many different types of information as we could. At the same time, we wanted to make sure that these materials were designed to be as flexible as possible: we were less interested in creating a unified “curriculum” than in providing solid tools on which others could build and which could easily grow more powerful over time.²

Our experiences developing software for use with the TLG made one need painfully obvious. Everyone who has done serious work with any existing system for accessing the TLG knows that we locate Greek words by scanning for strings of letters which they contain. The study of information retrieval offers two complementary measures which can illustrate the drawbacks of string searches: “recall” and “precision.” Suppose, for example, we look for all words containing *πεμπ-* when we wish to find instances of the verb *πέμπω* (“to send”). Everyone who performs such searches quickly realizes that while *πέμπω* may match *ἔπεμπε* (“he was sending”), *πέμποιτε* (a present optative: “may you [pl.] send”), and *πέμπομεν* (“we send”), the same string will miss *πέμψετε* (“you will send”), *ἔπέμθης* (“you [sg.] were sent”), and *πέπομφα* (“I have sent”). String searches regularly miss relevant words, and their “recall” is thus less than 100 percent. On the other hand, the key *πεμπ* also retrieves forms such as *ἐπεμπίπτω* (“to fall upon besides”), *ἐπεμπηδάω* (“to trample upon”), and *πεμπτώβελον* (“a five-pronged fork”), which do not belong to the verb *πέμπω*. String searches thus not only fail to recall every relevant instance; they also tend to collect extraneous information and

² This decision attracted substantial criticism for a number of years because the Perseus database was correspondingly less approachable than many other tools: there was no single master narrative that gave our selection of materials an obvious unity. Recently, however, the emphasis on tools and raw materials has begun to bear fruit. Because we designed the database as cleanly as possible, we have, with part-time effort and in less than a year, been able to “port” most of Perseus to the World Wide Web. As a website, Perseus provides a growing set of links on top of which others can build more thematically structured content.

hence have a “precision” which is less than 100 percent. Recall and precision are conventional measures for the success of a retrieval system, and, even more than a decade ago, we thought that software could perform much better according to both of these measures.

The problems of string searches flow from the nature of Greek: Greek morphology is complicated, and the connection between a form on the page and its dictionary entry is not always obvious. Students of classical Greek have recognized this problem for millennia, but they have spent an equally long time learning well-formulated rules for constructing Greek words. If a machine could recognize οἶσете, for example, as a form of φέρω, then it could scan texts and match inflected forms with their dictionary entries, thus rendering the fundamental task of searching for Greek words much more effective. Put another way, while there are times when we want to search for strings (e.g., if we are interested in all the compound words containing a given stem), we often want to search for “words” instead, and we thus wanted a system that could recognize and take apart Greek words.

Starting in early 1984, much of my own technical work during the past decade has therefore gone into the development of a system that could cope with the vagaries of Greek morphology. Users of *Perseus 1.0* were, as a result, able to search for dictionary entries rather than strings: they could ask for φέρω (“to bear, carry”) and retrieve not only φέρεεις (“you [sg.] carry”) and ἔφερε (“s/he was carrying”) but οἶσω (“I will carry”) and ἤνεγκον (“I carried”) as well.³

Greek morphology is complex because it encodes a great deal of information in a given form, and inflected Greek words are generally much more informative than in the modern languages of Western Europe.⁴ But even Greek contains a certain level of ambiguity: αἰσχύνη can be either the dative singular of the noun αἰσχύνη (“shame”) or a present indicative form of the verb αἰσχύνω (“to shame”). But in theory, the system that we developed can provide scholars with 100 percent recall (i.e., searches that find every single possible instance of a word), 100 percent precision (i.e., searches that distinguish between forms that definitely belong to a given word and those which are inherently ambiguous), or a reasonable combination of both. LSJ 9 introduces problems of its own: as Ilja Pfeijffer pointed out to us, searching for μνν with a string search leaves out the alternate dialectal form ννν, but even our morphological searches commit the

³ A brief survey of the morphological analyzer can be found in G. Crane, “Generating and Parsing Classical Greek,” *Literary and Linguistic Computing* 6 (1991) 243–45.

⁴ English morphology is not computationally challenging: Gerard Salton, in *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*, ed. Michael A. Harrison (Reading, Mass., 1989), surveys the main issues in three pages (379–82); the C source code of the Unix spelling checker (which handled morphological analysis of English) was only about ten pages when we first looked at it more than ten years ago.

same error insofar as they follow LSJ 9 in treating these two dialectal entries as separate words. With more than a hundred thousand unique strings in *Perseus 1.0* (and more than a million unique strings in the TLG), we still find instances where the system misses a form, but each such missed form is a bug that, once identified, is easily fixed. Even in *Perseus 1.0*, we found that the mistakes were comparable in number to typographical errors in the first edition of most printed lexica, and the searching tools were a powerful advance over what was otherwise available. Those using the World Wide Web Perseus can click a button to e-mail problems to us, and they can be fixed soon after they have been reported. In actual practice, it would have been difficult for most scholars to locate more forms of any reasonably complex verb with string searches alone, and the determined scholar could always combine both strategies.

We thus needed a system that can recognize ὄσουσι, “they will carry,” as a form of the dictionary word φέρω, “I carry,” and that also could, of course, analyze an arbitrary number of compounded prefixes (e.g., συν-επ-εν-ε-πήδησαν, “they trampled upon together”). Ten years before we began work on the problem, David Packard had developed an elegant and (typically) efficient morphological analyzer (MORPH), and his system solved these basic problems.⁷ We therefore set out to design a system that would deal with a wider range of morphological phenomena. The earlier system had, for example, ignored diacritics and did not explicitly deal with Greek dialects. More recently, linguists have developed software to analyze the morphology of various languages (PC Kimmo, originally developed in Finland, is one of the most famous⁶), but it was not clear that these could handle the details of Greek with the precision that we demanded. They seemed designed to work with a given language at a particular moment in time rather than a language as it manifested itself in various dialects over the course of five hundred or a thousand years.

Our goal was to extract as much information from the inflected form as we could—to paraphrase the old Chicago stockyards boast, we wanted to use every part of the word except the squeal. Thus our system could use the differing accents to distinguish between παίδευσαι (a middle imperative), παιδεύσαι (an aorist optative), and παιδεῦσαι (an aorist infinitive).⁷ We also designed our system to recognize variations

⁵ David Packard, “Computer Assisted Morphological Instruction of Ancient Greek,” *Computational and Mathematical Linguistics: Proceedings of the International Conference on Computational Linguistics 2* (1973) 343–56.

⁶ Laura Karttunen, “Kimmo: A General Morphological Processor,” *Texas Linguistic Form 22* (1983) 165–85; this work provided the basis for Lingsoft, a company that Karttunen co-founded in 1986.

⁷ The wording of this sentence reflects a major difference between print and electronic publication. In the printed version of this paper, the Greek words appear in Greek with full accentuation, and the difference between these forms is apparent. In the World Wide Web version of the paper, readers can choose to see the Greek

between dialects. We spent time at an early stage of development, for example, making our analyzer recognize that while ἀρετή (“virtue, *aretê*”) could be Attic or Ionic, χώρα (“land”) was an Ionic form, because in Attic Greek a phonological rule would have replaced the -η ending with -α after a rho. One of the first Greek texts on which we applied the first version of the current morphological analyzer was the heavily Lakonian Damonon inscription from the fifth century ([Buck] no. 71). While most of our work has subsequently concentrated on the conventional literary texts that constitute the bulk of Greek literature, we were concerned that the system which we were developing would one day be able to manage a wider, and less predictable, range of Greek.

These functions require a great deal of software—we have roughly 16,000 lines of Programming Language source code in the various software modules of our morphological analyzer—and it took years to develop and refine a set of interlocking utilities that could manage the underlying information and perform the actual analysis. Nevertheless, a morphological analyzer requires a tremendous amount of data: confronted with a form such as οὔσεις, it needs to know that there exists a regular future stem οἶς- which can appear in the active and which “belongs” to the verb φέρω. We found that, while encoding a compact database (such as the forty-page verb list in the back of Smyth) provided our analyzer with enough information for developmental purposes, Greek morphology is too complex for such partial expedients to produce the results that we desired.

The need for data forcibly directed my attention towards entering print lexica on-line. Our initial interest was less with the definitions than with the morphological information which the dictionary articles contained. We wanted to extract from the on-line version of the print lexica as much morphological information as we could (“information extraction” is so commonly needed that it has now evolved into a growing subfield of computational linguistics). In May 1985 (before any serious planning had even begun on Perseus), I delivered lectures on the topic of entering LSJ 9 on-line at the University of Chicago and the University of California at Los Angeles. It was Professor Bernard Frischer of UCLA who suggested that work should begin not with the massive LSJ 9 but with the much smaller and more tractable *Intermediate Liddell-Scott Greek English Lexicon* (which is roughly one-fifth as large). Given our needs, this made excellent sense, because the more common a word in ancient Greek, the more likely it would be represented in the smaller Lexicon. And, of course, since common words tend by nature to be the most irregular, the *Intermediate Lexicon*, although a fraction the size of LSJ 9, contained many of the most challenging aspects of standard Greek morphology.

in a transliterated form that does not reproduce the accents. I thus rewrote the sentence to state explicitly that there are accents (whether visible to the reader or not) which distinguished between these forms.

Professor Frischer began work entering the *Intermediate Liddell-Scott* at UCLA in the fall of 1986. He also provided key support to Neel Smith and Joshua Kosman, then graduate students at the University of California at Berkeley, so that they could develop tools for analyzing Greek morphology. As the Perseus Project began to take shape in mid-1986, we collaborated with the Packard Humanities Institute to complete the job of data entry for the lexicon. At the same time, we took over the job of developing the morphological analyzer (although even now, ten years later, we continue to use a number of basic routines developed by Smith and Kosman). Over the next several years, we extracted almost fifty thousand verbal and nominal stems from the *Intermediate Lexicon*, while expanding the capabilities of our software to manage and exploit this information. While refinements to such a system (which is essentially an electronic, rule-based grammar of Greek morphology) can go on indefinitely, our work on the linguistic phenomena covered by the *Intermediate Lexicon* had reached maturity by the time that *Perseus 1.0* was released in spring 1992.

Thus, seven years, ten thousand lines of code, fifteen thousand endings, fifty thousand stems, and millions of analyzed words after initial planning in 1985, we turned our attention to the ninth edition of the Liddell-Scott-Jones *Greek-English Lexicon*. In the spring of 1993 we corresponded with Anna Morpurgo-Davies of Oxford University and received permission from those in charge of LSJ to begin plans for an electronic version of LSJ 9. A proposal submitted to the National Endowment of the Humanities received generous support, and we were able to begin laying for the foundations for a new electronic database of Greek language and lexicography in the summer of 1994. Our database of nominal stems has expanded to 96,000, and we can now analyze well over 90 percent of the words in the TLG. But where our focus had been primarily upon the morphological information in the *Intermediate Lexicon*, our experiences had taught us much about what could be done even with a print lexicon that had been transported into an unfamiliar electronic world. The miraculously rapid development of the World Wide Web allowed us to make available in September 1995 our working version of LSJ 9.

But what problems does an electronic lexicon solve? What problems does it cause? These tools are still very much in their infancy and no one can predict how they will develop over the coming tools, but even now, at what is still a very preliminary stage, we are in a position to make some observations about how the tools now at our disposal change the way in which we work with Greek texts.

LSJ and the Limits of the Single-Volume Book

Proclamations and anxieties about the end of the book—or, to be more precise, the printed paper codex—may give way to equally equivocal results. And certainly the codex will be with us for some time to come. At one extreme, we can identify codices which can expect a long and prosperous life for the indefinite future: the codex

is, for example, ideally suited to the "airport" novel, which is inexpensive, fitting snugly inside a pocket, is ready for use in almost any circumstances, and, above all, is designed to be read in linear fashion as events (if not the characters) develop and suspense rises. Scholarly texts and "essay" style academic publications are equally well suited to codex form when we choose to read more than, say, a page or so at a time. (When we are reading intensively—looking up every footnote in a crucial section of an article, for example—many readers may find the codex less useful than an electronic hypertext.)

At the other extreme, however, stand reference tools such as the great *Liddell-Scott-Jones Greek-English Lexicon* (LSJ 9). This massive single volume squeezes almost 40 million key strokes into 2,000 dense columns of print, but its very size calls attention to the limits of the codex. None of the problems with the form of this publication should be taken as criticism of those without whose heroic labors LSJ would not exist. Most of these problems have been obvious since the first editions of LSJ in the nineteenth century, and I suspect that they tormented Liddell and Scott themselves at least as much as anyone who ever used their lexicon. A further problem reflects a more subtle change in the status and authority of monumental lexica in the age of electronic searching. Even if the mechanical limitations of the codex can be overcome, simply entering LSJ and making it available in electronic form is only the first step in the development of an electronic lexicon.

Electronic documents, although they share some features with their print counterparts, have their own set of problems and possibilities. Electronic tools for the study of Greek language will surely evolve and the course of that evolution may be unpredictable in the long run, but we can already begin to think through some of the immediate consequences which the move from ink to bits may have.

First, the codex LSJ 9 is obviously too big. No one, I imagine, has ever been surprised to find that he had without knowing been carrying his copy of LSJ 9 in his coat pockets for the past several days—indeed, it would be hard to imagine any item of apparel, much less a pocket, other than a backpack into which LSJ 9 would fit.

Second, the codex LSJ 9, because it is so big and because it presses against the limits of a single book, needs to cram as much print on each page as possible. The dictionary entries are, however, represented as hierarchical outlines. Whether or not such a data structure is the best format for representing the meanings of words, these articles were designed to be scanned quickly so that the reader could locate as quickly as possible the relevant section. LSJ 9 is, however, so full of data, and the typesetting in some ways so rudimentary, that there is very little room for blank lines, indentation, bolding or other visual clues whereby readers could orient themselves. The outline of a complex article is hard to detach from the swarm of tiny letters and the print thus inhibits precisely that style of reading for which the lexicon entries were designed. The codex LSJ 9 has a

potentially long future in the academic world as a case study to illustrate what designers of user interfaces should avoid.

These first two drawbacks together combined to inspire an alternative course of action that was as drastic as it was natural. Simplified versions of the large lexicon were produced for the use of students. First, James Whiton produced *The Abridged Liddell & Scott's Greek Lexicon*, which appeared in numerous editions throughout the nineteenth century. (According to the Harvard library system, the 1876 version published by Ginn Brothers in Boston constituted the seventeenth edition.) In 1887, Henry George Liddell himself completed work on his own version of a student lexicon, *An Intermediate Liddell Scott Greek Founded upon the Seventh Edition of Liddell and Scott's Greek-English Lexicon* (Oxford 1889). Together the abridged, intermediate, and full lexica have become famous as the "little Liddell," middle Liddell," and "Great Scott" respectively. The *Intermediate Lexicon*, which is substantially larger than "little Liddell," is a largely mechanical reduction of LSJ 7. Middle Liddell was only designed to cover those authors "mainly taught in schools," i.e., the authors which "public school" children of late Victorian England read. Liddell was thus able to leave out many words from less prominent authors. Second, the remaining entries were themselves shortened: quotations are truncated or deleted altogether; citations which appear in full form in LSJ lose their "chapter and verse" (e.g., "II. 15.586" becomes "II.," "Hdt. 1. 157" becomes "Hdt.," etc.).

Generations of students have started by working with the "little Liddell" and/or "middle Liddell." The shift from these teaching lexica to the full lexicon is often felt as a watershed event in the development of a classicist.

If, as Henry Liddell himself used to stress, the production of a lexicon is among the most thankless, dreary, and necessary tasks of scholarship, the subsequent abridgment, by largely mechanical means, of that lexicon was surely even more thankless, dreary, and necessary. But necessary as such an abridgment may have been, the separation of "pedagogical" from "scholarly" materials is a concession to necessity rather than a positive action. Ideally, the large lexicon would be so designed that it would be of use to student and scholar alike.

Third, LSJ 9 grew so large that it was no longer feasible to update it directly. Between 1843 and 1897, Henry George Liddell supervised eight separate editions of the Lexicon. Almost a half century passed before the next full revision, LSJ 9, was completed in 1941. Since then, no one has even attempted to work with LSJ directly. In 1968, Oxford University Press produced a supplement to LSJ, which was subsequently printed as an appendix to LSJ 9. A second supplement, which supersedes the first, is in the final stages of publication even as I write. These supplements contain a good deal of useful information—corrections, additions, new words, and, in the case of the new supplement, a number of completely revised articles—but the resulting combination of base lexicon and appendix illustrates the weakest aspect of codex technology. LSJ 9 was al-

ready too big and too hard to read effectively; the need to flip back and forth between two different versions may make much badly needed information available, but it does so at the cost of making the lexicon even more cumbersome. In effect, the lexicon—or more accurately, the process of updating the lexicon—has collapsed under its own weight.

At this point, the coexistence of physically distinct teaching and scholarly lexica had an odd consequence. On the one hand, no one had ever updated Henry George Liddell's "middle Liddell." The middle Liddell which students still buy today is identical to the crisp new volume which the fifteen-year-old Winston Churchill (who hated Greek) might have encountered in 1889 at Harrow. No one ever felt moved to bring either little Liddell or middle Liddell in line with either the eighth or the ninth edition. Both teaching lexica not only reflect our understanding of established Greek texts during the reign of Queen Victoria but were also composed before the sands of Egypt returned to life the work of Menander, gave us the crucial Aristotelian *Constitution of the Athenians*, and contributed significantly to the remains of many central authors. If the separation of scholarly and pedagogical materials was a sad necessity in the 1880s, the ossification of the student lexicon has become a source of embarrassment.

On the other hand, a research lexicon that appeals primarily to professional scholars has a much smaller audience than any teaching tool. The smaller number of potential purchasers reduces the amount of money which a new edition of the lexicon can raise and thus makes it more difficult to finance the considerable work needed to create a full update. Separate lexica for students and researchers thus fragmented an already small market and made it, in the long run, more difficult to serve the needs of both groups.

But if LSJ is too large, this problem ironically flows from the fact that it is also too short. LSJ is bursting at the seams because it contains only the merest fraction of the information that we might potentially wish to have about Greek words—most Hellenists have, at least occasionally, cast their eyes wistfully at the dense coverage of the *Thesaurus Linguae Latinae*. Readers should be able to begin with a broad overview of a word's meaning and then push down into more detailed information, ultimately reaching the detail of a specialized lexicon. So much Greek text survives, however, that no one has seriously considered a monumental lexicon like the *Thesaurus Linguae Latinae* as feasible—this is one reason why classicists twenty years ago moved to create a databank of texts such as the TLG rather than a new lexicon.

The mechanics of publication have even contributed to the near extinction of smaller, specialized lexica which individual scholars can, with great labor, produce. These lexica can be among the most useful tools that we have. Changes in scholarly fashion are partially responsible for the virtual disappearance of such works. Contemporary English departments, with their predilection for articles and books, have set the standard for publication, and the traditional philological

genres—critical editions and commentaries as well as lexica—have suffered: lexica are very difficult to produce and are unlikely to win young scholars first jobs, tenure, or promotion. At the same time, the mechanics of traditional scholarly publication have drained the life blood of these works: every undergraduate taking a course on Pindar should have a copy of William J. Slater's *Lexicon to Pindar*. This lexicon itself is an admirable codex, roughly 500 pages and quite easy to handle. Nevertheless this standard tool cost (when I last checked a year or so ago) \$175. A publication which, if sold for \$30, could reach many undergraduates and virtually all advanced students of Greek poetry thus finds itself relegated to research libraries and a tiny set of individual scholars.

The fourth problem has less to do with the size than with the nature of LSJ. Dictionary entries are not narrative prose. They were not written with the expectation that readers would normally read them from beginning to end. In most cases, a reader studying a particular author wants information about the usage of a particular word in the text. Is there some particular nuance to this term in this period and genre? Is the grammatical usage typical or unusual? Are there other texts in which this term appears which would provide a richer context? In other cases, a reader may wish to survey an overview of the usage for a particularly important term (e.g., *aretê* or *hubris*). In each case, even if we assume that the entries in LSJ are up-to-date and authoritative, the brief quotations which LSJ (or any lexicon) provides to illustrate a word are never sufficient. We always need to consult the source texts themselves so that we can evaluate these citations in their full context. Thus, even the intrepid classicist who has hauled his codex LSJ 9 onto an airplane has no way to read these entries as they were intended to be read, because the source texts are separate from the lexicon. This may not much concern the scholar ensconced in a personal library accumulated over the years, but it seriously limits the value of this lexicon to those without access to such a resource. If we want to expand the study of Greek, this is a serious drawback.

Electronic Documents and the Limitations of the Codex

The electronic format addresses each of these problems. First and most obviously, the storage capacity of hard disks is, as this article goes to press in 1998, already so vast that the cost of keeping LSJ 9 on-line is not great.⁸ Even with a full set of bloated indexes

⁸ It is perhaps worth recording for the historical record that in 1982 the first hard disk which the Harvard classics department, for example, purchased to store the TLG was the largest generally available: a 660 megabyte drive from Control Data Corporation. This device cost \$35,000, was approximately the size of a washing machine, arrived in a wooden crate (rather than a cardboard box) and required specialized systems programming (modifications to the Unix kernel) before it could be attached to the main computer. Its maintenance contract cost more than \$3,000 per year. As I write this article, 12 gigabyte drives can be purchased for around \$350 (a per megabyte cost 2,000 times less than in 1982), their disks are 3.5 inches

and a panoply of subsidiary data, LSJ 9 would not take up more than 80 megabytes of data. The eager classicist could easily store this lexicon and a wide range of Greek texts on a notebook computer. The source data can be distributed on a CD-ROM that is far lighter and less expensive to produce than the print counterpart.

Second and somewhat more surprising, the electronic medium has allowed us to make the lexicon more readable. Ink smeared on paper still has far better resolution than any standard computer display, and most people find it easier to read most documents on paper than on a screen. One classicist participating in an on-line discussion group who works extensively and comfortably with on-line materials recently contrasted the "repellent experience of reading from a screen" with that of reading a book. But LSJ 9 is a special case. The type is so small, the print styles so uninformative, and the page layout so unhelpful that most of those with whom we have spoken report that the electronic versions of the articles are easier to read than their print counterparts.

There are reasons for this: first, the print can be made larger on the screen than on the page. Second, we do not have to worry about cramming as much print into each square inch as possible. We are thus able to separate different definitions with blank lines, and this added white space allows readers to scan through a large article much more quickly. Before we began work, we knew that LSJ entries, with their hierarchical structure, would lend themselves well to "outline" viewers, and we had anticipated spending a good deal of effort trying to make articles from LSJ fit neatly into existing tools for viewing outlines. At the very least we planned (and indeed still plan) to implement an automatically generated "middle Liddell" view for each article: a computer program can easily filter out words from authors "not commonly read in schools," temporarily withhold quotations, and truncate citations. Since a filter cannot exercise the same careful judgment which Liddell lavished upon his abridgment, the consequences are easily remedied because the full version of the dictionary article can easily be retrieved.⁹ We did not, however, anticipate how much more useful these articles would become with technically trivial changes in formatting and without such automatically generated "outline views."

in diameter; they arrive by overnight mail in light cardboard boxes, and they plug directly into standard computers. Instead of maintenance contracts, they come with warranties. The quantitative shift is so great—disk storage is almost 200 times cheaper now than it was fourteen years ago—that the role of textual databases can undergo a qualitative shift.

⁹ Thus, Rodman Reynolds, a graduate student at the University of Washington, writes: "The best part about having the LSJ-9 on-line, from my perspective, is not that it spares me from having to come up with the *Sitzfleisch* to flip through the pages myself, but that, owing to the fact that space requirements in electronic media are defined in a totally different way from paper media, it is possible to display the entries and citations of that monster lexicon in an uncramped and readable way. After wearing my eyeballs to the bone by reading LSJ's fine print, a clear and roomy display of those entries on the Perseus site was a thrillingly welcome sight."

Third, the most basic searching abilities can be enormously helpful. Readers confronted with a vast and complex article can quickly search for those definitions, if any, which contain references to the author which they are currently reading: someone reading Thucydides can, for example, search a long dictionary entry for "Th." and quickly locate each citation to that author.

The superficial facelift to LSJ 9 has had an immediate and, to us at least, gratifying consequence with serious implications for future work. We wanted to see how the new formatting changed the way in which our readers viewed the "scholarly lexicon" and the "student lexicon." We conducted a simple experiment over the course of several months. We presented the lexica in two different ways and then noted the degree to which our users chose one lexicon or another. While we have no way of determining precisely who our users are and while some faculty members have written to let us know that they are using the philological tools on our Web site, all the evidence at our disposal suggests that students working their way through class assignments currently account for the lion's share of usage. Those reading our on-line texts were thus precisely the audience for whom the *Intermediate Lexicon* had been designed and for whom LSJ 9, in its print form, was too hard to use.

All Greek texts are linked to a morphological database: readers can thus call up a list of morphological analyses for most inflected forms in our texts (e.g., οὔσατε is the second-person plural future indicative active of φέρω, "you (pl.) will carry") and then call up a dictionary entry for that word. We also made both LSJ 9 and middle Liddell available on our World Wide Web server.

In the first phase of our study, we forced all users to call up the entry in LSJ 9 first. We were afraid that students would be intimidated by LSJ 9 and that they would immediately choose the smaller lexicon. We wanted to see how often readers, once they saw the LSJ 9 entry in its more readable electronic format, would still feel the need for a more compact version. Once readers had access to the full lexicon, they could click on a link to the more compact dictionary entry in the student lexicon: thus, the reader daunted by a long and complex entry could easily shift to the abridged and more manageable version. We chose to lead readers through LSJ first because we feared that students would choose the smaller lexicon from habit, and we wanted them at least to glance at the fuller entries before going to middle Liddell.

After several months, we shifted to a second phase. We changed the interface, placing links to both LSJ 9 and middle Liddell directly on the Web page with the list of morphological analyses. We were curious to see how many readers would now go directly to the intermediate lexicon and avoid the larger version.

We were thus pleased to find that, in the first phase of our study, our readers only called up the middle Liddell entries around 30 percent of the time. In other words, most of the time, the on-line

LSJ dictionary entry was clear enough that the student version was not needed.

When we changed the format and allowed readers the option of going either to LSJ 9 or to avoid LSJ 9 and go directly to middle Liddell, we expected that this ratio would drastically shift. We have been surprised to find that LSJ 9 remains far more popular than middle Liddell and that, in fact, LSJ 9 is still used far more often: in the time since we changed the interface, LSJ 9 is still used almost twice as often as the middle Liddell.

Preliminary conversations with students using the on-line lexica here at Tufts and e-mail from students elsewhere give substance to our statistical impression that the electronic format has fundamentally changed the value and function of the two lexica. One student of Greek, for example, said that, in working with books, she had always used the *Intermediate Lexicon* in preference to LSJ 9. The small, dense print and the large size of the articles made LSJ 9 almost unusable for her purposes. In the electronic environment, however, the blank spaces and the fact that text citations are highlighted (in blue, on her machine, although different users can configure this highlighting differently) have made LSJ much easier to read. Furthermore, now that she finds it easier to read through the LSJ 9 entries, she reported that she was much better able to appreciate the greater detail of this lexicon. In a course on Aristophanes, she found that, very often, the passage that she was reading was specifically cited in the LSJ 9 entry. The *Intermediate Lexicon*, by contrast, does not give precise citations, reporting only that a particular meaning shows up in Aristophanes rather than providing the full citation (e.g., Aristophanes *Clouds* 327) as does LSJ 9. This student reports that she now refers to the intermediate lexicon only at times when technical glitches make an article in LSJ 9 unavailable. As our formatting improves and these technical glitches become less common, she has found that she has less and less use for the student lexicon in print or electronic form.

The effect of the electronic medium upon citations deserves emphasis. A lot of effort went into shortening or excising quotations of actual source texts, while text citations were changed from, for example, "II.15.358" to simply "II." Where this had made the print version of the middle Liddell smaller, more manageable and easier to read, in the electronic lexica, the loss of this information is far more problematic. For those reading the traditionally popular Greek texts, LSJ provides extremely dense coverage. LSJ 9 contains, for example, more than 200,000 citations of specific passages included in *Perseus 2.0*. Since there are around 3.4 million words of Greek in *Perseus 2.0*, LSJ 9 specifically refers to one out of every seventeen Greek words in each *Perseus*. In an average text, the reader can expect that LSJ 9 will specifically gloss one word on roughly every two lines. In the texts most widely read (Homer, the tragedians, Herodotus, Plato, etc.), the citations are particularly dense. Our student, who

has been reading standard texts such as Aristophanes and the Homeric Hymns, came to expect that LSJ 9 would, as often as not, refer to the passage before her when the language was particularly strange or noteworthy. In LSJ 9, the full citations made it much easier for her to determine whether LSJ was glossing a given passage than the much more general note in middle Liddell that a usage appears somewhere in Aristophanes or the Homeric Hymns.

While it may be too early to draw any definitive conclusions, all the preliminary evidence at our disposal suggests that the transfer from print to electronic format has reversed the standing of the student and research lexica. Using the simplest tools and merely exploiting the most basic features of an electronic text, LSJ 9 has gone a long way towards rendering its "teaching" counterparts obsolete. We have now the chance to create a single reference tool that serves scholar and student alike. For the scholar, there are, at the very least, economic benefits: the more people use the lexicon, the lower the potential cost for each individual user, while the aggregate income from sales can be larger and thus subvent more work on the lexicon itself. For the student, working as soon as possible with state-of-the-art tools is pedagogically more sound, and the greater depth of the research lexicon gives scope to curiosity and exploration which the smaller lexicon could not support.

The third problem is obviously much more tractable in an electronic environment. Once a lexicon is on-line, it can be edited and updated continuously. At present, our work on LSJ 9 consists primarily in formatting and in correcting those errors which appeared during data entry. As soon as e-mail reaches us about problems in an entry, we are able to correct the problem. Each night at midnight, the computer checks to see whether any of the source files for LSJ 9 have been modified during the previous twenty-four hours. If so, the entire database is reformatted and a fresh copy, with fixes in place, is available on our Web server an hour or so later. The mechanics of electronic production make updating the lexicon a trivial task. The changes could comprise a single accent error in entry or hundreds of pages of new material. Once the data had been vetted by some appropriate editorial board and was ready, the new material could be included in the lexicon within an hour.

The electronic environment makes it possible not only to update LSJ 9 but also to manage different versions of the same work—something fundamental to all scholarly publication. The flexibility of electronic texts brings with it dangers as well as advantages. James O'Donnell, who is writing a response to this collection of essays, will probably raise several issues that will make me rethink what I have written. Suppose that I then return and modify my own text to deal with his suggestions. A subsequent reader of O'Donnell's piece who wished to see what I had actually said would find before her a text different from that with which O'Donnell worked. Similarly, a scholar encountering a reference to an LSJ entry in a publication

that dated from the reign of LSJ 8 would have to consult that version of the lexicon to make sure that he was reading the same version of the dictionary entry. In the case of LSJ, this is a cumbersome task but it can be done (if one's library has not thrown out the earlier LSJ as being "obsolete"). If I simply edit my on-line text, there may be no trace at all of the old version.

But rapidly changing texts are an old problem with which software designers have wrestled for more than a generation: most large programs are developed by teams of programmers, and the same code will, over time, pass through many hands. "Source code control programs" allow those editing the same text file—the "text" can be a program written in C++ or an edition of the *Agamemnon*—to keep track of the differing versions of the text as they change over time. A programmer can easily go back to the version of the program that was in use six days, six weeks, or six years ago.

Similarly, suppose a classicist cites a dictionary entry on $\lambda\eta\theta\eta$, for example. If the date on which that citation was made is recorded, the subsequent reader could always go back to the dictionary entry as it appeared on that given day. It is unlikely that the change in our lexicographic entries would ever be so rapid as to tax the abilities of standard systems available today to keep up with them.

Most computer scientists would probably recognize the 1968 Supplement to LSJ 9 as a standard "diff file" (i.e., a list of modifications to be made to an original source text, in this case LSJ 9). Their only question would be what software was used to integrate these modifications into the original source text. (A program could probably implement most of the modifications automatically, but it is unlikely that the formatting in the LSJ supplement would be sufficiently regular to make this task fully mechanical).

We have, however, no intention of updating LSJ 9 itself. First, the format of these entries is clearly not ideal for an electronic world: if it were, we could, for example, easily have written a program that would have extracted all the morphological information encoded in LSJ and converted it into a format that our morphological analyzer could exploit. Instead, it took over a year of work to extract this information from middle Liddell, and we can automatically extract at most 80–90 percent of the relevant information from LSJ 9. Extracting syntactic and semantic information promises to be even more intractable. These dictionary entries were designed to be read by human beings. Any modern lexicon must be designed for use both by people and by the software that people will use to extend the value of that lexicon.

Second, if electronic lexica must provide data to programs as well as to people, it is not at all clear that the format of LSJ 9 really serves the needs of those reading electronic texts. Long as the entries in LSJ 9 are, for example, they are still shorter than they might otherwise be—the print volume, bursting already at the seams, could not absorb much more paper, and the entries were already cluttered enough in a lexicon designed for daily use by scholars. In

an electronic format, it is relatively straightforward to hide or conceal information: the main article might display four or five examples, while the curious reader could then call up as many as the lexicographer had energy to include. There is, in theory, no reason that an electronic LSJ could not serve as the "front end" to a more comprehensive resource like the *Thesaurus Linguae Latinae* or to a collection of specialized lexica. William J. Slater's *Lexicon to Pindar* could coexist with LSJ: those studying Pindar (or Bacchylides or any lyric poet) would thus have the option to consult Slater also, and there is no physical reason why many such tools could not be assembled. Simply placing the most important print lexica on-line so that readers could choose to read about a word in a variety of sources would be a major step forward.

Third, an electronic lexicon is valuable not only for its contents but also because it can serve as a clearinghouse for additional information. Each dictionary headword may have its article in the lexicon, but the headword can also provide a hook on which to hang other information: as more and more journal articles are entered on-line, the lexicon can accumulate cross-references to those on-line articles. Someone consulting the LSJ 9 article on σωφροσύνη, a politically loaded word commonly translated "self-restraint," might see that a recent article in the *American Journal of Philology* contained an extensive discussion of its use in Plato. Now that a comprehensive lexicon is on-line, smaller projects that would have been buried as journal articles or slim monographs (works on, for example, rhetorical or specific medical terminology) can now reach a much wider audience by being linked to the larger lexicon. Contributions could range from analyses of particular usages of a word in one text to full-scale analyses of complex terms in a wide range of Greek texts (e.g., an electronic successor to Helen North's 1966 monograph on σωφροσύνη). The rich body of ideas about crucial terms in Greek could thus be collected and made available in a single place.

Fourth, the day of the monumental lexicon may itself be over. Few who have used any massive, multi-volume lexicon—and this is as true for the magnificent *Chicago Assyrian Dictionary* as the *Thesaurus Linguae Latinae*—have not sighed inwardly when picking up early volumes of the series that badly need revision that they will may never receive. The first volume must therefore be perfect and complete, but the process can take so long that none of its authors live to see the completion of the series—a devastating consequence for small fields trying to establish themselves. Until its lexicon is completed, Sumerian—one of the richest and least exploited languages of the ancient world—will remain a fringe subject, accessible to far fewer people than its immensely interesting content deserves.

In an electronic world, however, there is no reason to build a monumental volume "A" before proceeding on to "B." Instead, the lexicographers might begin by covering a core group of representative texts—perhaps a scientifically determined sampling of the language,

perhaps a comprehensive corpus illustrating a crucial period, perhaps a combination of the two. Instead of producing a volume with comprehensive documentation of the letter "A" (and primarily useful only to specialists), the lexicon covering the initial corpus would appear. Instead of turning next to "B," more texts would be added, and a new edition released. Indeed, it is not even clear what would constitute a new revision: if the lexicon is maintained on-line, its contents could be updated continually, with problems in old entries fixed, new citations added, and, from time to time, whole dictionary entries revised.

There are drawbacks to this approach. It is surely more efficient for a team of workers to assemble their citations at the start and then to write their dictionary entries. If the first lexicon covers a corpus that is too small, then adding more data will reveal weaknesses in many entries and force the lexicographers to spend much of their time writing entries from scratch. On the other hand, corpus linguistics is a fairly well-studied field and enough is known to determine the appropriate size for an initial lexicon.

Whatever the drawbacks, the traditional method has a major flaw: a field can wither and die if it remains inaccessible too long or is not able to maintain its relevance to students and faculty of other, potentially related fields. A lexicon that is less comprehensive, but appears after ten rather than fifty years, can stimulate interest in the field, attract study from outside the field, and in so doing create (or at least protect) those positions within the field on which progress depends. There is, in other words, a correlation between the number of people reading a language and the amount of labor available to develop its lexicographic tools.

In Greek, we are profoundly fortunate to have at our disposal a lexicon which, whatever its drawbacks, provides us with an extraordinarily solid foundation on which to build. The basic lexicographic information—what cases does a given verb generally take, what are its usages in the classical period—as been collected. We can look to a more incremental approach rather than wholesale (and often unnecessary) revision: the long-awaited new supplement to LSJ is an example of just such fine-grained improvement.

More is not, however, necessarily better, and a bigger, more complex, lexicographic database may not be more useful. The formal, hierarchical format in which we write dictionary entries is easy to read, but our colleagues in cognitive sciences have produced plenty of evidence that such neat hierarchies have little to do with the way in which we represent and manage words in our minds.¹⁰ We as a field need to think long and hard about how we structure our knowledge of both Greek and Latin. Now is an admirable time to convert our print lexica to electronic form and make them as useful as pos-

¹⁰ For a clear survey of the issues, see Jean Aichison, *Words in the Mind: An Introduction to the Mental Lexicon* (Oxford 1994).

sible, but we are in no position to begin massive new lexicographic projects that will take decades, if not generations, to complete. Whatever master plan we may produce is likely only to hinder us in a few years' time when we better understand what can and cannot be done in this "brave new world."

But if we do not at present have at our disposal any model for an electronic lexicon that will satisfy us for decades, we stand all the more in need of new, exploratory work that will probe the limits of the new medium. Theorizing and predictions based upon our own expectations can take us only so far. The field needs to be able to examine new electronic notes and articles about the Greek language, since only such concrete examples will allow us to see what directions are and are not fruitful to pursue.

The fifth drawback of the print lexicon—however many citations it may contain, it does not contain the source texts from which these citations are drawn—becomes a much more tractable problem in an electronic environment, because most of the literary sources cited in LSJ and a growing share of the non-literary (e.g., inscriptions and papyri) have been entered on-line. We can write programs to convert citations from simple text strings to active links into the source texts: clicking on "Thuc. 2.38" can call up the text of Thucydides' *History of the Peloponnesian War*, book 2, chapter 38. Citations abbreviated to save space (e.g., "idem" and "ibid.") can usually be interpreted: instances where a single number appears in a dictionary entry are trickier (e.g., how do you make sure that your program knows that "46" picks up, for example, on "Thuc. 2.38" and refers to "Thuc. 2.46"?), but clever programs can decode most of the 500,000 citations in LSJ 9.

Here again we have gained some initial experience in the World Wide Web version of Perseus. We were not at first able to work with a networked copy of the TLG, and, even if we could, there was no framework whereby the TLG could be made generally accessible on the Web. The 3.4 million words of Greek in *Perseus 2.0* pale before the 70 million currently within the TLG,¹¹ but they represent many widely read authors and provided us with a good starting point. As mentioned above, LSJ 9 cites passages in *Perseus 2.0* texts more than 200,000 times. We thus set about converting citations to Perseus texts into active links which readers could pursue. Readers of the on-line LSJ 9 can thus often call up the source texts cited and thus pursue their investigations beyond the lexicon. The lexicon, *Perseus 2.0* texts, and various indexes could all fit on the 2.5-inch hard disk of a 1996 notebook computer.

Electronic systems have thus allowed us to begin ameliorating five problems which the codex poses for a massive lexicon. While sanguine predictions about continued technological development are hazardous, it is hard to believe that we will not soon be able to

¹¹ See the TLG World Wide Web site: http://www.tlg.uci.edu/~tlg/TLG_Info.html.

carry lexica, TLG, and various searching aids on a five-pound notebook computer. Whether this takes another year or decade should make little difference to scholars who measure their field's history in millennia. Combining this much information on a desktop computer should not be a serious issue for scholars (or dedicated students, for that matter). At the time of this writing, gigabyte hard drives cost less than some newly published 500-page scholarly books.

Search for Words Rather than Strings

We can therefore now see some of the most immediate consequences of transferring a complex reference work such as LSJ from print to electronic form, but the changes described clearly represent no more than a first step. Each of the five problems which I have addressed flows from the characteristics of print, and print has thus set the agenda for most of what I have covered so far: if the electronic document executes these steps more gracefully than the codex, the codex nevertheless choreographed the dance. A lexicon is truly electronic (and not simply a print artifact map onto bits) only insofar as it supports tasks which have no analogue in the print world. While any Greek lexicon will continue to support the reading of classical Greek, the individual tasks by which the lexicon supports this larger goal will surely change. No one can, at this time, predict with any confidence how radical or superficial these changes will be—indeed, dramatic revolutions may cause turmoil but bring little real change, while subtle changes, invisible to those who experience them, may have much more profound consequences.

We can, however, flesh out one, very rudimentary dimension which the electronic document possesses and on which we have worked. I began this paper with a rapid survey of the work that we had done on the electronic representation and analysis of Greek morphology. The need for morphological data motivated our earliest efforts with electronic lexica more than ten years ago, and much of our interest in LSJ 9 has arisen from a single question. If we record in a print lexicon that φέρω is a verb that means “to carry” (and many other things beside), the human reader who has lavished time and energy on mastering classical Greek can associate φέρω with inflected forms such as ἔφερον, φέρομεν, etc. Fully electronic texts—not simply ASCII transcriptions of print but dynamic composites of text and software rather than paper and glue—change the relationship between text and reader.

Morphologically Sophisticated Searches for Words and Phrases

First, scholars can now search Greek texts more effectively than they previously could. They can not only ask for all forms containing -φέρ- but they can also, as already noted, ask for a dictionary entry such as φέρω. This obviously provides us with another useful tool, but it has other consequences besides simple searches. Common words tend to be both short and irregular, each of which is

problematic for string searches. The shorter the word, the shorter the string by which we can find it and the lower the precision of string searches: searching for ἀγ- to locate forms of ἄγω is annoying because it retrieves forms of ἀγαθός, ἀγγέλλω, etc. Irregularity is even worse: to locate forms of ἄγω, we need to search not only for ἀγ- but for ἠγ-, ἀξ-, ἀχθ-, etc.; only scholars who are both desperate and thorough can locate true horrors of Greek morphology, such as the irregular verbs εἶμι (“to be”), εἶμι (“to go”), and ἵμι (“to throw”) in this fashion. We are now in a much better position to investigate such words. And, of course, if we can locate one word (rather than a string), then we can take the next step and search for sets of words within a given distance of one another. Our Web site offers as an example of this a search for ποιέω (“to do”) near εἶδ (“well”) so as to retrieve examples of the idiom “to do well,” “to benefit.”¹²

Calculating the Frequency of Words Rather than Strings

Once a computer program can reliably associate πέμπετε with φέρω it can not only retrieve all possible instances of φέρω but can keep track of how often different authors use a given term. Any such system must recognize the limitations on its data. We can use context to rule out analyses which are not likely in a given text: thus, πέμπετε might be either a present indicative or an imperative form, but it would be, in a speech of Demosthenes, an unaugmented imperfect form (unless it occurred in a quoted passage of poetry—hence the need for tagging our texts so that software can recognize such a linguistically crucial “context shift”). Often, however, there is no reliable way for software to disambiguate a form. In some cases, teams have gone through and disambiguated each form in a text: e.g., if they encounter a form such as αἰσχύνη, they check the context to determine whether this form is, in this passage, a noun or a verb, and, if a verb, whether active or middle/passive.¹³ Nevertheless, no one is going to perform this task manually upon tens of millions of Greek words in the near future.

Any morphologically sensitive retrieval system must thus account for the ambiguity inherent in Greek morphology. Our search tools report both a maximum and a minimum number of citations: if we

¹² In this case, however, there is one significant step that we have not, as of this writing, taken: we do not index the several dozen most common forms in Greek (many of these are, in fact, indeclinable forms such as καί so that in this set inflected form and dictionary entry are often the same). Since a number of systems (including Pandora, which evolved parallel to, but separate from, Perseus) can perform very good string searches on Greek texts, we devoted relatively little attention to this problem in Perseus itself. To locate instances of καί δὴ καί, however, a system must be able to scan the text file itself.

¹³ The morphologically tagged texts of the Septuagint, developed in the 1980s at the University of Pennsylvania under the direction of Robert Kraft, are an example of what can be done: gopher://ccat.sas.upenn.edu:3333/11/Religious/Biblical/LXXMorph.

were estimating the true frequency of αἰσχύνω, the maximum number would include instances of αἰσχύνη because this form could belong to either αἰσχύνη or αἰσχύνω; the minimum number would, however, ignore forms such as αἰσχύνῃ which do not definitely belong to αἰσχύνω but would tally up forms definitely belonging to this word (e.g., αἰσχύνετε).

Simply reporting the raw number of citations is not, however, useful, because some texts are obviously much larger than others. The verb φέρω shows up 261 times in the surviving plays of Aristophanes and 463 times in the dialogues of Plato, but the verb is actually more than three times as common in Aristophanes because the corpus of Aristophanes is so much smaller than that of Plato. We thus divide the number of times a word occurs by the number of citations to calculate the number of times a given word shows up per 10,000 words in a given author. This "relative frequency per 10,000 words" allows readers to see the true frequency of a term in one author versus another. Any system should report the relative frequency for both the maximum number of instances (i.e., if every possible form does belong to a given word) and the minimum number of instances.

We can thus see at a glance that, when the relative sizes of their texts is taken into consideration, Herodotus, for example, uses a given word ten times more often than does Thucydides. We could, of course, calculate this by hand using a string search (or even a print lexicon, if such a tool exists for a given author), but how many such hand calculations do we make? The rapid calculations of the machine allow us to explore many different words as we search for gross patterns. Of course, if we simply draw conclusions from the raw summaries of a machine and do not examine the evidence carefully, we cannot expect to produce useful results. Such facility will facilitate its share of sloppy work.

I doubt, however, that we will lose more than we will gain. It is well known, for example, that Herodotus refers to women roughly ten times as often as does Thucydides,¹⁴ and the marginalization of women has become a well-known trait of Thucydidean narrative. Every single usage of the term γυνή ("woman") in Thucydides has been discussed in the literature, and the patterns of usage carefully analyzed.¹⁵

But if we widen our focus beyond terms for "women," the resulting pattern proves a good deal more complex.¹⁶ Thucydides not

¹⁴ See C. Dewald, "Women and Culture in Herodotus' *Histories*," *Women's Studies* 8 (1981) 93-127.

¹⁵ E.g., T. E. J. Wiedemann, "Thucydides, Women, and the Limits of Rational Analysis," *Greece and Rome* 30.2 (1983) 163-70; D. Harvey, "Women in Thucydides," *Arethusa* 18.1 (1985) 67-90; P. Cartledge, "The Silent Women of Thucydides: 2.45.2 Re-Viewed," *Nomodeiktes: Greek Studies in Honor of Martin Ostwald*, ed. R. Rosen and J. Farrell (Ann Arbor 1993) 125-32.

¹⁶ A detailed exploration of this complexity occupies much of the central section of G. Crane, *The Blinded Eye: Thucydides and the Invention of History*, ed. G. Nagy (Lanham, Md., 1996).

only marginalizes women, but brothers, sons, and all kinship terms. The elision of women is a product not simply of Thucydidean misogyny (although misogyny does, I think, play its part) but of the historian's general tendency to minimize the role of the family and place the *polis* at the center of history. Women disappear not simply because they are women but because they are associated with the *oikos* and thus have no more place in Thucydides' narrative than do the densely intertwined, personalized male ties that likewise fade into the background. Others may challenge the interpretation to which I allude here, but the ability to compare the frequency of many words rapidly made it possible to see at once that the relative frequency of women belonged to some larger pattern within the narrative. Of course, each frequency could have been calculated by hand, just as we can read texts from beginning to end in search for occurrences of single words, but time is limited, and none of us can pursue every lead. In the end, the ability to see relative frequencies (a simple example of what our scientific colleagues term "data visualization") provides us with a new tool. If it is useful, then reviewers will assume that we have used it and criticize those who fail to exploit it. If we base judgments on numbers without looking directly at the evidence, our colleagues will ferret this out.

Searching Definitions and Headwords in the Lexicon

Any sizable document can itself be searched, and these searches add a new dimension to two elements of the on-line Greek lexica. First, searching the English definitions of either middle Liddell or LSJ 9 (i.e., locate all words that contain "wealth" or "coin" in their definitions) converts these tools into a rough, but effective, English-Greek lexica. We had from the first wanted to search the definitions in LSJ 9, but we were frankly surprised at how useful even middle Liddell proved to be. An hour browsing the 35,000 words in this student lexicon would repay many scholars who were investigating words for looms and weaving or for "hearing" or clothing terms, however wide their knowledge of Greek and close their familiarity with ancient semantic tools such as the *Onomasticon* of Pollux.

Likewise, searching the headwords allows us to track down forms that would be hard to locate in a conventional string search. Asking for all words that end in ἄγω or ἄγομαι will locate the compounds of ἄγω. The morphologically intelligent searching system can then retrieve inflected forms: thus, we would search headwords to locate ἀπάγω, while the search system would then locate ἀπάξετε (which a string-based scan for -αγ- would not detect).

Combining Links to Perform More Complex Tasks

So far I have described our ability to search for Greek words, to calculate the frequencies with which these words appear, and to locate interesting words by searching definitions or headwords within

a lexicon. Each of these tasks is useful in itself, but, if combined, they can provide more powerful vehicles with which to explore the Greek language.

Suppose that we are working on issues of gender in the ancient world: we might then search the definitions of the lexicon for various terms germane to this topic: Ross Scaife and Suzanne Bonafas, in their World Wide Web site on gender in antiquity, suggest terms such as “suitor,” “betroth,” “courtesan,” “weaving,” and many others.¹⁷ Each of these terms is sufficiently specific so that it will yield fairly usable results. The morphologically sensitive word searches then allow us to push this inquiry into the second (and absolutely crucial) stage of looking at forms of these words as they appear in the actual source texts.

Sometimes, however, we need to explore more general terms. A search for words containing “woman” in their definition yields 119 LSJ dictionary entries. If our interest centers on Plato or Homer, then it makes no sense for us to examine this entire list exhaustively.

But if we have on-line tables of word frequencies, a program can go through the 119 words that our search for “woman” turned up and filter out those which do not appear in the author that we are currently studying (in the case of Plato and Homer, the relevant terms go from 119 to 33 and 28 words respectively).

Consider a task that is even more daunting by traditional standards. Imagine that we are interested in the rise of abstract vocabulary and in the use of nouns ending in *-σις* as a reflection of this trend. We can search for all dictionary headwords that end in *-σις*: a search in LSJ 9 brings up 3,606 such words, and the word frequency tool calculates that these make up between 2.5 percent and 5 percent of all words in the 3.4 million text database of Perseus. The variation is high because forms such as *ποιήσεις* could belong to either the *-σις* noun *ποίησις* (lit. “making”) or the verb *ποιέω* (“to make”), but the overall picture is clear. We can now narrow our search to Aeschylus (between 1.2 percent and 4 percent) and Plato (between 4 percent and 8 percent). Even with such a “noisy” search as this where the variation is so high, we can still see clearly the difference in style, and we can use such a tool to trace the degree to which prose authors begin to exploit the suffix *-σις* for abstract nouns.

The limitations of such work are obvious. Searching for dictionary entries that end in *-ος* will give us a good idea of how frequently omicron-stem nouns and adjectives appear. Neuter singular nouns (such as *γένος*) are, by comparison, sufficiently infrequent that their presence does not grossly distort the results of this search—“instruments” report a certain level of “noise,” and classicists will soon enough develop a feel for how useful one query or another will be. The scholar interested in the neuter singular nouns like *γένος* might have little

¹⁷ [Http://www.uky.edu/ArtsSciences/Classics/dioperseus.html](http://www.uky.edu/ArtsSciences/Classics/dioperseus.html).

use for this electronic tool because searching for headwords that ended in -ος would primarily retrieve omicron-stem nouns and adjectives: the o-stem nouns and adjectives thus swamp the relevant data.¹⁸

If we go one step further and combine a translation with our Greek text, the situation becomes even more intriguing. The student with no knowledge of Greek language—whether the nineteen-year-old sophomore or the fifty-year-old full professor of political philosophy—can in some cases work surprisingly well with the Greek words.¹⁹ Consider, for example, the usage of the terms τύραννος (“tyrant”) and βασιλεύς (“king”). In Euripides, these two words are virtually synonymous: the playwright calls the same figures “kings” or “tyrants.” He may be doing so pointedly, suggesting (as do several speakers in Thucydides) that rulers are rulers whatever their origin and that power alone matters.

“Greekless” students can use the computer to locate all instances of τύραννος in Aeschylus, look at the Greek that has been highlighted, and then look over at the facing English translation to see if they can see which English word corresponds to the Greek. In most instances, they find the relevant word. A bit of effort will reveal two things. First, English translations commonly anticipate Euripidean usage, blurring the distinction between these two Greek words and translating them indifferently as “king,” “ruler,” “sovereign,” etc. Second, Aeschylus never blurs this distinction: by comparing the Greek to the English translations, students can quickly see that Aeschylus always applies the term βασιλεύς (and its derivatives) and never τύραννος to Agamemnon, while he always calls Aigisthos, the usurper, τύραννος and never βασιλεύς. The professor of political philosophy with little if any Greek cannot develop a definitive case, but he can spot trends of this type and present a fairly well-defined query to a colleague who does know Greek. The undergraduate in a course on Greek drama may, if the lesson is properly established, never look at a translation in quite the same way and can acquire a much better sense for Greek as a distinct language and system of values than could the most eloquent lecture alone.

From Unidirectional to Bidirectional Links

Once we can associate any form with its possible dictionary entries, we can also reverse links that had traditionally pointed in one

¹⁸ The morphological database that we have built provides a way out of this problem. The morphological analyzer needs to understand that γένος is neuter and ἄνθρωπος is a masculine o-stem. This information can then be used to allow scholars to search for third-declension neuters such as γένος, and avoid forms such as ἄνθρωπος.

¹⁹ See, for example, the large list of links provided by Suzanne Bonafas and Ross Scaife at <http://www.uky.edu/ArtsSciences/Classics/diopseus.html>; G. Crane, “Composing the Culture: The Authority of an Electronic Text,” *Current Anthropology* 32 (1991) 293–311, describes how undergraduates, many of whom knew no Greek, studied the concept of wealth in Pindar by locating and examining the usage of Greek words.

direction only. Recall, for example, the fact that LSJ 9 contains more than 500,000 citations to specific passages of Greek literature, and most of these are concentrated in the texts which students have traditionally most commonly read: there are over 200,000 citations to the 3,400,000 Greek words in *Perseus 2.0* (one LSJ citation for every seventeen words in core Greek authors included in Perseus or roughly one citation every other line). These citations are sufficiently dense that they cover many (perhaps most) odd usages in a given "canonical" text.

Each of these citations constitutes a link from the lexicon to the text, but these links are unidirectional: they point from the lexicon to the text, but there is no way ahead of time for someone reading a given text to know whether LSJ 9 cites a particular word in a particular passage without flipping the pages of the lexicon and then scanning through the tiny print of what may be a very long entry. The labor involved in hunting through the lexicon is sufficiently great and the chances of finding a specific answer are sufficiently low to make this mechanical exercise not worthwhile, both for professional classicists and especially for the harried student for whom the information would do the most good.

Consider the problem of someone reading the text of Thucydides 2.39.1. The LSJ entry on the preposition εἰς specifically cites the usage ἐς πλείονας οἰκεῖν. The LSJ entry on εἰς (and any commonly used word) is, however, quite long (two full columns of small LSJ print), but the probability that any particular instance of εἰς appears in it is low. Reading through the dictionary article takes a long time, the chances of success are low, and return for such labor thus probably does not justify the effort. Few classicists are likely to recognize that the specific ἐς of Thuc. 2.39.1 actually gets mentioned in the εἰς article.

How is the reader to know whether a particular form in a particular passage is discussed in the lexicon? If the full passage is quoted in the lexicon, a computer program can, of course, compare text and citation. To some extent, full quotations make up for differing inflections: the headword for the dictionary entry in LSJ is εἰς while Thucydides uses the variant ἐς, but since the dictionary article quotes "ἐς πλείονας οἰκεῖν Th. 2.39," a computer program can match the LSJ citation with its corresponding source text.

In a running commentary (like that of Jebb on Sophocles), a program can look at the quotations in the commentary, compare them with the source text, and highlight the sections of the text which Jebb actually quotes. Jebb's edition may, of course, differ from the more recent TLG text, but, if text and commentary are in sync, standard string comparisons can usually make this connection. An editor could fix the remaining failed links (probably 5–10 percent of the whole) by hand.

With LSJ, the relationship between text and lexicon is, however, often more complex. Citations are often somewhat loose (they sometimes substitute, for example, indefinite pronouns for actual words,

thus making trickier the job of the pattern matcher. But what happens when the text includes a complex inflected form (e.g., ἐπελθόντες, an aorist participle “having attacked” from the verb ἐπέρχομαι) but the lexicon does not repeat this particular form? The source text in Thucydides reads τὴν τε τῶν πέλας ἐπελθόντες. The citation in LSJ reads τὴν τῶν πέλας ἐ. Th. 2.39. First, the lexicon thus silently excises the τε, which means that the string matcher has to know both that τε is the kind of word that gets left out and that the accent on the τὴν is also different in source text and lexicon citation (the absence of the enclitic τε changes the accent on τὴν from acute to grave and thus introduces another small, but annoying variation). Second, the inflected form ἐπελθόντες has been abbreviated to ἐ., while the LSJ cites only the chapter as a whole (2.39) rather than its section as well (2.39.2). How, then, is the reader to know that LSJ comments upon the inflected form ἐπελθόντες in this passage?

There is a computationally easy solution to this problem. The citations in the entry for ἐπέρχομαι are collected and turned into an electronic index for all the dictionary entries that cite “Thuc. 2.39.” The reader calls up “Thuc. 2.39,” the computer scans through each inflected form of this section, and then matches forms against dictionary entries. It recognizes ἐπελθόντες as a form of ἐπέρχομαι, sees that the article for ἐπέρχομαι, cites “Thuc. 2.39,” and then builds a link from the inflected form ἐπελθόντες to the specific section of the LSJ entry that cites Thuc. 2.39 (actually, “Th. 2.39” in the original LSJ with its parsimonious abbreviations). This algorithm is somewhat loose (it ignores the full citation), and thus every usage of ἐς in Thuc. 2.37 will point to the citation for ἐς πλείονας οἰκεῖν, but the imprecision is normally not a problem and a more elaborate scanning system can be produced if need be. The significance of such a function is easily underestimated. There are, of course, some cases where virtually no reader would find a citation in LSJ 9—surely only the most hardy have dug their way through the lengthy print LSJ 9 article on the preposition ἐς, to see that it cites the usage of this word at Thuc. 2.39.1, and it is not clear whether the time spent finding this piece of information is worth the effort. Viewed another way, a task that would not have been worthwhile before now becomes a reasonable course of action. The individual benefit from looking up one citation at any given time may be small, but, if readers change their habits and repeatedly look up more passages and give more thought to the meanings of words, then the cumulative effect can be substantial. It would be easy to dismiss this link from word to lexicon as a mere labor-saving device that allows us to do standard tasks more quickly. I do not think that this would be fair—in practice, I believe that at the very least readers of Greek will make much better use of the material in LSJ, locating citations which they might otherwise not have checked and saving time which they can spend reading the source text.

But, of course, links from a text into LSJ 9 are only one special case of a more general task. The same underlying system whereby we create links from source texts into LSJ 9 would work for almost any standard reference tool that finds its way into the electronic format. The reader could ask not only which words in a given passage were cited in LSJ 9, but which showed up in Denniston's *Greek Particles*, in Smyth's compact *Greek Grammar*, or in the voluminous Kuehner-Gerth. The mechanics are in fact sufficiently powerful that we could link a source text to many different on-line documents, including books and journal articles. In effect, the electronic texts becomes not only an object to be read but at once a map which directs the reader to other passages which cite particular words and phrases. If our goal is to support not only professional classicists (who are obliged to spend a great deal of time "keeping current" and familiarizing themselves with what others have written on Greek) but also readers from outside our field, then overlaying links on a source text clearly has substantial advantages.

A moment's reflection will make clear that the ability to highlight words and/or phrases cited in LSJ 9 (or Denniston or a selection of journals during the past ten years) provides in very rough fashion something of the commentator's task. We may stand far less at the mercy of our commentators' interests and predilections, just as TLG word searches have given us greater independence from the judgments of LSJ lexicographers. But just as searches of the TLG have, if anything, often increased respect for LSJ 9 and the demand for good dictionary entries,²⁰ this wealth of links will, if anything, increase the value of judicious commentary: widely studied passages will have enough links into other reference works and articles to overwhelm the reader. Just as access to automated text searches is no replacement for good dictionary entries, automatically generated "backlinks" will make the value of a good commentary even clearer. Lexicographer and commentator will surely have to rethink their tasks in light of these new electronic tools and the reference works which they produce in the coming years may begin to look very different from those to which we have grown accustomed, but they will play a crucial role. The coming century may well find technologically sophisticated philologists flourishing as they create tools that open up our field in new ways and to new audiences.

²⁰ In February 1996, on the Classics-I Internet Discussion list, I raised the question of whether "TLG searches had reduced the importance of LSJ"; almost all of those responding confirmed that the access to TLG only increased their desire for good, comprehensive, and up-to-date articles on given words. Ilja Leonard Pfeijffer, a young scholar from the Netherlands who had just completed a commentary on Pindar, noted: "My experience with TLG has increased my respect and admiration for LSJ. On the whole, they have done an excellent job. One may frequently find oneself drawing different conclusions from the material they cite, different interpretations of the semantics of single words, but their collection of citations is, with relatively few exceptions, representative of the actual usage and distribution of the word at hand; nearly always they have included the most relevant occurrences."

From the Inflected Form to the Lexicon

I have spent a good deal of time with tools for research and exploration, but I would like to conclude with a set of links that are somewhat simpler, but may ultimately be more important because they potentially increase the set of people who can actually read Greek texts: readers confronted with inflected forms in a Greek text can call up their possible morphological analyses and thence shift to their dictionary entries. Students of the language can begin reading a wide range of texts at an earlier stage and with greater speed. On the one hand, the mechanical labor of flipping back and forth between the pages of a dictionary is effectively eliminated: point at ἔφηρες, learn that it is the second-person singular imperfect indicative active of φέρω: click on φέρω and consult its definitions. Those who use *Perseus 1.0* or even *2.0* may find that the machine still takes a few seconds to make these links, but the lookups will sooner or later become so fast as to seem instantaneous. The time spent moving from text to lexicon will begin to approach zero, a transformation with, I believe, few, if any, negative consequences.

But if the flipping of pages has little value, the discipline necessary to learn, and learn well, that οἴσεται is a form of φέρω or to master with precision the principle parts of Greek verbs is more problematic. The beauty of Greek as a discipline has rested in part in the degree to which our students have needed to familiarize themselves with a daunting quantity of information so that they may then explore “truth and beauty” in Plato, Sophocles, or Homer. That some of us have tended either to stress the “truth and beauty” while fudging a bit on the more abstruse elements of Greek, while the rest of us have been hard put to match in imagination our purely mnemonic achievements is less important than the fact that our discipline challenges us to cultivate both sets of skills.

In practice, I, at least, have not found that ready access to electronic information has compromised student learning any more than did the lexica, glossaries, or student editions of earlier generations. In part, the traditional methodologies hardly produced uniform results; I spent years gazing at traditionally trained senior classics majors who would earn honors in Greek and Latin and before whom aorist passive forms, for example, dissolved into a shapeless blur on the page. While the most successful students learned a great deal, many (I suspect, most) students have had little to lose, and much to gain, by a change in methods.

It is up to us as instructors to adapt our teaching and our expectations. I am not at all sure whether I will or should stress the memorization of principal parts as much as did my teachers, but such time as my students may save from this task they must now devote to learning their basic paradigms better, reading more Greek, and acquiring a wider vocabulary. One of my colleagues seized upon the technology to raise his standards: since students could always

use the machine to determine the possible analyses of any Greek form in their assignments, he no longer accepted statements that his students were unable to identify a word. Students could throw up their hands before the obscurities of Greek syntax, but they had no excuses if they had not mastered the morphology.

However much this new technology may benefit traditional students, its greatest potential benefit lies elsewhere. Not only is Greek very hard to learn, but few indeed of even our best students ever have the time at their disposal to pursue their earlier interests later in life. Law school and the hundred-hour weeks of the young lawyer, the apprenticeship as a junior executive at Widget Technologies, and a decade lived out of suitcases and in business hotels, or even the dissertation and subsequent “tenure publications” on Kant leave little time for Plato, so that, as careers stabilize and withered curiosities sprout anew, a return to Greek is rarely practical. We recognize that few of our students will ever again see Kalypso’s island through the eyes of Hermes or ponder the rough grandeur of an Aeschylean chorus, but we do not have to accept this condition. If the new technology allows even a few people to read Greek who might otherwise have turned to some less demanding task, then its benefit will have been immense.²¹ Whether or not we few scholars perform our work more or less efficiently pales by comparison—we have already dedicated ourselves to the study of antiquity. Each sixteen-year-old-high school student in Kansas, forty-seven-year-old captain in the Australian navy, or fifty-five-year-old historian of science who reads a line of Greek gives to the shades a life which they would not otherwise have enjoyed. As always, technologies—for reading as for any other task, whether based on ink and the press or on bits and silicon—may be powerful or weak but they are meaningful only insofar as they touch the body or fire the mind.

Tufts University
CW 91.6 (1998)

GREGORY CRANE
gcrane@perseus.tufts.edu

²¹ This is already happening and we often get letters such as the following: “One of my fondest recollections of St. John’s College in Annapolis, Maryland was learning ancient Greek. Twenty years later I am truly delighted to find your Perseus project. I thank you for your efforts. I expect to spend many hours at your web pages”; “Working as a science editor and 20 years out of classics, I’ve only just discovered the online version. You who work daily with the languages can hardly imagine the joy of pulling up the opening of the *Agamemnon* and finding the Greek still gloriously readable”; “I teach Greek at a small college pretty far out in the sticks. The resources one would like are not richly available in this area. There is, for example, a big Liddell and Scott in the library . . . but you can only access it during library hours . . . in the library. Now it appears that I have an even richer resource on my table at home . . . log on and tap a bookmark.”