

# Concordances, Collocations and Lexical-based Language Analysis : Implications for Computer-aided Pedagogy

Preston Perluss, University of Paris-IV

## Abstract

Current personal-computer technology allows every computer user to perform statistical and concordance-type studies of written documents. How can teachers best benefit from such technology? What methods and hypotheses appear the most useful to develop educational materials using these techniques? Will they bring about an overall improvement in foreign language teaching? Such will be the object of consideration of this paper. The fundamental role played by a "collocation" will be illustrated through a wide variety of examples. The contextual study of words will draw our attention to their often unpredictable behavior and shed light on the weaknesses besetting "standard grammars" and their associated pedagogies.

## Preamble

How often has a student asked about the meaning of a given word – let's say "welfare"? Without hesitation the teacher will reply, "it depends on the context." This contextual dependence raises pedagogical questions to which modern technology — in the even modest form of personal computers — might be of some avail. Personal computers offer the potential for enormous change in both classroom teaching strategies and student apprenticeship techniques. The most appropriate use for our high-tech tools awaits discovery and development. Herein I merely seek to display the range of present-day possibilities. For my own research I have used a Macintosh Duo 280c but kindred software exists for systems running DOS Windows. The lexical analysis examples presented here have all been produced by the software application Hyperbase© available from Professor Étienne Brunet, à l'UFR Lettres 98 bd Herriot 06204 Nice. A version for DOS equally exists.

Several hypotheses concerning language learning and language teaching have recently come to the fore : first, the study of disembodied sentences, and *a fortiori* words, torn from context, provides little means of advancing a learner's use (and not mere knowledge) of language which is inherently context dependent.<sup>1</sup> Thus, the need arises for a coherent way of reconstructing context while at the same time analyzing some particular linguistic phenomenon, so to discern regularities in usage.

Moreover, the very idea of a word might well merit reconsideration when we take into account several facts concerning both speech and writing<sup>2</sup>. To dodge the problem of specifying a definition for "word", researchers use the term "lexie" or "lexeme" to describe the functional unit of language "memorized in competence"<sup>3</sup> that is, the base unit used by native speakers. Without begging any questions, it seems reasonable to accord credence to the idea that people do not systematically reconstruct all they say or write from a mass of items commonly referred to as words<sup>4</sup>. There exist larger blocks of language, learnt as wholes, knit together by hidebound convention. And while the productive features of language allow us to create novel

---

<sup>1</sup>Support for this view arises from considerations of synonymy, for one example see [Carter and McCarthy: 1988 p. 29]. Michael Lewis pleads strongly for context-based teaching [Lewis:1993 p.80-81]

<sup>2</sup>Bally [Bally: 1951] in his classic treatise draws attention to the conceptual obscurity surrounding the definition of the term "word"; much more recently Tournier [Tournier: 1991] makes a similar observation

<sup>3</sup>"Competence" for those unfamiliar with syntactic grammars is a fundamental term used to describe a speaker's knowledge of the language as a mechanism for creating all possible sentences of that language. See [Brown, Malmkjær and Williams : 1997]

<sup>4</sup>See [Nattinger & DeCarrico: 1992] for a detailed review of facts and theories.

utterances from our linguistic inventory, *both the inventory and these same features must first be acquired through contextual familiarity*. Finally, a large part of our commonplace intuitions about words derives from a skill we all take so much for granted we are no longer aware that it is acquired: I refer to literacy, our ability to read<sup>5</sup>.

My goal, when bringing to bear computer-aided lexical analysis, is to develop a general approach (I avoid saying method) which will allow us, as teachers, to analyze texts written for specific purposes and thus create classroom materials while simultaneously coming to grips with the lexical niceties of English, in particular collocational unity.

If faithfully presenting realistic context to our students is our primary concern, then we must draw on true-life materials. From these materials we must extract the essential lexical structures to teach the students in the given time frame which is allotted to us. We must go for the essential. Thus, we must find the essential. First, I wish to introduce a technical term "**vocable**" which may appear at first superfluous, but will (I hope) little by little gain plausibility: by "vocable" I mean any series of letters separated by blanks or punctuation marks in a text. Obviously, this concerns the written or transcribed language. The word "word", I insist, hides a lush undergrowth of references and I wish to by-pass these for a moment.

What we need to establish within the framework of a given text or discursive domain, are the most commonly-used expressions whether we consider them to be institutionalized utterances, functional blocks (called "polywords" by Michael Lewis), or merely compounds formed from sequences of vocables.

More importantly, and more generally, we need to establish a high-fidelity sample of what is written and said. And we need to display this to the students who will observe, formulate interpretations and implement their hypotheses through application to other situations and contexts. But first and foremost, to understand some particular segment of the language, the learners must see the segment in context, at work, in a real linguistic situation. If we can exhibit a certain number of situations the learner will all the more readily generalize from each occurrence. We must forever bear in mind this fact. *No one sentence can confirm and refute an interpretation of the various uses its component parts might have*. But when you display a larger array of possibilities to the learners, the learners themselves are ideally placed to appreciate the nuances and possible variations to which a given vocable might be subject. We appeal to an inductive approach to language learning rather than merely a deductive approach wherein either the teacher or a dictionary provides a rule which is then exemplified and finally practiced only to be embarrassingly (for the teacher) overturned by some example further on. A final remark should also guide us: the actual language appears more chaotic and unpredictable than the simplified structures we all tendentiously teach in the classroom. Yet students must learn to structure that chaos.

## Some examples

Consider a text. Most of what follows applies to limited corpora. In fact, it seems all the more reasonable to limit the scope of lexical analysis to texts or conversations chosen for some specific purpose. We must stress the importance of context and the context may well be the text itself (the expression "cotext" applies here) or a given conversation whose intentional circumstances are fully explored. First, we shall seek to extract those vocables which occur the most frequently in a text and which allow us to predict something of the overall information context of the text. The larger the sample or corpus, the greater the risk that the most frequent vocables be those with the lowest informative value precisely because they occur in such diverse contexts. Recall Lewis's law: the greater the number of contexts in which a vocable is found, the greater the need for a context to understand that vocable.

Figure 1 (below) displays a typical exercise based on the extraction of a low-information value vocable from a text. We have made a concordance of "how". By **concordance** is meant the simultaneous listing of all occurrences of a given vocable or series of vocables in a specified corpus.

Figure #1

---

<sup>5</sup> See Jack Goody [Goody 1973] for considerations on how writing influences culture and a culture's idea of itself.

Consider the use of "how" in the following sentences. "How" is often given as the English equivalent for the French "comment"; but what does this vocable mean in a given context? Are "how" and "comment" merely used in questions? In what other contexts do they occur?

P2 3a| raises the question of how any society can function  
P2 3c| a limited number of themes : how the mechanization of work has been  
P3 5b| of machines determines how men think , act and dream . ¶ In  
P3 5c| with ever greater subtlety how demand , supply and price interact  
P4 8h| had not left their mark on how workers behave both on the job and  
P9 18b|them to decide what work to do , how to do it and even when to do it .  
P9 18d|greater autonomy in their work . How these demands will be reconciled

"How", thus, replaces the relative expression "the way in which" in the above contexts.

Figure # 2, a window from the application Hypertext®, lists the most frequently occurring vocables in a given text by order of frequency. What is the subject of the text and by what chain of thought do you reach your conclusions? The reader can make an educated guess on the basis of the primarily nominal clues provided. The *lexical base* of a text is composed of the most frequently occurring informative vocables. Obviously, frequency must be interpreted as being relative to the size of the text. The text under study contains approximately 8924 separate vocable occurrences (certain of these are punctuation marks) with only 1754 vocable forms of which some 900 occur only once (the latter are called, in linguistic jargon, the text's hapax).

Figure # 2

rang	frq	mot	rang	frq	mot	rang	frq	mot
1	793	the	31	31	)	61	19	control
2	380	of	32	31	(	62	18	work
3	308	.	33	29	such	63	18	technology
4	295	,	34	29	planning	64	18	so
5	266	a	35	29	information	65	18	other
6	217	to	36	29	factory	66	18	only
7	200	and	37	29	an	67	18	most
8	177	in	38	28	tool	68	17	was
9	153	is	39	28	production	69	17	resource
10	113	be	40	27	products	70	17	process
11	109	-	41	27	;	71	17	must
12	101	for	42	26	one	72	17	made
13	83	can	43	26	each	73	17	its
14	80	by	44	25	at	74	17	company
15	79	on	45	23	cutting	75	16	were
16	78	that	46	22	not	76	16	controlled
17	63	are	47	22	many	77	15	when
18	62	manufacturing	48	22	have	78	15	they
19	61	as	49	21	time	79	15	robots
20	56	part	50	21	needed	80	15	operations
21	53	machine	51	21	materials	81	15	has
22	53	it	52	21	assembly	82	15	could
23	49	computer	53	20	would	83	15	but
24	44	system	54	20	workers	84	15	been
25	44	parts	55	20	product	85	14	which
26	43	or	56	19	tools	86	14	also
27	38	from	57	19	screen	87	14	about
28	37	design	58	19	robot	88	13	up
29	35	with	59	19	into	89	13	through
30	31	machines	60	19	data	90	13	three
-----								
91	13	processing	121	9	their	151	8	before
92	13	inventory	122	9	program	152	8	back
93	13	if	123	9	plant	153	8	areas
94	13	"	124	9	percent	154	8	aircraft
95	12	some	125	9	linked	155	7	workpiece
96	12	small	126	9	line	156	7	wood
97	12	photograph	127	9	however	157	7	where

Lexical analysis provides a rapid insight into not so much the precise argumentation used in a text, but a skeletal representation of its language. The essential lies therein. The analysis helps both the student and the teacher. In a classroom situation, whereas the initial study of a text or discipline will embrace the nominal structures of the thematic language, the teacher's fundamental ambition must be to find a suitable number of verbs associated with these nominal structures. Let's say that the nominal structures form the abstract description of the discipline and the verbs put these abstractions into specific situations. The verbs, in a sense, rely on the intentional context of the speakers and listeners. When designing a course syllabus most of the language-based exercise material will be found within the text corpora furnished by the students or clients.

Figure #3 provides a good example of text-specific collocation: What nominal structures co-occur with assess? What compounds are formed? "Assessing" and "risk" co-occur as do "assessment" and "information". The trainees will be called upon to note the compound structures such as "supervisory assessment" as well as "assessing the impact of..." In this example, readers should notice that the concordance is based on a form "assess" whose morphology varies. The verb, noun and participle are all used. Can the student and teacher apply their skills to draw general conclusions concerning the uses of the different forms?

**Figure #3**

P1 1a| a framework for supervisors to assess information about the  
P1 2a| the framework provides for the assessment of supervisory information  
P1 3a| ( e . g . for the purpose of assessing market risks and amines ) .  
P1 3c| derivatives should therefore be assessed with sufficient frequency to  
P1 4a| have identified as important for assessing the risks arising from firms'  
P1 4b| conceptual methods for assessing the risk exposures related to  
P2 5a| the framework discusses ways to assess the concentration of credit risk  
P2 5d| covers row approaches for assessing market risks . ¶One is to  
P2 5d| allow independent supervisory assessment of an institution' s market  
P2 6a| of information important for assessing the impact of derivatives on  
P2 6b| discusses the importance of assessing information on both  
P2 6d| for supervisors to begin assessing the nature and scope of an  
P2 8c| consultations have led to an assessment of the information necessary  
P3 9b| be accessible to supervisors to assess the risks of derivatives and  
P3 10g| conceptual methods for assessing the risk exposures related to  
P3 10j| that supervisors can use in assessing the impact of derivatives on  
P3 11g| not specified , allowing for the assessment of information obtained  
P3 11g| information may be obtained and assessed through on - site examinations  
P3 12c| example , that for purposes of assessing an institution' s market risk  
P3 12e| aggregation , consolidation and assessment of information across a  
P3 12f| with regulated firms how best to assess information that provides a  
P3 12g| ¶ Data on derivatives should be assessed with sufficient frequency and  
P3 12l| , where appropriate , for assessing the impact of derivatives on  
P4 13a| with the aim of achieving the assessment of understandable and  
P4 13f| needed for supervisory assessment . ¶Each of the four broad  
P4 13g| risks , supervisors should assess qualitative information about  
P5 17a| to an effective supervisory assessment of the credit risk inherent  
P6 22e| may also be necessary for assessing liquidity risk . ¶In addition  
P6 22e| inter - als v ould be helpful in assessing funding risk . ¶37 .  
P6 23e| positions . Supervisors should assess information on market risk by

Here, in figure #4, we return to the text whose lexical frequency hierarchy was given in figure 1. Note the compounds in which machine occurs.

#### Figure #4

P1 1d| by the industrial robot , a machine designed to replace the  
P1 3d| , building materials and machine tools . ¶It is in these  
P2 4a|shall frequently refer to is the machine shop . There metal parts are  
P2 4a| on a lathe . ¶The same set of machine tools can serve to make a great  
P2 4b| the efficient organisation of machine - shop operations difficult .  
P2 4c|or numerically controlling , the machine tools themselves were  
P3 7a| to determine how a cutting machine , such as a lathe , must be  
P3 7a| the capacity of the cutting machine , the material from which the  
P3 7b| the machinist set up his machine according to drawings supplied  
P3 7b| ; when numerically controlled machine tools were introduced , the  
P3 8a| a part and in programming a machine tool to make it illustrates  
P3 8c| for processing on a particular machine were kept only by the  
P3 8e| of materials , the control of machine tools and other single -  
P5 14c| of labor , materials , machine time and other resource  
P6 18b| card can be imprinted with a machine - readable code such as the  
P7 20c| earliest numerically controlled machine tools were programmed by means  
P7 20c| tape . Each instruction to the machine was represented by a pattern of  
P7 20d| digital computer mounted on the machine . ¶A modern computer -  
P7 20e|the computer control enables the machine to cut metal automatically to  
P7 20e| , the program can prevent the machine from cutting too deep into the  
P7 20f| in some cases it can signal the machine operator to change or sharpen  
P7 21a| - numerically controlled machine tools are linked by a hierarchy  
P7 21a| direct - numerically controlled machine tools . Typically each machine  
P7 21a| machine tools . Typically each machine is controlled by a  
P7 21b|mainframe computer to any of the machine tools in the network . ¶In  
P7 21c| about the status of each machine , the volume of its production  
P7 21c| controllers . As many as 100 machine tools can be connected in such  
P7 21d| must still be moved from one machine to another by manual methods .  
P7 21d| direct - numerically controlled machine tools are further linked by a  
P7 21f| , the fraction of each shift a machine spends cutting metal can be as  
P7 21g| - numerically controlled machine tool standing alone the cutting  
P8 22a| products . A worker operating a machine tool manually can note a defect

Here in figure 5 we have an exercise using a more grammatical vocable "one". Find the fixed expressions using "one". The idea of grammaticalized lexis gains currency through the use of such methods: it is not by merely learning the formal grammar of the vocable "one", that a learner may acquire the vocable's bewildering uses: disembodied analysis daunts more than it describes.

### Figure #5

How do we use the vocable "one"? Does any expression come to mind? Consider the following concordance of "one". What can be said about the lexical unit in which it occurs? Are there sentence frames involved or polywords?

P1 1c| turns out that manufacturing is one of the most difficult sectors of  
P1 1d|to replace the production worker one for one . ¶Actually the direct work  
P1 1d| the production worker one for one . ¶Actually the direct work of  
P2 5a| as it passes from one person to another . ¶The time  
P2 5e| and the routing of a part from one process to another on the factory  
P3 8e| . ¶By linking the six areas one can achieve what Joseph Harrington  
P4 10b| - view drawing . ¶In return one of the most important benefits to  
P4 10c| - - ray - tube terminal . In one method , called finite - element  
P4 11b|custom valves from six months to one month . A manufacturer of molds for  
P5 15e| , for moving the part from one operation to the next  
P6 P7 One of the most important benefits of a  
P7 21d| must still be moved from one machine to another by manual  
P8 22d|by hand , with each worker doing one small step of the job and passing  
P8 24a| with other machines . ¶One of the most important applications  
P9 25d|¶Computers can then be linked to one another and to the central data  
P9 27b| workers on the first shift , one worker on the second shift and )  
P9 27b|worker on the second shift and ) one on the third . ¶Such small  
P0 28a| technology , why would one expect the technology ever to be  
P0 29d|force on a collar , the axle and one of the collars are enlarged ( c )  
P0 31c| bulkheads can be machined at one time . Each cutting head of the  
P0 32a| a later stage in the assembly . One advantage of the computer  
P0 32b| that carries workpieces from one machine to the next The  
P0 33c| is represented schematically in one of the positions it assumes

Figure 6 provides the teacher with the raw material for various exercises — exercises which one might not have otherwise contrived, had we not performed the correspondence for “time(s)”. What verbs collocate with “time”? Which multi-vocable blocks appear useful for expressing time relations between the present and past? What compound nouns are formed from “time”? Students can be led to discern the difference between “time” used as a nominal base in compound nouns and when used as a modifier in such compounds. Finally, the learners (and teachers) can compare the plural and singular forms of “time”. How often do we teach “time” with an “s” attached? — as a verb? Or in fixed expressions such “as many times as you like”? Does this list exhaust the manifold potential of the vocable “times”?

### Figure #6

P1 1g| part to spend 95 percent of the time required for its manufacture  
P2 4c| themselves were introduced some time ago and have been widely adopted .  
P2 5a| one person to another . ¶The time required for a memorandum to  
P2 6d| it could be recalled at any time . ¶  
P4 11b| ¶Another company - reduced the time needed to design custom valves  
P4 12b|- process characteristics as the time required for the setup of the  
P5 14a|manufacturing process at a given time . ¶For example , a manufacturing  
P5 14c| of labor , materials , machine time and other resource elements that  
P5 15d| assembly of a product , on the time needed for manufacturing each part  
P5 15d| part ( including not only the time spent actually working on the part  
P5 15e|working on the part but also the time needed for setting up machines ,  
P5 15e| at each station ) , on the lead time needed for purchasing parts from  
P6 17c| it must take into account the time needed for shipment . It can , be  
P6 18a| . ¶For example , a worker' s time card can be imprinted with a  
P6 18b| working hours recorded by the time clock are automatically assigned  
P7 19b|workers rather than spending his time in efforts to relieve shortages of  
P7 20e| around severalaxes at the same time ; the computer control enables the  
P7 21g| tool standing alone the cutting time may be as low as 10 to 30 percent  
P0 31c|bulkheads can be machined at one time . Each cutting head of the machine  
P0 35c| on careful analysis of the lead time and manufacturing time for each  
P0 35c| the lead time and manufacturing time for each part and on an accurate

---

P3 9d| form it can be changed as many times as necessary without the major  
P5 13e| can cut down on long waiting times and eliminate most of the costs  
P0 29f| along the applied force is 100 times the distortion in other  
P0 33a| Setup times are so reduced that such a system



Figure 7 provides an example of a multiple context cloze, easily constructed with concordancing software. Such exercises, often used in the past to draw on students' memories so as to test and reinforce prior material, can be adjusted to actually teach new contexts and new applications for well-know vocables. Multiple context clozes remove the same form from diverse contexts and thus stress the range of application a given vocable or group of vocables (polyword or even an institutionalized utterance) might have. In the following example the missing vocable "order" has an overwhelming frequency in the polyword "in order to" but that frequency does not exclude other types of occurrence such as "a single order for".

### Figure #7

P2	5c	concerns had become onerous . In	___	to get work done on lime the
P2	6c	of a cathode - ray tube in	___	to " draw , " or enter geometric
P4	11d	it was originally intended . In	___	to design a new part and plan how
P5	15b	automatically generates an	___	for the wood on August 4 , four
P5	15d	uncertainties in demand . ¶In	___	to introduce manufacturing
P6	16c	. In the Kanban system the	___	for a part to be made at one
P6	16d	set in motion by a single	___	for . finished products at the
P7	19b	relieve shortages of parts , to	___	repairs for machines and to
P0	33d	precise geometry of the part in	___	to program the cutting tool , can
P0	35a	is needed must be done in	___	to determine when the component
P0	35e	back - scheduled deadlines in	___	to conform to the master schedule

Figure 8 furnishes a highly informative example of one of the most common ways to express logical consequence: by means of the three vocable unit (polyword) "as a result". Students will probably notice the presence of the comma after the polyword, providing further proof of the polyword's innate independence. As well, note the cases where "of" links the polyword to a complex noun group which expresses the cause. The examples have been taken from two different texts.

### Figure #8

dw	7d	agricultural sector . ¶As a result , prices for these products shot
dw	8b	Israel devoid of anything . As a result they probably had negative
dw	9f	rather than the usual two . ¶The results have not been long to
dw	11a	The result of these conflicting trends in
dw	14a	this situation and that , as a result , the . ¶account deficit is
mg	17a	of 1997 and into 1998 . As a result , we expect the yield curve to
mg	21d	t income during 1996 ) . ¶As a result , we are sceptical regarding the
mg	25c	by the markets . ¶As a result , Steve also remains relatively
mg	26d	fully discounted , and , as a result , sees German bond yields rising
mg	26e	for some months yet , and , as a result , sees official interest rates
mg	31a	in corporate issuance . ¶As a result , the flow of funds is likely to
mg	31c	in corporate cash flow . ¶As a result , we still expect a balance of
mg	32b	- in both cases , and , as a result , expect yields to rise ( in the
mg	37c	( especially ) the banks . ¶As a result , it is difficult to find
mg	39c	a dramatic profit recovery as a result of its recent investment in new
mg	40b	of at least 15% , mainly as a result of the continued recovery of B&Q
mg	42a	where growth is strongest . As a result , we think it has a higher share
mg	48a	unlikely to deteriorate on the result of the US election . It was less

Figure 9, a simple listing of the occurrences of "it" within two successive paragraphs, serves as an example for students to demonstrate that "it" seldom changes reference within a single paragraph. Another interesting exercise consists in displaying all the contexts containing "it" within a text. The various roles fulfilled by the vocable "it" provide food for thought as to the nature of textual cohesion and coherence. The same, of course, might be done for all relative pronouns.

### Figure # 9

Because the design is simple to alter in electronic form **IT** can be changed as many times as necessary without the major effort of redrawing . The design is accessible to everyone who must work with **IT** as soon as **IT** is electronically filed, so that manufacturing functions such as the planning and scheduling of production can be started earlier.

This method of scheduling is called distribution resource planning, and **IT** must take into account the time needed for shipment. **IT** can , be employed to generate shipment dates for various products; **ITs** output can become input for a manufacturing - resource - planning system.

As a further example of the possibilities arising from concordance-based investigations of language, examine the respective contexts in figures # 10A & 10B for “those” and “these”. If someone asked you cold out what functional characteristics tell them apart, I fear that the “extra-linguistic”, distance-from-speaker-to-reference-object criterion may indeed be adduced. Reality justifies a totally different criterion (a primary criterion, however, which is not without counter examples): in text “these” serves to anaphorically tag noun groups while “those” refers to some set of persons introduced in the text. Note that although “those who” is an intensely powerful collocation; “these who” is not at all a collocation, I would go so far as to affirm its impossibility.

#### figures # 10A & 10B

P2 3a| A disquieting feature of these dynamic internal shifts in the  
P3 6b| some of the more important of these misperceptions and the inadequate  
P4 7g| and take proper account of these developments , the performance of  
P5 10c| operating at full capacity , these delaying tactics have in certain  
P6 11g|1940 the proportions employed in these kinds of occupation were  
P8 15a| These last considerations are a bridge  
P9 18d| autonomy in their work . How these demands will be reconciled with  
P9 18g| of workers by inflation makes these employees receptive to union  
P0 19d| foreign investments , excluding these fees , came to \$32 . 8 billion ,  
P0 21d|at or below the poverty level . ¶These trends can be disregarded only by

P1 1f| unemployment destabilized those societies , that national policy  
P3 6e| added to the personal wealth of those in power . Only in retrospect has  
P5 9b|authority and legitimacy . ¶Only those economists who believe everything  
P6 11f| workers ( narrowly defined as those who provide services primarily to  
P6 12a| members of the work force : those between 21 5 and 29 years of age  
P6 12b| increase in the proportion of those in the 25-to-29 age group who  
P0 21c| and the large fraction of those families that live at or below  
P0 24b| , mining and construction ) and those in the service sector , defined  
P0 24c|is narrowly defined to designate those workers who provide services

## In conclusion

The full pedagogical value of this kind of lexical analysis remains to be exploited. The method of learning a language through collocational study reproduces, in a more condensed version, our own apprenticeship with written language. Both students and teachers may avail themselves of this new technology to learn the lexical mysteries or regularities which each text harbors. Ultimately, every learner should have access to a library-like data-base which can be analyzed accordingly, that is by performing concordances upon selected parts, when the need arises. Lectiel at the École Normale Supérieure de Saint Cloud<sup>6</sup> is one example of such a library. The *Trésor de la Langue Française* has such a data base (FRANTEXT) which may be consulted from work-stations in various university libraries; unfortunately, individual taxpayers cannot access this data base from home. In the future, students in every discipline will perform corpus analyses on the literature in a specific domain in order to learn the vocabulary thereof. More generally, the concordancing techniques will continue to evolve with the addition of statistical tools to refine search methods. Concordancing and lexical analysis offer a solution to one of the most thorny problems in language learning: the mastery of collocations, networks of associations much too vast for any single dictionary to embrace. Thus, in the near future, alongside the dictionary will sit the work-station ready and able to flesh out all too abstract definitions.

### Biodata

**Preston PERLUSS** majored in mathematics and minored in linguistic philosophy at San Francisco State University. In 1985, he obtained a magistère in teaching French as a foreign language from the University of Paris IV-Sorbonne. Since then, he has been teaching and conducting research in lexicology and discourse analysis while preparing a doctorate at the University of Paris-IV in 18th century Parisian urban history.

---

<sup>6</sup>The system of computer-aided learning as presented by Mr Arnaud Pelfrène during the colloquium on *Usage des Nouvelles Technologies dans l'enseignement des Langues*, Université de Technologie de Compiègne, 13-14 March 1997.

## Bibliography

### General works on vocabulary, lexicology, and “lexical philosophies”

- Bally, Charles. *Traité de Stylistique française*, 1951 Paris, Klincksieck
- Brown, G., Malmkjær, K. and Williams, J. *Performance and Competence in Second Language Acquisition*, 1997, Cambridge University Press.
- Carter, Ronald and McCarthy (eds.). *Vocabulary and Language Teaching*, 1987 London, Longman. (collection of articles by various authors dealing with all aspects of the lexicon).
- Goody, Jack. *La Raison Graphique : la domestication de la pensée sauvage*, 1979, Paris, les Editions de Minuit. (Describes the deep modifications the introduction of literacy exercises on our culture and cultural perception)
- Lewis, Michael, *The Lexical Approach*, 1993, Hove, Language Teaching Publications (a pedagogical manifesto of great insight and persuasive force).
- Nattigner, James R., & DeCarrico, Jeanette S. *Lexical Phrases and Language Teaching*, Oxford University Press
- Perluss, Preston, "Lexical-based Language Learning", in *Prism*, vol. 1 Summer 1997
- Sinclair, J. M.(ed.), *Looking Up*, 1988, London, Collins.
- Tournier, Jean, *Précis de Lexicologie Anglaise*, 1991, Paris, Nathan

### Works on Corpus linguistics

- Barlow, Michael, homepage internet site :<http://www.ruf.rice.edu/barlow/corpus.html>
- Clear, J. "From Firth principles : computation tools for the study of collocation" in M. Baker, G. Francis & E. Toganini-Bonelli(eds) *Text and Technology : In Honour of John Sinclair*, 1993, Amsterdam, John Benjamins.
- Clear,J., Fox, G. Francis, G., Krishnamurthy, R. and Moon, R. "Cobuild : the state of the art", 1996 *International Corpus Linguistics*
- Garside, Roger, Leech, Geoffry & McEnery, Tony, *Corpus Annotation: Linguistic Information from Computer Text Corpora*, 1997, London, Addison-Wesley.
- Wichman, A, Knowles, G., McEnery, T. & Fligelstone, S., *Teaching and Language Corpora*, 1998 London, Longman.