Text and Discourse Analysis and Their Implications for Use in the Preparation of EAP Materials[1]

Suraprom Puangmali
Chulalongkorn University Language Institute

Introduction

This paper is a report on work being done at the Chulalongkorn University Language Institute (CULI) as part of a research project financed by the University. The task is a two-year project. It was begun in October 1976 and is expected to be completed in 1978.

Early in 1976, a seminar and workshop on text and discourse analysis was held. It was attended by a group of interested English-language instructors from a number of faculties. A team of some 25 analysts were then recruited from among these instructors. A series of meetings followed for the purpose of working out a viable analysis system. The textual materials used in the present analysis were sampled from written academic texts. However as a long-term research objective of CULI, spoken texts will also be dealt with in subsequent projects.

Definition of some terms

EAP, English for academic purposes, is a catch-all term for the type of English course which caters to the needs of specific groups of students of, generally, higher education institutions. The courses are said to be "discipline-based" (Strevens, 1977, 91); sometimes they are known as "subject-specific" or "subject-oriented". This type of English course is a sub-division of ESP, English for special purposes, a broader term for the kind of English which is either *occupational* or *educational* and which "....can with advantage be deliberately matched to the specific needs and purposes of the learner" (*ibid.*, 89). To illustrate, EAP for Chula students will be based, partly, on a needs analysis to be carried out on the basis of a questionnaire returned from the faculties. On the basis of an initial and incomplete needs analysis, reading for technical information ranked the highest of the four language skills in terms of emphasis.

'Text' refers to the physical representation of 'discourse'. A text can be perceived through orthographic spelling or phonological realization. Ideally, it also incorporates, in order for a discourse to be coherent and successful, the syntactic (or structural) and lexical components in their appropriate sequence. Discourse, then, is the final product of human language communication. If a number of sentences are put together in an unorganized way, the whole body of running text does not necessarily create a discourse. Conversely, a discourse does not need to consist of many utterances. It can be represented by only one sentence, a paragraph, a chapter, or even a whole book. Generally, it has unity and logical progression, as well as a definite purpose, and its material is organized in such a way as to be appropriate to a particular audience (Nilsen and Nilsen, 1975).

Purpose and Rationale

Since English seems to be one of the most important skills with which our students can attain their respective educational goals, effective communication through English should receive the first priority in learning/teaching. An important role in developing these skills is played by the study of the formal (syntactic, lexical and phonological) component of verbal communication. However, in addition to the mastery of the formal component of English, students must also learn to come to grips with its communicative purposes, as well as organizational techniques used by the writer/ speaker in creating discourse.

But how do students express communicative purposes and what kinds of organizational techniques do our students need for successful communication in their respective disciplines? How are these purposes and techniques expressed in linguistic forms? The literature at present has not much to offer and what there is has little direct application for language pedagogy, although the works done in this area, especially in discourse analysis, have given many insights into the intricacies of academic communication. Further, most of these works have been on EST, English for science and technology, another sub-division of ESP. A major syntactic study of scientific texts was done by Huddleston (1971). He investigated 27 texts on biology, chemistry, and physics using a transformational grammar framework. His study focused on such areas as mood, transitivity, voice, and relativization. Cowan (1974) carried out a lexical and syntactic analysis of some texts from medical journals. A master list of words was drawn from over 100,000 continuous texts, whereas the information on syntax, taken from 1,500 sentences, consisted of frequency counts of major syntactic constructions and frequencies of groups of structures occurring with other structures.

In studying scientific discourse, Selinker, Lackstrom and Trimble have shown how the rhetoric of science affects the use of language in areas such as articles,

tenses, statives, passives (1975), and presuppositions (1974). In these papers they discuss problems in these areas that are likely to baffle the foreign learner. The work by Jones (1974) and Mountford (1975) represents a very serious inquiry into the process of scientific discourse. They proffer two viable theoretical models for discourse analysis but they must be subjected to validation by the use of more data. And, again, the models have to be converted into practical pedagogic undertakings.

Owing to the above and due to the fact that we have no information on EAP other than EST, we felt we had to resort to our own resources. When we had text and discourse analysis in mind we thought it would provide valuable training for the analysts. Moreover we would have the chance to familiarize ourselves with texts of various academic disciplines which will be useful in our selection of materials and in materials writing. Through the analysis of selected texts, we hoped to discover:

1) a taxonomy of organizational techniques which an author in a discipline uses;

2) how the various communicative acts are accomplished by the author through language use. In other words, we wanted to find out how he manipulates syntax and lexicon in his acts of defining, classifying something, arguing for or against a case, etc. We intended to make up a list of communicative purposes or functions, together with their expressions in sentences[2];

3) what linguistic (syntactic and lexical) features form the common core of EAP and their relative frequencies of occurrence;

4) whether each individual discipline has features peculiar to itself, and, if any, what they are.

For (3) and (4) this meant that we would do a statistical study of EAP grammar and lexicon using selections from the same sources as those for discourse analysis but they would be from different passages. This is because discourse analysis is very time-consuming and the number of passages analyzed will be necessarily too small and inadequate for the syntactic and lexical study. This study will form a basis for a pre-program evaluation of our students' linguistic competence.

A word of caution is appropriate here. Although we will put a great deal into it in terms of time and effort, our research project has a pragmatic pedagogical goal. It must not be thought of as a piece of truly rigorous linguistic research.

Procedure

The project included an analysis of academic texts[3] in three areas: vocabulary, structure, and discourse. The approach adopted can be described as *eclectic* and *pragmatic.* A brief description of each area follows.

1. Vocabulary Analysis (VA)

The purpose of VA was mainly to obtain a common frequency list of words across the disciplines represented by the texts. The list will also show the grammatical category (as noun, verb, adjective, etc,) of each word, and the meaning with which it occurred. The list will be used for the selection of test items and vocabulary control in writing teaching materials.

The approach is semantic rather than structural. Occurrences of the same word with the same meaning will be tallied. Homographs (words with identical spelling but different meanings) will be listed as separate items.

We classified the words in a selection into grammatical, non-technical, and technical words (or units). The first were not listed. The classification of the last two groups was aided by consultation with general dictionaries, appropriate specialized dictionaries, and subject specialists.

The criteria for analyzing and counting were similar to those described by Puangmali (1976, 41). Briefly, we did not count inflectional endings unless their presence adds a different meaning to the items. Again, semantic consideration came into play. Items such as *give rise to*, *on the other hand* and *bring about* were counted as single items.

The books used were recommended by the staff of those faculties which are likely to need EAP. The selection of passages was made such that 2,000 items would be drawn from each book. Each selected page had to yield a minimum of 50 items.

Since the compilation of the data from the analysts would be computerized, we had to record the data on coding forms, the information of which would be converted into a computer-analyzable format. The information was punched on computer cards which will be further transformed and finally processed. The information to be recorded on coding forms was analyzed and entered manually. The information included the lexical items, their grammatical categories, locations (references to the sources from which they were taken) and an indication of whether they are technical or non-technical. The lexical items recorded so far numbers almost 30,000.

2. Structural Analysis (SA)

The purpose of SA was to derive a frequency list of structural patterns used in the sentences sampled, at random, from the textbooks. The list will be used primarily as a basis for the selection of grammatical items for a pretest before EAP materials are made.

The approach used was essentially *structural*. The list is classificatory: the sentences were dissected into their appropriate grammatical categories, listed under the main headings of *sentence complexity, complex noun phrases, nominalizations, adverbials,* and *tense.* The analysis focused on those types of structures which were expected to add significantly to the complexity of sentences for foreign language learners.

These categories were organized in a tabular form. The selected sentences were copied just above the table and the relevant data was slotted into the appropriate columns in the table in coded form. One hundred sentences were chosen from each of the textbooks.

Due to the difficulties involved in developing a system of analysis which meets the practical needs of the Institute, the actual analysis has only just begun to get underway. There is not yet enough data analyzed to obtain significant frequency counts.

3. Discourse Analysis (DA)

The purpose of DA was to capture those features used by the authors to obtain the cohesion of texts and the coherence of discourse, i.e., the grammatical and lexical characteristics of textual cohesion and the organizational, or rhetorical, devices contributing to discoursal coherence. Two grammatical cohesion features we were interested in were anaphoric and cataphoric reference; the lexical information relevant to our instructional objectives was summation, inclusion, synonymy, and definition (at the word and phrase levels). In connection with coherence, we wanted to compile a list of organizational strategies used in creating a whole discourse, a conceptual paragraph (Lackstrom, Selinker, and Trimble, 1975, 252). and a physical paragraph *(ibid.).* Therefore, our DA is basically classificatory in nature.

In terms of approach, we have been indebted to a number of people. The credit for our cohesion study system goes to the materials adopted and prepared by Ken Moody and David Charles for use in their classes, which formed part of the Three-Month Specialized Advanced Course on Curriculum Development and Syllabus Writing for Programmes of English for Special Purposes offered in 1975 at RELC, Singapore. Our system for analyzing discourse per se (in terms of organizational devices) was pragmatic and was heavily based on the Selinker-Trimble-and-Trimble model (1975). Our study of these communicative acts which are realized at the sentence level looks at the functional role of a sentence in relation to the overall context of the discourse.

The excerpts were selected from the beginning, the middle, and the end (but not the first and last chapters and not from the first or last section of the selected chapter) of the books. They were also self-contained (discussing and exhausting a topic). Their length ranged from two to three pages, or a little longer

The excerpt (discourse) was broken down into conceptual paragraphs, which are composed of one or more physical paragraphs, each of which in turn consists of, generally, a number of sentences. The analysts identified the overall communicative purpose of the excerpt, the conceptual and physical paragraph, and the individual sentences. All this information went into tables, and was later tallied categorically. Separate forms were devised to record the information on anaphora, cataphora, and lexical analysis. The Appendix shows some of the most frequent categories and sub-categories of cohesion and coherence.

Problems and Limitations

We were aware of the time constraint imposed upon us, so we set a manageable size of corpus data for DA. We further realized that although our data varied considerably, the data used for the first year was not very extensive. We certainly hope that, subsequently, we will be able to add to the present body of data. We will do the same with VA and SA, even though we are slightly happier about the size of their data bases. We would not want to jump to any conclusions simply on the basis of this data alone.

Another problem we had was the rather large number of analysts. This caused some difficulty of standardizing the results for each of the analyses. Perhaps VA has had the least problem: the criteria for selecting items, though in many ways *ad hoc*, were relatively more clear-cut than those for other analyses. Further, all the raw data was edited one more time before being keypunched.

In developing a system for SA we faced the difficulty of recording the frequencies of occurrence of those structures that are expected to give our students trouble and of capturing the overall complexity of the sentences and their major structural components. At the same time we had to be sure we did not devise an analysis system that was too complex and difficult for the analysts to apply correctly. The structural area which caused us the most problems was that of adverbials. In particular, we have found it very difficult to analyze and properly categorize the types of the adverbials we find, Due to the complexities involved with this structural category we expect it to be a problem throughout the analysis.

The worst problem of all seems to lie in the area of DA, where different terms may have been used by individual analysts to describe the same communicative purposes or acts. For example, in the Appendix under Total Discourse and Communicative Act, *explain* and *expound*, *describe* and *state*, and *explain*, *expand* and *explicate* are very close in meaning. This has to be re-examined and the compilation of the results adjusted.

Applications

We expect to have as a result of our text and discourse analysis two kinds of lists of data. One would be that which has features (linguistic and discoursal) common to all or most of the disciplines from which representative texts have been investigated. This may be called a list of *common-core* features. The other would enumerate features *distinctive* of each of the disciplines in terms of their type and frequency of occurrence.

The vocabulary lists and the structural lists will be compared with other lists which are available, and then they may be modified if need be. The lists of communicative purposes and communicative acts would certainly be improved through subsequent analysis of more data and, immediately, through some subjective evaluation such as the consideration of the philosophy and methodology underlying the discipline, a taxonomy of students' abilities to perform certain functions required by the discipline (Jones and Roe, 1975, 3) etc. Here is where the role of the specialist teacher is prominent and essential. Language teachers, through their professional experience, can also feed additional linguistic input into the results of the analysis. Further they are helpful in translating the learning objectives provided by the specialist teacher into concrete teaching materials and methods.

The whole set of data will then be compared with the contents of the Foundation English syllabus. Any items which coincide with those in Foundation should not and will not be incorporated in EAP. This new set of data will later be evaluated in terms of difficulty, probably through a pretest and, possibly, through direct assessment by the target population, who in general will be post-Foundation students. The outcome of the evaluation will determine the final set of learning items.

If the materials for EAP are to be largely discourse-based, it might be useful for us to organize EAP, first, into a number of major *core* courses and then later create *discipline-based* courses as they prove necessary. The core courses could be skills-oriented. The skills to be developed should eventually contribute to discoursal ability, or communicative competence. For instance, we might have a core course for the biological sciences, another for the physical sciences, and the other for the social sciences. Textual materials for a core course would then have to be related to a number of allied disciplines whereas those for a subject-specific course would be far more restricted and, preferably, authentic.

Conclusion

I would like to close this paper by bringing up and responding to several criticisms of discourse analysis, and the teaching of communicative purposes, or

functions. First, those who are discourse-analysis-minded have often been accused of jumping onto the bandwagon since the advent of DA in the language-teaching field. In some cases this may have been so. But the criticisms should not rule out the need for DA by language teachers since, in addition to the generally held view that in efficient language-learning one must know what functions linguistic forms perform, DA provides the teachers with, first, a chance to familiarize themselves with samples of the *rhetoric* of the academic discipline and, second, *authentic* contexts with which to handle teaching items.

The second criticism, or, rather, question, is what one does with a taxonomy of communicative functions or notions, and that the teaching of such functions would probably increase the learning load of the student. In my view, the teaching of functions should not lead to a significantly heavier learning load. In a largely structural syllabus, attempts have been made to relate structures (forms) to functions (meanings), but only in a somewhat non-systematic way. With this approach, one problem has been that one often is not aware of a multiplicity of meanings for a structure. On the other hand, by proceeding from functions to forms one capitalizes on one's already developed cognitive abilities and expresses one's ideas using the different forms of the new language. So this type of learning, beginning with the functions rather than the forms, becomes more meaningful because language forms usually do not exist in a vacuum. Even if this kind of teaching would be more burdensome for the learner, I still believe that it would be worthwhile. The history of foreign-language teaching in this country is much too familiar with the saying that one who is well versed in the grammar more often than not cannot communicate well enough in the foreign language in real-life situations. This is simply because the chance for one to be cognizant of the functions of the forms of the language and perfrom the functions in the language has not been stressed within the structurally-oriented language programs one has studied under.

APPENDIX

Some of the Most Frequent Categories and
Sub-categories of Cohesion and Coherence

COHESION

Anaphora	Cataphora	Lexical Analysis
−reference	−demonstrative	−summation
	−lexical	−synonymy

COHERENCE

Total Discourse	Conceptual Paragraph	Physical Paragraph	Communicative Act
−describe	−describe	−describe	−describe
−present	−explain	−explain	−explain
−explain		−state	−state
−expound			−contrast
			−expand
			−explicate

REFERENCES

Cowan, J.R., "Lexical and Syntactic Research for the Design of EFL Reading Materials," *TESOL Quarterly*, 8:4 (December 1974), 389–399.

Halliday, M.A.K. and Ruquaiya Hasan. *Cohesion in English.* English Language Series, Title No. 9. Kandolpn Quirk, general ed. London: Longman, 1976.

Hunddleston, Rodney D. *The Sentence in Written English: A Syntactic Study Based on an Analysis of Scientific Texts.* Cambridge: Cambridge University Press, 1971,

Jones, Keith, "The Role of Discourse Analysis in Devising Undergraduate Reading Programmes in EST (English for Science and Technology)." Paper presented at the Ninth Regional Seminar, SEAMEO Regional English Language Centre, Singapore, 23–27 July, 1974.

Jones, Keith and Peter Roe, "Problems in Designing Programmes in English for Science and Technology Overseas (EST)." Paper presented at the SEAMEO Regional English Language Centre: 10th Regional Seminar, 1975.

Lackstrom, John E., Larry Selinker, and Louis P. Trimble "Grammar and Technical English," "*The Art of TESOL: Selected Articles from the English Teaching Forum, Part Two.* Washington, D.C.: English Teaching Forum, 1975, 250–260.

Mountford, A.J., "Discourse Analysis and the Simplification of Reading Materials for English for Special Purposes," Unpublished M. Litt. thesis: University of Edinburgh. 1975,

Nilsen, Don L.F. and Alleen Pace Nilsen. *Semantic Theory: A Linguistic Perspective.* Rowley, Massachusetts: Newbury House, 1975, 47–54.

Puangmali, Suraprom, "A Study of Engineering English Vocabulary," *RELC Journal*, 7:1 (June 1976), 40–52.

Selinker, Larry. R.M. Todd Trimble, and Louis Trimble, "On Reading English for Science and Technology: Presuppositional Rhetorical Information in the Discourse." Paper presented at the Regional Seminer on the Teaching and Learning of English for Scientific and Technological Purposes in Southeast Asia held at RELC, Singapore. 21–25 April, 1975.

Selinker, Larry, Louis Trimble, and Robert Vroman, "Presupposition and Technical Rhetoric," *English Language Teaching*, 29:1 (October 1974), 59–65.

Strevens, Peter. *New Orientations in the Teaching of English.* Oxford: Oxford University Press, 1977.

28

Notes

[1] I thank Dan Brooks, of CULI, for reading the frist draft of this paper and for making comments on it. He has also provided the information about the grammatical categories for the structural analysis and the material regarding the structural analysis problem, which has been included in this paper.

[2] We have been able to identify and sample the communicative purpose of a passage or paragraph and the smaller units (communicative acts, generally represented by individual sentences) which contribute to that purpose. For example, it was not difficult for us to demonstrate that a comparison and contrast of two terms may involve an act of defining (the terms) and citing examples. We could also make a statement as to how the acts are sequenced. But it seems improbable at this stage that making a generalization about the structure and interaction of the constituents of a discourse (a passage, paragraph, etc.) will be an easy task. Halliday and Hasan (1976) make a similar observation, "...But it is doubtful whether it is possible to demonstrate generalized structural relationship into which sentences enter as the realization of functions in some higher unit. as can be done to all units below the sentence....we cannot.... list a set of possible structures for a text, with sentence classes to fill the structural roles...." (10). This might be partly because, due to a time constraint, our corpus of textual data was not very substantial. In addition the basic concepts and methodology of the individual disciplines as reflected in the sampled texts are, seemingly, so varied as to prevent us from reaching a satisfactory description of discourse structure which will be generalizable for a discipline.

[3] The passages were selected from basic and specialized textbooks used in the Faculties of Economics, Engineering, Political Science, and Science.