INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700  800/521-0600
Idiomaticity in basic writing: Formulas and idioms in the writing of some multilingual and Creole-speaking community college students

Sonomura, Marion Okawa, Ph.D.

University of Hawaii, 1993
IDIOMATICITY IN BASIC WRITING:
FORMULAS AND IDIOMS IN THE WRITING OF
SOME MULTILINGUAL AND CREOLE SPEAKING
COMMUNITY COLLEGE STUDENTS

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAII IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

LINGUISTICS

MAY 1993

By

Marion Okawa Sonomura

Dissertation Committee:

Patricia A. Lee, Chairperson
Derek Bickerton
George W. Grace
Donald M. Topping
Suzanne E. Jacobs
ACKNOWLEDGMENTS

I would like to express my sincerest appreciation to the student writers who allowed me to interview them and to use their papers for my dissertation study. A special mahalo is also due my classmate in composition research, Linka Corbin-Mullikin, for her encouragement and for introducing me to Guy Nishimoto, who, in turn, solicited the cooperation of Martine Aceves-Foster and Sandra Kelley-Daniel for me. To these teachers at the Kapiolani Community College, who were sympathetic to my needs and who assisted me to the fullest extent possible in the gathering of writing samples, I am most deeply grateful.
ABSTRACT

Errors in the writings of Kapiolani Community College (Honolulu, Hawaii) basic writing students, who are native speakers of English As a Second Language (ESL) or of Hawaii Creole English (HCE), were examined for idiomaticity, defined by a criterion of native-like appropriateness and naturalness. The purpose of the study was to learn how the demands of idiomaticity, in contrast to grammaticality alone, affected the academic acceptability of the writings. Native-like idiomaticity was seen to be a more stringent demand on student writing than mere grammaticality.

Specific problems of the basic writers could be discussed expeditiously only after errors were classified according to idiomatic types and a convenient terminology for idiomaticity was developed. Phrasematic expressions comprise three general categories, (1) idioms, which are semantically noncompositional and ambiguous, having literal counterparts, (2) collocations, which are identified by the high degree of predictability of concatenation of their constituents, and (3) formulas, which are common usages.
The ESL basic writing students produced more than twice as many total errors as did their HCE speaking classmates, and two significant differences were found in the error patterns of the two groups of basic writers: (1) ESL speakers made more errors of both kinds, grammatical and idiomatic, than did the HCE basic writers. (2) Whereas more of the ESL writers' errors were grammatical errors than were idiomatic errors, those of the HCE basic writers were more idiomatic than grammatical. Moreover, HCE writers produced more colloquially acceptable writing errors than did ESL writers.

The results of this study suggest that the requirements of idiomaticity are so prevalent that we cannot continue to ignore them, even if we must deviate from the established grammar-lexicon view of language in order to accommodate them. Not only language description, but such related disciplines as language teaching and learning demand the incorporation of idiomaticity into a fuller view of language which includes language use.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgements</td>
<td>iv</td>
</tr>
<tr>
<td>Abstract</td>
<td>v</td>
</tr>
<tr>
<td>Explanation of symbolic conventions</td>
<td>xi</td>
</tr>
<tr>
<td><strong>CHAPTER 1. INTRODUCTION</strong></td>
<td>1</td>
</tr>
<tr>
<td>1.1. Idiomaticity and basic writing</td>
<td>1</td>
</tr>
<tr>
<td>1.2. Topic</td>
<td>3</td>
</tr>
<tr>
<td>1.3. Description of the study</td>
<td>4</td>
</tr>
<tr>
<td>1.4. Delimitation of the study</td>
<td>6</td>
</tr>
<tr>
<td>1.4.1. Idiomaticity instead of grammar</td>
<td>6</td>
</tr>
<tr>
<td>1.4.2. Limited to written data</td>
<td>9</td>
</tr>
<tr>
<td><strong>CHAPTER 2. JUSTIFICATION</strong></td>
<td>12</td>
</tr>
<tr>
<td>2.1. Need for studies focused on idiomaticity</td>
<td>12</td>
</tr>
<tr>
<td>2.2. Observations of the need from various fields</td>
<td>12</td>
</tr>
<tr>
<td>2.2.1. First language acquisition</td>
<td>13</td>
</tr>
<tr>
<td>2.2.2. Second language learning</td>
<td>15</td>
</tr>
<tr>
<td>2.2.3. Bilingualism</td>
<td>18</td>
</tr>
<tr>
<td>2.2.4. Pidgin-creole studies</td>
<td>20</td>
</tr>
<tr>
<td>2.2.5. Morphology and syntax</td>
<td>21</td>
</tr>
<tr>
<td>2.2.6. Semantics</td>
<td>23</td>
</tr>
<tr>
<td>2.2.7. Pragmatics</td>
<td>26</td>
</tr>
<tr>
<td>2.2.8. Psycholinguistic studies</td>
<td>27</td>
</tr>
<tr>
<td>2.2.9. Neurolinguistic studies</td>
<td>28</td>
</tr>
<tr>
<td>2.2.10. Performance errors in speech</td>
<td>29</td>
</tr>
<tr>
<td>2.2.11. Error studies in writing</td>
<td>31</td>
</tr>
<tr>
<td>2.2.12. Spelling</td>
<td>32</td>
</tr>
<tr>
<td>2.3. Past research in student writing</td>
<td>34</td>
</tr>
<tr>
<td>2.3.1. Error studies</td>
<td>34</td>
</tr>
<tr>
<td>2.3.2. Written-spoken contrasts</td>
<td>35</td>
</tr>
<tr>
<td>2.3.3. Basic writing</td>
<td>37</td>
</tr>
<tr>
<td>2.3.4. Black English</td>
<td>38</td>
</tr>
<tr>
<td>2.3.5. English as a Second Language</td>
<td>38</td>
</tr>
<tr>
<td>2.3.6. Hawaii Creole English</td>
<td>39</td>
</tr>
</tbody>
</table>
2.4. Potential benefits of the study of idiomaticity .................. 41

CHAPTER 3. IDIOMATICITY .................................................. 43

3.1. The development of theories of idiomaticity ....................... 43
3.2. Avoidance of idioms .................................................. 46
3.3. Common definitions of "idiom" ...................................... 50
  3.3.1. Dictionary definitions ....................................... 50
  3.3.2. Traditional grammar definitions ............................ 52
  3.3.3. Linguistic definitions ...................................... 54
    3.3.3.1. Structuralist attempt .................................. 54
    3.3.3.2. Tagmemic views ....................................... 56
    3.3.3.3. Transformational-generative accounts ............... 57
    3.3.3.4. Stratificational contribution ....................... 59
    3.3.3.5. Post-generative departures .......................... 60
3.4. Collectors and etymologists ...................................... 64
3.5. Early theorists .................................................... 65
  3.5.1. Jespersen 1924 .............................................. 65
  3.5.2. Smith 1925 .................................................. 66
  3.5.3. Murat H. Roberts 1944 ..................................... 69
3.6. Recent theorists .................................................. 71
  3.6.1. Soviet phraseologists ...................................... 71
    3.6.1.1. Bakhtin 1929 and 1953 ............................... 71
    3.6.1.3. Mel'čuk 1960 ......................................... 76
    3.6.1.4. Amosova 1963 and 1966 .............................. 77
    3.6.1.5. Reichstein 1973 and 1974 ............................ 78
  3.6.2. Bar-Hillel 1955 ............................................. 79
  3.6.3. Hockett 1958 ............................................... 79
  3.6.4. Malkiel 1959 ............................................... 80
  3.6.5. Hymes 1962 ................................................ 81
  3.6.6. Katz and Postal 1963 ..................................... 82
  3.6.7. Pike 1967 ................................................. 82
  3.6.8. Chafe 1968 ................................................ 83
  3.6.9. Weinreich 1969 ............................................ 84
  3.6.10. Fraser 1970 ............................................... 85
  3.6.11. Makkai 1972 .............................................. 86
3.6.13. Coulmas 1981 ...................................................... 88
3.6.15. Strässler 1982 ........................................................ 91

CHAPTER 4. PROBLEMS FOR A THEORY OF IDIOMATICITY ...... 100

4.1. Problems of definition of idioms and idiomaticity .......... 100
4.2. Identification of idioms ............................................. 101
  4.2.1. Criteria for inclusion of a form as idiom .............. 102
  4.2.2. Single words and multiword idioms ..................... 104
  4.2.3. Logic, grammars, and universals ....................... 105
  4.2.4. Terminology for idioms and idiom-like phrases ....... 109
4.3. Some aspects of idiomaticity .................................... 123
  4.3.1. Dimensions of regularity ................................... 123
  4.3.2. Degrees of regularity .................................... 125
  4.3.3. Form, meaning, and function ............................ 126
  4.3.4. Grammar and lexicon ..................................... 127
  4.3.5. Figurative meaning and literal meaning ............... 129
  4.3.6. Formulaicity ............................................... 131
4.4. Some supralinguistic factors affecting language use ....... 132
  4.4.1. Neurolinguistics and language processing ............ 132
  4.4.2. Speaker versus hearer perspectives .................... 133
  4.4.3. Automaticity ............................................. 134
  4.4.4. Familiarity and frequency ............................... 135
  4.4.5. Cohesiveness ............................................ 137
  4.4.6. Repetition and redundancy .............................. 138
  4.4.7. Affective strength and prosody ......................... 138

CHAPTER 5. IDIOMATICITY AND BASIC WRITING ..................... 140

5.1. The study ............................................................ 140
5.2. Purpose .............................................................. 140
5.3. Design ................................................................. 143
5.4. Data collection ..................................................... 147
5.5. Analysis procedures ............................................... 148
5.6. Results of the analysis .......................................... 150
5.6.1. Error counts .................................................. 150
5.6.2. Types of errors ................................................ 153
5.6.2.1. Accidental and meaningful errors .............. 154
5.6.2.2. Grammatical errors ................................. 163
5.6.2.3. Lexical and morphological errors ............ 176
5.6.2.4. Compositional and stylistic errors ............. 178
5.6.2.5. Logical and semantic errors ................... 183
5.6.2.6. Pragmatic errors ..................................... 188
5.6.2.7. Phrasematic errors ................................... 196

CHAPTER 6. CONCLUSION ........................................... 228

6.1. Interpretation of the findings ............................ 228
6.2. Implications of the findings ............................ 234
6.2.1. Theoretical implications ............................. 234
6.2.2. Practical implications ................................. 235
6.3. Some preliminary conclusions .......................... 238
6.3.1. The relationship of idiomaticity to grammar .... 238
6.3.2. Psycholinguistic hypotheses ......................... 240
6.3.3. Additional observations .............................. 241
6.4. Suggestions for further study .......................... 242

Notes ................................................................. 245
References ........................................................ 247
EXPLANATION OF SYMBOLIC CONVENTIONS USED

Asterisks before examples indicate unacceptability.

Boldface within the text is used for emphasis of special terms or definitions.

Capital letters are used for grammatical category labels, sometimes in abbreviated form (ADJ = adjective, NP = noun phrase, etc.)

Italics are used for citation of linguistic examples. They are preferred over and used without quotation marks where quoted materials are cited as linguistic examples.

Question marks before examples indicate questionable acceptability.

Single quotation marks indicate glosses.

Underlining is used sparingly for emphasis where they aid the reading. Used with italicized examples, they may indicate the specific parts of the example to which attention is being called, or they may group graphically segmented terms into presumably unanalyzed phrasal units. In Chapter 3, continuous underlining is used to highlight the various idiom definitions presented to facilitate comparison of definitions.

Square brackets are used to enclose some examples where use of parentheses is confusing.

A pair of empty square brackets [], used in examples, indicates a missing word where one would ordinarily be expected.

No attempt is made to alter quoted material to conform to the above conventions.

Letters and numbers preceding cited examples are the researcher's identification codes and may be ignored.
CHAPTER 1

INTRODUCTION

1.1. Idiomaticity and basic writing.

Writing produced by students enrolled in remedial writing classes displays a surprising diversity of structures which diverge from the expected norms of academic writing. Commonly termed "errors", these non-conforming structures may be grammatical anomalies, mistakes in lexical selection, stylistic inconsistencies, breaches of poetic or prose convention, or something which writing instructors may mark "awkward" or perhaps label "unidiomatic". Just what are these frequently occurring, elusive, unnamed forms which are not appropriately classed with any of the other types of errors but which are nevertheless unacceptable? Do they constitute a unitary phenomenon of some sort, identifiable by some specific criteria? How are they different from grammatical errors? These are some of the initial questions which motivated this study, and for which the term "idiomaticity" was used as a tentative focal point. Idiomaticity is used here in a much broader sense than is
encompassed in the usual meanings of the word "idiom". Briefly stated, idiomaticity is what makes the expressions in a language sound appropriate and natural in that language.

The study of idiomaticity in student writing required a far-ranging preliminary exploration into the research literature of various subfields of linguistics, bilingual education, writing, and the teaching of English composition. From this search, a blurred picture emerged of an uncomfortable fit between the generally accepted view of language and the demands of idiomaticity in language use. The view of language in question is one Pawley (1985b:85-86) referred to as the "grammar-lexicon model"--one which assumes that a language is fully describable in terms of a lexicon of single words and a grammar for combining those words. No accounting is made of the multiword complexes which carry unitary meaning, the frequency of occurrence of certain collocations compared to their possible paraphrases, the acceptability of certain ungrammatical and illogical sequences, or the fact that many theoretically possible combinations are never actually used.
A careful analysis of student writings with attention to problems of idiomaticity made clearer the necessity of viewing language in such a way that facts of language use can be accounted for. Facts of idiomaticity, then, should no longer have to be set aside as anomalous, as has long been common practice, but attempts should be made to account for idiomaticity within a description of language which includes language as used. To begin to account for idiomaticity, the existence of multiword units must be recognized and the problem of how they should be accommodated in a language description must at least be acknowledged, if not solved. Only after we have discovered, or devised, a productive way of talking about idiomaticity will we be able to make progress toward establishing native-speakerlike idiomaticity as an explicit target for student writers, whereas now, such a goal is only vaguely assumed.

1.2. Topic.

Although the title of this dissertation is stated in positive terms--Idiomaticity in basic writing: formulas and idioms in the writing of some multilingual and creole-speaking community college
students--what is really discussed in this study is lack of idiomaticity. The writings of unskilled writers are conspicuously unidiomatic when judged by the standards of written English expected of college students. Idiomaticity is presented in this dissertation as the ultimate goal for student writing, meaning that, beyond grammaticality, the best of student writing must also be expressed in language which is fluent and natural in the special idiom of academic written English. The topic of this study may be more accurately, though less economically, expressed as "Lack of idiomaticity in basic writing: a study of the failure of some multilingual and creole-speaking basic writing students to use expected formulas and idioms in their community college writing."

1.3. Description of the study.

113 papers written by 67 students enrolled in four basic (remedial) writing classes at Kapiolani Community College in Honolulu were collected and analyzed for sentences which were unacceptable as academic writing. The unacceptable sentences were then sorted according to types of errors which they contained, using
a list of categories tentatively drawn from the available literature on formulas, idioms and idiomaticity. Errors identifiable as grammatical were also noted so a quantitative comparison of idiomatic and grammatical errors could be made. Compositional and stylistic errors as well as errors of lexical selection fell somewhere between, so they were tentatively included in the definition of idiomaticity.

Writers were identified as belonging to one of two groups: 37 were non-native speakers of English (ESL) and 30 were native speakers of Hawaii Creole English (HCE). (One basic writer who was a native speaker of Black English dialect and another who was a native speaker of general American English were excluded.) Since there were more ESL basic writers, they submitted more papers, a total of 60, than did the HCE basic writers, who submitted 53 papers. However, the HCE speakers tended to write somewhat longer papers, so the number of pages produced by each group was almost equal--119 ESL and 117 HCE.

The sorted lists of deviant sentences were studied and used as a basis for comparing the idiomaticity of written language used by
the two groups, for forming some generalizations about idiomaticity and lack of idiomaticity in basic writing, for identifying types of idiomatic expressions in English, for describing some specific difficulties basic writers have with idiomaticity, and for discussing some problems for the formulation of a theory of idiomaticity which should undergird the teaching of idiomaticity to basic writing students.

1.4. Delimitation of the study.

1.4.1. Idiomaticity instead of grammar.

Error studies, until very recently, have almost without exception been focused on grammar. The struggles of student writers with number agreement, with the use of tense and mood, with agreement of pronouns and antecedents, and with the use of comparatives and superlatives occupy the major portion of the current literature on common errors. The common practice of researchers faced with idioms is to exclude them. For this reason, a study of student writing with a focus on idiomaticity was felt to be
overdue. To ignore the selection of "The Way of My Life" as the title of a paper instead of "My Way of Life" because they are both technically grammatical is to lose sight of important differences in acceptability to native speakers. Or, conversely, to insist a student write a number of classes is instead of are because the verb must agree with the singular subject number is to ignore the native speaker's normal way of expressing the same idea, in this case, the native-speaker's sense of idiomaticity in the use of semantic, or notional, rather than grammatical number agreement.

Tentatively, idiomaticity will be defined rather informally as the qualities of an instance of language use which render it appropriate, normal and natural to native-speakers of the language. For this study, idiomaticity will be judged against an intuitive sense for a written standard. Idiomaticity will be assumed to be concerned only with the form given to content and not extend to include sociolinguistic factors of speaker and hearer relationship, relative status, politeness requirements, and the like. Phonological factors are taken into account only as they affect writing. Idiomaticity is viewed as being constrained by conventions of
language as surely as grammar is. As such, idiomaticity involves the
givens of a language and not merely the options, implying that native
speakers are able to judge whether a form chosen is normal or
strange, usual or unusual, if not right or wrong. While most
definitions of "idiom" focus on the decoding problem, idiomaticity is
concerned with difficulties of encoding as well. We are not only
concerned with how a nonliteral phrase (*kick the bucket = die*) is
interpretable but also with how a literal phrase (*someone's at the
doors*) is known to be the usual and natural expression for a
particular situation.

Idiomaticity is thus not limited to what are usually called
"idioms" but involves a heterogeneous set of linguistic phenomena,
including fixed idioms, formulas, exceptions to grammatical rules,
lexicalized compounds, collocations, colloquialisms, irreversible
binomials, phrasal verbs, conventionalized figurative language such
as frozen similes and metaphors, proverbs and sayings, etc. The
question of the size of items to which idiomaticity applies will be
discussed further in Chapter 4, but for the time being, this study
excludes single words and multisentential sequences. A simple
heuristic for identifying a phrase as idiomatic would be to ask, "Would a non-native speaker know that the expression is a conventional way of saying what it says in English in this context?" Admittedly, application of this test yields a huge quantity of phenomena which would be labeled idiomatic. The advantage of using this test, however, is that it captures nearly every kind of linguistic error produced by students which are not describable as grammatical in usual ways.

1.4.2. **Limited to written data.**

The choice of written English as the object of this study was motivated in the first place by a personal interest in writing and the teaching of English composition to non-native speakers of English, and, secondly, by a recognition that although the written idiom places special demands on non-native learners of English, little has been done to increase our understanding of idiomaticity in writing.

Because the view that gives priority to spoken language has long held ascendancy in the linguistic discipline, the decision to
collect written data for a study of idiomaticity invites censure. The following facts, however, may provide sufficient justification for the use of written data in the study of idiomaticity: (1) When idiomaticity is the focus, written language is a different idiom from spoken language and facts gleaned from the study of the spoken idiom are not automatically valid for the written. Anyone who has read a transcript of extemporaneously spoken language does not need to be convinced that written language is not merely a graphic record of the spoken. (2) There is no reason why written language cannot be the object of study, even though observations about written language may not be valid evidence for statements about the spoken language. When directives are issued that spoken data is to be preferred over written, the concern is usually that evidence from written data is being used to make invalid generalizations about language, or languages, as a whole. Attempts will be made in this dissertation to carefully discriminate between statements made about English in general and about written English in particular.

Finally, it is important to emphasize that the "written standard" by which idiomaticity was judged in this study is not an
undisputed, thoroughly codified instrument of authority of any kind. A few words have been written, and much more can yet be said, about the nature of language standards and standard languages (Joseph 1989). Many also recognize that all languages have an abundance of varieties, induced not only by geographical and social differences, but also by the diverse functions which languages perform. There is, unfortunately, no room to discuss thoroughly the complex topics of language "registers" or of standardization here. Let it at least be made clear that what is considered to be acceptably idiomatic in the relatively formal registers of academic writing can differ vastly from what is idiomatic in the colloquial registers.
CHAPTER 2
JUSTIFICATION

2.1. Need for studies focused on idiomaticity.

Pressure to study idiomaticity in student writings comes from two directions. First, evidence is accumulating from many fields that phrasal units such as idioms and formulas are important factors in language. Second, idiomaticity has been set aside rather than being dealt with for so long that there is a serious deficit in our understanding of the workings of language which inconveniences learners.

2.2. Observations of the need from various fields.

A representative sampling of observations which have led to a recognition of the necessity of accounting for idiomaticity will be described briefly. These observations come from research in a number of subfields of linguistics and from related disciplines. In general, these observations share the common characteristic of
demonstrating the behavior of multiword groupings as unitary wholes which, in some way or another, need to be treated as exceptional because current theories will not accommodate them. At the same time, those who have looked more closely at these multilexical units have observed that the obvious solution of entering such phrases in the lexicon as lexical items is overly simplistic and will not suffice.

2.2.1. **First language acquisition.**

Until recently, a prevailing assumption was that children learn language by first acquiring a small store of individual "words", more or less equivalent to an adult's word. At a later stage, children are said to begin combining two words into simple syntactic structures which subsequently grow more complex as the child matures (Tager-Flusberg 1985). Although it had often been informally observed that all children did not necessarily proceed in learning language beginning with single words (as usually defined) before producing any phrases, it was not until recorded data from actual child speech were carefully analyzed that the earlier assumption was seriously
questioned. In 1972, Moskowitz (1972 and 1980) discovered evidence of unanalyzed units being produced alongside rule-combined counterparts in the phonology of children. In the same year, Clark (1972) also concluded:

... imitated items, whether single words or longer stretches, are not fully analyzed by the child for syntactic features or syntactic structure at the time they are imitated, but syntactic understanding proceeds in the course of production, usually after a lengthy period during which the utterances have been practiced as unanalyzed wholes (Van Lancker 1975:140).

In noting that 14-month-old Minh was producing in his one-word stage such multilexical units as look at that, what's that, and open the door, Peters proposed that "while words and morphemes may be the ultimate units in a logical and economical description of a language, the actual units used in speech production may be different" (Peters 1983:2). Although it is probably true that unitary items need to be acquired before relationships between them are formed, boundaries which identify the units are not at all obvious to a new learner, and what the child perceives as units do not necessarily correspond to the adult's.
While some have seen formulaic phrases as dead-ends in the language creativity process (Krashen and Scarcella 1978), Peters agrees with Wong Fillmore that formulas become a way into the language by providing material for analysis: "the analysis the learners perform are on those things which are most available to them--the well-practiced and familiar expressions they find in their own speech repertories" (Wong Fillmore 1976:301). Children, whether learning a first or some additional language, may first store phrasal wholes in their lexicons as well as single words, but

... items in the lexicon are subject to analysis by the rules as they are induced and those items that yield to such analysis may lose their status as unitary items of storage. This in fact must be the main process used by children initially breaking into the language system (Peters 1983:15).

2.2.2. Second language learning.

While Moskowitz, Clark, and Peters focused their studies on first language development, the statements made by Wong Fillmore about the role of formulas in language learning were made on the basis of her studies in second language learning by children and
appear to apply equally well to either case. In her study of five Hispanic children learning English in primary school, Wong Fillmore observed that imitating "formulaic speech" constituted an important strategy for the most successful language learners. Typical formulas she recorded were those which allowed children to participate in activities which provided further opportunity for learning--Lookit, lemme see, gimme, you know what?, I wanna play, Whose turn is it?, Knock it off (1979). These are initially learned by rote as unanalyzed wholes.

... what children have to work with are expressions which make sense to them already, at least in the context of use. They know how an utterance can be used, even if they do not always know how it segments, what all of its individual components are, or how they function. Eventually they manage to sort out the parts of these utterances and to discover what other items might substitute for them (1976:301).

Hakuta discussed the role of prefabricated language in the learning of English as a second language by five-year-old Ugisu. "What she did was learn whole expressions that proved useful in her interactions with friends, utterances like Don't do that!, Do like this, This is mine, That's not yours, Do you want this one?, and I know
how to do it " (1986:126). What is interesting about the last expression is that when Uguisu began constructing an internal grammar for questions, she began to deviate from the correct prefabricated form she had memorized earlier and was now saying I know how do you spell 'Vino' and I know how do you write this. Hakuta presents this as a dramatic example of evidence that the child eventually analyzes the components of whole prefabricated utterances and begins using the the parts in a productive way (1986:127-28).

Hakuta noted that well over half of Uguisu's earliest utterances were of the prefabricated sort, which accorded with reports from a number of other second-language acquisition researchers referred to in his discussion (127). Hakuta concludes that linguistics preoccupied with Chomskyan goals may find prefabricated language uninteresting, but that

It is the researchers concerned with the generalized functions of language for whom the existence of prefabrications is of interest. ... Why do speakers of American English say a lifetime ago but not an extended time ago, or somewhere else but not sometime else? We could attempt to build formal rules that will generate the
allowable expressions, but Bolinger's interpretation seems persuasive: 'We have not heard it done. We have no memory of it' (Bolinger 1976:4; Hakuta 1986:129-30).

2.2.3. **Bilingualism.**

Code-switching and mixing are frequently attested phenomena of bilingual speech. Various attempts have been made recently to describe code-switching behavior according to grammatical rules, in particular, to identify points where code-switching can be predicted to occur, or, at least, to state where switching cannot occur.

Interestingly, predictions made on the basis of judgments of bilinguals (Lipski 1978) were shown not to hold up against actual data recorded from conversations (Pfaff 1979). According to the judgment of bilinguals, it should be difficult to switch inside a prepositional phrase (PP) and impossible to switch between an article and a noun (NP). However, Pfaff found in her taped corpus that PP internal switching was far more common than switching at PP boundaries. Moreover, switches between article and noun were more common than at NP boundaries. In general, actual data shows
that code-switching occurrences do not conform to the boundaries of grammatical structures (Appel and Muysken 1987:121-28; Romaine 1989:115-120).

 Constraints based on linearity (Poplack 1980, Woolford 1983, Sobin 1984) and dependency (Di Sciullo, Muysken and Singh 1986, etc.) have been proposed and reformulated. However, none of the proposals has yet been found to be consistent enough with empirical observation to be accepted by many scholars.

 An approach based on idiomaticity has yet to be formally proposed. The data collected in the present study indicate that a plausible argument might be made that at least some constraints on code-switching may be idiomatic. It appears, for instance, that switching can occur at the boundaries of idiomatic phrases and mixing can occur internally wherever a conventionalized phrase contains a filler slot, such as in the playful switch to Hawaii Creole English in *the sooner the mo betta.*
2.2.4. **Pidgin-creole studies.**

In a dissertation by Perlman on Hawaiian pidgin and creole English, Perlman excluded from his analysis all idiomatic phrases which contained the indefinite article *a* or *an*, such as *a little bit, a little while, quite a few, quite a lot*, etc. because they were exceptions to the HCE speakers' normal grammar which allowed only *wan* as the indefinite marker, or zero. Exclusion is the standard procedure in grammatical analysis for handling idiomatic expressions. However, inclusion of these phrases reveal a pattern of increasing use of *a* and *an* indefinite articles in idiomatic phrases before they enter the speaker's vocabulary in non-idiomatic constructions.

While this is only a single example, it demonstrates that idioms and formulas may be the way new grammatical forms enter a language in process of formation, and suggests that the study of idiomaticity in pidgins and creoles may be critical to understanding the process of grammar formation.
2.2.5. **Morphology and syntax.**

A defining characteristic of many, though not all, idiomatic expressions is their syntactic irregularity in that they are ungrammatical and cannot be generated by the regular rules of grammar. The idiomatic status of other phrases, which may be completely grammatical, is discernible by syntactic restrictions or by syntactic privileges which make them similar to single lexemes. Pawley (1985a) noted that phrases are "lexicalized" when their constituents are invariable (*Who do you think you are?* but not *Who do you think you will be?*), are transformationally defective (*The bucket was kicked by John, *Jennifer's hair was let down by her), are a derived compound with no underlying structure (*It's easier said than done*), use the definite article on first mention (*Call the doctor/the priest/the fire brigade*!), are sentential modifiers (*let's-take-no-more-of-this-nonsense feeling*), or are syntactically irregular (*long time no see*).

Reichstein (1973) systematized his view of "phrasal collocation" in this way, according to Strässler's summary:
... phrasal collocation [sic] have two different qualities: regularity in usage (=usuality) and irregularity in their structural semantic organisation. The latter has two components, idiomaticity and frozenness, where idiomaticity is the semantic irregularity, frozenness the syntactic one, i.e. the absence of regularity in paradigmatic and syntagmatic transformations (Strässler 1982:23-24).

Vachek (1976) noted the existence of "quotational compounds" citing never-to-be-forgotten, out-of-this-world, etc. He also listed examples of prepositions which govern whole clauses (He will go to where I was last year, I do not object to what you say, etc.), phrases which take derivational endings (the I-don't-know-ish expression on his face) or inflectional endings (the man I saw yesterday's hat), phrases to which affixes are attached (the ex-Prime Minister), and group genitives (Jacob and Esau's quarrel).²

Weinreich succinctly summed up the problem for theory:

... to a linguistics that is preoccupied with productivity in the strongest, Chomskyan sense, idiomaticity represents a basic theoretical stumbling block; for under the rubric of idiomaticity we are concerned with complex structures that can be recognized and analyzed but not naturally generated by any explicit machinery so far proposed" (1969:23).
Chafe identified four peculiarities of an idiom which make idiomaticity an "anomaly in the Chomskyan paradigm" (1968:109-112): first, its meaning is not an amalgamation of the meanings of its parts; second, idioms exhibit transformational deficiencies; third, some idioms are not syntactically well-formed; and fourth, the text-frequency of a literal counterpart to an idiom is extremely low compared to that of the idiom.

2.2.6. Semantics.

The main problem of idioms for semantics is that idioms often run counter to the basic assumptions of "the principle of compositionality, which states that the meaning of a complex expression is determined by the meaning of its constituents and the manner in which they are combined" (Ladusaw 1988:91). An apt definition of idiom is "a complex expression whose meaning cannot be derived from the meanings of its elements" (Weinreich 1969:26).

Semantics has also long struggled with the fact that the illogical is often acceptable in the natural use of language. Despite the fact
that literacy proponents since the beginning of Latin grammar schools have labored valiantly to establish logical principles as a basis for grammar, illogical usages have continued to "plague" even the writings of the most erudite. Sturtevant (1947:12) credited Huxley with the sentence, "No event is too extraordinary to be impossible" and Woodrow Wilson with the similarly illogical "No man is too big to decline the presidency." These may be performance errors, but other illogical forms which become idiomatic are not avoided except by the most inflexible; for instance, the following exemplify types of constructions which are widely used but have been considered illogical in some sense of the word by some grammarian or other: _This computer is faster than any computer I've seen_, _All the members present raised their arms_ (Quirk 1965:217), _I couldn't care less; I can't seem to find my keys; I don't believe he's coming; That long nose of his_ (Jespersen 1926:141), _I am friends with him_ (Shakespeare in _Henry IV_, cited by Wheeler 1965:9); _I have not the least objection in life to a rogue being hung_ (Thackeray, cited by Jespersen 1926:147), _I should have liked to go there_ (Quirk 1965:217).
Beyond these, many idiomatic expressions are identifiable by their logical violations, involving not the main terms of the proposition but the operators which relate them. This is especially true with idiomatic negation which sanctions the use of double negatives in many colloquial dialects of English and even demands it in dialects which require negative harmony. A less common kind of double negative is the sort of sentence in which a negative is neutralized (*I wouldn't be surprised if it didn't rain*. Quirk 1965:217). While logical errors in negation are extremely frequent in spontaneous speech, other "illogical" negatives are conventionally accepted, fully idiomatic expressions in English, such as the use of *don't believe* and *can't seem to* above, which are not typically regarded as errors.

Besides the examples above which exhibit a variety of syntactic and semantic structures, other "illogical" constructions which usually get by in English use involve a process of syntactic blending, which will be further discussed under "Performance errors" below (2.2.10).
2.2.7. Pragmatics.

The problem noted by pragmatics scholars which points to the necessity of recognizing idiomaticity in the study of pragmatics is that there are implicatures which are not motivated by conversational principles such as Grice proposed (1975), but are simply attached by convention to both single lexical items and multilexical units. Especially difficult to handle are those involving discourse and social deixis because syntax is sensitive to the implicatures attached to them in languages which use them (Levinson 1983:128; e.g., tu/vous, 129). Strässler, in attempting to handle idioms within a pragmatic theory, noted that in idioms, the information beyond that provided by truth-conditional semantics are conventional implicatures inherent in any idiom structure: "the difference between idioms and their literal counterparts again lies in the additional information conventionally implicated through idioms" (1982:117). So far, however, nothing further has been presented beyond labeling idiomatic meaning "conventional implicature".
Besides conventional implicatures, indirect speech acts also share some characteristics of idioms which has caused some scholars to propose an idiom theory for indirect speech acts (Sadock 1974 and 1975; Green 1975).

2.2.8. Psycholinguistic studies.

Some of the most recent empirical research in idiomaticity has been conducted in the psycholinguistics field by Cacciari and Tabossi (1988), Cacciari and Levorato (1989), Gibbs and others (Gibbs 1986 and 1987, Mueller and Gibbs 1987, Gibbs and Nayak 1989, etc.), Reagan (1987), Schraw, Trathen, Reynolds, and Lapan (1988), and Schweigert (1986). Through various experiments with children and adults, these scholars have investigated comprehension and production of many types of idiomatic expressions. Taken together, these experiments show that idioms, while sharing many characteristics of ordinary language, exhibit some special behaviors and require some special strategies for comprehension and production, in general, supporting suggestions offered by studies in neurolinguistics.
2.2.9. Neurolinguistic studies.

Perhaps the most concrete evidence of the psychological unity of some multilexical forms comes from neurolinguistic research. In a 1975 paper, Van Lancker extended the notion proposed by Jackson (1932) of an "automatic-propositional" dimension in language use, based on her intensive studies of clinical evidence that some kinds of language are more impaired in aphasia than others. Less impaired was language which is holophrastic, low in information content but high in affective content, highly frequent, memorized, semi-automatic and ritualized. These findings lend support to intuitive and popular perceptions that there is a great difference between propositional creativity and nonpropositional formulaic uses of language. As early as 1924, Jespersen speculated that free expressions involve a different kind of mental activity from formulas (19), but evidence has been documented only in recent years, thanks to increased interest in neurolinguistic research, that the brain may indeed have at least two different ways of producing and processing linguistic material (Heny 1985).
2.2.10. **Performance errors in speech.**

The field which provides some of the most convincing indirect evidence for the special status of idiomaticity in language processing is that which studies performance errors in speech. Both normal and abnormal speech have been of interest for the information they contribute to our understanding of the workings of our minds in language production and reception. Slips of the tongue (Aitchison 1987; Cutler 1982) and syntactic blending (Wheeler 1887, Curme 1931, Sturtevant 1947, Stemberger 1982, Cohen 1974 and 1987) are common problems of normal speakers, many of which suggest that multilexical units are stored in the mind just as are single lexical items, ready for retrieval as wholes when needed. For instance, when Prince Edward tongue-slips in an interview about his old school, "It is difficult to use *capital punishment* in any institution" (Aitchison 1987:18) in a context in which it is obvious he meant *corporal punishment*, a suggestion arises that the words *capital* and *corporal* are both strongly related to *punishment* in mental storage and are possibly even stored as whole phrases.
Hesitations, pauses and fillers have also been studied for the clues they give to the organization of words in the mind. We would like to know, for example, why a person might say, "I've er, I've er, I've er given it some thought ..." in which the hesitation vocalization falls between have and given, when phrase structure grammar would not consider I have to be a construction whereas have given would be. It would appear that phrases are grouped differently in usage or performance than they are in our current grammatical descriptions.

Examples of syntactic blends are copious in spoken language data. As early as 1887, Wheeler described the "syntactical phenomenon (contamination)" as "two equally good expressions for the same thought, but cast in entirely different form, occur simultaneously to the mind and are merged in one" (reprinted in 1965:8-9). Cohen collected almost 2,000 of them, mostly from his own alert listening, from 1970 to publication date (1987), including such examples as Why did you do that for? analyzed as a blending of Why did you do that? with What did you do that for?).
Blending can occur with a wide variety of structures. Some of the purportedly illogical idiomatic phrases above (2.1.6) may be analyzed as syntactic blendings.

2.2.11. Error studies in writing.

The study of error in writing has usually been focused on grammar. However, in her 1977 study of errors in the writings of "basic writers", Shaughnessy found "blurred patterns" and "consolidation" errors to be common. An example of a "blurred pattern", similar to what is called a "syntactic blend" in linguistics by analogy with lexical blend, is *If they decrease in a great number At least I can say is that I will have a college degree*, which blends *at least I can say that* with *the least I can say is* (49-50).

What is interesting about the consolidation errors (coordination, subordination, and juxtaposition) of these adult basic writers is that frequently used formulas and common sequences (underlined) are usually kept intact despite confusion with the
grammatical rules for combining sentences or parts of sentences, as in *But many colleges have night classes so you could have worked and gone to college also pay for your education although some other programs to help pay on some where you don’t pay or some where you don’t pay at all so you were lazy* (53). Although these errors may be described as syntactic errors, insights are lost when the idiomatic nature of the most common phrases are ignored.

2.2.12. **Spelling.**

A problem which makes the distinction between idiomatic phrases and lexical items somewhat arbitrary is the fact that many standard word boundaries have been established by accident of history. This is very obvious in the inventory of discourse markers where we write *none the less* and *in addition* as separate words but *nevertheless* and *furthermore* as single words and *inasmuch as* as a mixed phrase. The lexicalized nature of many phrasal idioms are sometimes made transparent in the innocent misspellings of children and new writers of English who tend to render multilexical units as
single words, such as *alot* and *lotta*. One student's paper I examined contained the word *ichather* (for *each other*). The syntactic behavior of many kinds of multilexical units does not differ from that of similar unilexical ones, and in this sense, according to Hockett, multilexical idioms may be considered a subset of idioms which also includes all single words. Hockett, in fact, felt that consistency required him to grant every morpheme idiomatic status (1958:172).

This sketchy survey through the literature of various fields which touch on idiomaticity was made to emphasize the point that we cannot continue to set aside and ignore idioms forever. However complex idiomaticity may seem to be at this point, there is so much of it in every language that much is being left out of language descriptions, not to speak of explanations, when we exclude idiomaticity altogether. Sorhus (1977) claimed that as much as 41% of her corpus of recorded speech was formulaic. If this number is indeed fairly representative, the effect of the neglect of idiomaticity by linguists will certainly be felt in language teaching and in the teaching of rhetoric and writing, where the inclusion of instruction in idiomaticity is rare and sporadic.
2.3. Past research in student writing.

2.3.1. Error studies.

Errors have always concerned teachers and we could trace error studies back to the dawn of formal education, but the systematic study of errors received renewed emphasis in this century in the sixties, when the assumption that "where languages differed most, learners would make frequent errors" led to the contrastive analysis of languages for second language learning. Viewed optimistically as the most promising solution to "first language interference" errors, contrastive analyses were made of all the most productive grammatical patterns of numerous pairs of languages. As usual, idioms were set aside as too idiosyncratic to be widely applicable.

The assumption that errors arise where grammatical rules of the two languages diverge the most eventually proved to be untenable (Saville 1971, Dulay and Burt 1973, and Schacter 1974) and contrastive analysis has lost its former impetus. Nevertheless,
important contributions have continued to be made in the study of errors from cognitive orientations. Kroll and Schafer saw errors "as clues to the linguistic and cognitive processes that function unobserved ... as windows into the mind" (1978:142-43). They recommended use of error analysis as a diagnostic tool upon which writing instruction could be based (244). The main object of interest to these error researchers continued to be violations of grammatical rules. Shaughnessy's Errors and Expectations (1977) was the first which paid considerable attention to error phenomena which were not strictly classifiable as grammar in usual ways, but which included common phrasal patterns ("blurred patterns" and "consolidation" errors). To date, the study of errors in idiomaticity has barely begun.

2.3.2. Written-spoken contrasts.

In reaction to the centuries-old bias in favor of written language, modern descriptive linguistics in the early years insisted on the primacy of spoken language, claiming that written language is a mere record of the spoken. Recent recognition by a minority of
scholars that spoken and written language are substantially different modes which both have their own appropriate uses marks a significant step in the direction of recognizing differential idiomaticity for each mode.

Although the volume of literature on grammatical contrasts is considerable, what the contrasts between written and spoken language mean for the study of idiomaticity overall has not yet been tackled. A few works which have paid substantial attention to idioms and formulas with reference to written and spoken language differences go back to the earliest years of this century. The sources of our most ordinary phrasal expressions, Smith (1925) noted, are what were once the idioms of substandard speech. Indeed, Roberts claimed, grammar itself is fossil idiom (1944:300), being the conventionalization of what was initially idiom in speech to the extent that it becomes acceptable even in writing. Barber (1964) discussed many idiomatically acceptable though literarily condemned phenomena in terms of language change, since forms which are in the process of change are judged "common errors" before the change is widely accepted. Chafe concluded a summary of research on "The
relation between written and spoken language" (Chafe and Tannen 1987) with a reference to Pawley and Syder (1983) which presents the thesis that natural selection operates in language. Pawley and Syder catalogued a number of distinguishing features from conversational language data and suggested that "the many differences represent adaptive variation to the very different conditions under which conversation and literary production typically take place" (Chafe and Tannen 1987:390). The differences catalogued from conversation (Pawley and Syder 1983a:554-56) are phrases and expressions we generally term colloquial or idiomatic, unacceptable by standards of edited English, but common in the casual speech of even well-educated native speakers of English.

2.3.3. Basic writing.

Various sources of writing problems, or more specifically, errors, of basic writing students were vigorously sought, especially during the seventies--linguistic, social, developmental, educational, etc. But in every case, the writer was assumed to be deficient. The irregularity of the language itself was only rarely recognized as a
source of difficulty for students, and even when the demands of "print code" or written language were identified, they included only such problems as were deemed requirements of grammatical rules. The idiosyncratic demands of idiomaticity are generally overlooked.

2.3.4. **Black English.**

Here, again, great advances have been made in the recognition of differences between Black English grammar and standard English, but little attention has yet been focused on distinctions between Black English idiomaticity and what is properly idiomatic in standard English, or, in writing for academic purposes.

2.3.5. **English as a Second Language.**

Second language learners have traditionally been seen as having the learning of grammar and an adequate portion of a simplex lexicon of the individual words of the new language as their goal. A short-lived period of phrase-pattern teaching, based on the
most productive grammatical patterns and excluding the idiosyncratic idioms, was ultimately given up as unproductive. Very little systematic teaching of idiomatic patterns and formulas was ever attempted except for common conversational routines of greetings, leave-taking, classroom procedures, asking of directions in tourist-oriented courses, and the like. The formal presentation of idioms is usually delayed until well after the student is considerably advanced in English, although idioms inevitably arise frequently in informal discussions from the very beginning. Most recently, student demands have led to the production of bilingual dictionaries in many languages which list English phrasal noun compounds and phrasal verbs with idiomatic meanings. Idiom dictionaries are also much more available now than in the past but only include a fraction of the idioms a student is likely to encounter through daily interaction.

2.3.6. Hawaii Creole English.

Since the writing of grammars for creoles is a fairly recent enterprise, idiomatic expressions have usually been set aside while
grammar writers focus on the most regular systems in the language. In fact, because less central structures have not as yet been described thoroughly, the differences between grammar and idiom have not been clearly identified or dealt with explicitly.

Rapidly changing and unwritten languages present a particular challenge for idiomaticity because the boundaries of idiomaticity and grammar are apt to be especially difficult to define in these cases. When is an idiom chunk used productively enough to merit promotion to grammatical status? Roberts claimed, "The best field of study for the idiomatologist lies in the subtle frontiers between almost identical languages and cultures. For experience shows that the idiom is the first element to change, when two languages begin to diverge from each other" (1944:306). As suggested by Perlman's study of the articles referred to in section 2.2.4, it might also be so that idioms are the first element to change when two languages begin to converge, which would make it profitable to include idiomaticity in the study of the decreolization of Hawaii Creole English.
2.4. Potential benefits of the study of idiomaticity.

Sentences written by ESL- and HCE-speaking students enrolled in remedial writing classes have frequently been noted by writing instructors to contain many errors not easily described or explained grammatically. Attempts to account for some of the non-grammatical errors in terms of idiomaticity seem promising.

ESL and HCE speakers impressionistically differ in the use of formulaic and idiomatic expressions, but no systematic study has yet been attempted to account for the differences observed. Information gained from such work would potentially contribute:

- to our understanding of the role of idiomaticity in language use,
- to a meaningful identification and classification of linguistic signs, including multiword signs,
- to the formation of a systematic theory of idiomaticity,
- to our knowledge of how human beings think and process language in the brain, and
- to our general knowledge of how languages work.
In addition, the work of analyzing students' writings in terms of idiomaticity has practical implications for the teaching of writing to native speakers faced with unfamiliar literary standards, to speakers of widely divergent dialects of English, and to non-native speakers of English who have the complex burden of learning the idioms of colloquial as well as of written English.
CHAPTER 3
IDIOMATICITY

3.1. The development of theories of idiomaticity.

Whether there were theories of idiomaticity before this century depends on what may be considered to be "theories". In the "rigorous and explicit" sense in which linguistics has used the term in recent years, one would be hard-pressed to claim any such theories of idiomaticity were ever formulated, and, in fact, not much is known about even the most informal kind.

Following a brief survey of definitions, this chapter will begin with a description of a collection by Smith (1925) and a treatise on idioms by Roberts (1944). Then a leap is made, over some sporadic and mostly pedagogical explorations, into the late fifties from which point the theoretical development begins to appear somewhat cumulative. To date, no attempt at developing a rigorous, formal theory can be said to have been successful or widely accepted. At best, scholars of idiomaticity have succeeded in clarifying the
problems which idioms and their like pose for a general theory of language. It is generally in this loose sense that the notion of "theory of idiomaticity" is being used throughout this chapter. The number of scholars who are expressing concern for finding a way to handle idiomaticity in a more systematic way appears to be growing.3

A summary of the development of theories of idiomaticity from the late fifties to the early seventies is contained in a dissertation by Strässler (1982), who appears in turn to have been guided by Weinreich's (1969) and Makkai's discussions (1972). Briefly, Strässler reviews the leading Soviet phraseologists, Vinogradov (1946, 1947, 1953, and 1954), Amosova (1963 and 1966) Reichstein (1974 and 1975) and Mel'čuk (1960). He then turns to the non-generative approaches of Hockett (1958) and the tagmemicist, Healey (1968). He mentions Bar-Hillel's "Idioms" (1955) and Malkiel's "Irreversible Binomials" (1959). Finally, Strässler reviews the attempts to handle idiomaticity either within or with reference to the transformational-generative framework, Katz and Postal (1963), Weinreich (1969), Chafe (1968), Fraser (1970), and Newmeyer (1972).
Earlier, Makkai, from whose work Strässler has drawn, had also summed up in somewhat greater detail the works of approximately the same period (beginning with Hockett) in his 1965 dissertation, but he later also included, in the 1972 publication of his dissertation, Pike (1967) and Balint (1969) among the tagmemecists and McCawley's (Quang 1971) refutation of Fraser (1970) with the generativists. Summaries and comments on these works will be repeated in section 3.6 below for convenience.

Going back still further, many references to English idioms were made by writers on the English language around the turn of the century, according to Logan Pearsall Smith (1925), and a number of collections, mostly focused on etymologies, were written, but apparently little attempt was made to develop any kind of theoretical framework for idiomaticity. The one exception is perhaps Murat H. Roberts (1944), who consciously attempted to establish a theoretical foundation for the study of idioms in a broadly philosophical paper, "The Science of Idiom: a Method of Inquiry Into the Cognitive Design of Language." Theories may vary greatly in rigor and explicitness, but since the least that might be expected of a
theoretical discussion is that it attempt to define the object of investigation, some standard definitions of recent periods will be briefly surveyed below. Before going into definitions, however, some speculation will be indulged in as to why idioms seem to be so vaguely defined.

3.2. Avoidance of idioms.

The usual practice of linguists in the past has been to avoid idioms altogether. At least since Humboldt noted that language makes infinite use of finite means, linguists have considered it their task to discover the most general patterns and rules for linguistic structures, presumably assuming at the same time that exceptional forms may be accounted for after that task is accomplished. Since idioms were generally viewed as exceptional rather than pervasive phenomena of language, such notables as Bloomfield (1926 and 1933), Harris (1951), Chomsky (1957 and 1965), Martinet (1962), and Lyons (1968) conspicuously avoided idioms.
Moreover, inasmuch as idioms are basically semantic entities, Bloomfield had earlier indirectly recommended the avoidance of idioms by recommending avoidance of the study of semantics, claiming it entailed the pursuit of the sum-total of all knowledge (1933:74-75). Roberts, in fact, viewed everything in natural language as idiomatic, which would have indeed made the "science of idioms" little less than epistemology (304-5). While some scholars eventually returned to writing principally about semantics (Ullman 1957 and 1962, Ziff 1960, Lyons 1963 and 1977), they continued to avoid idioms. Even in presenting a lengthy study on "irreversible binomials", one type of phrasal unit which includes what would normally be termed idioms, Malkiel insisted on "steer[ing] clear of any reference to the ill-defined category of 'idioms' or phraseological formulas" (1959:115).

Probably the main reason idioms have been avoided in this century is that idioms do not fit neatly into the framework that structural linguistics has created for the description of language. That framework assumes that larger segments of language are comprised of ever smaller segments, the smallest segment being
some kind of "prime" (Householder 1959) which can be analyzed no further. It would be convenient if the form-meaning primes were simple units like coins, a certain form connected with a certain value, which could be added together to form a structure whose meaning is a composite of the individual meanings of the primes plus whatever the rules for combining them added.

But idioms, by comprising complexes of forms which act like simple forms in some ways and like analyzable syntactic structures in others, greatly complicate description. They have unitary meanings and often fill category slots in invariant forms as simple lexical items do. Yet, like ordinary syntactic structures, they have inflectional and derivational morphemes attached to constituents in the usual way (*kicked the bucket*) instead of being attached to the end as would be expected if the phrase were treated as an unanalyzable lexical item (*kick the bucketed*). To make matters worse, many idioms are also syntactically deviant (*by and large, beyond compare, far be it from me*, etc.), and are "frozen", or invariable. Worst of all, their conventional meanings are often not deducible from their parts.
From the vantage point of language use also, idioms are no less recalcitrant, since the basic assumption about language use has been that users have a store of minimal units like morphemes in a lexicon in the mind which comprise the raw materials from which utterances are created anew as needed. The fact that idioms do not add up to what they are supposed to mean by this formulation is one major problem for theories of language use. Another is that compared to their literal counterparts the overwhelming asymmetry in text frequency of idiomatic phrases is not accounted for by this view. A third is that real evidence from psycholinguistic and neurological research as well as from experiences in first and second language learning contradict these assumptions. Knowing a language has traditionally meant to have a store of the morphemes of the language in mind and knowing the rules by which those morphemes are combined into meaningful sentences in that particular language. But, again, the prevalence of formulaic and idiomatic expressions in every language calls this assumption into question.

Yet another fact about idioms which apparently discourages potential investigators is that they are so heterogeneous, lexically,
syntactically, and semantically, displaying an apparent continuum of regularity on various dimensions. Furthermore, there are so many different genres of idioms and idiom-like phrasal expressions. All these complexities, of which a systematic description is attempted in Chapter 4, no doubt have contributed to the avoidance of idioms in the formulation of theories of language and language use in descriptive, structural, and generative linguistics.

3.3. Common definitions of "idiom".

3.3.1. Dictionary definitions.

Although the use of idioms and formulas has always been a part of language as far as we know, examples of the use of the word "idiom" in the Oxford English Dictionary (OED) only begin in the late 1500's. The earliest use cited is in the form *idioma* in 1575: Gascoigne's *"So woulde I wish you to frame all sentences in their mother phrase, and proper *Idioma*."* Donne used the form *idioms* in 1628, and the singular *idiom* was used by Davies, Dryden, and Godfrey in 1662, 1666, and 1674, respectively. Contexts of the examples
display a range of meaning as summarized from the OED below.

Very early on, the suggestion that idioms characterized the disfavored colloquial, rather than the cultivated and academically approved written styles of the language, is contained in the examples. The lexicographer, Samuel Johnson, in discussing idiomatic innovations of "illiterate writers" in the preface to his dictionary (1755), expressed regret that "colloquial licentiousness" sullied the grammatical purity of English ([xii]).

The OED offers, among others, the following definitions of "idiom" which reflect its derivation from the Greek *idios*, meaning "own, private, peculiar":

1a. The form of speech peculiar or proper to a people or country; own language or tongue.
1b. In narrower sense: The variety of a language which is peculiar to a limited district or class of people; dialect.
2. The specific character, property or genius of any language; the manner of expression which is natural or peculiar to it.
3a. A form of expression, grammatical construction, phrase etc., peculiar to a language; a peculiarity of phraseology approved by the usage of a language, and often having a significance other than its grammatical or logical one ("idiom", 1989).
Definition 3a defines most concretely the phenomena which result in the perception of the existence of 1a, b, and 2; that is, a language or dialect is a particular "idiom", as in 1a and b, and has its idiomatic character, as in 2, to the extent that it exhibits idiomatic expressions, constructions, and phrases, as in 3a. The term "often" in 3a is significant for this study, since this study is not limited to ungrammatical or illogical idioms.

A quick survey of a number of available dictionaries revealed entries similar to the OED definitions. Exactly what characterizes idioms and sets them apart from non-idiomatic phrases is hard to know despite these definitions, taken singly or as a composite.

3.3.2. Traditional grammar definitions.

Traditional grammarians, here meaning pre-structuralist or prescriptive grammarians, tended to look upon idioms as did Johnson, with distrust. They tended to view idioms as phrases which violated the laws of language in undesirable ways, either grammatical or logical. As Smith (1925) put it:
... Of these, the first kind, the ungrammatical phrases made acceptable by usage, are the most obvious, and in any old-fashioned book on good English will be found lists of these wild creatures of talk, nailed up, like noxious birds and vermin, by the purists and preservers of our speech ... grammar in this sense is the natural enemy of idiom, and continually preys upon it (178).

Although grammars reflecting the prescriptivist attitude toward idioms are now hard to come by, the vehemence with which defenders of the use of idiomatic language present their arguments attests to the earlier prevalence of the opposite view. We can infer from these arguments that the term usually designated multilexical entities (since the term was usually applied to phrases) which did not conform to prevailing standards of grammaticality and/or logic. Smith used "idiom" in this "narrower sense, meaning the idiosyncrasies of our language, and, above all, those phrases which are verbal anomalies, which transgress ... either the laws of grammar or the laws of logic" (168).

The noncompositionality of idioms as a defining feature was early noted by Henry Sweet (1889): "the meaning of each idiom is an isolated fact which cannot be inferred from the meaning of the words
of which the idiom is made up" (139). This feature has been the most generally accepted criteria for idiomaticity in this century, and is responsible for the early placement of the study of idioms in the domain of lexicology since it suggested that an idiom was only incidentally made up of more than one word, but that, having unpredictable meaning, it was otherwise like a word.

An early defender of idioms, Roberts (1944), declared:

The idiom ... is the quintessence of a language, since this idioma supplies the combinatorial government of the tongue in question. ... Idiom, in this generic sense of idioma, includes all the common definitions of the word. Whether the idiom is a figure of speech, an anomaly of grammar, or merely a group of words carrying as a whole a meaning not to be gathered from its component parts separately considered, the comprehensive definition will hold (300).

3.3.3. Linguistic definitions.

3.3.3.1. Structuralist attempt.

Hockett was perhaps the first linguist to attempt a formal definition of "idiom". Hockett's idiom was "Any Y (Y= 'any
grammatical form the meaning of which is not deducible from its structure') in any occurrence in which it is not a constituent of a larger Y" (1958:172). The consequences of this definition are complex. For one, Hockett admits:

... If we are to be consistent in our use of the definition, we are forced also to grant every morpheme idiomatic status, save when it is occurring as a constituent of a larger idiom, since a morpheme has no structure from which its meaning could be deduced (172).

Hockett's "idioms" thus are different from those as commonly defined in that they are not required to be multilexical. Other characteristics usually attributed to multilexical idioms, however, are shared by monolexical idioms, such as that they have non-idiomatic counterparts (e.g., white paper) they may share exactly the same form as another idiom (e.g., Statue of Liberty, the famous statue or a football play) (172), they need to be entered into lexicons separately, they have to be learned as a whole, and they are the raw materials out of which utterances are built (173).

How far Hockett's definition diverged from the prevailing ones is made more obvious by comparing it with a series of definitions
quoted or paraphrased by Malkiel (1959) when he advised readers to steer clear of "the ill-defined category of 'idioms' or phraseological formulas":

... These have been variously spoken of as sequences yielding imperfectly to routine grammatical analysis, as passages strikingly rebellious to literal translation ..., as semi-autonomous pieces of congealed syntax ... as word-groups whose aggregate meaning cannot be fully predicted even from thorough knowledge of each ingredient ... and, in stylistic or esthetic terms, as cliches, i.e., as combinations once suffused with fresh metaphoric vigor, but gradually worn thin by dint of use (115).

All of these definitions repeated by Malkiel require idioms to be multilexical, but focus on different aspects of idioms--syntactic irregularity, semantic irregularity, or stylistic characteristics.

3.3.3.2. Tagmemic views.

Pike called his phrasal unit a "hypermorpheme" and described it as composed of "a specific sequence of two or more specific morphemes" (1967:426). What we would normally consider to be idiomatic expressions are apparently a subset of hypermorphemes:

The semantic component of a hypermorpheme may include the (predictable) sum of the meanings of its
comprising morphemes ... but it may sometimes also contain an additional or overriding hypermorphemic meaning (or idiomatic meaning) which is not predictable from its included parts (427).

He compares *to step on the gas* which has a hypermorphemic special meaning with a literal hypermorpheme *to step on a worm*.

Healey's (1968) definition excludes monomorphemic lexemes as idioms, but his list, which is reproduced in the Appendix to Makkai's dissertation, includes compound lexemes as well as phrasal and clause units.

Balint (1969) defines idioms as "*a phraseological unit whose meaning cannot be arrived at from the separate meanings of the constituents of the unit*" (3). Compounds are thus excluded since they are not phrases. His criteria is basically noncompositionality.

3.3.3.3. Transformational-generative accounts.

Katz and Postal (1963) accepted noncompositionality as the main criterion for idioms: "The essential feature of an idiom is that
its full meaning, and more generally the meaning of any sentence containing an idiomatic stretch, is not a compositional function of the meanings of the idiom's elementary grammatical parts." They distinguished "lexical idioms"--"idioms that are syntactically dominated by one of the lowest syntactic categories, i.e. noun, verb, adjective, etc." from "phrase idioms"--"those whose syntactic structure is such that no single level syntactic category dominates them" (275-76).

Weinreich (1969) began his discussion with a definition of idiom which he presented as "the common understanding of an idiom as a complex expression whose meaning cannot be derived from the meanings of its elements" (26). He later developed more fully a stable terminology which included the idiom as a subset of phraseological unit, which is

... any expression in which at least one constituent is polysemous, and in which a selection of a subsense is determined by the verbal context ... A phraseological unit that involves at least two polysemous constituents, and in which there is a reciprocal contextual selection of subsenses, will be called an idiom.
This definition excludes phrases with unique constituents from the set of idioms—"luke warm, runcible spoon, spic[k] and span, kith and kin, hem and haw, cockles of the heart and so on"—because "ambiguity is an essential characteristic of true idioms" (44). He calls the excluded set "stable collocations", following Mel'čuk (1960).

3.3.3.4. Stratificational contribution.

Makkai does not actually give a short, clear definition of idiom in his 1965 dissertation (published in 1972), but he does provide definitions for his "lexemic idiom" and "sememic idiom" which are placed in two separate "idiomaticity areas" according to his stratified view of language. "Any polylexonic lexeme which is made up of more than one minimal free form or word ... each lexon of which can occur in other environments as the realization of a monolexonic lexeme is a lexemic idiom" (122). Like in Weinreich's taxonomy, phrases containing "cranberry morphs" such as spick-and-span are thus excluded from true idioms, but compounds, whether written as phrases (White House) or as single words (blackbird), are included. "A polylexemic construction whose aggregate literal meaning derived
from its constituent lexemes functions additionally as the realization of an unpredictable sememic network is a sememic idiom” (128).

3.3.3.5. Post-generative departures.

Chafe, in recognizing "Idiomaticity As an Anomaly in the Chomskyan Paradigm" (1968), proposed that semantics rather than syntax needs to be the central and initiating component of a generative grammar. He is able to accommodate idioms in his account of language using this tactic without making them exceptions. His semantics-based definition of idiom is presented in these terms:

**Idioms**, then, constitute a subset of semantic units characterized by the fact that they are subject to literalization rules ... literalization can be looked on as a process rather like symbolization; we might say roughly that **an idiom is a semantic unit which is first symbolized by other semantic units before these go on to be symbolized by phonetic units.** From a Chomskyan point of view, idioms seem to be syntactic arrangements whose meanings cannot be assembled from the meanings of their parts. If we reverse this view to begin with semantics rather than syntax, we can say something much more plausible: that idioms are semantic units like other semantic units, but that they require conversion into arrangements of other semantic units before they are further encoded into sound (121-22).
Chafe's definition includes grammatical as well as ungrammatical idioms, and accommodates transformational deficiencies naturally.

When examining the work of the late seventies and early eighties, it becomes more difficult to know whether new terms have arisen to describe the same phenomena as were subsumed under "idiom" earlier, or whether the object of investigation itself has been changed. New terminology began to be applied to idiom-like phenomena, such as, "formulas" (Pawley and Syder 1983a), "prefabricated language" (Hakuta 1986), "routines" (Hymes 1962, Coulmas 1979 and 1981), "formulaic expressions" (Tannen and Özbek 1981), "nonpropositional speech" (Van Lancker 1975), etc., while discussion continued on the description of such phrases as kick the bucket, red herring, and thank you. Some researchers, on the other hand, began to broaden the field to include many types of phrasal units not formerly considered idioms, such as those Mel'čuk (1960) referred to as "stability of collocation" and those containing "cranberry morphs" which Makkai (1972) called "pseudo-idioms". Others emphasized discourse markers or social and conversational
deixis. It became more difficult to see what all the different types of phrases had in common, if anything.

Pawley and Syder undertook to systematize the field, taking as their object the kind of language that is required "when what can appropriately be said (the sense or pragmatic force of the message) and how it is said (the form) are more or less rigidly set by social convention" (Pawley1992:22). Pawley and Syder (1983b) argued:

fluent and idiomatic control of a language rests to a considerable extent on knowledge of a body of 'sentence stems' which are 'institutionalized' or 'lexicalized'. A lexicalized sentence stem is a unit of clause length or longer whose grammatical form and lexical content is wholly or large[ly] fixed (191).4

Later, Pawley (1985b) defined speech formula as "a conventional pairing of a particular formal construction with a particular conventional idea or idea class" (88). The relationship of idiom to formula is that "All idioms are formulas but not all formulas are idioms (in the strict sense of a construction with an unpredictable meaning or irregular form); most are not idioms" (89). For Pawley, since idioms are just two particular kinds of formulas, namely, those which are semantically noncompositional or those
which are syntactically nonconforming, speaking "idiomatically" does not necessarily require use of "idioms", but does demand use of formulas of various types.

Fillmore, Kay and O'Conner (1988), in "Regularity and Idiomaticity in Grammatical Constructions: the Case of Let Alone", retain the term idiom for the constructs which Pawley renamed "formulas". It appears that for them anything idiomatic would be an idiom, of which there are many greatly varying types:

We think of a locution or manner of speaking as idiomatic if it is assigned an interpretation by the speech community but if somebody who merely knew the grammar and the vocabulary of the language could not, by virtue of that knowledge alone, know (i) how to say it, or (ii) what it means, or (iii) whether it is a conventional thing to say. Put differently, an idiomatic expression or construction is something a language user could fail to know while knowing everything else in the language (504).

They accept Makkai's (1972) distinction between decoding (undecodable by hearers unfamiliar with the idiom) and encoding (unencodable only) idioms, including among examples of encoding-only idioms answer the door, wide awake, and bright red. They
refer to both types as "idioms", whereas Pawley would have called
the latter "formulas" rather than "idioms". They also distinguish
between grammatical and ungrammatical idioms, between
substantive and formal idioms, and between idioms with and without
pragmatic point (505-6).

The problems of defining idioms and idiom-like constructions
will be discussed further in Chapter 4. Čermák (1988), in discussing
the search for criteria by which forms involved in idiomaticity are
termed and identified, pointed out that questions as to the autonomy,
specificity and adequacy of criteria need to be addressed.

3.4. Collectors and etymologists.

In this century, the earliest students of idioms were primarily
collectors interested in gathering idioms and seeking their
etymologies. However, one copious collection published in 1925,
which was the decades-long effort of Logan Pearsall Smith, went
considerably beyond listing.
Smith was not able to find any "complete collection" or "exhaustive treatment" of idioms and found his most useful information in a chapter on idioms by John Earle (1890), Abbott's *Shakespearian [sic] Grammar*, and several published dictionaries of English phrases and idioms, none of which were "at all complete" (168f). Smith, therefore, turned to the Oxford English Dictionary as his main source of information on idioms to produce what was probably the first extensive classification and discussion of the idioms of English, *Words and Idioms* (1925).

3.5. Early theorists.

3.5.1. Jespersen 1924.

From the evidence of his frequent references to Jespersen's works, Smith was apparently greatly stimulated in his theoretical notions by Jespersen, who could perhaps be considered the earliest theorist on idiomaticity in this century although he did not pursue the formalization of his theories. First, he recognized a basic distinction between "free expressions" and "formulas" (1924:18-24)
which "pervades all parts of the grammar" and which he speculated required different kinds of mental activity (19). Free expressions were seen as composed of words inserted as needed into a few basic sentence frames. Formulas were not constructed in the same way but were "felt and handled as a unit, which may often mean something quite different from the meaning of the component words taken separately" (18-19).

Jespersen seemed to be intensely concerned with the idiosyncracies of grammar and discussed at length and with copious examples the conflicts which occur in any real language between "the speaking community [and those] who very often prefer logical consistency to ease and naturalness" (238).

3.5.2. Smith 1925.

"English Idioms" is the longest essay of several contained in Smith's Words and Idioms (1925, reprinted in 1943) and contains the greatest number of examples of idioms. In it, Smith begged out of the theoretical task in need of attention:
An adequate study of our idioms, based on all this new material [in the OED], will no doubt be made one day by some one much more competent than I am to deal with the subject. In the meantime, however, there may be a place for a preliminary sketch, written by a reader of this Dictionary, who has made for himself a large collection of idioms, and who has arrived at certain conclusions about the idiomatic element in our speech, and its place in literary diction, which seem to him not devoid of interest (1943:169).

A sampling of his collection showing how he classified various phrases will be adequate to demonstrate his procedure. Smith begins with a mention of the idiosyncratic nature of prepositional usage in English, then discusses very briefly "terse adverbial phrases" (by chance, for once, in general, of course, etc.) (171), "phrasal verbs" ("keep down, set up, put through, and thousands of others", some with two prepositions, to take up with, to make up to, to get on with, to go in for, etc.) (172), "phrasal collocations, in which two words are habitually used together for the sake of emphasis" (at beck and call, to cut and run, down and out, dust and ashes, enough and to spare, etc.) (173).5

His subclassification of this large set includes "repetitions of the same word" (again and again, by and by, more and more, neck and
neck, etc.), those whose "emphasis is helped by alliteration (bag and baggage, to chop and change, humming and hawing, etc.), those which rhyme (art and part, fair and square, high and dry, etc.) (174), "contrast of two alternatives" (heads or tails, hit or miss, a jot or tittle, etc.), "inclusive phrases" (ever and anon, fast and loose, first and last, give and take, from head to heels, etc.), comparisons (as bold as brass, as cool as a cucumber, etc.) (175), and certain proverbs and proverbial phrases (all's well that ends well, better late than never, extremes meet, familiarity breeds contempt, etc.) (176).6

Thus he proceeded through an astounding quantity of material (some of the phrases unfamiliar today), grouping together similar phrases by various criteria. Along the way, he often discusses issues current in his time, such as the conflicting demands of prescriptive versus descriptive grammars and of colloquial versus literary language, many of which continue to be problems in our day.

Most of the remainder of his examples are classified according to source, such as the sea, soldiers, the chase, animals, houses and buildings, etc., the most prolific source being the human body.
The thesis Smith presents is clearly that idioms, far from being the undesirable violations some prescriptive grammarians would have us eradicate, are an inevitable and necessary part of language, "little sparks of life and energy in our speech ... diction deprived of idiom ... becomes tasteless, dull, insipid" (276).

Since Smith's purpose was to defend idiom in the best uses to which it is put in English, it was not his contribution to systematize the study of idiomaticity, despite his early interest in idioms and his recognition of the necessity of doing so. Nevertheless, his work is a valuable resource for its quantity of examples and a fitting starting point because of his discriminating selection, the insights displayed by his classification, and the appropriateness of his well-considered nomenclature.

3.5.3. Roberts 1944.

In 1988, Hopper advanced the notion that grammar is created by discourse. A similar idea was introduced years earlier by Murat H. Roberts in his 1944 publication "The Science of Idiom", which has
the surprisingly contemporary-sounding subtitle, "A Method of Inquiry Into the Cognitive Design of Language," especially in view of Chomsky's claim that linguistics is a branch of cognitive psychology.

In Roberts' scheme of things, "discourse, the *logos* of meaning to be expressed, stands in polar relation to *language*, the psychophysical mechanism which does the expressing through its dynamic aspect of utterance (*la parole*)" (295). Structuralists might prefer to call this *discourse* the 'content' or 'meaning' as opposed to the 'form' of an expression. "Language and discourse are the opposite hemispheres which conjoin to produce the complete sphere of communication" (299). Roberts goes on to say, "Idiom belongs primarily to discourse, grammar to language. But since idiom has created language, which is its subconscious deposit, idiom has created grammar. Grammar, in his [i.e., Roberts'] view is fossil idiom" (300).

Roberts reasons that all idioms originate as personal innovations by individuals (200) and that each idiom is a mental monument of history (304).
While the charge of scientism may be leveled against Roberts for his grandiose scheme for subsuming all of the study of natural language under the "Science of Idiom", Roberts' recognition of the importance of studying the idiom as a source of language change is a significant contribution worth following up.

3.6. Recent theorists.


Though little cited in linguistics, Bakhtin was a linguist and literary scholar of considerable reputation in the Soviet Union, having published several extended treatises on language in which he argued "for a method of analysis that recognizes the primacy of utterance-in-context and shows the dynamic relationship between system and utterance" (Bizzell and Herzberg 1990:926). In his 1929 publication "Marxism and the philosophy of language" (originally attributed to Voloshinov), Bakhtin complained that "the structure of
a whole utterance is something linguistics leaves to ... rhetoric and poetics" (Bizzell and Herzberg 1990:926 and 940-941). On the other hand, rhetoric has recognized only a very few genres and tropes, and does not employ the means of analysis available to linguistics and semantics. Bakhtin's attempt was to unite linguistics and rhetoric so language use could be systematically accounted for.

Because Bakhtin is not well-known in linguistics, liberal portions from translations of his writings will be quoted, not only to introduce his contributions to idiomatology, but also to show that many ideas which are often considered to be of recent origin, especially in pragmatics, discourse analysis, and conversation analysis, were actually propounded in the early years of this century.

Bakhtin was centrally concerned with "speech genres" which can be related to the notions of idiomaticity and discourse genre, and which obviously involve idiomatic expressions and formulaic routines.

... Only when social custom and circumstances have fixed and stabilized certain forms in behavioral interchange to some appreciable degree, can one speak of specific types
of structure in genres of behavioral speech ... Village sewing circles, urban carousels, workers' lunchtime chats, etc., will all have their own types (Bizzell and Herzberg 1990:940).

Years later, Bakhtin wrote "The Problem of Speech Genres" (1953) in which he explained, "Each separate utterance is individual, of course, but each sphere in which language is used develops its own relatively stable types of these utterances. These we may call speech genres" (945). Speech genres are given to speakers by the speech community (956-57).

... A large number of genres that are widespread in everyday life are so standard that the speaker's individual speech will is manifested only in its choice of a particular genre, and, perhaps, in its expressive intonation. Such, for example, are the various everyday genres of greetings, farewells, congratulations, all kinds of wishes, information about health, business, and so forth (957).

Bakhtin proposed that the utterance, as opposed to the sentence, should be taken as the real unit of speech communication, and a systematic description of language use should proceed from analyses of utterances. He objected to Saussure's (1916) definition of the utterance (la parole) as an "individual act" because Saussure ignores the fact that an individual's freedom to select forms and
combinations of forms are strictly constrained by the conventions of language use (958).

... Thus, a speaker is given not only mandatory forms of the national language [grammar and lexicon] ... but also forms of utterances that are mandatory, that is, speech genres. The latter are just as necessary for mutual understanding as are forms of language. Speech genres ... have a normative significance for the speaking individuum, and they are not created by him but are given to him. Therefore, the single utterance, with all its individuality and creativity, can in no way be regarded as a completely free combination of forms of language, as is supposed, for example, by Saussure [and others] ... They see in the utterance only an individual combination of purely linguistic (lexical and grammatical) forms, and they neither uncover nor study any of the other normative forms the utterance acquires in practice (958).

From this, idiomatic material would seem to be a matter of performance for Bakhtin if we consider utterance a real rather than abstract construct. However, the utterance, both written and spoken, must be conceived of as an abstraction and be a part of competence for its normative influence to be transmitted. Bakhtin saw the investigation of utterances in context as essential to any theory of meaning.
Unfortunately, because access to English translations is limited, complete reliance on Weinreich's (1969) and Strässler's (1982) summaries and commentaries for the information about Soviet phraseological studies presented hereafter could not be avoided. Nevertheless, awareness of this body of works is necessary for a proper perspective on the course taken by American scholars who followed Weinreich's lead.


Weinreich (1969), in his brief survey of the study of idioms and idiomaticity in America and Europe, noted the relative vigor with which scholars in the Soviet Union pursued the investigation of "phraseological units" or "stable collocations" within a rapidly developing field of lexicology. A bibliography by Babkin (1964) contains over 900 Soviet works on phraseology from 1918 to 1961, over a hundred dealing with English alone, according to Weinreich (1969:25). Strässler later noted that Häusermann (1977) claimed to have counted 7,000 items in 1974 (21).
Most of the works apparently used the theoretical paradigm and classification scheme devised by Vinogradov (1946 and 1947) (Weinreich 1969:25, f6), based on de Saussure's (1916) concept of "motivatedness". In this framework, according to Strässler (1982) who cites Vinogradov's 1953 and 1954 publications (157), phraseological units fall into three classes: (1) completely unmotivated phraseological collocations, (2) completely motivated but in metaphorical reinterpretation unmotivated phraseological units, and (3) completely motivated phraseological combinations, the components of which are only restrictedly linkable. Strässler offers no examples for (1), arguing that any idiom is motivated in some way for certain people. For (2), he cites *to hit the hay/sack* and *to pitch (strike) one's tent*, and for (3), *to take exception to* and *to run a risk*.


Weinreich apparently felt Mel'čuk's contribution to phraseology was the distinguishing of idiomaticity from stability of collocation, which is
measured by the probability with which the given constituent predicts the appearance of the other constituents. ... Idiomaticity, on the other hand, is defined as uniqueness of subsense. What Mel'čuk calls 'stability of collocation' is thus a high degree of contextual restriction on the selection of a monosemous dictionary entry; what he calls 'idiomaticity' is a strong restriction on the selection of a subsense of a polysemous dictionary entry (Weinreich 1969:44).

Strässler also approves of Mel'čuk's separation of idiomaticity from what he calls "frozenness", which is a function of frequency of usage. Nevertheless, he considers serious the problem that idiomaticity is still based on frozenness since they both depend on specific lexemes (26).


In choosing not to follow Vinogradov, Amosova emerges as his first significant critic. Despite his widespread influence, Amosova charges Vinogradov with using criteria that are too vague and interdependent (1963:31, cited by Strässler 1982:23). Amosova views words as having contextually actualized meaning (as did Grice, 1975, Chafe 1968, etc.), and any word which is only realized in a
fixed context belongs to a phraseological unit. Strässler gives as English examples blue movie and dear-John letter in which the words movie and letter are "key-words which actualise the specific meanings of blue and dear-John" (23). Idioms are phraseological units in which either constituent can be considered the key-word which actualises the meaning of the other, as in Strässler's examples, grey mare, red herring, and white elephant.


The theory developed by Reichstein, which abandons lexicology altogether, identifies two critical qualities of phrasal collocations: regularity in usage ("usualness") and irregularity in structural (syntactic) and semantic organization ("frozenness" and "idiomaticity").

The very complex classification Reichstein charted includes three different aspects, L (lexical), S (structural-syntactic), and C (communicative-grammatical), which can be combined in three different ways (LCS, LS, and CS), and three grades of frozenness
(absolute, selective and preferential). In addition, Reichstein
differentiated non-perfect (partially lexified) from perfect (fully
lexified) phrases. Placing these in a two-dimensional paradigm
creates eighteen category slots. Strässler (1982) takes issue with the
lack of differentiation of frozenness and idiomaticity in this scheme,
and questions the usefulness of the whole endeavor.


As early as 1955, Bar-Hillel foresaw the problems idioms
present for machine translation and subsequently, for the generative
movement which, although he could not have known it then, was
about to dominate American linguistics. The problems he anticipated
cannot be said to have been successfully solved as yet.


An idiom by Hockett's definition above (3.3.3.1 "Any Y [Y= 'any
grammatical form the meaning of which is not deducible from its
structure'] in any occurrence in which it is not a constituent of a
larger Y") (172) is like Householder's (1959) "sentence-grammar prime"--"these are the items which must be glossed and listed in the lexicon that accompanies the grammar" (235) or Conklin's (1962) "lexeme"--"meaningful form whose signification cannot be inferred from a knowledge of anything else in the language" (121). A consequence is that "the raw-materials from which we build utterances are idioms" (Hockett 1958:173), be they single morphemes or composites. The reason Hockett's idioms are not limited to multilexical units is that his distinguishing criterion is idiosyncrasy. The line between unilexical and multilexical entities, often drawn for convenience sake, seems arbitrary to Hockett because "An idiomatic composite form, like any single morpheme, has to be learned as a whole" (173).


While others like Smith above were intrigued by phrases which conventionally linked two terms, Malkiel was first to select them according to a syntactic common denominator, that of irreversibility. Any two terms linked by and ought to, in principle, be reversible,
but Malkiel presents a collection of binomials which always occur together in a fixed order. Although Malkiel avoids discussion in terms of idioms and formulas, he illuminates the field by his discovery of a syntactic criterion by which this one type of fixed phrase can be identified. In addition, this paper makes a considerable contribution as a demonstration of a microscopic examination of a seemingly tiny but consequential topic. Most important for theoretical reasons, he isolates a major problem for transformational-generative approaches in the transformational deficiencies which are inherent in irreversible binomials.

3.6.5. Hymes 1962.

Substitute the word "speech genres" for "routines" and the arguments which Hymes (1962) makes for the study of routines sounds very much like the pleas of Bakhtin (1953). Hymes insisted that a thorough account of its conversational routines is necessary in order to describe the verbal behavior of a community, including "identification of idiomatic units, not only greeting formulas and the like but the full range of utterances which acquire conventional
significance for an individual, group or whole culture" (Hymes, 1962:38f.).


The innovative solution to the idiom problem in Katz and Postal's attempt to handle idioms within the transformational-generative framework was to separate the lexicon into two parts, a lexical and a phrase idiom part, a tactic which Weinreich (1969) later refined.


Critical discussions of the tagmemic handling of idioms proposed by Pike (1967), Healey (1968), and Balint (1969) are contained in Makkai's published dissertation (1972). In Makkai's assessment, the strength of Pike's system and taxonomy is in its incorporation of cultural factors into linguistic theory. "Pike's system ... offers an organic, self-contained and mature theory of language in relation to human behavior ... it includes a serious attempt at
classifying cultural institutions of all sorts" (1972:44). Healey's and Balint's articles both appeared in the little circulated journal of the Summer Institute of Linguistics in New Guinea, Kivung. Their principal contribution is thus perhaps the influence they have had on Makkai in formulating his ideas on idiomaticity, Makkai having considered Healey's classified list valuable enough to be reprinted in the published (1972) version of Makkai's dissertation.


To Chafe, idiomaticity simply cannot be accounted for within the Chomskyan paradigm because that paradigm insists syntax is central and initiative. "An alternative model is one in which semantics is initiative; in which a semantic component generates grammatical structures which are then subject to conversion into phonetic structures" (1968:117). He lists the four features of idioms which make them anomalies in the Chomskyan paradigm, briefly, (1) noncompositionality, (2) transformational defectiveness, (3)
ungrammaticality, and (4) frequency assymetry, and demonstrates the failure with critiques of Katz and Postal (1963) and Weinreich (1969).

3.6.9. **Weinreich 1969.**

The most frequent criticism of Weinreich's discussion of idioms is leveled at his definition of idiom (above, 3.3.3.2, "A phraseological unit that involves at least two polysemous constituents, and in which there is a reciprocal contextual selection of subsenses, will be called an idiom.") Strässler (1982) objects to the putting of "the cart before the horse" in the suggestion that an idiomatic meaning is already one of the polysemous subsenses available to be selected (32). "Subsense-assigning can be done only ex post facto, after the meaning of the idiom is already known to the analyst", Makkai reminds us (1972:49). Čermák (1988) concurs: "It is not before the level of the combination of the constituents into a syntagma is reached that an idiomatic meaning arises, and its exhaustive decomposition into distinct parts is therefore impossible" (426).
Weinreich's paper is, nevertheless, a rich source of instruction in the incorporation of idiomaticity into an existing grammatical framework which demands rigor and explicitness. In addition, Weinreich considers idiomaticity in terms of other unproductive and semiproductive constructions and deals with such aspects of idiomaticity as familiarity and degrees of analyzability and freedom, which subsequent scholars (e.g., Van Lancker 1975) have found to be important notions for idiomaticity.

3.6.10. **Fraser 1970.**

Fraser's claim to have discovered a hierarchy of transformational frozenness for phrase idioms has not, on the whole, been convincing. McCawley (Quang 1971) promptly refuted Fraser with numerous counterexamples from vulgarisms. Makkai (1972) added to this list a considerable number of nonscatalogical idioms, showing that McCawley had unnecessarily narrowed his focus to one of at least eleven "neatly classifiable categories" of idioms which counter Fraser's claim. Whether or not idioms form an implicational series with respect to transformability, as Fraser's "Frozenness Hierarchy"
suggests, Fraser showed that idioms display degrees of frozenness from almost completely unrestricted to completely frozen, and McCawley's refutation reveals that Fraser's list separates idioms into different types, the "full idioms" being the most frozen.


Because of his commitment to the stratificational framework, considerable effort would be required to integrate Makkai's contributions into a non-stratificationist discussion and cannot be wholly successful. Makkai's work, nevertheless, is the most thorough classification and theoretical attempt available to date. The sheer quantity of his taxonomized data on phrasal verb idioms is substantial, but his extensive coverage of the various types of idioms within his definition is also a significant contribution. Much of his terminology has been retained except for those which depend on the stratificational framework, and his work is referred to frequently throughout this dissertation.
3.6.12. **Newmeyer 1972 and 1974.**

Newmeyer made an attempt in his 1972 paper to show that idioms are inserted at the level of shallow structure and argued that idioms should not be treated as units. But in 1974, Newmeyer reconsidered and decided that since no point where idioms are inserted can be motivated, either in deep or shallow structure, the "lexical integrity analysis" of Fraser and Weinreich cannot be correct. He concluded then that idioms are semantic units only, not lexical units, and "... the only difference in behavior between a lexical item used as part of an idiom or used literally should be those which follow from the meaning of the idiom" (342). The evidence he finds is "far more regularity to the behavior of idioms than is generally believed. Given the meaning of an idiom and the meaning of its literal equivalent, one can predict, to a surprising degree, its syntactic behavior" (327).

The consequences of the notion that idioms may be considered unitary as semantic entities while regarding them syntactically as separate lexical items requires further critical examination.
Whether or not he knew Bakhtin's works before writing his introduction, Coulmas was surely of a like mind on the issue of integrating social and cultural factors into the study of language and in recognition of the importance of conversational routines (like Bakhtin's "speech genres"). He rues the almost complete neglect of prefabricated linguistic expressions which came with the uncritical acceptance of the assumption that "almost every sentence has an occurrence probability of close to zero" (1981:1).

Coulmas, in his substantial introduction, agrees with Hymes (1962), Chafe (1968), Makkai (1972) and Bolinger (1976) that an adequate description of the sociolinguistic behavior of a speech community must include idiomaticity, routine, and collocatability as properties of expressions (Coulmas 1981:5). "Every normal member of a speech community can distinguish routine utterances from idiosyncratic ones" (16). If so, this is a part of every native speaker's competence which must be accounted for.
The 1981 volume edited by Coulmas contains many significant specific descriptions by a number of linguists of the use of prefabricated language in conversational routines from various languages of the world, and thus presents major empirical contributions to the study of universal idiomaticity.


In the fourth chapter of his Essay on Language, Grace (1981) classifies a number of unusual structures in English which are not normally accounted for in grammars. He groups these phenomena under the label "idiomatology" (46). This chapter was the main stimulus for my interest in the phenomenon of idiomaticity because of Grace's display of so many kinds of seemingly arbitrary and unmotivated restrictions in the actual use of languages, some of which mandate the use of illogical and semantically anomalous forms as well as grammatical exceptions. Grace presents several types of idiomatological phenomena: (1) exceptions to general rules (e.g., "Use of singular rather than plural number in expression of the type five-year gain, fifty-cent cigar") (47); (2) arbitrary restrictions (e.g.
irreversible binomials (Malkiel 1959), idiosyncracies in the use of prepositions, arbitrary collocational constraints such as in time telling, and clichés, and the restrictive selection of a particular interpretation of phrases which are potentially ambiguous, such as HCE *I slept late* meaning to get to bed late at night, and *wet my pants* in contrast to *wet my shoes* (Bolinger 1976:6; Grace 1981:48-49), (3) specific idioms (*kick the bucket*), and generalized idioms (productive patterns which are nevertheless noncompositional or illogical, such as *Didn't you know that?* requiring the same responses as *Did you know that?* (50); and (4) "certain patterns of usage which seem to require that the speaker assume a certain point of view (50)" such as "the peculiarities of the use of tenses in English and other languages (e.g. Robin Lakoff's examples: "The animal you saw was a chipmunk: see, there he is running up a tree. That thing rustling in the bushes over there will no doubt be a chipmunk: let's wait till it comes out" (Lakoff 1970:839; Grace 1981:50-51).

Perhaps some of Grace's examples can be accommodated in a very complex grammar but this is certainly not possible for all of them. The question of what is legitimately handled in a grammar is a
crucial one, the answer to which depends on what sort of general theory of language is adopted.

Chomsky's observation that "In general, it is to be expected that enrichment of one component of the grammar will permit simplification of other parts" (1972:12-13) would seem to be applicable when struggling with the problem of determining a boundary between the idiomatic and the nonidiomatic. The idiomatic seems to belong to the lexicon while the nonidiomatic conforms to syntax, but the lexicon is greatly complicated by the inclusion of idiomatic material. At the same time, the more of the idiomatic that is accounted for in syntax, the more complex the syntactic component becomes.


Strässler attempted to handle the idiom as a pragmatic phenomenon, as "a functional element of language" (11). While he is not alone in focusing on function (e.g., Čermák 1988:426), his conclusion that idioms display great regularity in functional
behavior, while an important observation about idioms, leaves many crucial questions about idioms unanswered. We are still in need of answers to how idioms are like and unlike ordinary words and phrases and how the unique characteristics of idioms may be accounted for in an adequate description of language as well as of language use.


In one of their two 1983 papers, "Two puzzles for linguistic theory", Pawley and Syder ask two important questions about nativelike expression: (1) How does a native speaker select from grammatically possible alternatives the sentence that is natural and idiomatic in his language, and (2) How is he able to produce fluent stretches of spontaneous connected discourse? (191). They argue that fluent and nativelike language depend on a speaker's knowing an enormous quantity of largely fixed expressions, which they called lexicalized sentence stems (191). They demonstrate with copious examples (1) that many grammatically possible paraphrases (here, of *I'm so glad you could bring Harry! and I want to marry you*) are
never actually used (195-96), and (2) that fluent sequences of speech are made up of conventionalized clauses and disfluencies occur at their boundaries, or whenever the speaker uses an "integrating" rather than "clause-chaining" strategy in constructing stretches of speech (199-208).

Finally, Pawley and Syder discuss at length the problems posed for linguistic description and theory by the existence of lexicalized sentence stems which are grammatically and semantically regular but are special only in being familiar. They relate the problem to that of semiproducive derivational rules and suggest that perhaps, as Chomsky said of those, "significant systematization is just not possible" (Chomsky 1965:192). Pawley and Syder agree:

For what really matters is not the economy of the description but its fit with what the native speaker knows of his language. If the native speaker knows certain linguistic forms in two ways, both as lexical units and as products of syntactic rules, then the grammarian is obliged to describe both kinds of knowledge; anything less would be incomplete (217).

In two other papers, published in 1985, Pawley attempts a systemization of the notions of "lexicalization" and "formulas". In
"Lexicalization" he identifies the differences between the
grammarians' (i.e., a linguist's) assumptions about the lexicon and
the dictionary-maker's practices. For the linguist, a lexicon is the
part of the grammar which lists all of the "lexical formatives"
(Chomsky 1965:84), or, the "basic irregularities" (Bloomfield
1933:274) of a language. Nothing should be listed which can be
regularly derived from items in the lexicon plus the rules in the
grammar because listing derivable forms would be redundant.

Dictionaries, on the other hand, usually include these forms
because dictionary-makers work on different assumptions and
principles with regard to the value of parsimony as compared to
convenience and practicality. "... dictionaries may include any
composite form if it is a common usage, i.e., if it is recognized by
members of the language community as a standard way of referring
to a familiar concept or conceptual situation" (1985a:101).

As his definition of speech formula Pawley gives, "a
conventional pairing of a particular formal construction with a
particular conventional idea or idea class" (1985b:88), making a
formula synonymous with a dictionary-maker's rather than a grammarian's lexical entry. In "Lexicalization" he suggests several criteria by which lexicalization of composite forms is determined: (1) institutional status in a culture, (2) resemblance to simple lexemes, (3) special syntactic restrictions, (4) ellipsis, (5) writing conventions, and (6) arbitrariness. Lexicalized constructions, or formulas, may be lexically specified in all cases (Every man has his price), in some cases (Long live NP!), or they may be unspecified (in time-telling, M PREP H, meaning, Minutes, PREPosition such as to or after, Hour).

A relatively small number of formulas are syntactic idioms (having irregular phrase structure, e.g., easy does it). Many thousands are semantic idioms (drop a brick as well as put one's foot in one's mouth, both meaning 'commit a social gaffe'). But most formulas are syntactically and semantically well-formed (tell the truth), although "many formulas have a grammar that is partly unique. There are special restrictions as to how constituents may be moved, inserted, expanded, etc. without destroying the formula" (90).
Since well-formed formulas often have literal counterparts of exactly the same formal construction, formulas cannot be identified by formal criteria, but are identified by the meaning and function with which the formal restrictions are associated. "Good morning is a formulaic expression only when it serves as a greeting rather than as a description of the weather" (90).

In 1992, Pawley summarized his view on formulaic speech in an article for the International Encyclopaedia of Linguistics. Here, he presents formulas as "the basic units from which extended formulaic discourse is built" (22). He distinguishes formulaic expression, also called "formula" in Pawley 1985b, from formulaic construction:

the pairing of a grammatical formula, specifying not a single expression but a class of expressions, with a discourse function. Who does Mary think she is? represents the grammatical formula \texttt{Who do-TENSE \textit{X} \textit{think pronoun} be-TENSE?}, used to express indignation at someone's presumption (23).

He also recognizes a "semantic formula", which is "a familiar or stereotyped grouping of ideas expressed by a short stretch of
discourse, but not necessarily by the same words (compare Pleased
to meet you and Delighted to make your acquaintance) (23).

Similar are semantically different constructions which serve
the same discourse function or pragmatic purpose in conversation--
Thanks!, That's very kind of you! and I do appreciate what you've
done (23).

Pawley notes that speech formulas are far from being marginal,
despite the tendency of modern linguistics to view them as such:

Paradoxically, the commonplace speech formula
seems to play a central role in the creative use of
language. In the production of extended discourse,
formulas are essential building blocks--ready-made units
which free speakers and hearers from the task of
attending consciously to each word. Thus freed, they are
able to focus on the larger structure and sense of the
discourse, or to nuances of wording or sound ... Formulaic
constructions provide schemas for saying new things
without breaking conventions of idiomaticity and good
style--something that grammar alone does not do (23).

Formulas, therefore, while not the same as idioms, have the
same function in language use as idioms. Idioms, in fact, are a subset
of formulas in Pawley's system (1985b:89).
This paper is an attempt to integrate idiomaticity into a generative model of grammar. It would require the grammar to "include principles whereby a language can associate semantic and pragmatic interpretation principles with syntactic configurations larger and more complex than those definable by means of single phrase structure rules" (501). An exhaustive discussion of the conjunctive phrase *let alone* and other semantically related phrases *much less, never mind, not to mention*, etc. leads to the conclusion that a grammar must be more than "a system of general grammatical rules and a lexicon of fixed words and phrases" (534).

The productive and highly structured nature of much in the idiomaticity of a language should not be ignored by consigning all irregularities to a list of exceptions. They propose that the generative machinery be altered to accommodate idiomatic structures first and still be powerful enough to be generalized to more familiar structures (such as have been represented by individual phrase structure rules in the past) (534).
Fillmore, Kay and O'Connor have thus chosen to complicate the syntax rather than the lexicon of the grammar. Some subsequent researchers have made the opposite choice (e.g. Brame 1984a and 1984b; Dyer 1989). The phrasal-lexical approach claims there is no separate grammar, that all rules are in the lexicon. Zwicky (1988) insists a sharp separation between lexicon and syntax must be preserved, but that the Fillmore-Kay-O'Connor insights must also be accommodated. Idioms and free constructions alike involve stipulation of form and even specific constructions at times.
CHAPTER 4

PROBLEMS FOR A THEORY OF
IDIOMATICITY

4.1. Problems of definition of idioms and idiomaticity.

The purpose of this chapter is to summarize the major problems which must be confronted in any attempt to formulate a theory of idiomaticity. No attempt will be made to formulate such a theory, as an attempt of that magnitude is not the purpose of this dissertation. Based on the works discussed in the preceding chapters, problems of identifying idioms and idiom-like phrases will be investigated. An attempt will be made to develop a stable terminology specifically for use in discussing the basic writings analyzed for this dissertation. Finally, various aspects of idiomaticity which need to be considered in any theoretical discussion will be suggested, again taking advantage of the findings of authors of previous studies on idiomaticity. It is hoped that placing this chapter before the treatment of the student papers will make it easier to talk about the students' writings with uniform terminology.
assumption of a shared background of information about the complexities of idiomaticity.

4.2. Identification of idioms.

The first task in identifying idioms is to establish some practical working criteria by which idioms can be identified. Thereafter, decisions about the boundaries of the set must be made, for which a number of factors external to the set must be considered. Finally, a terminology must be developed which would be useful, at least for immediate purposes of analysis and discussion. The terminology presented in section 4.2.4 is suggested very tentatively since it is recognized, first, that the terms suggested are not mutually exclusive in all cases; secondly, that several cross-cutting terms which have been used previously by others are useful classifications of particular segments of idiomaticity, although they did not work neatly into this terminological scheme. Finally, it is very doubtful that the field of idiomaticity is covered exhaustively by the suggested terminology, despite my best efforts.
4.2.1. **Criteria for inclusion of a form as idiom.**

Traditionally, idioms have been identified by their peculiarity of meaning: "the meaning of each idiom is an isolated fact which cannot be inferred from the meaning of the words of which the idiom is made up" (Sweet 1889:139). Since the advent of generative syntax, considerable attention has also been focused on syntactic peculiarities of idioms, especially on transformational limitations of idioms (Katz and Postal 1963, Weinreich 1969, Fraser 1970, etc.). However, neither peculiarities of semantics nor of syntax are necessary or sufficient criteria for the identification of forms involved in idiomaticity. A full accounting of idiomaticity apparently cannot be made without considering such realities as frequency and familiarity, and perhaps even affective factors, prosody, redundancy, triteness, etc. which are manifestations of conventionalization.

Selection of a criterion for this dissertation was aided by employing a test case. The most difficult problem of establishing identificational criteria was to find a basis for separating a literal phrasal unit from a literal free expression since in form and meaning
they appear identical. In this case, frequency of usage could be the only directly observable basis for claiming conventionalization, but to actually collect enough data to establish accurate frequency statistics is practically impossible and has never been done. Instead, a compiler of idiom dictionaries, for example, depends on an intuitive ability to recognize a familiar phrase when the collector encounters one. Other researchers have applied various indirect means of finding out whether certain combinations of words carry unitary meaning for chosen subjects (Makkai 1972; Mackin 1983). Pawley suggested the terms "laymen's lexemes" for forms identified by the criterion of being common usages and "grammarians' lexemes" for forms identified by the criterion of not being fully predictable by the rules of grammar and semantics because he found these criteria yielded different lists (1985a:103). Since even grammarians' lexemes are lexemes only when they are common usages, familiarity due to frequency in usage seemed to be the basic criterion for idiomaticity rather than noncompositionality. It was necessary for the researcher to rely on native-speaker familiarity with the idiomaticity of English to glean material for discussion from the corpus of student writings.
4.2.2. **Single words and multiword idioms.**

One way to view idiomaticity is to recognize it as having to do basically with noncompositional meaning. Thus, everything from single morphemes to full discourses would be included within the topic. Hockett’s (1958) inclusion of every morpheme in his definition of idiom is a consequence of his choice of idiosyncrasy, or noncompositionality, as the basic criterion for his definition of idiom. On the other hand, one might alternatively choose to view idiomaticity as having to do primarily with unitary phrases larger than words but within the bounds of a sentence. Meanings of the phrases could be literal as well as noncompositional by this view.

From the start of my research, I chose the latter view of idiomaticity because I was not primarily interested in noncompositionality, but I was deeply interested in the phenomenon of conventionalization of phrases, some of which were absolutely ordinary in syntax and meaning.7 Literal phraseological units appeared to cause non-native writers at least as much trouble as figurative idioms, if not more, yet they seemed to have received little
attention until recently. Single word idioms were excluded from this study by necessity of limiting the corpus to manageable size, even though it is recognized that they are the same kind of phenomena as phrase idioms in terms of noncompositionality of meaning.

4.2.3. **Logic, grammars, and universals.**

Logic is sometimes cited as a criterion for identifying a form as an idiom, i.e., a conventionally accepted phrase which is not strictly logical (in an ordinary, non-technical sense) is often considered to be an idiom. *I don't believe he'll come* and *I can't seem to find my keys*, etc. have been termed illogical when taken literally, yet are fully acceptable, at least in colloquial English. The question of what logic has to do with language, however, has not been answered fully by philosophers or linguists, and just how far logic is actually applied to linguistic structures in natural language is extremely problematical. According to systematized laws of logic, conjuncts are freely permutatable so $A + B$ should be equivalent to $B + A$. By this law, *I got all wet and fell into the pool* should be equivalent to *I fell into the pool and got all wet*. Is *What was your name?* illogical in any sense,
or does it have nothing to do with logic at all? Should we say *All I did was rode in a car* or *All I did was ride in a car*? Does the choice depend on what is logical? However we choose to define logic, it is generally the case that idiomatically acceptable forms are not required to be "logical" in any sense. At the same time, being illogical is not a necessary condition for idiomaticity either, as most idioms are perfectly logical, at least in some sense.

Grammaticality as a criterion for identification of an idiom is not as straight-forward as might be hoped, either. Ungrammatical phrases which are acceptable are usually considered to be idioms, but whether a phrase is ungrammatical depends on the status of the grammar. Therefore, what is considered idiomatic by this criterion depends, not only on what is considered acceptable, but also on what is considered to properly belong to the grammar. Is the *both of us* grammatical? How about *Let him do what he wants*? Since language is dynamic, grammar is perpetually in the process of change. That idiom appears to lead in that change has been noted at least as early as 1944 by Roberts: "The best field of study for the idiomatologist lies in the subtle frontiers between almost identical languages and
cultures. For experience shows that the idiom is the first element to change, when two languages begin to diverge from each other" (306). Grace (1987) agreed that "rather than in the most rigidly conventionalized parts of the language [the grammar] ... the entering wedges of much linguistic change are to be found there [in the 'ways of talking']" (124). Idiomaticity appears to be a crucial factor in the argument between two major camps which Hopper (1988) has distinguished in linguistics regarding the nature of grammar: those who hold that an "a priori grammar" underlies language and those who see grammar as emergent from discourse.

Whether any principles of idiomaticity are universal or whether idiomaticity can be described only specifically for each language and culture is yet to be determined. The idiomaticity of a particular language is by definition, of course, language specific. It would appear, however, that, like grammar, some principles governing idiomaticity may be universal, while others are specific. For instance, it seems all languages have rule-governed patterns, semi-productive patterns, and idiosyncratic idioms and formulas. And just as they all have grammars, they all have exceptions to
grammatical rules, as far as anyone knows. Is it possible that grammatical rules develop from the patterns of idioms which begin as memorized exceptions? When use of the pattern of the idiom is extended to other words the pattern may then go through a period of semipruductivity and may eventually become fully general and productive, at which point we might say that a new rule has emerged from the idiom pattern.

It is reasonable to speculate that all languages have such processes in progress at all times, but the existence of writing systems, standardization of languages by governments, universal education, and so forth, would all have some effect on the processes of language change and on the idiomaticity of the language. Speakers of nonstandard dialects of a language and non-native speakers also have an effect on the idiomaticity of a language, particularly when they are numerous in a speech community. Since idiomaticity is a function of frequency of usage, a large number of people using a particular idiom, however "wrong" by the imposed standard can have an impact on the standard.
4.2.4. **Terminology for idioms and idiom-like phrases.**

After more than two years of futile attempts at creating a classification with mutually-exclusive categories which accommodates at least the three scalar dimensions of lexical, syntactic and semantic regularities or irregularities of all types of idiomatic phrases, I have contented myself with the following terminology for the practical purpose of getting on with my dissertation. The terminology at least allows me to discuss the student papers which I have studied with some consistency and with a minimum of confusion about what I am talking about.

**Terminology:**

**Idiomaticity**--the property of an expression which imbues it with its peculiar (special) quality and which renders the expression natural in the language. Idiomaticity is a function of familiarity and frequency of usage. An expression having proper idiomaticity is judged by native speakers to be expressed in the usual, expected
way. Of the set of grammatical phrases which can express a conventional idea, only a very small subset will be idiomatic, or the appropriate conventional way the idea is expressed.

By regular formation of -ity nouns from -atic adjectives, the word "idiomaticity" is assumed, in one of its senses, to refer to a quality derived in turn from an attribution something like "constituting or containing an idiom, or idioms". Indeed, some theorists, such as Reichstein, have used "idiomaticity" to refer to the semantic irregularity of phrasal idioms (1973). However, as the term is used here, the "idiomaticity" of an expression does not depend on its constituting or containing an "idiom" as ordinarily and narrowly defined. For instance, a phrase may be "unidiomatic" because it is ungrammatical, or because it contains a faulty lexical selection, or for various other reasons having nothing to do with idioms per se.

The units within a sentence which affect the idiomaticity of the sentence may be single words or combinations of two or more words. Because single words are supposedly handled as entries in the
lexicon which are used according to rules for selection, the choice was made to focus the study of idiomaticity on multilexical groups, or phrases. (Otherwise, the size of the study would have been prohibitively large.)

The phrases which are involved in idiomaticity are not limited to those defined syntactically as phrases. To distinguish them from syntactic phrases, the term "phrasems" (borrowed from Greciano 1986) will be used in reference to phrases discussed in terms of idiomaticity. While the term "phraseological unit" for essentially the same phenomenon is better established, especially in the field of lexicography, "phrasem" was selected for convenience because of its brevity. The longer term will occasionally be used synonymously, out of habit or for variety.

phrasem--"polylexical, more or less fixed, sometimes figurative verbal signs" (Greciano 1986). "Phrasem" is the most general term for the superordinate class involved in idiomaticity. A group of words forming a construction which is not a phrasem, but is instead a nonce phrase constructed by regular syntactic rules, is called a "free
expression" (Jespersen 1924). A phrasem is always polylexical and more tightly bound as a conventional semantic unit than a free expression. It may but does not have to have nonliteral meaning. It also may or may not be syntactically unusual, and may or may not have its lexical content fully specified. The superclass "phrasem" is composed of the subclasses "idioms", "collocations", and "formulas".

Some terms which have been used to represent roughly the whole field of phrasal phenomena here called phrasems, or various portions of it, are "idiomatic", "conventional", "fixed", "formulaic", "frozen," or "prepatterned expressions", "phraseological" or "phrasal units", "phrasal wholes", "formulas" (or, less often, "formulae"), "chunks", "prefabricated language" (or, "pre-fabs"), "nonpropositional language", "ready-mades", "routines", "stock phrases", "unanalyzed units", and no doubt, others. Our phrasem covers the same territory as was apparently intended by Jespersen's "formulas" when he divided phrases into two kinds, "formulas" and "free expressions" (1924:18-24).
"Phrasematic", rather than "phraseological", will be used as the adjectival form of "phrasem", since the previously used word has already established specific meanings and usages which are not identical with the newly intended ones.

**idiom**—a phrasem which selects a nonliteral alternative from options of a literal, compositional interpretation (which combines words by general syntactic rules according to their inherent lexical and semantic specifications) and nonliteral, noncompositional ones. Thus, every idiom is necessarily ambiguous out of context, each such unit having a literal counterpart by definition. (There are, of course, also the single-word idioms which are not phrasems.)

A "full idiom", the sort of phrase most commonly identified as an "idiom", is a subset of the class of idioms. Other subsets of idioms are "euphemisms", "frozen figurative forms" (such as conventionalized hyperbole, frozen metaphors, similes, proverbs, sayings, etc.), "semi-idioms" and "tournure idioms".
Idioms:

**euphemism**--a phrasem which deliberately exploits potential ambiguity by substituting inherently ambiguous terms for offensive unambiguous ones *[pass away, go to the bathroom, wet one's pants]*

**frozen figurative form**--such units as conventionalized hyperbole (exaggeration) *it kills me*, frozen metaphors *the cat's meow*, frozen similes *soft as silk*, proverbs *A stitch in time saves nine*, and sayings *Two heads are better than one*

**full idiom**--an idiom whose meaning cannot be inferred from the ordinary meanings of its constituents; i.e., one which has no constituents which contribute its ordinary meaning (Newmeyer 1974) *[red herring, kick the bucket; butter up]*

**semi-idiom**--an idiom which is not a full idiom, that is, one in which at least one constituent contributes its ordinary lexical meaning while the rest of the constituents do not *[white lie; swallow one's pride; stir up trouble; float a loan]*

**tournure idiom**--a "turn of the phrase", overlaps with full idiom but consists of at least three words (Makkai 1972:148-155) *[kick the bucket, fly off the handle, come off it]*
collocation, or, stable collocation—a phrase whose constituents occur together exclusively, or with a highly restricted set of variations. "The stability of a collocation is measured by the probability with which the given constituent predicts the appearance of the other constituents (Mel'čuk 1960, paraphrased by Weinreich 1969)." The main difference between a collocation and an idiom is that a collocation is never ambiguous between literal and nonliteral interpretations while an idiom always is. A collocation may be an extremely rigid "unique collocation", a "restricted collocation", or a "common collocation".

Collocations:

unique collocation—at least one member of this type of collocation never occurs in the language in any other environment [kith or kin; tit for tat; spick and span; filthy lucre; bated breath; leave in the lurch; (be) on tenterhooks; to and fro]

restricted collocation—at least one member of this type of collocation occurs in a tightly restricted set of semantically related environments, often only one other [nary a
common collocation—all members may sometimes occur in other environments but occurrence in a tightly restricted set is extremely frequent [raging/pounding surf; wild animals; assets and liabilities; cold water flat; Latin and Greek; two wrongs don’t make a right]. The members of common collocations are not necessarily monosemous, but their most ordinary meanings are usually selected; i.e., the interpretation of the collocation is literal rather than figurative, even when there is more than one possible literal interpretation. Common collocations are distinguished from free constructions only by frequency of occurrence which causes the combination of terms to become conventionalized, stabilized, familiar, expected or required.

formulas—phrasems which are neither idioms (because they are not necessarily ambiguous, often having no current literal counterpart) nor collocations (because they may be ambiguous). Formulas are the "everything else" category, and therefore may bear further analysis
and synthesis upon some basis which is not yet apparent to me.

Formulas are highly conventionalized and are tightly bound to context. Most formulas are better described in terms of stability of function than by meaning.

Subsets of formulas include "conventional implicatures", "routines", "arbitrary conventional usages", and "ungrammatical conventional usages".

Formulas:

conventional implicatures--discourse deictics or "cohesive devices" like connectives, social deictics, etc. [by the way, in any case, your honor]

routines--mixed phrasems which constitute part of the inventory of "appropriate things to say" in social intercourse such as greetings, apologies, rejoinders, thanks, etc. [how do you do, never mind, much obliged, excuse me].

arbitrary conventional usage--conventionally accepted phrases which violate logic or contradict reality when interpreted literally [can't seem to; don't believe/think; try and go;
neutralized negatives, e.g., *I wouldn't be surprised if it didn't rain*, double genitives, *that long nose of his*

**ungrammatical conventional usage**—conventionally accepted phrases which violate general grammatical rules [*be that as it may; trip the light fantastic; by and large; long time no see; no go; it's me; there's + NOUNplural* (e.g., *there's lots*); *more well-informed/most well-dressed, beyond compare; a few/lot are, ten foot pole, many a boy* (Shaw 1975)]

Other, crosscutting terms for phrasems:

These terms refer to types of phrasems which are clearly distinguished by syntactic form and which cross-cut the categories of idioms, collocations, and formulas.

**binomials**—are phrases which are composed of two terms joined by a conjunction, usually *and*. If they may not be switched around they are *irreversible binomials* (Malkiel 1959). They may be idioms [*tooth and nail*], collocations [*spick and span*], or formulas [*here and there*]. There are also a few **trinomials**, usually with fixed order—idioms [*hook, line, and sinker*],
collocations \textit{[blood, sweat, and tears]}, and formulas \textit{[no ifs, ands, or buts about it]}.

\textbf{compounds} consist of two adjacent nouns which are treated as a unit. Examples: idiom \textit{[lady killer, card shark]}, collocation \textit{[alligator shoes]}, formula \textit{[women doctors]}. Common collocations composed of words of other major categories such as Adj + N, etc. are also labelled compounds by some authors (e.g., Makkai 1972).

\textbf{phrasal verbs}--phrasems consisting of a verb followed by a particle which carry unitary meanings. Phrasal verbs may be ambiguous idioms (like \textit{put up} which can mean 'provide lodging', 'preserve fruit', and a number of other things (Makkai 1972:135-36) besides its literal meaning), unambiguous collocations like \textit{go away}, or they may be formulas like \textit{give in} which has no literal analytical interpretation but has a conventional unitary meaning.

\textbf{prepositional verbs}--similar to and may be considered a subtype of phrasal verbs. They differ in that the form which follows the verb is a meaning-bearing preposition rather than a particle (which functions to combine with the verb to form a
unitary meaning with the verb but bears no analytical meaning). Prepositions must have objects (a following NP) so a prepositional verb is always transitive whereas a phrasal verb may be intransitive. Furthermore, unlike a particle, a preposition cannot follow its object NP (*look a word up* but not *look a tree up*).

**support verbs**—"families" of phrases which are composed of a very common main verb with other words added to form a unit. They overlap phrasal verbs and prepositional verbs but also include phrasems which do not have prepositions or particles: idiom *[come hell or high water]*, collocation *[come home, come uppance]*, formula *[come here]*

**lexicalization**—Any phrase is "lexicalized" to the extent that it conventionally acts as a lexical unit, or, is wordlike. Lexicalization is characteristic to varying degrees of all phrasems, and is not a dimension on which phrasems vary in type. Even a fully lexicalized phrasem, however, may not behave as an ordinary lexical item and may not permit such productive inflections as are applied automatically to ordinary lexical items. If inflections were added to
the end of phrasems as is usual with ordinary lexical items, we would expect *He kick the bucketed* and *They memorized the song by hearts.*

One of the difficult problems in the handling of phrasems in a grammar is to figure out how phrasems may be lexicalized, that is, be treated as unanalyzed units semantically, and yet permit or prevent ordinary syntactic processes to operate on internally occurring parts, so that, for instance, verbs within the phrasem may be inflected for past tense [*he kicked the bucket; we shot the breeze, etc.*] or the phrasem *by heart* be prevented from being inflected for plural.

A second problem is to deal with the transformational deficiencies of phrasems. One possible solution is to recognize phrasems as semantic units but not syntactic ones (Newmeyer 1974). Another is to say the phrasem is "literalized" (filled in with the words and syntax) after it is represented semantically (Chafe 1968, 1970). Syntactically, some phrasems act like ordinary analyzable syntactic structures while others do not. As semantic units, each phrasem has its own specifications as to syntax. This results from
each phrasem having its own unique history on its way to lexicalization. (See Pawley 1985a for a thorough discussion of lexicalization.)

**Lexification**—A phrasem is "lexified" by being filled with specified lexical items. The minimal phrasem has at least two words. (Even though we accept Hockett's definition of idiom which entails that a monomorphemic word is also an idiom in itself, single word idioms are never phrasems, however extraordinary or dependent their meanings may be.) Neither of the essential two words needs to be specified, however, since there is a type of formula which requires no specified words—the semantic formula. "A semantic formula," in Pawley's scheme, "is a familiar or stereotyped grouping of ideas expressed by a short stretch of discourse, but not necessarily by the same words (compare *Pleased to meet you* and *Delighted to make your acquaintance*)" (1992:23). The semantic formula, which we may say is **unlexified**, thus pairs a not-yet-lexified semantic unit with a discourse function. With this one exception, all phrasems are at least **partially lexified** by virtue of containing at least one invariant lexical item [*hit the hay/sack; motley crew/crowd; the X-er the Y-er*;
lightning hits/strikes, etc.] Many are fully lexified, meaning every lexical item in the expression is invariable [by heart; in spite of; blind date; white elephant]. "Lexification" is thus distinct from "lexicalization".

4.3. Some aspects of idiomaticity.

4.3.1. Dimensions of regularity.

The main problem in trying to classify the various types of idiomatic and formulaic phrases is that the most prominent peculiar characteristics of the phrases may be sometimes syntactic, sometimes semantic, sometimes functional and sometimes lexical, while at the same time they must, of course, also function in the other dimensions as well. If we choose to classify phrases by syntactic form we can easily identify compound nouns, noun phrases of other kinds, phrasal verbs, and syntactic blends. If we prefer semantic criteria, we may identify conventionalized figurative expressions, full and semi idioms, and tournure idioms. Colloquial expressions, conventional implicatures, conversational formulas,
euphemisms, listing, rejoinders, repetition, topicalization and tags are basically identified by pragmatic function while connectives and partitives are distinguished by grammatical functions. Or we might focus on classification according to lexical or morphological characteristics and isolate binomials, unique-member collocations, conventionalized uses of articles, comparatives and superlatives, gerunds and infinitives, plurals, prepositions, and tense marking.

Were we to start with any of these groupings, however, each class would still have to be further subclassified according to their less salient characteristics: for example, compound nouns, most prominently identifiable by the syntactic feature of being comprised of two nouns, include members along a continuum of semantic opaqueness (army uniform being completely transparent, crocodile tears somewhat less so, and duck soup completely opaque), members which serve noticeable pragmatic functions (family jewels is a euphemism, etc.) and members which do not, and members which have peculiar lexical or morphological forms (women doctors forms plural by marking both words of the compound, etc.) and others which are perfectly ordinary, lexically. In fact, setting aside
pragmatic functions, there are still at least three dimensions along which all phrases may be analyzed—lexical, syntactic and semantic.

4.3.2. Degrees of regularity.

The regularity of a phrase lies along a continuum in each dimension. A phrase may range from fully-lexified (all words specified) to unlexified (as in the case of "semantic formulas" according to which pleased to meet you and happy to make your acquaintance are the same formula) and from fully lexicalized (conventionalized as a unit) to almost unlexicalized (close to a free expression) in the lexical dimension, from syntactically regular (hit the sack) to syntactically irregular (be that as it may), and from semantically regular (bright red) to semantically irregular (red herring). It is therefore fruitless to try to assign each phrase to some cell in any kind of paradigm, as one is tempted to do as long as classification seems to be a first step in making order of any jumble of items. This effort may be practically futile because, for one thing, the paradigm would be extremely complex, requiring a minimum of three dimensions, and more importantly, a grid would not represent
the graduated, or scalar, nature of the regularities and irregularities accurately. Reichstein (1973), in fact, developed a complex grid which had eighteen cells. Strässler complained of the lack of utility of Reichstein's scheme. Indeed, the question of what such a classification scheme gains us must be critically pondered. Furthermore, while we can recognize that regularities in each dimension range along continua, why would it be necessary, or is it even possible, to place each type of phrase somewhere along each continuum?

4.3.3. Form, meaning, and function.

Phrasems are form, meaning and function complexes, just as are words and sentences. The feature which makes them different from free expressions is their conventionalization, which may have varying effects on the several dimensions of the phrasem: (1) the phrasem may become more or less lexically filled and fixed, (2) it may become syntactically frozen and transformationally defective, (3) it may become semantically noncompositional, but (4) it always becomes more tightly bound to a grammatical or pragmatic function.
Conventionalization may be defined as the binding of a form with a function, and when a phrase becomes conventionalized its association with a particular function becomes more stable.

4.3.4. Grammar and lexicon.

A basic theoretical question posed by the demands of idiomaticity, or by the existence of phrasems in languages, is whether language can be accurately characterized by a model which consists of clearly separated parts of grammar and lexicon. Pawley summarized the assumptions about the lexicon and its relationship to grammar which has been held by most linguists since the 1930s as follows (abridged):

(1a) A language description is conceived of as a grammar.
(1b) The lexicon is viewed as a component of the grammar.
(1c) ... any one form-meaning pairing should be specified only once in the grammar, whether it be in the lexicon or by the syntactic rules which apply to lexical items.
(1d) If a given form-meaning pairing cannot be predicted by the productive rules of grammar, it must be listed in the lexicon; if it can be predicted, it does not belong in the lexicon.
(1e) ... the lexicon consists essentially of single words and morphemes (Pawley 1985a:99).

Pawley supports these generalizations with a sampling of quotations from American linguists between 1933 and 1976, two of which are as follows:

(2a) The lexicon is really an appendix of the grammar, a list of basic irregularites (Bloomfield 1933:274).
(2d) The base of the grammar will contain a lexicon, which is simply an unordered list of all lexical formatives. ... The lexicon is a set of lexical entries, each ... [pairing] a phonological distinctive feature matrix [with] a collection of specified syntactic features (Chomsky 1965:84; Pawley 1985a:99-100).

Chomsky and Halle (1968) have maintained Bloomfield's original principle: "Regular variations are not matters for the lexicon, which should contain only idiosyncratic items" (12). Aitchison, echoing Brown, blames theoretical linguists' excessive focus on syntax which caused them to seriously "underestimate the complexities of the lexicon" (Brown 1984:10):

... in recent years they have regarded syntax, which involves combinations of words, as more important than the words themselves ... Only recently has this viewpoint been challenged, and much that was ignored or placed elsewhere in a grammar is now being shifted back into the lexicon (Aitchison 1987:25-26).
Phrasems, because they are multilexical units which often, but do not always, conform to regular rules, present difficulties similar to those of semiproducitve derivational processes. In 1960, Lees proposed derivation of nominalizations by syntactic transformations, but more recent work has treated derived nouns and nominals as basically lexical, including Chomsky (1972) and Jackendoff (1975) in "Morphological and semantic regularities in the lexicon". The extreme position is taken by Dyer (1989) whose claim, as summarized in the Linguistics and Language Behavior Abstracts (1990) is that

... linguistic knowledge is fundamentally formulaic. All linguistic expressions are organized as generalized language use events--each retaining in a phrasal/lexical memory one or more hypotheses concerning the syntax, meaning and relevant context of use of each expression. As a result lexical memory is fundamentally phrasal in nature, where phrases consist of patterns for mapping structure to meaning in context. There is no separate grammar in phrasal/lexical approaches; all syntactic structures are encoded in the lexicon (LLBA 24:8).

4.3.5. Figurative meaning and literal meaning.

Another question, pertinent to the class of idioms, which have noncompositional meaning, is whether the unusual meanings of idioms are dependent in any way on simultaneous existence of literal
meanings. Several recent series of experiments dealing with the mental processing of idioms seem to show that this is not the case. Swinney and Cutler (1979) carried out experiments which supported a hypothesis of shorter processing time for idioms than literal strings, suggesting that idioms are not parasitic on literal meanings. Cacciari and Tabossi (1988) found that for predictable strings, only idiomatic responses were received from subjects. Likewise, Burbules, Schraw and Trathen (1989) found that "interpreting idioms depends on retrieval of a specific conventional meaning. In this sense, idioms are hardly 'figurative' at all" (LLBA 23:1395).

Gibbs and others found similar effects with several experiments (Gibbs 1986, Mueller and Gibbs 1987, etc.). Although psycholinguistic experiments can only observe mental processing indirectly, the results of these studies seem to indicate that idioms are stored and retrieved as wholes with their idiomatic meanings rather than calculated and synthesized from intricately polysemous parts. The notion that idiomatic meanings are "secondary" meanings then seems questionable. We can establish, nonetheless, that it is
part of the competence of native speakers to perceive certain meanings to be more usual, ordinary, or "primary" than others for most common words which have idiomatic alternatives.

4.3.6. Formulaicity.

Boundaries between formulas and free expressions are especially hard to define when we are trying to classify expressions which are neither syntactically nor semantically anomalous. In other words, idioms are easier to classify than formulas; and common collocations are almost impossible to separate from unique and restricted collocations. The reason is obviously that literal formulas and collocations are formally exactly like free expressions, and the only characteristics which distinguish them from free expressions are such unmeasurable qualities as familiarity and frequency of usage, semantic unity, and strength of bonding with pragmatic functions. Free expressions, in other words, can become formulaic when they are used over and over to convey some stable meaning and for a uniform pragmatic purpose, with no change in form.
4.4. Some supralinguistic factors affecting language use.

Idiomaticity requires the identification of "supralinguistic" factors which operate over and above language-as-is and governs language-as-used because idiomaticity is a quality of language use. These factors reflect principles which govern real entities, like brain activity and speech performance, rather than abstract entities, like words, sentences, or grammatical rules. The final section of this chapter speculates on what real-life factors, linguistic or extralinguistic, need further study in order to make a full accounting of idiomaticity in language use.

4.4.1. Neurolinguistics and language processing.

At present, it appears that information provided by psycholinguistic and, in particular, neurolinguistic studies promises to greatly enhance our understanding of language processing in the brain, including the processing of what Van Lancker (1975) called "nonpropositional" language (See section 2.2.9). If it can indeed be established that the brain is capable of processing language in at
least two different ways, we stand to gain much by distinguishing the characteristics of language processed in one way as opposed to the other. Current grammatical theories attempt to describe only propositional language, which is assumed to be assembled according to an algorithm using unilexical items from an inventory, or to model the assumed algorithm. Evidence is growing that attention must be given as well to nonpropositional language, which may be heuristically processed rather than following an algorithm.

4.4.2. Speaker versus hearer perspectives.

Hockett's suggestion that linguists try to overcome a bias toward viewing language always from the speaker's perspective in favor of a hearer's viewpoint may be helpful in studying idiomaticity. Hockett (1987) noted that from the hearer's perspective, it is not always necessary to identify separate words to understand a sentence or phrase (37) and that a phrase may constitute a perceptual gestalt (51). The distinction made by Makkai (1972) between encoding and decoding idioms is sensitive to the difference between speaker and hearer perspectives.
4.4.3. **Automaticity.**

Automaticity refers to pattern habituation. Animals, including human beings are capable of performing repeated sequences of acts in rapid succession without conscious attention to each act. The role of this capacity in performance of nonlinguistic tasks extends also to speech—it allows fluent and rapid speech performance with little conscious attention to each segment produced. Just as many aspects of such activities as driving a car, typing, playing the piano, and shaving can be automatized, many of the individual skills involved in speech, both physical and mental, can be largely automatized.

Pawley and Syder (1983b) (See section 3.6.16) suggest that fluency in speech depends greatly on knowledge of a huge inventory of fixed expressions. Their analysis of recorded speech shows that hesitations and disfluencies occur between fluent stretches comprised largely of fixed phrases. A key to native-like fluency in speech thus appears to be in the automatization of many useful phrases through frequent use.
4.4.4. **Familiarity and frequency.**

Frequency of use and familiarity to users are, of course, related in that the more frequent the usage in a speech community, the more familiar to users in that community a bit of language is likely to be. This relationship is exploited by researchers because obtaining actual frequency counts has often been found to be impossible.

Mackin testified to the impossibility of establishing any meaningful order of frequency by counting with two personal anecdotes. First, he found that even his most diligent reading turned up only one occurrence of most idioms (except for sentence connectives, *in fact, of course*, etc.). Second, he described a study of which the object was to record all instances of the phrase *red herring* which occurred in the presence of a particular group of investigators, working separately. In a matter of only a few months they found an astounding abundance of the phrase—no less than fifty occurrences, in fact. "Yet, in all my years of collecting from written sources I had not come across more than three or four occurrences of this expression," Mackin wrote (1983:vi).
Familiarity is rarely recognized as a formal characteristic of words or phrases in language, probably because of its gradient nature and the difficulty of measuring it. Yet, the study of idiomaticity seems to require recognition of, and, perhaps, a more specific formalization of the notion of familiarity.

Weinreich (1969) recognized that, if familiar phrases which are not idiomatic in the classic sense of being semantically noncompositional are excluded from special treatment in a grammar, they must be somehow identified as familiar. His suggestion was that a system for marking words and phrases as familiar be developed.

Van Lancker (1975) concluded that "the features of cohesiveness, redundancy, familiarity, frequency and affective strength are associated with occurrences of nonpropositional speech" (147). Familiarity and frequency have been discussed above. The rest will be discussed below.
4.4.5. **Cohesiveness.**

Although the first things which come to mind when relating cohesiveness to idiomaticity are probably the conversational deictic devices (*in the first place, let me backtrack a little, anyway, etc.*), the imperative of cohesiveness is really much more pervasive in language use. Indeed, an overriding demand of all linguistic expression in any real context must be relevance or appropriateness, which may be seen as a requirement that the expression fit or cohere to the context.

Considerable attention has been applied in the teaching of composition to how cohesion, or coherence, is achieved in writing in the absence of the immediate contextual support which would be present in oral conversation. Among linguists, Chafe and Tannen (1987; Tannen 1984) and Gumperz, Kaltman and O'Connor (1984) have been actively engaged in similar studies. It is clear, upon close examination of how cohesion is achieved both in speech and in writing, that much depends on formulaicity and idiomaticity. Sorting out the formulas and idioms which aid cohesiveness in discourse in
both communication modes, discovering which they share and which ones are peculiar to each mode, and so forth, would be valuable practical contributions which attention to idiomaticity could make.

4.4.6. Repetition and redundancy.

Translators have perhaps been first to notice that "Repetitions and other such word groupings acquire extra meaning through conventions that change from language to language" (Stringham 1981:73). A literal translation of Lincoln's "of the people, by the people, for the people" may be greeted by some as unnecessary redundancy. That speakers of different languages or different versions of a language often react differently to repetition is a fact of idiomaticity. Repetition is perhaps a kind of syntactic idiom. Such phenomena as this are worth investigating further.

4.4.7. Affective strength and prosody.

Besides appropriateness, naturalness, fluency and coherence, Pawley noted that prosody and artistry "(aptness of rhyme, simile,
"contrast, novelty, etc." depend on knowledge of speech formulas. He states, "The special advantage of formulas as analytic constructs is that they allow a unified treatment of such diverse elements" (1992:24). Van Lancker, likewise, proposes that the affective feature of nonpropositional speech needs closer linguistic scrutiny to discover what role emotional intensity and vividness play in human language ability (1975:147).

The features briefly discussed above do not necessarily constitute an exhaustive list of factors strongly associated with idiomaticity. They merely suggest some of the main lines of investigation which may be worth following in the formulation of a fuller, multifaceted view of idiomaticity.
CHAPTER 5
IDIOMATICITY AND BASIC WRITING

5.1. The study.

A description of the study is given in section 1.3.

5.2. Purpose.

Broadly stated, the purpose of this study was to find out more about what is needed beyond a knowledge of grammar and lexicon as envisioned in current linguistic frameworks to use the language like a native speaker. In particular, the study focused on the concept of idiomaticity, defined as *the qualities of an instance of language use which render it appropriate, normal and natural to native-speakers of the language.*

The focus of the study was directed toward information that would have practical implications for the teaching of written composition to non-native English speakers and speakers of Hawaii
Creole English. Therefore, the writings of students who were judged through placement tests, previously administered by the English department, to possess inadequate skills for college writing tasks were collected and analyzed. The purpose of the analysis was to discover to what extent the problems encountered by students involved idiomaticity, and to identify specific types of phrases which were particularly troublesome to each group of students.

A common assumption about "knowing a language" is that it consists of knowing (1) an adequate portion of the lexicon of the language (single "words") and (2) the grammatical rules for combining those words into sentences. At least as early as Sweet (1889), we have accepted the general division between grammar and dictionary. As Jespersen put it, "Grammar deals with the general facts of language, and lexicology with special facts" (1924:32).

However, recent linguistic research has revealed that a very large portion of linguistic production, which accounts for its idiomaticity, consists of highly conventionalized groups of words which are not assembled word by word as needed, but are learned
and used as wholes. "There seems to be no place for this concept of idiomaticity in the current grammar-lexicon models of languages" (Pawley and Syder 1983b:221), it appeared, so a way to talk about idiomaticity was sought from various sources. This study began with the hypothesis that control of a large store of these prefabricated chunks is essential for full comprehension and for fluent, nativelike production of a language (Pawley 1985a, 1985b and 1992, Pawley and Syder 1983a and 1983b, etc.). Therefore, it was expected that many problems in the students' writings would be with these conventionalized phrases.

Nativelikeness as a criterion for language use is a stringent demand, yet writing adequately for academic purposes requires it. The application of this criterion resulted in a much broader range of structures which are claimed to be idiomatic than has generally been accepted. A definition of idiomaticity which makes adequate distinctions between idiomatic and non-idiomatic and which includes all forms meeting the criterion of enhancing nativelikeness was needed.
Furthermore, the identification and classification of idioms and formulas appeared to be a necessary step toward making the nature of idiomaticity discussible, so a preliminary attempt was made to formulate a workable taxonomy which would at least be useful for discussing the students' problems. Since nativelike ability in the use of idiomatic language in writing is already demanded of students, it is hoped that even a small increase in our ability to discuss idiomaticity may be a valuable contribution to scholars interested in idiomaticity as well as to non-natives who hope to attain an adequate level of skill in the use and comprehension of an additional language.

5.3. Design.

Since this study was embarked upon without the aid of any kind of established theoretical framework for idiomaticity, the study was designed to make as broad a survey as possible of the sorts of problems that a particular group of student writers might encounter. Rather than zeroing in on a particular type of construction, or a limited set of types of constructions, the intention was to scrutinize for idiomaticity all phrases written by the chosen set of students.
during one semester in remedial English classes, idiomaticity being tentatively defined as "naturalness" or "nativelikeness". The criterion applied was, "Is this the way it would usually be written by skilled writers who are native speakers of English?"

It was immediately obvious that two sources of lack of idiomaticity would be grammatical errors and lexical selection violations. It was decided that although idiomaticity is affected in major proportions by these two factors, they would not be appropriately discussed in this paper, since the focus here is on problems outside of grammaticality and lexical selection.

It must be admitted at this point that the dividing line between grammatical and nongrammatical errors is not as clear as might be hoped, the distinction being dependent entirely upon what is considered to be in the grammar. Since it is beyond the purposes of this study to choose between competing theories of grammar available, some decisions made here are necessarily tentative. Generally, an unnatural locution was considered to be outside the realm of grammar if I could not see how it could be accounted for
without redundancy within a generative model of grammar which includes a lexicon of unilexical items (single "words") and autonomous syntactic and semantic components.

The distinction between a single lexical entity and a complex one was not as straightforward as expected either. Historical accident has frequently determined the orthographic conventions which combine or separate morphemes so that what is "one word" is often somewhat arbitrarily determined.

The choice to view the whole vista of idiomaticity broadly rather than to examine a chosen portion microscopically runs counter to current trends in linguistic research. This decision was made because of the relative "newness" of the ideas which point to the necessity of giving attention to idiomaticity. A concern was that if the forest could not be seen for the trees, the trees would appear trivial and uninteresting. Idioms and everything like them have in the past been brushed aside by the false perception that they are unusual and infrequently occurring exceptions, probably because when specific idioms are examined, they in fact do not occur
frequently, just as is true of specific lexical items. It is only when phrasems of all types are viewed together that their astounding pervasiveness is perceived.

The design of this study, in summary, was simple: Remedial writing students' compositions would be collected and analyzed for idiomaticity. Every expression (single words as well as groups of words) which does not sound native-like to the researcher would be identified. All sentences containing phrasal errors would be copied on separate lists, tentatively labeled according to types of phraseological units found in the literature on idioms perused beforehand. These lists would then be studied to discover what traits characterized each type of phraseological unit and a description of the unit would be attempted. The lists would then be revised to include only those phrasal errors which conformed to the description.

All errors would be counted, those produced by speakers of English as a Second Language (ESL) and by Hawaii Creole English speakers (HCE), separately. Grammatical and single lexical errors
would also be counted from the original student papers, and numbers of kinds of errors would be compared.

Finally, an attempt would be made to describe the most common problems that students have with idiomaticity when writing for academic purposes. If any patterns of differences in the errors committed by ESL writers as compared to HCE writers were detected, attempt would be made to describe those patterns and to suggest what factors may account for such differences.

5.4. Data collection.

First drafts of students' compositions were collected. The researcher met with each student and noted whether the student was a native speaker of HCE or a speaker of ESL. The papers of students who were neither were eliminated. Some ESL students, though not native speakers of HCE, were fluent speakers of HCE, having arrived in Hawaii before adolescence. These students were initially classed with the ESL speakers, but some adjustments were later made on the basis of their writing performance.
Recordings were made of each student reading his/her paper as an additional check on whether the student's control of orthography was adequate to accept a paper as a valid specimen. Some mismatches cropped up, for example, a student who wrote taking by surprised read it as "taken (or takin'?!) by surprise" and another who read correctly "strange character" and "large-framed", even though she had written them incorrectly as stranged character and larged-frame. Another student orally inserted the missing as in each of two places where he had left them out in writing. These orthographic errors were not counted as errors.

5.5. Analysis procedures.

Each paper was carefully read by the researcher, initially from beginning to end to get an overall understanding of the content, and afterwards in more detail, paying attention to phrasal expressions. Each phrase which was not expressed in a manner the researcher felt to be native-like and natural was underlined and paraphrased in more natural form to determine the source of the problem. A label for each error was then written in the left margin of the paper.
If the error which involved a whole phrase was of the sort which could not easily be labeled, it was marked "unidiomatic" or "phrase" and a specific determination of type of idiomatic error was made as sentences containing such errors were copied on to lists, which allowed grouping of similar types of phrases for which labels had not as yet been chosen. A residue of unclassifiable or infrequent errors were combined in a category called "miscellaneous".

Initially, close to fifty different labels for types of errors were used, including some which named grammatical error types, some which named types of idioms or formulas, and some which named processes which caused errors (such as "syntactic blending"). The difficult task of refining this mixed list is an on-going process which cannot yet be considered completed.

Tally sheets were made for ESL and HCE separately. Grammatical, lexical and idiomatic errors were tallied with the help of an assistant. Tallies were then added up and totals were recorded for each type of error.
5.6. Results of the analysis.

There are two facets to the analysis of the data collected, (1) the error counts, and (2) the discussion of types of errors involved in lack of idiomaticity.

5.6.1. Error counts.

Although the quantity of writing produced by each group of basic writers was fairly equivalent (119 ESL and 117 HCE pages), the ESL writings contained more than twice as many errors as HCE writings (2180 ESL, 1052 HCE errors). The biggest differences were in grammatical errors. In order of quantity of ESL errors, use of articles topped the list with 205 ESL errors of which 183 were purely grammatical and 22 were interpretable as idiomatic errors. The 44 HCE errors were all grammatical. High error count categories are as listed below:

<table>
<thead>
<tr>
<th>Kind of error</th>
<th>ESL</th>
<th>HCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>articles</td>
<td>205</td>
<td>44</td>
</tr>
<tr>
<td>wrong words</td>
<td>182</td>
<td>75</td>
</tr>
</tbody>
</table>
There were only eight error-types which gave HCE writers more trouble than ESL. All actually happen to be colloquial types of errors and, therefore, all could have been included in the category labeled "colloquialisms". The counts are as listed below:

<table>
<thead>
<tr>
<th>kind of error</th>
<th>ESL</th>
<th>HCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>colloquialisms</td>
<td>115</td>
<td>162</td>
</tr>
<tr>
<td>Hawaii Creole forms</td>
<td>74</td>
<td>91</td>
</tr>
<tr>
<td>topic fronting</td>
<td>23</td>
<td>24</td>
</tr>
<tr>
<td>jargon</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>tags</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>&quot;all-purpose&quot; which</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>reason why ... is because</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>there's/there is + plural</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Totals</td>
<td>222</td>
<td>309</td>
</tr>
</tbody>
</table>
The total number of grammatical errors and idiomatic errors are compared below. Because some kinds of errors, like the articles above, are divided between grammatical and idiomatic, some errors of judgment have probably been made, even though considerable care was exercised in studying and sorting the lists. Where it seemed justifiable, as will be explained in the detailed discussions later, whole categories were automatically assigned to one or another. Even so, a few indeterminate errors remained.

<table>
<thead>
<tr>
<th></th>
<th>ESL</th>
<th>HCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ungrammatical</td>
<td>1170</td>
<td>404</td>
</tr>
<tr>
<td>unidiomatic only</td>
<td>964</td>
<td>629</td>
</tr>
<tr>
<td>indeterminate</td>
<td>46</td>
<td>19</td>
</tr>
<tr>
<td>Totals</td>
<td>2180</td>
<td>1052</td>
</tr>
</tbody>
</table>

The most important general tendency to notice, besides the fact that ESL writers produced more than twice as many errors as HCE writers, is that more ESL errors were grammatical types than were idiomatic errors, while more HCE errors were the idiomatic kind than were strictly grammatical errors.
5.6.2. Types of errors.

All that is required for a phrase to be idiomatic is that it be a natural and usual way to say what is being said. To claim a phrase is idiomatic is not to say that it is an idiom. An idiom is only one special kind of phraseological unit of a set of such units. Of course, speakers do not only memorize whole phrases to use when needed, but creativity (the open-ended productivity) of human language allows users of the language to draw items of various sizes from a lexical store and combine them in accordance with grammatical rules in a limited variety of ways. The requirements of idiomaticity, however, constrain the ways in which the product resulting from the combination of lexical items by general grammatical rules can actually be used in a given context. In addition, the converse also holds--idiomaticity allows multilexical forms which are ungrammatically constructed to become acceptable by convention. Idiomaticity, being a matter of conventionality, is a function of frequency of usage. Thus, phrases can be unidiomatic in various ways, as long as those ways cause the phrase to sound unnatural, unusual, or non-native.
5.6.2.1. Accidental and meaningful errors.

A few words are needed here about performance errors. Some errors are clearly accidental mistakes which are recognized by the writer as error, and which do not reflect the writer's competence (in the Chomskyan sense). These are appropriately ignored if it can be determined that such is the case. Having the students read their papers aloud was the simplest means available for culling such accidental mistakes from the corpus studied.

In contrast to accidental errors, the errors we are interested in are meaningful ones because the writer does not perceive them to be errors. They cannot be brushed aside as "only performance" because they reflect the writer's competence. Admittedly, it is not always clear whether an error is one of performance only or whether competence is involved, especially when gleaned from written material. Undoubtedly, some errors of assignment will be made, but, on the whole, "purely performance" errors will be identifiable by their arbitrariness. The researcher suspects an accidental error when the error is of the sort that has an obvious explanation in
terms of orthography or phonology, for example *We went to to beach* and *the speaker was well-burst*, or when it is unlike anything ever heard before. Meaningful errors are usually plausible--the reader can believe such an error can happen. They also usually make sense--the reader can conjecture how such an error can occur. And, perhaps surprisingly, most meaningful errors are interpretable--the reader usually knows what the writer intended. If this were not so, this study could not have been done at all without long hours of discussion with each writer.

One category of errors which is very significant because of the frequency of its occurrence is "syntactic blending". Syntactic blends have been named by analogy to lexical blends, which are amalgamations of two words into one (such as *smog* from *smoke* and *fog*). Although lexical blends have traditionally been considered to be a part of parole because they often begin as slips of the tongue, it is actually only at this point of origination that the blend can be considered performance-only. Once the blend is transmitted and conventionalized it becomes a part of the language and a user has added it to his competence. Syntactic blends are no different. A
syntactic blend may originate as an "anacoluthon", when a speaker begins a sentence one way and ends it in an incompatible way, but common ones can become conventionalized for certain groups of speakers or speech communities.

Whether a syntactic blend is an error or not is often a matter of history: Is this a nonce-form for the speaker or did he receive it whole from others in his speech community? Is Why did you do that for? (a blend of What did you do that for? and Why did you do that?) an idiomatic expression in the writer's community? Such difficult problems have not been solved for this study. It appears that either extensive counting of frequency of usage would be necessary or familiarity tests would have to be devised and administered to an adequate sample of native speakers. The researcher has had to fall back on native-speaker intuition and experience with non-native speakers to aid in analysis of the data.

Syntactic blends are a rich source for future study with a focus on idiomaticity since attempts to account for them within current linguistic frameworks, ignoring the phrasal-unit phenomenon, have
not been found to be satisfactory (Fay 1982). The blends which are obviously based on commonly used phrasems are especially illuminating. The following is from an ESL student's paper:

(1) M6c Well, both have undergone a series of growing up

Native speakers well-acquainted with the phrasems a series of changes (or even, to undergo a series of changes) and a/the process of growing up will immediately recognize the blend of these two phrasems in the writer's sentence. Following are several other ESL examples with the familiar phrasems out of which the blend is constructed suggested below the students' sentences.

M11 War means nothing but a heartbreaker.

[War is nothing but a heartbreaker]

[War means nothing but heartbreak]

M13 He has a long mustache that has never shaven for a long years.

[for a long time]

[for many years]
K16  However, KCC is on the hill next by a Diamond head ...


G9  But only few percentage of Americans do go to church every Sunday.


G11a That's exactly the way our parents expected us to do.


G15a ... most of them are willing to give up what they have in Hong Kong and to start a difficult life on a strange place in any time of their age


Fay has observed that the one thing syntactic blends have in common is that the phrases out of which the blend is created usually have the same meaning (1982:163). The blends collected from the students' writings generally bear this out, although occasionally the resulting blend does not seem to have the same meaning, as in these HCE examples:

M14b She often complain about not having to save money anymore because ever since I move in to her apartment her salary is barely enough for the two of us.

[not being able to save money]
[not having money to save]
[?not getting to save money]

K2 I really don't mind writing if it takes me to become a nurse.

[if it takes that for me to become a nurse]
[if it leads me to becoming a nurse]

Fay distinguishes substitutions blends, in which a word or phrase from one "target" intrudes into another (If I ever get my hold
on them), and splice blends, "defined by the concatenation of either part or the whole of one target with part of the other" (Who is it that?) (1982:168). Cohen (1974) identified four basic kinds of substitutions which, through the blending process, become sources of idiomaticity: substitution of (1) antonyms, (2) synonyms, (3) words from the same semantic field, and (4) phonetically similar words. Some substitution blends in the student papers may have been overlooked, unfortunately, because they may have been identified as "wrong word" errors and excluded from the data, especially if a phrasem was not recognized during analysis as either a source of the error or as the target. In practice, it is not easy to place each blend squarely into one or the other of Fay's or Cohen's categories, many of the errors being interpretable in alternate ways, as indeterminate or as complexes. Therefore, no attempt was made to count the different types for this study, although such an analysis would be well worth doing. Altogether, there were 53 blends extracted from the ESL writings and 28 from the HCE.

Blends created by HCE writers generally were not as idiosyncratic as those composed by ESL writers. As can be confirmed by
examining the examples above, ESL blends are clearly errors, certainly unidiomatic and usually ungrammatical as well. On the other hand, most HCE blends had a familiar ring to them—they are commonly used among HCE speakers and some are even heard in the casual speech of other native speakers of English as well. Some reflection is often required to determine what criteria a reader is applying when identifying an HCE blend as an error.

M5a the areas that have been destroyed by the earthquake will take millions of dollars in order to look like how it was before
[to look like it looked before]
[look how it looked before]
[to be how it was before]

M9 things that causes drug users to do drugs are peer pressure from the people they are around with
[people they are around]
[people they run (or hang) around with]
M26 The past of Nick Adams brings him to being the person that he is now, as well as the present and what he is faced up against.

[(be) faced with]

[(be) up against]

K10 At McKinley High School there was actually no coffee shop where public goes to eat for lunch.

[go for lunch]

[go to eat]

[go to lunch]

[?go to eat lunch]

G3a For example, the same two companies that are far apart can be able to fax images to each other on how the final product should look like.

[of how the final product should look]

[of what the final product should look like]

The importance that syntactic blends, which are normally considered a type of blunder, have for the study of idiomaticity is
that some types of blends seem especially prone to happening, and the more they happen the more familiar they become--i.e., they tend to become idiomatically acceptable on the basis of frequency and familiarity. The HCE examples above seem to be examples of the sort of "entering wedges" which precede grammatical change. Pawley and Syder (1983) have catalogued a number of vernacular structures from New Zealand speech data (but having similar counterparts in American vernacular English as well) which participate in the process identifiable as language change in progress. Closer attention to syntactic blends committed by HCE speakers may reveal a portion of grammar emerging in similar manner.

5.6.2.2. Grammatical errors.

Generally, a phrase which violates general grammatical rules is unidiomatic as long as the phrase resulting from that violation is not a conventionalized idiom. That is, except for those specific idiomatic expressions which are ungrammatical but happen nevertheless to be conventional idioms in our language (*be that as it may, every now and then, it's me, that long nose of his, many a boy is, etc.*), any
expression which is ungrammatically constructed will simultaneously be unidiomatic. Lack of idiomaticity due to ungrammaticality needs no further discussion except to make this one point: that just as it is often difficult to determine the grammaticality of a phrase, so is it often difficult to agree upon the idiomaticity of a phrase which is on the borderline. The source of the problem is complex. Some of the more obvious reasons we might suggest are: (1) we do not yet have grammatical descriptions which we agree describe perfectly the actual grammars we use, nor do we have perfect descriptions of idiomaticity, (2) the actual grammars we use are not the same in every detail and neither do we use the same memorized phrases, (3) grammars, lexicons, and phrasal units are always in the process of change, (4) there are probably always some details of grammar which are indeterminate for any speaker at any point in time, (5) the means we employ to infer grammar from performance and intuitive evidence may be imprecise or invalid.

The counts of strictly grammatical errors as compared to errors which involved idiomaticity shows that one of the main reasons for lack of idiomaticity of student writing is its ungrammaticality.
As a demonstration of the ambiguities between the strictly ungrammatical and the grammatical-but-unidiomatic, examples of the basic writers' problems with comparison and the use of comparatives and superlatives will be presented. These problems can be considered strictly grammatical only by adopting a prescriptive orientation toward the grammatical rules for comparatives. Attempts at describing actual usage of skilled users of the language reveal that rules are much more complex than those prescribed and that description is facilitated by recognition of phrasal units and formulas in the inventory of the unpredictables in the language. The errors of unskilled writers serve to highlight those areas where knowing a general rule is not helpful.

Since a standard exercise for English composition classes is the "comparison and contrast paper", students very early find themselves struggling with the idiosyncrasies of English comparatives and superlatives as well as with other peculiarities of making comparisons. The prescriptive rules governing comparatives seem fairly straight-forward: in addition to the absolute forms, adjectives and adverbs have different forms to indicate relative
degree. The comparative degree, which is used to compare two entities, is indicated by the suffix -er or by a preceding more or less while the superlative degree, which is used to compare more than two, is indicated by the suffix -est or a preceding most or least. More/less and most/least are used for adjectives with more than two syllables, and for adjectives and adverbs ending in -ly. There are also a few suppletive forms like good, better, best and bad, worse, worst which must be committed to memory.

Although prescriptive rules may purport to reflect actual rules, attempts at describing practice reveal that much of actual usage is either arbitrary or guided by several factors not easily accounted for in most descriptions. For that matter, the prescriptive rule above which refers to syllables is not easily incorporated into most current formal grammars which insist on autonomous syntactic and phonological components. Prime candidates for idiomatic treatment are the two-syllable adjectives which are divided between those which take -er/-est and those which follow more/most. Barber (1964) listed several forms "in the process of change": common, cloudy, fussy, quiet, cruel, subtle, clever, profound, simple,
pleasant. He noted in addition that some one syllable words are beginning to take more/most: crude, plain, keen, and forms of the type more well-informed and most well-dressed exist alongside better-informed and best-dressed (131).

Van Lancker (1975) names "familiarity" as "one of three or four factors determining whether the terminational method (adding -er, -est at the end of the word) or the periphrastic method (placing more or most before the word) is used" (142). Admitting familiarity as a factor in the choices speakers make recognizes the idiomatic nature of uses of comparatives and superlatives in English.

In addition to the number of syllables and familiarity, the phonetic shape of the end of the adjective may have an influence on choices. More clever and more clear may be preferred by speakers over cleverer and clearer, for example, for the phonological or physiological reason that they are easier to pronounce and hear. It is doubtful that the choices users make between the alternatives offered by comparative forms can be adequately accounted for strictly in terms of morphological or syntactic rules.
A total of 59 errors in comparison were made in the student papers collected for this study, 17 by HCE-speaking writers and 42 by ESL writers. Of the HCE errors, 8 were the sort that could be considered simple grammatical errors while 9 involved idiomaticity. 23 of the ESL writers errors were grammatical while 19 were idiomatic. Idiomaticity seems to have been almost as much a problem for these writers as was grammar. The most common error counted as grammatical was double-marking of comparatives—*more wider, more cooler, more closer, more quieter, more quicker,* and *more better.* Using periphrastic form with single-syllable adjectives was also counted as grammatical error rather than idiomatic, in an effort to accommodate the prescriptivist viewpoint: *most safe, most wild, more clear* (two instances), and *more strict.*

Phrasematic problems with comparison and comparatives were more varied. Some students failed to complete a comparison with a *than* clause, as is required in edited English where the objects of comparison are not immediately obvious. It may be that in casual English, there is greater leniency in the optional omission of overt expression of *than* clauses which are recoverable by the hearer from
context. It is also possible that this use of more without a following than clause is an idiom in HCE, where it seems to mean nothing more than much, or, something like largely, generally, or for the most part. If asked "More than what?" the best answer is probably "more than not".

ESL

M1a Both Hawaii and Samoa depend more in the ocean for living and much more about our natural resources.

M1a The customs are more alike, but Hawaiian has lost their custom and culture.

HCE

G5 Second of all, the Big Island is more of a slower pace.

G5 Honolulu is more of a fast pace.

G5 Its a place more for tourist.

The construction of sentences using than caused other problems also. The ESL student who wrote the following may not
have known that *than* requires a preceding comparative form, or was unfamiliar with uses of the phrase *rather than*.

G4a Instead she told them that it's good to run the business by family than others. [it's **better** to run the business by using family members than by hiring others, or, it's good to run the business using family members **rather than** hiring others]

Colloquially, phrases introduced by *unlike* seem preferred over *than* constructions in sentences like:

G11a So, I always tried to overcome my shyness and [be] a little more assertive unlike my sister.

Students' propensity to select the trendy over the more conservative alternatives often required in academic writing can be attributed naturally to their greater familiarity with the current colloquial preferences. These choices thus reflect the idiomaticity of the variety of English spoken by the student's peer group.
Other errors, however, seemed to exhibit the students' unfamiliarity with required phrasems. Students who produced sentences like the following seemed to be unfamiliar with the phrasem of the pattern as ADJ/ADV as, for instance, as tall as, as tired as, as quickly as, etc.):

ESL

K17a KCC does not as McKinley has that much of proms and night dances.

G9 Although United States are twice bigger that doesn't mean that they are better because the population is about the same.

Besides twice as big as or two times as big as, there is also a possible choice of two times bigger, of course, in which case the writer would not have had to use the as ADJ/ADV as phrasem at all.

HCE

M5 ... but obviously thinking about the girls he would marry was not quite interesting as his trout fishing.
M18a On the other hand, if you're poor you have to work as harder you can.

G6 For they were not as strong or physically big in size like the ball players today.

Experience in hearing these types of errors spoken by students and the preservation of these errors by the student writers in the oral reading of their papers confirm that these are not accidental errors of graphic representation.

Attempting to use the compound words well-mannered, well-behaved, well-known caused two students problems. The status of the comparative forms of well-PARTICIPIAL compounds may be no more stable today than when Barber (1964) noted ambivalence concerning them.

ESL

M6c Today's youth are less mannered and liberated, whereas, the youth of the early sixties were more disciplined and well-behaved.
M6c On the other hand, youth in the early sixties were more mannered, more disciplined, and well behaved.

M10 Samoa is one of the most well-known island of the South Pacific.

Use of such words as different, difference, same, compare, contrast, etc. require knowledge of particular particles or prepositions which must accompany them in various contexts. Different from is not used in the same context as different than and difference between, and neither are same as and similar to. In addition, the uses of other, each other, one another, both, either, neither, etc. cause difficulty for non-native speakers when attempting to make comparison statements in writing. The following examples show a variety of alternatives chosen by the basic writers which somehow went awry.

ESL

M2a So how come people find it hard to afford living in Hawaii contradictory to how easy they can make a living in CA.
M2a California as a state is one of the top rich states among the other american states.

M2a ... California has the largest number of tourist any other Nation might have.

M15 But there was some difference from here.

M15 It was exactly same price with diet pepsi.

M15 I tried to find the differences of goods.

K8a Both, KCC and My High School have difference in Campus Appearance.

G9a [last year] Aloha studium had an average of about 45,000 people per game, Whereas about 32,000 people compare to this year.

G17 Everyone lives his or her life differently from one another, especially the people who live in communist country comparing to the people who live in free world.

M12b Their differences include the neatly and sloppy.

HCE

M14a They talked about how the places the son had been were like no place in American compared to America.
The school procedures between K.C.C. and Pearl City High School is very different.

My High School & KCC do not look alike from the activity they do in school. [are not alike in]

The writer of M14a crossed out the phrase which actually expressed what he had intended, and by using compared to instead, produced an expression to which native-speakers would give an entirely different interpretation.

A final interesting piece of evidence of the consequences which the noncompositionality of phrasems has on student writing is displayed in an HCE writer's creation of the phrase underlined:

but the team with the most mental toughness might be able to pull out a victory.

Why not "the team which is mentally the toughest" or "the toughest team mentally" or "the most mentally-tough team"? An explanation in terms of idiomaticity would claim that the phrase
mental toughness is a current conventionalized phrasem which is syntactically frozen and cannot be decomposed, or its constituents recomposed, in this student's idiom.

It is hoped that this discussion on comparison has succeeded in demonstrating not only that idiomatically acceptable usage in making comparisons involves knowledge of various phrasems as well as of grammar, but more importantly that there is yet much work to be done in clarifying the relationship between ungrammaticality and lack of idiomaticity.

5.6.2.3. Lexical and morphological errors.

Another way a phrase may be rendered unidiomatic which will not be discussed beyond this paragraph is that an expression is unidiomatic in that it contains an isolated "wrong word". (Those lexical selection errors which involve conventionalized collocations will, however, be discussed later.) Illustrations of isolated wrong words are: the use of surrounded in the sentence Samoa are surrounded with very tall mountains where covered would have
been semantically correct, *banished* where the writer meant *vanished* (*With in six months, her excitaments or attitude towards her job just banished*), *composition* where *component* would have been appropriate (*Two makers didn’t indicate the main composition which makes different taste*).

Instances of a writer using the wrong word form (i.e., morphological errors) are: *understanding* where the form *understandable* is required, as in *They had a slight argument but that is understanding if you both don’t agree on one thing*, and the nonce creation *adventural* in *The adventural young son started his trip*. Morphological errors are extremely common and are a major cause of lack of idiomaticity in a phrase, but since they involve single words they are defined out of this study, except for those which are conventionalized, for instance, those "wrong word form" errors which involve a choice between a gerund or infinitival construction.

Errors in lexical selection were especially copious in the ESL writings, there being 182 ESL wrong word choices as compared to 75 for HCE writers. A similar disparity is evidenced in the use of wrong
forms of morphologically complex words. Whereas HCE writers used 17 wrong forms, ESL writers used 75, many actually nonce creations. Semiproductivity in the morphology of synthetic parts of the language seem to cause new learners plenty of grief.

5.6.2.4. Compositional and stylistic errors.

Perhaps the most controversial type of unidiomatic expression included in this study are those phrases which violate idiomaticity by reason of not conforming to demands of composition and/or style. A sentence like Actually, basketball is trying to be used to bring all the poor communities closer together (meaning An attempt is being made to use basketball to bring all the poor communities closer together) may be analyzed as containing a grammatical error in that a selection restriction on try which requires it to have an animate agent as subject is violated. A teacher may comment, "Who is trying? Can basketball be said to try?" However, alternatively, basketball is trying to be used can be analyzed as being analogous to basketball is being used in which the subject basketball is not the agent but the affected (i.e., it is a passive construction).
The reason such an expression appears in a student’s composition is that this sort of phrase is used in normal colloquial speech, possibly due to the latter analysis. That is, colloquial language permits $AGENT$ try to use $AFFECTED$ to be passived as $AFFECTED$ be trying to be used (by $AGENT$), while the standards of Edited American English do not. Whether this is a problem of grammar, style, logic, or idiomaticity depends, of course, on how these aspects of written language use are defined. Lack of idiomaticity caused by violation of rules of composition and style were included in this study because they reveal the inability of basic writing students to distinguish colloquial idiomaticity from the idiomaticity of the academic, or written, registers. (For a discussion of the relationship of idiomaticity ["speech genres"] to style, see Bakhtin 1953).

The following ESL sentence exhibits one kind of error which some English composition teachers consider to be a matter of style. It can also be discussed in terms of idiomaticity:
If they can, most of them are willing to give up what they have in Hong Kong and to start a difficult life on a strange place.

Why is the second underlined to, if not outright ungrammatical, at least unidiomatic in some way here? Could it be because willing to is an infinitive phrasem, similar to a phrasal verb, and as a multi-lexical unit cannot be decomposed, nor have copies of its parts repeated? Many native speakers of English would probably feel that the sentence is not right somehow, but if we can offer the new learner of English a more specific explanation as to why, they may avoid the temptation of assuming it is just a matter of style.

It is most interesting that the error types which HCE writers committed more than ESL writers were all colloquial types (see section 5.6.1). In addition to the general category "colloquial" which included a variety of errors, such as inappropriate use of deictics, quotations, shifts in point of view, direct address, slang, etc., separate listings were also made for such colloquial usages as Hawaii Creole English ("Pidgin English", as Hawaii's residents call it), topics, tags, jargon, the use of which as an "all-purpose connective"
(Shaughnessy 1977), the use of there's or there is with a plural complement, and the use of the stigmatized redundancy the reason why ... is because ... As it turned out, HCE speakers used all these colloquialisms inappropriately in their academic writing more than ESL writers, and these colloquial errors were the only ones which were committed more often by HCE than by ESL writers (309 HCE to 222 ESL).

The trials students endure in learning to write academically can partially, but only partially, be attributed to the "natural" process of language change. It has long been a truism in linguistics that language change is inevitable, that there will always be syntactic rules in the process of change whose status in the language is indeterminate as to grammaticality or acceptability. On the other hand, it has only recently been acknowledged that as modern advances in communication have accelerated the adoption of "standard" languages throughout the world, the direction and nature of language change has been greatly complicated. The impact of such inventions as writing, the printing press and its successors, educational institutions and governmental support of universal
education have all left their mark on language and have undoubtedly
at times turned the tide of language change (See Joseph 1987 and
Coulmas 1989 for fuller discussion). Although there is no room here
to delve into this profoundly complex topic, it is critical that teachers
recognize that formal written language has its own idiomaticity,
especially that special register of written language which is expected
of writers in an academic setting. In a "Calvin and Hobbes" cartoon
which appeared in the Honolulu Star-Bulletin recently, the comic
strip writer, Bill Watterson, comes close to the truth of how the
demands of academic writing must appear to basic writers:

Calvin: I used to hate writing assignments, but now I
enjoy them. I realized that the purpose of writing
is to inflate weak ideas, obscure poor reasoning,
and inhibit clarity. With a little practice writing
can be an intimidating and impenetrable fog! Want
to see my book report?
Hobbes (reading): "The dynamics of interbeing and
monological imperatives in Dick and Jane: a study
in psychic transrelational gender modes."

That the HCE speaking basic writers made so many errors of
the sort which are acceptable as colloquial language but unacceptable
only by application of an "edited English" standard is evidence that
these students are not lacking in a means of linguistic expression for
their own oral purposes. They need only to learn to use the college writing idiom, which means conscious attention to the many special conventions demanded of written language—coherence, explicitness, conciseness, appropriate tone, formality, avoiding redundancy and triteness, parallelism, and many others. The requirements of "style" associated with the written registers are integrally related to idiomaticity, and any presentation of them without reference to formulas and the like cannot hope to be complete.

5.6.2.5. Logical and semantic errors.

Unskilled writers may be affected in two different ways by the assumption that language ought to be logical and semantically compositional. In one case, they may have trouble composing sentences which are logical, or "make sense". Apart from accidental errors, problems crop up in the assigning of proper thematic roles (languages can be very particular about what can be an agent, for example, as demonstrated in section 5.6.2.4), in choosing words which are semantically compatible (including what can be included in the same list of more than two items), and in expression of logical
relationships of conjunction, disjunction, entailment, and especially negation.

Examples from student's attempts to use conjunction appeared to be appropriate ones through which to discuss semantic and logical problems. We might examine first the sentence And he wore battered hat and black bag carrying clothes, books and articles. The semantic incompatibles conjoined are of course wore and bag, since a bag is not something worn. In addition, the strangeness of the item articles in the list of contents of the bag may be because the word is a supernymic one. It is a more general word which includes clothes and books.

Such errors, not always easy to identify as semantic, are sufficiently familiar to frequent readers of student papers, and will not be further exemplified since they are not essentially idiomatic.

Other errors of conjunction, also not demanding of idiomatic explanation, may have better pragmatic than semantic explanations. For instance, I'm the only child and we have a cat (juxtaposition
pragmatically implies identity or close relationship) and the

describing of a man as unique and unusual (involves scalar
implicature, i.e., simultaneous occurrence at two points on the same
scale is impossible). Still others are simply false to fact, as ... in south
California Sandia it warm and sunny day and night. A final example
of errant conjunction which does not require special idiomatic
explanation is the confusing of universes of discourse, as in this
sentence which describes a character in a story: Nick is a nice, young
man with a little sleeping problem, that he and I believe will go
away with time.

The errors in the use of logical operators which will exemplify
idiomatic errors below are different. Note, by the way, that sentence
connectives (which will be discussed in 5.6.2.6) were analyzed
separately from negation and the operators and, but, and or when
used to relate simple terms, which were grouped together. The total
number of errors, idiomatic or other, are:

<table>
<thead>
<tr>
<th></th>
<th>ESL</th>
<th>HCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>negation</td>
<td>14</td>
<td>7</td>
</tr>
<tr>
<td>and, but, or</td>
<td>48</td>
<td>17</td>
</tr>
</tbody>
</table>
Nine of the 14 ESL errors in negation involved idiomaticity, while six of the seven HCE errors did. Five of the idiomatic negation errors, three ESL and two HCE, were in the adding of the tag \textit{or not} at the ends of sentences containing conditional (\textit{if} or \textit{whether}) clauses (\textit{to see if school is hard or not}; \textit{I am going to call and tell them if I am coming back or not}). Like this tag, most of the errors in negation were in the use of forms colloquially acceptable:

\textbf{ESL}

M6c We maybe changed in outlooks and age \textbf{but values not}.

M15 It tasted me \textbf{not so good}.

G4 Peter hates when his girl friend is late for \textbf{not even} 10 minutes.

\textbf{HCE}

M5 \textbf{Both of them couldn't} sleep so they figured they should talk ...

G2 \textbf{... the teachers are always not} there for you at all times.

Of the errors in the use of logical operators, seven involved idiomaticity, for instance, the misuse of or ignorance of "correlatives", formulas for correlating conjuncts. Each example below requires a
particular formula like *one is X and the other Y, ranging from X to Y*, *which is X and which Y*:

M1a Samoa are divided into two group, One is Western Samoa and American Samoa.

M6d This country had undergone a series of problems ranging from its political and economic instability.

M10 It is very hard to tell which is an American Samoan and a Wester Sam.

This section began with the claim that the demands of logic and compositionality affect student writers in two ways. The foregoing was a discussion of the problems students have in attempting to compose sentences which conform to those demands. In the other case, students may find themselves at odds with English conventions themselves which are not "logical" and do not "make sense". Some of the conventions which may well seem illogical to non-native writers of English actually take the form of rules of grammar. For example, "grammatical" and "notional" concord can come into conflict. Should the novice writer use *is* or *are* following *a number of Noun plural*? To
add to the confusion, the student is apt to hear his native-speaker associates use "proximal" concord in casual speech, where the verb is chosen to agree with the form of the closest noun preceding it. That articles are a "compulsory category" in English, forcing students to distinguish between definite and indefinite, mass and count nouns, as well as singular and plural, seems arbitrary to many students whose native languages do just fine without them. Furthermore, except that concrete nouns tend to be count while abstract nouns are usually mass, which specific nouns are considered to designate countables can be highly arbitrary, causing student writers much grief. ESL writers produced a total of 98 errors of number agreement and HCE writers, 56 errors.

5.6.2.6. Pragmatic errors.

Some types of errors committed by students may be considered pragmatic errors. A few which did not involve idiomaticity, but referred to such devices as conversational and scalar implicature, were presented in the discussion of semantic errors. Others of these do involve phrasems of various sorts.
Euphemisms, for example, are idioms with a pragmatic function—to avoid social offense. Euphemisms are very scarce and only two errors in their use occurred, and only in the ESL writings:

M20 He is partly at fault so he must take care of his responsibility, he shouldn't have done it, if he can't keep the child that they conceived.

Done it, obviously meaning 'had sexual intercourse', is introduced abruptly here without antecedent.

K8 She will be a mother soon ... She hold is a baby girl.

The appropriate euphemism for 'be pregnant with' which the writer did not seem to know is, of course, be carrying.

Most of the phrasems classed under "formulas" have some pragmatic point. Conventional implicatures, which involve discourse deictics (such as connectives, floor holders, rejoinders, etc.) and social deictics (politeness formulas and the like), and routines, which include apologies, thanks, greetings, and farewells, are the formulas which have been discussed most extensively as pragmatic. In
addition, conversational implicatures also take advantage of the existence of formulas and the expectations which come with their use. Because this study focused on a written corpus, very few examples of conversational phrasems were gleaned. Most pragmatic phrasal errors were in the use of discourse connectives or were colloquialisms inappropriately used in writing. Since colloquialisms were discussed as errors of style, or register, in section 5.6.2.4, the following examples will be of errors in the use of connectives, chosen from a total of 58 ESL errors, and 37 HCE errors.

The problem of sorting idiomatic from non-idiomatic errors in the use of connectives seemed insurmountable at first glance. Taking one case, how was I to decide whether the overuse of but was a problem of idiomaticity? Should I have said that but is unilexical so it is irrelevant to this study? Obviously not, since overuse of but results from the non-use of more appropriate alternatives, nevertheless, none the less, however, on the other hand, etc., some single words, others phrasems. Since it was clearly meaningless to treat unilexical ones differently from multilexical ones, all errors in the use of sentence connectives were counted as idiomatic errors.
Novice writers tend to avoid complex sentences altogether. Even one semester in college, however, is enough to convince most students that they must attempt to lengthen their sentences, at least some of the time. Basic writing instructors frequently observe that the earliest strategy adopted by many beginning writers is to pick a favorite connective like and, but, or even as, and to use it liberally. The following ESL basic writer seemed to like but, which is perhaps an improvement over the copious use of and, which many students remember being corrected for in childhood:

M1a But I would like specific compare Hawaii with American Samoa. ... The green color of the Islands just like a green pastor, where the sheppered took the flock to fed. But Hawaii's beautiful nature color are fail away. Men has to water the green to kept the green grow. But in Samoa, everything is green. The beaches here in Hawaii are so beautiful just like beaches in Samoa. But both the water are warm.

The costom are more alike, but Hawaiian has lost their costom and culture. But we still upheld our costom and culture.
Another ESL writer preferred *as*:

M2 *As* the son came back he didn't have too much to say but I have missed Home.

M2 The father advise was one of the main important steps toward the trip preparation *as* he thought.

M2 *As* the father advised his son to buy a suitcase instead of a backbag and to take a raincoat with him.

M2 England was like no place in america *as* he said.

M2 He phoned his father from Europe *as* he discribed the Notre Dame, the Eiffel Tower, Edinburgh, Blair Castle, hotel rooms, meals, people ...

Apparently, students employ this strategy until they begin to add a variety of connectives to their verbal repertoire. The writers of M1a and M2 had both been taught English only in school in their native lands and apparently had not gone much beyond sentence level instruction. Students with more experience with longer discourse display familiarity with phrasal connectives. Their papers, both ESL and HCE writings, still contain many errors, however. They
continue to exhibit the use of simple conjunctions where phrasal formulas are more appropriate and attempts at idiomatic connectives which often go awry. *As a result* was tried by many basic writers, but not always successfully. The first two below were by ESL speakers, the last by an HCE speaker:

M12 If suddenly a natural disaster destroys the partner's prosperity, he or she can't rely upon the partner; they divorce. *As a result*, divorce affects the children.

G13 Because it costs less to import food to Las Vegas, food are very cheap at that city. *As a result*, people often eat first and then they would start gambling.

M5a [following a paragraph which describes the San Francisco earthquake of 1906, ending with *City Hall, libraries, courts, theatres, and restaurants had disappeared*] *As a result*, earthquakes are very dangerous.
Not like, not just, and not only (that) were used often as sentence initial connectives by ESL writers, influenced, no doubt, by colloquial usage:

M18  Not like married women that they have broad understanding in to baby's.

M29  Not just because of English is my second language and not used to it.

K8a  Not only that here in the cafeteria, it so nice I can't imagine the view.

K9a  Not like the RHS teacher they only have their own classroom,

Nevertheless, or else, even though, by the time, even if, first of all, on the other hand, so that, at a time when, at one time, along with, in which, as well as, and for example were among the phrasal connectives used erroneously in the papers. In addition, students used such unfamiliar (to the researcher, at least) phrases as in other time (for at another time) and as like before (for as usual, as in the past, as accustomed), which may be nonce creations or may be newly accepted HCE. The HCE phrase but yet also occurred several times.
As a final example of problems with connectives, whether unilexical or phrasematic, a rather long passage produced by an HCE writer will be used to illustrate a common sort of problem which occurs when writers are pressed to combine short sentences and to use discourse transition markers. This writer appears to have a stock of connectives which he picks from with only a vague feel for their effects on the direction which the discourse then takes:

G3  Obviously, basketball is still being tried in other countries. Actually, basketball is trying to be used to bring all the poor communities closer together. Because basketball requires only five players per team, for basketball will take less time to complete a game.

On the other hand, baseball is already being played around the world, and baseball has a lesser chance of being successful in poor areas. Because baseball requires more than nine players per team, and takes longer to complete a game.

Therefore, the pace of these two different games will vary. In this case, basketball is a game that is much quicker. Because of the distant the players travel from end to end of the
court. For example, a player getting the rebound can throw the ball to his teammate running towards the opponent basket, for basketball requires only one minute to score six baskets.

Whereas, baseball requires a lot of time to score points on the board. Because baseball depends on the strength and mental toughness of each team. For example, [etc.]

Here again, it is clear that the difference between unilexical and multilexical form is unimportant. There are connectives which function in just the same way in either form. What is important is that discourse deixis in the written mode as well as in the oral involves a very special repertory of linguistic forms which must be part of every skilled writer's stock.

5.6.2.7. Phrasematic errors.

Because it is impossible to discuss all phrasems fully in one dissertation, only two of the many types of phrasematic errors found in the student papers will be discussed at length using the terminology and conceptualizations presented in Chapter 4, namely,
(1) collocations and (2) phrasal and prepositional verbs. Since the latter category is one of the cross-cutting ones, however, discussion of these phrasal and prepositional verbs will include all three major types of phrasems--idioms, collocations and formulas. Also, a representative type from the group of formulas, namely, the connective, has actually already been discussed immediately above in section 5.6.2.6.

No other subtypes from the category "idioms" was chosen for fuller discussion because there were very few full or figurative idioms, or attempts to use them, in the student writings. This accords with observations commonly made by teachers and researchers that new users of a language avoid using idioms in writing. Even with native speakers, only 51% of students aged 18 to 22 scored in the Competence Range on a test for knowledge of idioms, according to the study, "Performance of University Students on the Fullerton Subtest of Idioms". Results of the tests showed that as chronological age increased, the percentage of students attaining the "Competence Range" increased (Brasseur and Jimenez 1989).
It is not surprising, then, that students who do not feel fully competent as writers would not be prolific in the use of idioms. Attempts to use or manipulate idioms, as mature language users do, sometimes prove disastrous. Among the ESL papers was one in which the student's attempt to vary the moral, "Slow and steady wins the race" (Aesop's fable) went completely unrecognized when he closed his paper with "Also, slow and steady wins the my English problem". The teacher suggested two corrections of this sentence: (1) "I should plan for slow and steady wins", and (2) "Planning for quick and great gains in too short a time is part of the problem with my English."

Collocations warrant more than a mention. Generally, collocations have to do with adjacent pairs--what can occur together with what. However, the term "collocation" is used more restrictively here, to name a specific class of phrasems, consisting of "unique collocations", "restricted collocations", and "common collocations". No unique and restricted collocations were found in the student papers, but such collocations are actually quite rare in our language anyway. We will therefore be looking at uses, attempted uses, and non-use of
common collocations where they would be expected. At times collocations may not be adjacent, modifying phrases or words sometimes intervening between collocates. This is especially frequent when the common collocations are noun/verb collocates, but they are included if they are a common unitary usage. (The words which constitute the pair will be underlined in all examples.)

Phrasems under the heading of common collocations take various syntactic shapes. One easily recognized set are compound nouns, composed of two adjacent nouns which form a semantic unit. Some grammarians define compound nouns to include all and only forms made up of two lexical items (of whatever categories they may be) which are combined into one lexeme which functions as a noun. Since we have defined unilexical forms out of this study, we are including in this discussion only compounds made of two separate lexical items functioning semantically as a single noun.¹⁰

The term "compound" is used here to refer to noun + noun units only and collocations of words which belong to other categories are called "common collocations" without creating any additional
terminology for subclassifications. There seemed no necessity for naming all the other possible collocations for the time being.

Since compounds are one of the cross-cutting categories, not all compounds are collocations, of course. Many are idioms and a very few fall into the subset of formulas (see examples in section 4.2.4).

Of close to a hundred noun + noun compounds which appeared in the student writings, all but eight were correctly used. Many were surprisingly sophisticated combinations: *winery operation, job opportunities, consumer goods, accountant controller, transportation system*, etc. Students apparently acquire these compounds as wholes as they are introduced to the relevant concepts. The newly learned compound, unfortunately, may not always be used correctly at first, as evidenced in these examples by two HCE writers:

M25a Asians and Asian-American students would have a *motivation factor* or make sure they obtain the education in order to be a success.

K19a The class size, physical and amount of *student ratio* is the same.
Occasionally, a novel compound crops up: *shoplifting people*, *California state*, *abortion operation*, *appreciation word*, *business companies*, and *business parents* (where the student meant *working parents*) sound strange to me. Or, at other times, students come close but add the wrong derivational morpheme to a correct or similar stem: *working habits* instead of *work habits*, for example, or *tendency count* for *attendance count*, and *place test, replacement test*, and *placement testing* for *placement test*. *Shoplifting people* (*people who shoplift*) and *appreciation word* (*word of appreciation*) represent a more general problem of semi-productivity.

Some forms were on the borderline for me: *?infrastructure programs* (unfamiliar) and *?heart problems* (*heart trouble* seems more idiomatic), for examples. We can well empathize with students' struggles with idiomaticity when even experienced native-speakers find it difficult to judge on the acceptability of seemingly indeterminate combinations.

Similar problems occur with other collocations besides compound nouns. ESL students committed 47 errors in collocations
while HCE students made 13 such errors. If idioms are subject to the vagaries of fashion, common collocations are probably even more so, and what may be a delightfully fresh collocation today may be an annoying cliché on every lip tomorrow, only to disappear completely the next month, or year, or century. For instance, it seems to have become trendy to return to saying *make one crazy* where only *drive one crazy* had been acceptable for years.

Most composition teachers have probably at times pondered whether to accept a semantically appropriate, though unidiomatic, collocation, or to insist on the more familiar, though possibly trite, form. Similar difficulties were experienced in listing unidiomatic collocations. Take: *the sky twingle up high*. Spelling and grammar aside, is the collocation of *sky* with *twinkle* all right, or should it be *stars twinkle*? How about *cheap stores, satisfy my stomach, town library, lazy side, and cook us pancakes*? Discount stores, *satisfy my hunger, public library, lazy streak, and make us pancakes* seem more idiomatic to me, but should students be expected to conform to my standard of idiomaticity? One daring student created all of the
unusual collocations, or unusual uses for familiar collocations, in the sentences below:

M6b Insurgency, as famously uttered by the media, has been a major issue in the Philippine government.
M6c ... they always respond appreciation word in gratitude ...
M6d What are our endmost part in this domineering condition?
M6d Young person who will represent us must be knowledgable, well burst and can comprehend in making constructive solutions.

In some cases, papers assigned on the same reading material yielded interesting evidence of the collocations which were lacking in the students' verbal repertoires. Several papers, for example, exhibited novel alternatives where the most expected phrase would have been something like didn't say anything or said nothing:

M1 she did not even said any words.
M8 ... but she one time responded by saying no talking.
M8a She cannot think of anymore talking ...
The expression *have an abortion* was also unfamiliar to many students, who produced such alternatives as *goes for an abortion operation, take abortion, get the operation, have the operation, have the abortion* (on first mention, when definite article would otherwise not have been expected), and, most misleading of all, *do the abortion*, which has an entirely different meaning altogether.

"Support verbs" are involved in several faulty collocations. These are very common verbs like *come, do, get, go, make, stay, run, take*, and so forth, which may be combined with other words to form a phrasem with a unitary meaning. Being well aware that the most common verbs in the language are the most likely to participate in idiomatic collocations, Makkai used a list of 100 of the most common verbs to build phrasal verbs by adding adverbs or particles, and tested them for productivity (1972). The lexicalized combinations of support-verb-plus-other-words which are actually utilized in English are known to all native-speakers, but speakers of ESL must learn them gradually as they increase their total lexical store. Meanwhile, they make mistakes like the *do the abortion* example above. Others which occurred were *get any solution* (instead of *find any solution*)
experiences he intended to do (instead of have), and do the easiest way out (instead of take).

HCE speakers, on the other hand, usually have adequate control of an ample supply of support verb collocations, but some HCE support verbs may not be identical with those considered to be standard American usages and, therefore, required for academic writing. Hence, the peculiar "local" flavor of cook pancakes instead of make pancakes, and Hurry, it's time to eat breakfast as compared to it's time for breakfast or it's time to have breakfast. Examples of support verbs used in phrasal verb phrasems will be presented in the next topic of this section.

Circumlocution was sometimes the last resort when a required collocation was not available, as in staying in their own house (for staying home) and making a discussion of the possibilities that they can go through (for discussing, or considering their options).

The type of unitary phrase in English which is most numerous is what has been called a verbal phrase by Vachek (1976), a phrase-
verb by Householder (1959), and a phrasal verb by Smith (1925), Makkai (1972) and others (Strässler 1982:16). "A phrasal verb is recognized by the structure Verb + Adverb" (17) and has a unitary meaning. A couple of examples given by Makkai are give in and take off. Others have preferred to call the preposition look-alike following the verb a "particle" (as in Particle movement) rather than an adverb when it functions as part of the verb. Whole dissertations have been written to describe phrasal verbs (Sroka 1975, etc.) and the collection of phrasal verbs into dictionaries has been undertaken by a number of scholars (Cowie and Mackin 1975, Courtney 1983, Sinclair et al 1989, and Spears 1993).

Phrasal verbs may function transitively or intransitively. The two examples above are of intransitive phrasal verbs since they do not require NPs to follow as objects. Strässler gives the intransitive examples to get up and to show off. They are easier to recognize than transitive phrasal verbs which have the same form as both prepositional verbs (another kind of phrasal unit) and free constructions of verbs followed by prepositional phrases. We might
grasp the differences by contrasting the three structures, using the words *look up* as constants:

- **look up a word**: phrasal verb, Verb + Particle + NP
- **look up a street**: prepositional verb, Verb + Prep + NP
- **look up a tree**: free construction, Verb + Prep + NP

Phonological, syntactic, and semantic criteria have been proposed to distinguish (1) phrasal verbs from prepositional verbs and (2) phrasal verbs from similar free constructions, but some doubt has been expressed as to the necessity of distinguishing (3) prepositional verbs from similar free constructions (Quirk et al 1972).

Strässler distinguishes phrasal verbs from prepositional verbs (Verb + Preposition + NP) by claiming, first, that the particle following the verb in a phrasal verb is semantically empty (*to rely on* and *to object to*) while the preposition following the verb in a prepositional verb is meaning-bearing. Secondly, Strässler notes that, syntactically, phrasal verbs permit permutation of the particle
and the object NP (to look up a word and to look a word up) while this is not possible with prepositional verbs. In the second point, Strässler concurs with Quirk et al (1972).

Phrasal verbs are distinguished from prepositional verbs by Quirk et al (1972) on the bases of phonology and syntax:

(a) The particle in phrasal verbs is normally stressed and, in final position, bears the nuclear tone, whereas the particle in the prepositional verb is normally unstressed and has the 'tail' of the nuclear tone on the lexical verb:

\[
\begin{align*}
\text{He called 'up the man} & \sim \text{The man was called 'UP} \\
\text{He 'called on the man} & \sim \text{The man was C'ALLED on}
\end{align*}
\]

(b) A syntactic difference is that the particle of a phrasal verb can often stand either before or after a noun, whereas it can only stand after a personal pronoun. ... prepositional verbs take ... personal or ... relative pronouns after the preposition, and admit ... an inserted adverb; phrasal verbs, on the other hand, have particles which can be separated from the verb by ... a noun or ... a pronoun (815-16).

A transitive phrasal verb may also be distinguished from a free construction of the same form, V + Prep + NP. Using Strässler's semantic criteria, we can contrast the form up in look up a chimney with look up a word (containing a phrasal verb). We see that the semantic content of up indicating 'direction away from the center of
the earth' is contained in the first while it is not present in the
second. In the second, up simply functions to give the phrase look
up a unique semantic specification as a unit, separate from look or
any other word. Syntactically, look up a word and look a word up
are acceptable, but look up a chimney cannot be permuted to look a
chimney up while retaining the same literal meaning. Generally, free
constructions seem to have the same syntactic constraints as
prepositional verbs. On the other hand, use of the pronoun it or
relative pronouns with this particular free construction seems
dubious, although insertion of an adverb is permissible:

?Shine a light in the chimney before you look up it.

?The chimney up which you look could be dangerous.

Look carefully up a chimney.

According to Quirk et al, prepositional verbs may be
alternatively analyzed as free constructions. "The two analyses can
be regarded as different, but equally valid and complementary ways
of looking at the same structure" (819). This may be true for
phrases with literal meanings, but it seems preferable to regard a
phrase which has a nonliteral meaning as a prepositional verb. Furthermore, viewing even literal constructions of this sort as free constructions obscures the perception of their semantic unity as well as the fact that they are familiar and frequently occurring units.

The difference between a prepositional verb and a free construction can be grasped by contrasting, as Makkai (1972) has demonstrated, the phrases *grandma came down with the suitcases* with *grandma came down with the measles*, and *the general went through with his tanks* with *the general went through with his plans* (146). Strässler's criterion based on emptiness of meaning in the particle can identify a phrasal verb, but the main distinguishing criterion for a prepositional verb seems to be semantic unity—in the first example, *came down with* as a unit means 'contracted' and in the latter, *went through with* means 'executed'. Testing for unity of meaning, preferably (though not necessarily) by finding a lexically encoded (one word) equivalent, can most often be applied. The contrasting test is not always usable since literal counterparts of a phrasal verb are not always available, as in the cases of #3 of Makkai's list below.
Makkai, who did not separate phrasal verbs from prepositional verbs, noted that verbs which occur with adverbs may be grouped into several different types in terms of literalness versus unpredictability of meaning (here referred to by the terms "an idiom" or "idiomatic meaning"):

1. The combination is a polylexemic literal constitute (e.g., *go away* as in *when did he go away?*)
2. The combination occurs both as a literal constitute and as an idiom (e.g., *come up* is a literal constitute as in *has he come up from the basement yet?* but it is an idiom in *what's come up?* meaning 'what has happened?')
3. The combination occurs only as an idiom (e.g., *give in* as in *he gives in too easily.*)
4. The combination has several idiomatic meanings in addition to occurring as a literal constitute; e.g., *put up* is a literal constitute in *put up those books on the shelf, will you?* and idiomatic in *I'll put up the Browns overnight* meaning 'accommodate, provide lodging'; a different idiom in *mother puts up perserves every fall* meaning 'preserves fruit'; again a different idiom in *they put me up to it* meaning *'gave me the idea'*; again different in the complex form *I can't put up with this much longer* meaning 'tolerate'...
5. The combination does not occur as a literal constitute but occurs in several idioms; e.g., *work up* does not occur as a literal constitute ... Idiomatically it occurs in the senses 'muster' as in *I couldn't work up the courage*, 'prepare or digest mentally' as in *work up this text for tomorrow*, and 'be distraught' as in *I was all worked up about it*...

Related to this Verb + Particle type of phrasal verbs are those which end in a particle but which contain nouns, adjectives, or adverbs, such as *take advantage of, feel sorry for,* and *do away with.* Here again, the whole phrasal verb seems to have a unitary meaning like 'exploit', 'pity', and 'exterminate', often expressible in a one-word alternative but not always, because some phrasal idioms develop precisely because the language lacks a convenient one-word expression. And, again, the alternate analysis which regards the particle as an ordinary preposition is possible but not without resultant loss of sense of the unitary nature of the phrasem.

The foregoing examples were all transitive, a following NP being obligatory, as are most examples of this type. Similar in form but having different internal structures are such idioms as *tell someone off* ('insult someone', in which the semantic object *(someone)* is the same as the formal object. Intransitive examples are comparatively scarce and are more complex: *beat (one's) brains out* ('think hard') seems to be semantically intransitive but its external form is transitive.
Another group of phrasal verbs, containing mandatory and antecedentless *it*, is included among Makkai's tournure idioms: *to break it up*, *to cut it out*, etc. All Makkai's examples exhibit unpredictability of meaning, placing them soundly in the class of idioms.

After all these distinctions are made, though, it must be admitted that since what prepositions can be used with specific verbs and objects, even for free constructions, is often arbitrary or idiosyncratic, the differences between phrasal verbs, prepositional verbs, and verb + prepositional phrase free constructions cannot really matter much to students learning English. In any case, they must learn by experience and memory anyway what verbs go together with what particles or prepositions, and what object NPs affect those choices.

In the discussion which follows, phrasal verb is used as a cover term as Makkai used it, with no distinction being made between real phrasal verbs which are followed by particles and prepositional verbs. However, free constructions of verbs followed by
prepositional phrases which do not have a unitary meaning and which are not common uses were excluded. Verb + infinitive phrases are mentioned briefly and finally, idiomatic uses of prepositions are discussed at the end of the section.

An observation frequently made informally by writing teachers and confirmed by statistical counts is that ESL basic writers make frequent errors in the use of prepositions. If particles are not clearly distinguished from prepositions this impression will be especially exaggerated, but the data gleaned from this study shows that errors in the use of both particles and prepositions are indeed common in ESL writing. A total of 48 phrasal verb errors were counted from the ESL papers while the HCE writers produced far fewer--17 errors. Where HCE basic writers do make them, the errors are more likely to be conventional in HCE.

One very common type of problem encountered is that the student omits a particle required for the meaning intended. This may occur because the writer does not recognize that the meaning is encoded in the form of a phrasal verb in English, or, does not know
the exact form of the phrasal verb idiom, or, the idiom itself has
taken a new form without the particle in the student’s dialect, as is
sometimes the case with the HCE manifestations:

ESL

M24 American students parents often are too busy to take care []
their children.
M8a ... he argues [] her and tries to get a good answer of whether
she can do the abortion or not.
M29 I have learned English based [] grammar, rather than
conversation.
K9 I went [] three years school there.
K17a Every class is limited [] about one and half hours in regular
school schedule.
K17a ... the costs of foods in KCC are similar [] the cost in restrants.
K18a They don't have ability to take care [] themself.
K20 My father passed [] long time ago.
G9a UH accomplished something they had been dream [] for a long
time,
Both schools are windy and you have to put tons of hairspray on. It is mainly about a forty years old father telling about his fifteen years old son's who travelled to trips to all the places he had never been before. He (son) already been to places he (father) had never gone before. They talked about how the places the son had been were like no place in America compared to America. Because his son had gone to so many places which whom the father had never been before, he (father) actually thought he (son) was the first man to land on the moon ...

The occurrence of *put* without the following particle *on* can be said to be lexically encoded in some versions of HCE. It carries a meaning something like 'to apply' in reference to cosmetics, condiments on food (*"When I eat hamburger, I put mustard and catsup"*), etc., just as does the phrasal verb *put on* in Standard American English (SAE). In a grammar describing such an HCE dialect, though, there would presumably be no necessity of listing
put in an idiom list separately from other lexical items, since put
would merely be polysemous and have one meaning be 'to apply',
whereas listing the SAE phrasal verb of the same meaning separately
would be essential because of its multilexical form, the particle being
mandatory according to the idiomatic demands of the language.

In the examples from M14a, the writer interestingly uses been
to and gone to to mean 'visit' when the phrases occur before the
object NPs, but uses plain been and gone when they follow the
objects. With so few examples it is difficult to know how consistent
the author's usage may be elsewhere, or whether other HCE speakers
share the same rule, but, at least here, the idiosyncratic rule is
followed consistently.

Another common error of basic writers is the following of a
verb with a particle when none is required. The writer seems to
assume that a phrasal verb is involved when in fact a simple verb
expresses what the writer means:
ESL

G4a When mom's business started to grow-up, Paul and Peter wanted to hire more employees, but mom didn't want to.

G11 ... yet foreign residents make up of very low percentage of total population

G17 most people will find out that in communist country there is nothing to fight for [By context, find makes better sense here.]

HCE

G1a I figured out the cost of my smoking habit to be approximately $450 for one year.

M14b In conclusion, it is very upsetting not having to voice out my explanation to my parents.

The use of figured out ... to be in the above sentence may be the result of syntactic blending of the alternative paraphrases: (1) I figured the cost to be and (2) I figured out that the cost was ... In the second example, perhaps the idiom come out with suggested the use of out with voiced since come out with one's opinion would be
semantically equivalent to *voice one's opinion*. This may also be considered an instance of syntactic blending.

A third kind of phrasal verb error is one in which the wrong particle is selected:

**ESL**

M1a Both Hawaii and Samoa *depend* more *in* the ocean for living and much more *about* our natural resources.

M8 The man suggested that she *goes* for an abortion operation and the girl seems not to *care about* this. [In the story, the girl clearly dislikes the idea of having an abortion, therefore the writer should have used the phrasal verb *care for*]

K14a Because, here at KCC, you can go home anytime, go to the bathroom without *asking permission to* the teacher ...

**HCE**

M26c And who are therefore *limited to* their knowledge of vocabulary and simple mathematics. [*limited in or by was apparently intended from evidence of the entire context*]
Phrasal verbs are not limited to one particle per verb. In English, two particles often follow verbs in forming various phrasal verb idioms, as in *put up with* and *be in for*. Quirk et al call these "phrasal-prepositional verbs" and describe them as combinations of phrasal verbs and prepositional verbs (1972:816). Following are several examples of a two-particle phrasal verb used incorrectly. In K17a’s sentence the use of the extra particles resulted in a totally unintended meaning being conveyed, *sign up for* being a different idiomatic expression from *sign up* and plain *sign*. HCE writers made far fewer of these errors, and those which they made consisted of much subtler meaning differences:

ESL

K17a students hold their yearbook and ask their friends, classmates and teachers to *sign up for* their yearbook. [*sign* was intended]

M10 People of Western Samoa are still *holding on*[] the Samoan custom.

G11 yet foreign residents *make up of* very low percentage of total population. [*make up* was obviously intended. The *of* may have come from the passive idiom *be made up of*]
Their discussion led up to little arguments. [where led to seems to suffice]

Finally, phrasal verbs used in place of a single word can sound less literate, or even very odd. In the sentence by G11 above, for example, the word constitute in place of make up of would likely be considered an improvement, although make up would do. Two more sentences by the same writer include the phrasal verb change into which is less awkwardly rendered become, a phrase get an experience of where experience alone would have fit comfortably, and takes over where controls would have worked.

It's difficult for [a] homogeneous society like Japan to change into [an] open society.

However, in another ten years, when the young generation who get an experience of living or studying in foreign countries takes over the society, Japan could be changed more successfully.
Considering how numerous phrasal verbs are in English, phrasal verb errors, as a proportion of the total number of errors in the corpus, actually are surprisingly few. One reason for this may be that new learners who are unsure of the exact connotations and nuances of meaning contained in these tricky phrases tend to avoid them. Another obvious reason is that writing standards encourage avoidance of phrasal verbs whenever a single-word, Greek- or Latin-based equivalent can be found--it is deemed better writing style to say *apply* than to use *put on*, etc. Although I have no statistics to bear me out, my impression is that phrasal verbs are far more plentiful in colloquial language and in informal writing. Fluent native speakers use them freely and sentences which employ them sound natural and idiomatic:

K1 I am interested in Hotel Management and plan to go into that field, so I plan to work my way up. I like to read a variety of books in my spare time. I also like going to the beach.

The HCE speaker above exhibits no difficulty with them, but the newest students of English frequently seem to have problems
with Verb + Infinitive constructions, which look much like phrasal verbs because the infinitive to is homophonous and homographic with the adverb, particle, and preposition to. The difference is that the infinitive is followed by a verb whereas the others are followed by noun phrases. The most common problems are with like to and want to. (An example with willing to was used near the end of the discussion on stylistics in 5.6.2.4.)

ESL

M1a But I would like [] specific compare Hawaii with American Samoa.

K14a Students here can graduate anytime they want [].

K18 I like [] watch movie.

Why should we talk about these in terms of idiomaticity when they could simply be grammatical errors? The student does not know that the syntactic rules require the infinitive to be expressed in constructions where a verb follows another verb or where the verb copy is deleted, as in K14a's sentence above (Students here can graduate any time they want to graduate is supposedly the
underlying source of the expression). One answer is that these particular Verb + Infinitive phrases are exceptionally common phrases and their frequency of occurrence identify them as phrasems of some sort. Another is that they sometimes carry a unitary sense which is not predictable from knowing the lexical items without the infinitive. In the case of the words *like* and *want*, the meanings of the verbs happen to sufficiently convey the combined verb + *to* meaning, but consider the words *have* and *get*. No amount of knowledge of the meanings of these words allow us to predict the meanings of *have to* and *get to*.

M14b In conclusion, it is very upsetting not having to voice out my explanation to my parents.

The differences in meaning between *have to* ('must') and *get to* ('be allowed to') are truly idiomatic in the sense of being unpredictable, and this author obviously meant 'not being allowed to voice her opinions was upsetting' which would colloquially be expressed as getting to voice one's opinions rather than having to.
Verbs which can participate in constructions with infinitives constitute an idiosyncratic list (agree, aim, attempt, claim, dare, decide, hope, learn, mean, promise, refuse, want, wish, etc.). Other verbs cannot take infinitive clauses as complements and must instead have -ing participles follow (admit, avoid, consider, dislike, keep, practice, resent, suggest, etc.). Then there is a third list which may take either infinitive or participial complements (dread, forget, hate, intend, like, prefer, begin, plan, try, etc.) (Selected examples from Quirk et al 1972:836).

English idiomaticity, in fact, specifies which prepositions are to be used with many particular adjectives and verbs. Student errors which indicate unfamiliarity with the required pairs are extremely copious and a few examples from one paper should suffice to demonstrate the prevalence of the problem:

M6b Insurgency...has been a major issue in the Philippine government.  [for]
M6b What are the socio-economic effects toward the nation as a whole?  [on]
M6b ... thus, lessening the job opportunities of the people. [for]

M6b Parents are always worried with their children, whener they hear [] casualties or deaths. [about]

The situation is actually even more complex than that certain verbs and adjectives require certain prepositions. Sometimes it depends on what follows the preposition, for instance, agree with someone, agree to a proposal, agree on a procedure, etc. (Elsbree, Altizer and Kelly 1981:226). The student, who might have taken the pattern from something like scattered around the room, must have been surprised to find the following corrected:

M12b He puts his books and papers around his desk and leves it there after studying.

After laboring a considerable time over distinguishing strictly grammatical from idiomatic errors in the use of prepositions, I have decided to count all prepositional errors as idiomatic. My reason was mainly that it was impossible to find a clear case of wrong, missing, or intrusive preposition which should have been correctable by a
non-native on the basis of general grammatical rules and a lexicon. Further study may prove me mistaken in this, of course.

Although practical grammars like Quirk et al's (1972) include lists of lexical items among their rules, current formal paradigms do not allow such mixing of components. Where the facts of the language can only be accounted for by including lists among the rules, or conversely, by including rules in the lexicon, it is reasonable to admit that we are dealing with idiomaticity. Available options are to either seek a formal apparatus which can handle the facts or make adjustments to the current model. Pawley and Syder suggest that the requirement that grammars be maximally parsimonious cannot be maintained if the facts of language use are to be adequately represented, because it may simply be the nature of human minds that we know common phrases in two ways--as formulas and as free expressions. Despite the resulting redundancy, it may be necessary to include the constituents of a phrase in the lexicon in addition to listing the phrase whole (Pawley and Syder 1983b:216-20).
6.1. Interpretation of the findings.

Before making some general interpretive statements, a brief summary will be made of the interpretive remarks made in section 5.6.2. of the preceding chapter. In that section, various types of errors which were identified in the analysis of the data were classified (accidental, grammatical, lexical, etc.). For each of the types, it was intended that those errors which were identified as being basically idiomatic would be discussed, using examples from the student papers. But, since the huge quantity of material made it necessary to be very selective in displaying and discussing examples, only one or two types of phrasal phenomena (syntactic blends, comparison formulas, negation formulas, etc.) were chosen for discussion of each error type (except phrasematic errors, for which several were chosen). Effort was made to select a phrasal type which was the most appropriate for showing how idiomatic factors affected or motivated each error type, and no phrasal type was used twice.
This procedure was not intended to mislead readers into assuming there is a one-to-one matching of phrasal type with error type, but was done for the sake of brevity.

First, (5.6.2.1) syntactic blends, usually viewed as simple blunders, suggested that even apparently random mistakes and those not obviously idiomatic may have their sources in idiomaticity. In each case, they could be shown to result from the blending of two semantically similar and equally familiar conventionalized phrases, suggesting that blends constitute meaningful psycholinguistic evidence of the holistic nature of phrases in mental processing.

Second, (5.6.2.2) the problem of sorting the ungrammatical from the grammatical-but-unidiomatic was demonstrated with a discussion of student errors in making comparisons and in the use of comparatives. This showed that not only are some very specific formulas required in making comparisons, but that the idiomatic features of frequency of use and familiarity seem to be involved in the choice of what syntactic form comparative markers should take. If we look more carefully at "grammatical errors", considering the
possibility that idiomaticity may be involved, we are very likely to find many more instances of grammatical errors which are actually the sort of error that no amount of grammatical knowledge would help the student avoid.

Third, an explanation of why lexical (5.6.2.3) and stylistic errors (5.6.2.4) were excluded from this study was given. The cases where idiomaticity is involved in lexical selection and stylistics were to be covered in the presentation of phrasematic errors.

Fourth, (5.6.2.5) some examples of errors in the use of negation were discussed, and it was suggested that idiomaticity often has little obvious relation to semantic compositionality or logic. In addition, conjunction errors demonstrated that phrasal correlative formulas were often required for idiomaticity in the use of logical operators. It was intended that this convince readers of the necessity of reconsidering how semantics might be affected by the demands of idiomaticity.
Fifth, (5.6.2.6) it was noted that pragmatics and idiomaticity are closely related since formulas and idioms usually develop to serve pragmatic functions, and euphemisms were used to illustrate that point. Then, because conversational formulas are scarce in written work, sentence connective errors were presented as evidence that written discourse deixis is heavily idiomatic. Here, again, the selected examples were from only one type of phrasem which has a clearly pragmatic function, and many other types remain to be examined closely according to idiomatic criteria.

Finally, (5.6.2.7) compounds and other common collocations, some support verb phrasems, phrasal verbs, prepositional verbs and idiomatic prepositions were discussed, and examples from student writing demonstrated a variety of ways in which focusing on idiomaticity illuminated many formerly obscure problems in student writing. Much remains to be studied.

The strongest general statement that I believe can be made with confidence at the conclusion of this study is that there is much more of idiomaticity involved in language use than grammarians and
linguists have generally acknowledged. In fact, beginning with the assumption that "everything is idiomatic", as was suggested to me by Dr. George Grace in an early conference, allowed me to accept appropriateness and naturalness as the common denominator of idiomaticity, which was adopted for this study from Pawley and Syder (1983a and 1983b). Employing the standard of naturalness by which to evaluate students' errors resulted in the discovery of idiomatic sources to far more varieties of errors than ever expected.

To summarize the earlier interpretation of the statistical results:

(1) ESL writers made more than twice as many errors (2180) as did HCE writers (1052).

(2) More of the ESL writers' errors were grammatical than were idiomatic, while the opposite was true for HCE writers. In fact, about two-thirds of the HCE writers' errors were idiomatic rather than just grammatical (629 idiomatic as compared to 404 grammatical).
(3) ESL writers, nevertheless, produced more total errors in the idiomatic category (964) than did HCE writers (629).

(4) The only categories in which HCE writers produced more errors than ESL writers turned out to be all "colloquial" ones. The total sums of errors labelled "jargon", "tags", "topic", etc. were 222 (ESL) and 309 (HCE).

These results are not surprising. They confirm informal impressions that ESL writers have the expected problems with grammar while HCE writers' basic problems are "something else, perhaps dialectal." While this general difference may not be obvious to non-HCE-speaking readers, those who are well-acquainted with the HCE idiom have long been aware of it. "The problem is that they write just like they talk," was one teacher's assessment of the basic writing of HCE speakers. As this study reveals, this comment is not far off the mark, since the HCE students used many colloquial forms which are not acceptable in written texts. It is important to notice, nevertheless, that ESL writers have more trouble with idiomaticity as well as with grammar than HCE writers.
6.2. Implications of the findings.

6.2.1. Theoretical implications.

It is hoped, first, that this study lends weight to the contention that it is important to study idiomaticity because there is so much of it in use, as Weinreich (1969) noted, and yet it is not possible to discuss idiomaticity in the prevailing framework which (1) cleanly separates grammar and lexicon, and which (2) considers matters of language use to be irrelevant to language description. The assumption of the grammar-lexicon model is that the grammar contains only rules (in terms of classes and categories to which they apply) and no lexical items, while the lexicon contains only lexical items (with a specification of their semantic and syntactic features) and no rules. The study of idiomaticity seems to reveal that (1) it is common for languages to have lexified structures which require mention in grammars, and/or (2) that lexicons are probably far more complex than has been assumed and may even require inclusion of rules of some kind.
Idiomaticity also demonstrates the importance of taking usage into account in description, which means such factors as frequency, familiarity, affective strength, cohesiveness, redundancy (Van Lancker 1975:147) and the like are relevant to description. Further study of idiomaticity may reveal still other factors of language use not yet identified.

The study of idiomaticity leads to the asking of many questions which may be otherwise nonsensical: How do human brains handle multilexical but conventionally unitary forms? Can the lexicalized forms be semantically unitary but syntactically analytical? What is the normal way of putting together the things we say? How much is memory involved, and how much is generalization involved? These questions could be multiplied many times over and finding the answers would be only the beginning.

6.2.2. Practical implications.

In a sense, the first "practical" implication is theoretical. The results of this study imply a need for the formulation of a theory of
idiomaticity to undergird the teaching of idiomaticity to writing students. Basic writers and second language learners have serious problems with the idiomaticity of the written register. Recognition of the demands placed on students by this idiom would be a starting point for discovering what needs to be taught. Identifying students' problems may be a first step, for example, pinpointing receptivity and productivity difficulties which the students experience:

Reception problems--(1) non-recognition of phrasems produced by others, ranging from just simply not being aware that a fixed expression is a conventional one, to attempting to decode idioms which are not decodable, and taking something literally which was meant nonliterally, (2) inability to understand what a phrasem means, (3) misinterpretation of a phrasem (making a wrong guess as to meaning), and so forth.

Production problems--(1) ignorance of phrasematic expressions appropriate to particular situations, ranging from not knowing there is a conventional thing to say and a conventional way to say it, to not even knowing anything should be said at all, (2) inability to select
the right phrasematic expression, (3) use of a conventionally non-literal idiomatic expression intending its literal meaning, (4) inability to encode a phrasematic expression correctly in every detail (e.g., *shoot the bulls*), etc.

An indirect implication of this study is that the teaching of idiomaticity would not be identical to the teaching of grammar. It suggests that if idiomaticity is a matter of memory rather than rule-application, specifics rather than generalizations should be taught. This may mean that students must be exposed to and allowed to experience as many correct models as possible. The fact that recognizable forms, correct by HCE colloquial standards, abound in the HCE writers' papers, and are present in the ESL papers also, implies that practiced forms will be repeated. Furthermore, this study suggests that HCE writers may not be taught best in identical manner as ESL writers, since their problems are not identical. If forms as well as patterns must be taught, it appears that non-native speakers may benefit from practice with generally applicable patterns, while HCE speakers may need more practice with the special phrasems used in academic writing.
6.3. Some preliminary conclusions.

6.3.1. The relationship of idiomaticity to grammar.

A metaphorical abstraction may make it easier to conceive of the relationship of idiomaticity to grammar:

Grammar (including its lexicon) is one end of a continuum of conventionalization which we are here calling idiomaticity. The extreme pole on the opposite end is idiosyncracy or nonce creation, which occupies a place on the idiomaticity scale once it is repeated and passed on.

The more often the form is repeated, the farther it gets from the nonce end and the closer it gets to the grammar-lexicon end; that is, after enough repetitions, the creation becomes conventionalized. At some early point in the process, it is included in a list of multilexical units, which is similar to the lexicon of single lexical items, except that its parts may be somewhat more susceptible to analysis.
When the pattern of a conventionalized form is extended to uses with words other than those in the original lexification, the pattern has become semiproductive. If the rules underlying the new semiproductive pattern do not fit in the grammar, the pattern is included in a list of semiproductive rules.

A pattern is fully productive when all forms of the same type conform to the pattern. A productive conventionalized pattern forms a grammatical rule when a fully productive pattern is fit into a grammatical paradigm. Idiom then has become grammar, or, as Roberts put it, "grammar is fossil idiom" (1944:300).

Grammar is thus a subset of idiomaticity. Idiomaticity includes all that is peculiar and natural to a language, or, is conventional to it, and its grammar is one of those things. Its lexicon, which is a part of its grammar, is another. Its list of multilexical units and the list of semiproductive rules are yet other subsets. These two lists fit between the lexicon and grammar on an idiosyncratic to general continuum, but they are prior to both lexicon and grammar on the
idiomaticity continuum, since the grammar-lexicon is completely conventionalized.

6.3.2. **Psycholinguistic hypotheses.**

Units of our mental lexicons range in size from single morphemes to polymorphemic combinations. "We form our utterances from these basic units according to a few general rules and very many specialized and specific restrictions" (Bailey 1983:x)

Our brains have at least two ways of processing and producing language. One is **algorithmic**--repeated operations according to a few basic rules which are the objects of Chomskyan research--and the other is **heuristic**--shortcuts, or rules-of-thumb, taking advantage of experience or former learning; i.e., having already assembled a stretch of language from basic units according to required algorithms, or having once received such a segment already formed which suits our need, we simply use it again as we remember it, without having to reconstruct it from scratch. We use what we have a memory of having heard before (Bolinger 1976:1).
The units we call up are semantic or idea units (Chafe 1985: 106-8), conceptual in nature. The concepts they represent are unitary regardless of the number of lexical items which constitute them.

6.3.3. Additional observations.

Semantic units do not necessarily correlate with syntactic units, therefore, mobility and modifications to syntax, both internally and externally, are possible with these semantic units. As entries in a complex lexicon, phrasems exhibit varying degrees of frozenness—they are internally analyzable, modifiable or movable to varying degrees, by interaction between syntactic, semantic and pragmatic rules operating on the rest of language and historical accidents. Thus, limitations to productivity and transformational deficiencies are historical products of such interaction.

The syntactic variations of phrasems are relatively unimportant to attempts at finding general rules of syntax, and so forth. That there are degrees of lexical, semantic, and syntactic
regularity and frozenness in multilexical units is a trivial fact about them, due to each phrasem having had a unique history, just as is true of unilexical items in the lexicon. As with single lexical items, multilexical units with their privileges and limitations simply have to be memorized.

It is part of a native speaker's competence to be able to both decode and encode phrasal semantic units in the conventionally established ways of the speech community to which the speaker belongs. This competence is based on experience and memory as well as on innate properties of the brain.

6.4. **Suggestions for further study.**

Besides more psycholinguistic and neurolinguistic studies which may help uncover clues to how human beings process language, a variety of studies of idiomaticity in different directions could contribute much to our understanding of language and its use. Of course, potential studies can easily be dreamt up. Some obvious topics in need of further study have been mentioned in the previous
chapters, for instance, syntactic blends and each of the phrasem categories. A few other possibilities which occurred to me as I proceeded in this investigation, and which still have appeal follow.

Cross-cultural/cross-linguistic studies may be revealing. For example, Vachek (1976) claims Mathesius thought English was particularly rich in stable phrases and attributed that quality to its analytical structure. Such observations want serious investigation.

The diachronic study of idiomaticity could clarify its role in language change. How close to the mark is Roberts' claim that grammar is fossil idiom (1944:300)? Are idioms indeed the first to change when languages diverge? How about when they converge, or even collide, as in language contact?

Finally, an undying hope remains that there may yet be found some practicable way of obtaining frequency counts and of quickly producing frequency lists of idiomatic expressions of all kinds. Even a limited collection, say, of common phrases within a particular rhetorical genre would be helpful if it results in an accurate record of
relative frequencies in the usage of particular phrasems at a particular time.

Perhaps the chances would be improved for carrying out the foregoing types of studies if applied linguistics incorporated all research in language use, instead of being limited to English as a second language as it tends to be today. Idiomaticity would then be a natural part of applied linguistics.
NOTES

1. (Chapter 1, page 9)
   This accords with Makkai's discussion of "idioms of encoding" (1972:25), a concept which identifies a useful distinction from the more obvious "idioms of decoding".

2. (Chapter 2, page 22)
   Vachek, by the way, writes in this same paper, "... there is a number of ..." (408) but "... a relatively high number of important papers on problems of written English have appeared of late ..." (409), observing grammatical concord in the former and notional concord in the latter.

3. (Chapter 3, page 44)
   See Makkai (1972:25) for works dealing with idiomaticity in the intervening years.

4. (Chapter 3, page 62)
   Pawley discussed "lexicalization" further in Pawley 1985a.

5. (Chapter 3, page 67)
   Smith apparently did not notice as did Malkiel (1959), that this set of phrases consists overwhelmingly of pairs of terms which may not be reversed.

6. (Chapter 3, page 68)
   Malkiel, in his 1959 paper, followed a similar format to this section of Smith's publication.
   All page numbers refer to the 1943 republication of Smith's 1925 work.

7. (Chapter 4, page 104)
   Perhaps the term "phraseology" seems to be more descriptive of the object of my interest than "idiomaticity", but that term is also unsatisfactory in other ways since it already has established uses in lexicography, and it does not carry the meanings of peculiarity or naturalness which are needed.
8. (Chapter 5, page 140)

The OED defines "idiomaticity" as "The quality or state of being idiomatic" and "idiomatic" is defined as: "Peculiar to or characteristic of a particular language; pertaining to or exhibiting the expressions, constructions, or phraseology approved by the peculiar usage of a language, esp. as differing from a strictly grammatical or logical use of words; vernacular; colloquial." I hope my briefer definition is generally equivalent to this although stated less specifically.

9. (Chapter 5, page 166)

Here is one fact of English which makes the arbitrary division between phrasal and single-word idiomaticity undesirable, since we would not want to treat periphrastic comparatives separately from inflected ones.

10. (Chapter 5, page 199)

Again, this is admittedly somewhat arbitrary, since we are relying on orthographic conventions, which render some compounds as single words and others as two words, to determine which forms to include. Orthographical conventions do, however, accord with certain phonological and semantic principles which may be valid for distinguishing single from pairs of lexemes, so the boundary we are adopting cannot be wholly arbitrary.

11. (Chapter 5, page 207)

Unfortunately, his examples of to look up and to look after are unclear (Are the phrases in #4 meant to be examples of phrasal verbs or prepositional verbs?) and inconsistent since he uses the same phrase (it is unclear which) as an example of a prepositional verb in one instance and as an example of a phrasal verb in another.
REFERENCES


254


McCawley, James D. 1971. [See Quang Phuc Dong]


_________. 1974. The regularity of idiom behavior. Lingua 34. 327-42.


_________. 1985b. On speech formulas and linguistic competence. Lenguas Modernas 12.84-104.


