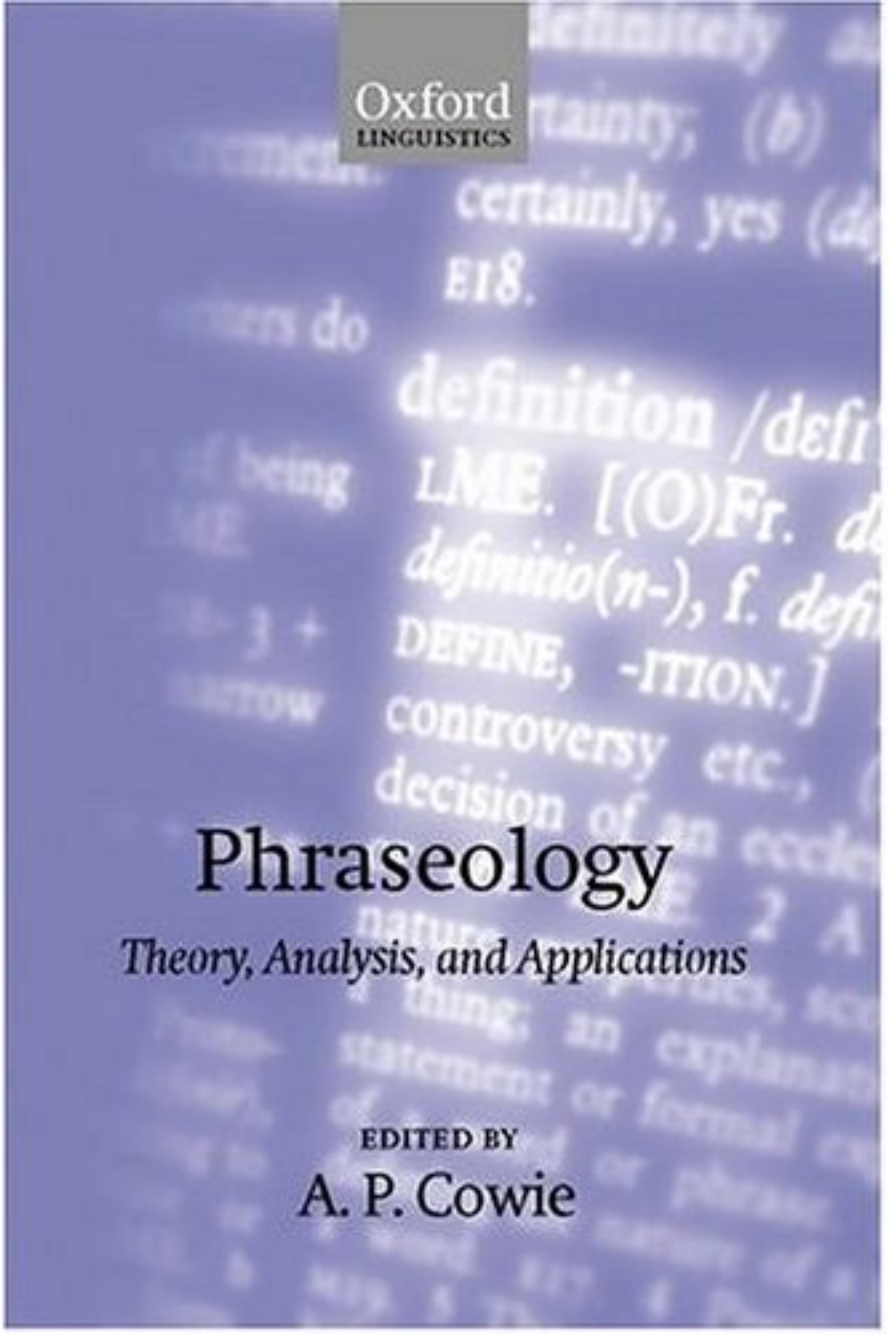

Oxford
LINGUISTICS
Phraseology
Theory, Analysis, and Applications
EDITED BY
A. P. Cowie

PHRASEOLOGY

Theory, Analysis, and Applications

Edited by A. P. COWIE

CLARENDON PRESS . OXFORD 1998

Oxford University Press, Great Clarendon Street, Oxford OX2 6DP Oxford New York

Athens Auckland Bangkok Bogota Bombay Buenos Aires Calcutta Cape Town Dar es Salaam Delhi Florence Hong Kong Istanbul Karachi Kuala Lumpur Madras Madrid Melbourne Mexico City Nairobi Paris Singapore Taipei Tokyo Toronto Warsaw and associated companies in Berlin Ibadan

Oxford is a registered trade mark of Oxford University Press

Published in the United States by Oxford University Press Inc., New York

British Library Cataloguing in Publication Data Data available

Library of Congress Cataloging in Publication Data Phraseology theory, analysis, and applications edited by A. P. Cowie. (Oxford studies in lexicography and lexicology Includes bibliographical references and index. 1. Phraseology. L Cowie, Anthony Paul. II. Series. P326.5.P45P48 1998 418-dc21 98-15867 ISBN 0-19-829425-5

1 3 5 7 9 10 8 6 4 2

Typeset by BookMan Services Printed in Great Britain on acid-free paper by Bookcraft (Bath) Ltd., Midsomer Norton

---

Acknowledgements

I am indebted to several colleagues and friends for specialist advice and practical support during the editing of this book. My first debt is to the series editors, Gabriele Stein and Noel Osselton, for invaluable comments on the structure and content of the book, as well as on numerous points of detail. I am particularly grateful also to Peter Howarth, who in addition to helping me put the manuscript into machine-readable form, kindly agreed to compile the consolidated bibliography. I greatly appreciate, too, Hilary Walford's meticulous work in preparing the typescript for the press, and express my thanks to her. Finally, I acknowledge with gratitude the secretarial support and constant encouragement provided by my wife, Cabu, throughout the preparation of this book. To her, a special measure of thanks is due.

A.P.C.

[This page intentionally left blank.]

Contents

List of Figures

List of Tables

List of Contributors

Bengt Altenberg Associate Professor of English, Department of English, Lund University, Helgonabacken 14, S-223 62 Lund, Sweden.
Natalya Bragina Senior Researcher, Pushkin Institute of the Russian Language, Volgina ul. 6,117485 Moscow, Russia
A. P. Cowie Lately Reader (now Honorary Reader) in Lexicography, School of English, University of Leeds, Leeds LS2 9JT, UK
Thierry Fontenelle Computational Linguist and Translator, European Commission, Translation Service, Development of Multilingual Tools, Jean Monnet Building JMO B2/14, L-2920 Luxembourg
Rosemarie Gläser Professor of Applied Linguistics, Language Centre, University of Leipzig, Augustusplatz 9, D-04109 Leipzig, Germany
Sylviane Granger Professor of English Language and Linguistics, Centre for English Corpus Linguistics, Catholic University of Louvain, B-1348 Louvain-la-Neuve, Belgium
Peter Howarth Senior Teaching Fellow, Language Centre, University of Leeds, Leeds LS2 9JT, UK
Igor Mel'čuk Professor of Linguistics, Department of Linguistics and Translation, University of Montreal, P.O. Box 6128, Montreal H3C 3J7, Canada
Rosamund Moon Editorial Manager, Cobuild, and Honorary Research Fellow, Department of English, University of Birmingham, 50 Edgbaston Park Road, Birmingham B15 2RX, UK

Elena Oparina Senior Researcher, Institute for Scientific Information on the Social Sciences, Russian Academy of Sciences, Nakhimovskij pr. 5121,117418 Moscow, Russia
Irina Sandomirskaya Independent Researcher, Garibal'di ul. 7411, 117313 Moscow, Russia
Veronika Teliya Professor of Linguistics, Institute of Linguistics, Russian Academy of Sciences, Bol'shoj Kislovkij per. 1-12, 103009 Moscow K-9, Russia

Abreviations

BBI	M. Benson, E. Benson, and R. Ilson, *BBI Combinatory Dictionary of English*
CIA	Contrastive Interlanguage Analysis
CMM	Concepts-Meaning Model
CR	*Collins-Robert English-French French-English Dictionary*
ECD	*Explanatory Combinatory Dictionary* (theory)
ECD	*Explanatory Combinatory Dictionary* (published work)
EFL	English as a foreign language
ELTJ	*English Language Teaching Journal*
ERPD	A. V. Kunin, *English-Russian Phraseological Dictionary*
EURALEX	European Association for Lexicography
ICE	International Corpus of English
ICLE	International Corpus of Learner English
ITL	Instituut voorTbegepaste Linguistiek
L1	first language
L2	second language
LDOCE	*Longman Dictionary of Contemporary English*
LF	lexical function
LOB	Lancaster-Oslo-Bergen (Corpus)
MR	machine-readable
MTM	Meaning-Text Model
MTT	Meaning-Text Theory
NLP	natural language processing
NNS	non-native speaker
NS	native speaker
ODCIE1	A. P. Cowie and R. Mackin, *Oxford Dictionary of Current Idiomatic English*, vol. i
ODCIE2	A. P. Cowie, R. Mackin, and I. R. McCaig, *Oxford Dictionary of Current Idiomatic English*, vol. ii
SEC	C. D. Kozlowska and H. Dzierz+anowska, *Selected English Collocations*
SLA	second-language acquisition
TESOL	*Teachers of English to Speakers of Other Languages (quarterly)*

1

Introduction

A. P. COWIE

1
PAST ACHIEVEMENTS AND CURRENT TRENDS

Following a steady growth of scholarly interest and activity over the past twenty years, chiefly in Western Europe, but also in the USA, phraseology has now become the major field of pure and applied research for Western linguists that it had, much earlier, for scholars in the former Soviet Union and other countries of Eastern Europe. Its coming of age has been marked by a number of international conferences, while its acknowledged importance to both theoretical and applied linguists is reflected in several large-scale research projects which have phraseology as their sole or principal focus of interest (e.g. Heid and Freibott 1991; Fontenelle 1992a, c; Heid 1992). [1]

Current concerns are by no means purely descriptive. Interest in the analysis of what are variously named 'phraseological units' (Ginzburg et al. 1979; Gläser 1986a), 'word-combinations' (Akhmanova 1974; Cowie 1994), and 'phrasal lexemes' (Lipka 1991; Moon in this volume) is accompanied by an increasing awareness of the prevalence of ready-made memorized combinations in written and spoken language and a wider recognition of the crucial part they play in first- and second-language acquisition and adult language production (Pawley and Syder 1983; Peters 1983).

There are also implications for linguistic theory. The notion that native-like proficiency in a language depends crucially on a stock of prefabricated units -- or 'prefabs' -- varying in complexity and internal stability is now set in critical opposition to the atomistic view, rooted in generative theory, that the workings of a language can be explained by a system of rules of general applicability, a lexicon

1Recent major conferences include the International Symposia on Phraseology held in Leeds (1994) and Moscow (1996) and the International EUROPHRAS Congresses held in Saarbrücken (1992) and Graz (1995).

largely made up of minimal units and a set of basic principles of semantic interpretation (Bolinger 1976; Pawley 1985; Fillmore, Kay, and O'Connor 1988).

Recognition of phraseology as an academic discipline within linguistics-the term itself, like the adjective 'phraseological', reflects Eastern European usage-is evident not only from vigorous and widespread research activity, but also from the publication of several specialized dictionaries reflecting one theoretical perspective or another (e.g. Cowie and Mackin 1975; Cowie et al. 1983; Sinclair and Moon 1989, 1995), and from the attention increasingly given to the subject in textbooks on lexical semantics (Cruse 1986), lexicology (Carter 1987; Lipka 1991), and vocabulary in language teaching (Carter and McCarthy 1988). Yet, despite the existence of Gläser's authoritative textbook (Gläser 1986a) and the availability of monographs devoted to the phraseology of specific languages (e.g. Fleischer 1982; Gréciano 1983; Corpas Pastor 1996), interlanguages (Howarth 1996), and functional varieties (e.g. Müller 1993), there is to date no book-length account in English of the various theoretical currents which inform present-day phraseological studies, nor one which takes account of the associated disciplines, such as computational analysis, language-learning, stylistics, and lexicography, to which those studies are making such a vital and invigorating contribution. The present volume, made up of contributions from several leading specialists, is an attempt to fill that gap.

Three major theoretical approaches are represented in this volume, either directly, or indirectly through description or practical application. 'Classical' Russian theory, with its later extensions and modifications, is probably the most pervasive influence at work in current phraseological studies and is unrivalled in its application to the design and compilation of dictionaries. In those of Igor Mel'čuk and his associates, especially, it has been developed and applied with great rigour and sophistication (Mel'čuk et al. 1984/1988/1992). Most of the contributors to this volume, whatever their present or original affiliations, use some version of the system of categories originally developed within that classical tradition. For all these reasons it is the first to be examined in detail.

A second strand can be described as broadly anthropological and is represented here by the contribution of Veronika Teliya and her colleagues, which proposes an extension of the Russian phraseological tradition to embrace the cultural dimension. They argue persuasively that this element must be elaborated in all its richness and complexity if the phraseology of a language is to be fully described and understood.

A further, highly productive, line of development in phraseology extends from the pioneering work of J. R. Firth, via the neo-Firthians Michael Halliday and John Sinclair, to many of the specialists currently engaged in the analysis of phraseology in computer-stored corpora of spoken and written English. Sinclair has played a leading role in developing corpus linguistics in Britain as a basis for both phraseological research and dictionary-making, and two contributions to this volume -- by Rosamund Moon and Bengt Altenberg -- are within the tradition he has done much to establish (Sinclair 1991). But, while making frequency-based analysis a central feature of their chapters, Moon and Altenberg also explore structural and pragmatic issues in some depth. Such flexibility in adapting an established approach to specific analytical and practical ends has proved extremely fruitful.

The chapters by Rosemarie Gläser, Sylviane Granger, and Peter Howarth are in the broadest sense essays in applied phraseology. The emphasis in Gläser's essay is stylistic: it sets out in detail the expressive resources of the phrasal lexicon as drawn upon in a variety of literary and non-literary texts. Granger and Howarth, by contrast, both provide comparative analyses of word-combinations in the writing of native speakers and foreign learners of English. Their aim is to identify the norms implicitly recognized by native speakers and to demonstrate and explain how the usage of foreign students departs from them.

The dictionary as a codified record of phraseological norms and as an indispensable aid to language-learning and teaching is never far from the surface of this survey. Three chapters are, however, wholly or largely devoted to this theme. Like that of Igor Mel'čuk, whose involvement in lexicography has already been referred to, Thierry Fontenelle's interest in dictionaries is both theoretical and didactic. Drawing upon a standard bilingual dictionary in machine-readable form, he extracts its very varied collocational material and then enriches the database with a system of 'lexical functions' (derived from the work of Mel'čuk) which specify the lexical-semantic relations of collocations. The particular advantage of creating such a database is that it enables the user to access specific items via general semantic indicators. Finally, and also bringing together theoretical and practical concerns, Anthony Cowie examines a number of dictionaries against the background of a historical account of British and Russian phraseological theory. His purpose is to show the increasing extent to which British and American dictionaries of idioms and collocations are influenced by Russian analytical models and to illustrate the growing cross-fertilization between previously separate traditions.

2
THEORETICAL PERSPECTIVES: THE RUSSIAN TRADITION

Russian phraseological theory, in the forms in which it developed from the late 1940s to the 1960s, was first mediated to non-Russianspeaking scholars by Klappenbach (1968), Weinreich (1969), Arnold (1973), and Lipka (1974). Its principal legacy is a framework of descriptive categories that is comprehensive, systematic, and soundly based. [2] Most of the early schemes and subsequent refinements (as in Gläser 1986a) are agreed in recognizing a primary division between 'word-like' units, which function syntactically at or below the level of the simple sentence, and 'sentence-like' units, which function pragmatically as sayings, catchphrases, and conversational formulae. Examples of the former are *in the nick of time, a broken reed, and break one's journey,* and of the latter *There's no fool like an old fool, The buck stops here,* and *You don't say!* One of the first Russian phraseologists to refer to this distinction was Chernuisheva (1964), whose sentencelike units (called 'phraseological expressions') included sayings and familiar quotations. [3] As Table 1.1 shows, the difference is still recognized by British specialists, such as Cowie and Howarth, who are influenced by Russian models. As the table also shows, there is a wide diversity of terms used to refer to 'propositions' and 'nominations'.

Further important distinctions were drawn within both the 'pragmatic' and the 'semantic' classes. In the former category, as we have seen, were sayings and catchphrases (for a detailed subcategorization, see Gläser 1986a, 1988a), but the early Russian work is especially known for its subclassification of 'nominations' or 'composites'. Here a major contribution was made by Victor Vladimirovich Vinogradov and Natalya Nikolaevna Amosova, the latter remembered especially for her view of 'phraseologically bound' meaning, which was to be widely influential.

Within the general class of nominations (for which he used the term 'phraseological unit'), Vinogradov (1947) drew a distinction between 'phraseological fusions' (also called 'idioms'), 'phraseological

[2] In phraseology, as in other fields within linguistics, it is not uncommon for individual scholars to apply different terms to the same category (or the same term to different categories). I have therefore thought it essential -- while not suppressing individual differences -- to provide a general framework (see Table 1.1), in which the terms used by any one individual can be understood in relation to those used by others.

[3] H. E. Palmer, in a pioneering classification of English word-combinations based on examples from existing dictionaries, confined his attention (in the 1933 *Interim Report*) to wordlike units. His subcategorization of sentence-like units was not made till 1942, and remains unpublished (Bongers 1947).

TABLE 1.1. Terms used for 'sentence-like' and 'word-like' combinations

Author	General category	Sentence-like (or pragmatic) unit	Word-like (or semantic) unit
Cheuisheva (1964)	Phraseological unit	Phraseological expression	--
Zgusta (1971)	Set combination	Set group	--
Mel'čuk (1988*b*)	Phraseme, or Set phrase	Pragmatic phraseme, or Pragmateme	Semantic phraseme
Gläser (1988*a*)	Phraseological unit	Proposition	Nomination
Cowie (1988)	Word-combination	Functional expression	Composite
Howarth (1996)	Word-combination	Functional expression	Composite unit

unities', and 'phraseological combinations'. The first subcategory was made up of combinations that were 'unmotivated' (or semantically opaque) and in general structurally fixed. English *spill the beans* fits this category very well. Beyond the fusion, Vinogradov recognized a partially motivated type, whose meaning could be seen as a metaphorical extension of some original neutral sense. An English example of this category is *blow off steam,* which has undergone figurative extension from a (still active) technical meaning. The boundary between unities and fusions is, of course, not clear-cut, but varies according to the linguistic and cultural experience of the individual.

Vinogradov's third category, the phraseological combination, is the most interesting -- just as, in practice, it is the most difficult to delimit. In the case of units consisting of two open-class words, such as *meet the demand,* phraseological combinations have one component used in a literal sense, while the other (here the verb) is used figuratively. Note,

though, that this specific sense of *meet* is determined by its context. (It is, according to Vinogradov, 'phraseologically bound'.)

The differences between Vinogradov's scheme and Amosova's need not concern us here -- they are explored more fully in Cowie's contribution to this volume. For Vinogradov, the figurative element in the third subcategory could be phraseologically bound by a single word or a limited set of words (in relation to *meet,* this set would include -besides *demand -- need, requirements,* and *request*). For Amosova (1963), the 'bound' sense must have a single determining word (*teeth in grind one's teeth, fall in break one's fall*). Such combinations -- to which she

gave the name 'phraseme' -- represented for Amosova the outer limits of phraseology.

It is precisely the area over which Vinogradov and Amosova disagreed -- the 'fuzzy' zone to which *meet the demand* and *grind one's teeth* both belong -- which is of particular interest to several analysts today (cf. Cowie and Howarth 1996). Like Vinogradov, they would regard both examples as falling within the same subcategory (to which the term 'restricted collocation' is now commonly applied). The contentious issue -- given that we are on the borders of phraseology-is the degree to which a collocation can be varied and still remain 'restricted'. Table 1.2 incorporates the tripartite schemes of Vinogradov and Amosova -- schemes which, despite evident differences in terminology, are still closely imitated in the late 1990s. [4]

As Table 1.2 shows, categories with a nominative function are ranged along a scale or continuum from unmotivated and formally invariable idioms to partially motivated and partially variable collocations. Beyond the latter are 'free' or 'open' combinations, whose make-up can be explained in terms of general restrictions on cooccurrence, and which lie outside the limits of phraseology altogether (Klappenbach 1968; Cowie 1981; Fernando and Flavell 1981).

Igor Mel'čuk occupies a special place in this classification. As both Tables 1.1 and 1.2 show, the categories he recognizes correspond closely to those of the classical models, and indeed can be traced back to them. Most significantly of all, collocations, which Mel'čuk defines with the rigour he brings to the entire categorization, are for him the focal point of the system. They constitute, after all, the absolute majority of 'phrasemes' -- i.e. word combinations of all types -- in the lexicon and, as the studies of Moon, Howarth, and others have demonstrated, in many genres and specialist fields as well.

Mel'čuk recognizes that the meaning of a two-word collocation includes 'intact' the meaning of one its constituents, but that the other component of its meaning is expressed by an element contingent on the first. This formulation, with its echoes of contextual 'binding', places Mel'čuk firmly within the classical Russian tradition. And, as he also perceives, collocations can be subcategorized according to the category of bound element (a support or 'light' verb in the case of *take a step,* a 'full' but contextually determined attributive sense in the case of *black coffee*).

Igor Mel'čuk is particularly associated with the development of

[4] Several of these schemes are compared by Howarth in his detailed critical discussion of 'nominations' and 'propositions' (Howarth 1996). His account is complemented by Corpas Pastor, who discusses taxonomies developed by Spanish and German scholars (Corpas Pastor 1996).

TABLE 1.2. Subcategories of word-like combinations ('nominations')

Author	General category	Opaque, invariable unit	Partially motivated unit	Phraseologically bound unit
Vinogradov (1947)	Phraseological unit	Phraseological fusion	Phraseological unity	Phraseological combination
Amosova (1963)	Phraseological unit	Idiom	Idiom (not differentiated)	Phraseme, or Phraseoloid
Cowie (1981)	Composite	Pure idiom	Figurative idiom	Restricted collocation
Mel'čuk (1988*b*)	Semantic phraseme	Idiom	Idiom (not differentiated)[a]	Collocation
Gläser (1988*a*)	Nomination	Idiom	Idiom (not differentiated)	Restricted collocation
Howarth (1996)	Composite unit	Pure idiom	Figurative idiom	Restricted collocation

[a] Melčuk also recognizes a so-called quasi-idiom, in which the meaning of the whole is derived compositionally from those of the parts, but where there is also 'an unpredictable addition'. An example is *bacon and eggs* (where both ingredients are often fried and the whole is traditionally served -- in the UK at least -- as a breakfast dish).

Meaning-Text Theory and the bulk of his contribution is taken up with an account of how collocations can be rigorously and systematically described in terms of 'lexical functions'. From the elaborate treatment which he provides, two elements of central importance can be singled out. In the simplest terms, a lexical function (or LF) is a general and abstract meaning, coupled with a deep syntactic role, which can be expressed in a wide variety of ways (i.e. have a great variety of lexical realizations) according to the lexical unit (the 'keyword') to which it is applied. In the case of the LF **Oper** (meaning 'do' or 'perform'), and a keyword such as the noun *support*, the appropriate realization is the verb lend. Since about sixty Simple

Standard LFs have been recognized, it is clear that the number of individual collocations that can be accounted for is considerable. It is worth noting also -- as Mel'čuk has demonstrated, and as Fontenelle has shown more recently -- that LFs have much to contribute to the design of collocational dictionaries, where a persistent weakness is the failure to specify adequately the semantic categories to which collocates belong (cf. Cowie, this volume).

The formal apparatus developed by Mel'čuk is a systematic means of indicating not only collocations but also the deep-syntactic patterns to which they conform. The LFs which apply in a given case may be numerous and their semantico-syntactic relations highly complex (as in Mel'čuk's treatment of the noun *analysis*). Here there are parallels with recent work in semantic frame theory, in which a keyword can be associated with a valence description, one which specifies the participants and circumstances involved and shows how those elements may be lexically and grammatically realized (Fillmore and Atkins 1994).

3
THEORETICAL PERSPECTIVES: THE CULTURAL ELEMENT

Drawing upon an exceptionally wide range of Russian examples, Veronika Teliya and her colleagues develop a conceptual framework for describing cultural data as represented in the meanings of multiword units. They argue that phraseology is a particularly fruitful point of focus for 'linguo-cultural' analysis. This is especially true of restricted lexical collocations, to which the analyst cannot do full justice unless cultural meanings are taken into account.

The analytical approach adopted here is situated within a broader anthropocentric paradigm whose central assumption is that every language, and especially its figurative meanings, are concerned with

the reflection and extension of the world-view shared by the linguistic community. The expressions by which a culture is implemented are passed from one generation to the next through linguistic and cultural norms of usage. Language -- and especially phraseology -- is thus a crucial mechanism contributing to the formation and reinforcement of a cultural identity.

Teliya and her colleagues suggest that there are various channels through which language is penetrated by culture. One is the cultural 'seme' (or 'semantic component'). Here, a cultural element constitutes part of the total meaning of the word or word-combination, reflecting encyclopaedic knowledge of the material or historical realia it denotes. A second channel is the cultural concept. Cultural concepts are abstract notions (such as English 'conscience' or 'honour') which construct the world-picture in a culturally specific way.

The major channel, however, is cultural connotation, since it draws on both semes and concepts. It has to do with the interpretative relationship between linguistic items and such symbols of non-verbal codes as stereotypes or myths. Connotation is especially noteworthy in restricted collocations and idioms. In the former, its activation is tied to the cultural information contained in the base of the collocation (often a noun), and the nature of the semantic specialization of the collocate (often a modifier). A recent example is *arkhitektorui/kapitanui perestroyki* ('architects/captains of *perestroika*'), where cultural connotation is linked to the metaphorical interpretation of political reform as construction or seafaring.

Cultural connotations can also arise from the interpretation of concepts or subconcepts. A subconcept such as 'loose woman' may be manifested in a broad spectrum of idioms, collocations, and proverbs, and the exact nature of the cultural connotations may be revealed only if an entire field or domain is studied. The specific images of 'wandering' or 'strolling' in a variety of Russian expressions derive from the same prototype to move about aimlessly' -- a concept that is linked to the folk tradition that a decent woman cannot walk about by herself. Such examples lend eloquent support to the view that the phraseology of a language is deeply marked by its cultural patterns.

4
PHRASEOLOGY IN WRITTEN AND SPOKEN CORPORA

Rosamund Moon and Bengt Altenberg have both been involved for some years in the analysis of large-scale written and/or spoken corpora. Moon, whose research -- as here -- is often linked to her work

as a lexicographer, adopts a descriptive approach which is multidimensional and theoretically flexible. She makes use, for example, of a framework of categories resembling in some respects systems that have been developed independently of corpus-based studies. She also demonstrates that corpus data can be used as fruitfully for investigating the behaviour of a set of items already assembled as it can for identifying such items in the first place.

The starting point of her study is the observation that while 'grammatical' collocations such as *at last, in time,* and *in fact* have very high frequencies in corpus data, the more colourful, stylistically marked expressions -- apart from occurring infrequently -- are often lexically or structurally manipulated in the contexts in which they appear. Furthermore, idioms in the strict sense *(bite the dust, spill the beans)* are exceedingly rare. These intuitive judgements Moon wishes to test against an eighteen-million-word corpus, examining correlations between frequency, form, idiom type, and discourse function. Her aims, like her categorization, are prompted partly by her concerns as a lexicographer -- they are linked to such questions as: 'How much evidence about this class of collocations can I expect to find?' and 'Do the variant patterns in which these idioms occur create special problems of retrieval?'

Especially interesting are the results of correlating frequency with phraseological type. Moon sets up three subcategories within her general class of 'phrasal lexemes'. These are 'anomalous collocations', two of whose subtypes (illustrated by *beg the question* and *on show/ display*) correspond closely to the 'restricted collocations' recognized in other studies; 'formulae' (which include sayings, proverbs, and similes); and metaphors (which can be transparent, semi-transparent, or opaque -- i.e. fully idiomatic). Certain findings bear out those of other analysts (e.g. Howarth 1996): that most metaphors, for instance, have very low levels of occurrence (here less than one per million), while the most commonly occurring items are likely to be anomalous collocations of some kind.

Moon also throws a good deal of light on the syntactic patterns spanned by phrasal lexemes. 'Predicates', consisting of a verb with its complementation, make up a remarkable 40 per cent of the total and there is a close association between predicates, on the one hand, and metaphors and anomalous (i.e. restricted) collocations, on the other. Examination of the corpus material also provides evidence of limited transformation potential and confirms, for instance, restrictions on the passivization of metaphors, though there is too little data for firm conclusions to be reached about idioms. What does clearly emerge is that many phrasal lexemes are frozen in particular transformations,

such as the passive. Such information is of particular value to EFL lexicographers, as it provides a trustworthy basis for guidance in areas where learners continue to experience great difficulty.

In an analysis spanning a broad range of categories, Moon discusses in greatest depth *lexical* phrasal lexemes -- that is, those consisting of two or more open-class words and functioning at or below clause level. An important part of Bengt Altenberg's description, by contrast, is concerned with the *pragmatic,* and specifically with the range of commonplace clauses which regularly occur in spoken English as signals of agreement, acknowledgement, thanks, and so on. It is worth noting, too, that -- whether clauses or phrases -- the kinds of combinations examined are often semantically transparent and structurally uninterrupted, so that they lend themselves well to the approach which Altenberg adopts towards the analysis of his spoken material, which is to focus on continuous strings of three or more words occurring at least ten times.

The description proceeds on two levels, grammatical and functional. The major categories of the grammatical classification are full clauses, clause constituents, and incomplete phrases, the bulk of the examples consisting of single and multiple clause constituents. Interestingly, the independent clauses typically function as responses, and more specifically as a range of speech-act types such as 'thanks', 'reassurance', or 'agreement', thus constituting a category which resembles Gläser's routine formulae.

Few of the full-clause examples seem to be entirely opaque and invariable, and even those expressions with a fixed core generally allow some -- admittedly limited -- expansion. Generally, what chiefly serves to define their phraseological character -- and that of dependent clauses also -- is their pragmatic specialization: even if they are not fully lexicalized, they represent conventional means of conveying specific pragmatic meanings.

The great majority of word-combinations in the sample, however, are not clauses but word sequences below that level, and, of these, the great bulk realize a succession of two or more clause elements (thus subject + verb). Altenberg classifies the sequences according to a positional scheme which specifies both the clause elements (e.g. adverbial, complement) which they realize and the textual functions (e.g. stem, rheme) which they fulfil. Here he refers to a division, which is methodologically as well as theoretically crucial, between a thematic starting point containing given information (and followed by some frame-setting element) and the propositional core conveying new information. The former element is made up of items taken from a limited stock of frequently utilized items -- which Altenberg's

methods are well suited to retrieving -- while the latter is more variable, less stereotyped, and thus less easily captured.

Although Altenberg's study focuses for the most part on combinations with a functional or pragmatic role, his findings echo several of those reached by analysts concerned with collocations and idioms at the propositional core of the utterance. The first is that prefabricated expressions pervade all levels of linguistic organization -- lexical, grammatical, pragmatic -- and affect all kinds of structures, from entire utterances to simple phrases. The second is that there are relatively few examples that are completely invariable or opaque. Like other contributors to this volume, Altenberg recognizes that phraseology is essentially concerned with the more or less conventional and the more or less free.

5
PHRASEOLOGY IN SPECIAL-PURPOSE LANGUAGES AND FOREIGN-LEARNER LANGUAGE

Rosemarie Gläser's work is a meeting point of many strands in present-day phraseological studies. She has contributed substantially to our understanding of the role of multiword units in specialpurpose language, and of the problems of translating idioms and collocations into a foreign language. Most significantly in the present context, she has thrown much light on the contribution made by phraseology to the stylistic expressiveness of literary texts. The system of phraseological categories associated with Gläser, and developed in a number of theoretical and descriptive studies, draws on the analytical schemes initiated by Russian phraseologists. However, it is more elaborate than most other systems and especially notable for its treatment of a range of propositional (i.e. sentence-length) expressions, including proverbs, maxims, slogans, and quotations.

Central to the expressive use of word-combinations is creative manipulation of their form. Deliberate variation, or even distortion, of an idiom or formula to achieve a particular stylistic effect is of course a common device in speech and writing, and is to be distinguished from variation within a multiword unit which is familiar and systemic (and recorded as such in phraseological dictionaries). Creative modification, and the associated play on literal and transferred meanings, are always tied to a specific context.

A further key notion in the stylistic analysis of texts is that of genre -here understood to embrace both literary discourse and technical or scientific discourse. The genres examined in this study cover a broad range, from popular science articles to literary texts.

The work of fiction is in fact the richest field for the deployment of phraseology for stylistic effect, since the literary author has access to the entire wealth of the language, and can draw on its expressive resources on various levels. Referring specifically to the wordcombinations used, Gläser illustrates the wide differences in overall effect that can be achieved, moving from a realistic contemporary novel employing informal conversational formulae to a richly complex work in German, Christa Wolf's *Kein Ort. Nirgends*, which she examines alongside its English translation. Gläser's close analysis gives ample proof of the stylistic resources of the phrasal lexicon and provides firm support for her appeal for a discipline of 'phraseostylistics'.

Sylviane Granger and Peter Howarth are concerned with the analysis of phraseology in the written English of advanced foreign students. Both are interested in identifying the phraseological norms implicitly recognized by native speakers; both are concerned with demonstrating and explaining how the usage of foreign learners deviates from such norms. They differ considerably, though, in the material they examine and the analytical methods they apply. Granger's work on phraseology is one outcome of the International Corpus of Learner English (ICLE) project, the original aim of which was to gather and computerize corpora representing writing by learners from a variety of mother-tongue backgrounds. Her methodology in the present study involves a comparison of native and non-native varieties of English, her hypothesis being that learners will make less use of prefabricated language (collocations and formulae) than native speakers. They will, in other words, use words 'more as building bricks than as parts of prefabricated sections'.

The first part of Granger's investigation focuses on restricted collocations, specifically on amplifiers functioning as modifiers of adjectives (thus, *bitterly cold, unbearably ugly*). This provides firm evidence of sharp differences between native and non-native usage. Examining the choice of individual amplifiers, for instance, Granger notes that *completely* and *totally* are significantly overused by the learners. The wide range of words with which these maximizers combine suggests that they are 'all-round' amplifiers, chosen because learners are ignorant of, or not prepared to risk using, a specific conventional adverb.

A striking feature of Granger's study is her enterprising use of elicited informant data to determine how far foreign learners have developed a sense of what constitutes a conventional ready-made collocation in English. This part of the investigation yielded particularly interesting results, demonstrating that the learners had a sense of

'salience' that was not only weak compared with that of the native speakers, but also partly misguided. For instance, far fewer learners than native speakers selected the stereotyped collocation *bitterly cold;* at the same time, a number of learners, but no native speakers, thought collocations like *fully different* to be normal.

Weighing the pedagogical implications of her research, Granger warns against the notion of basing EFL teaching programmes on a view of first-language acquisition according to which the child first acquires 'chunks' of language, then analyses them, and finally develops from them regular syntactic rules. This approach is not borne out by the little we know of the role of routine patterns in secondlanguage acquisition. Rather, the evidence seems to suggest that there are two learning strategies, one involving automatic speech and the other creative processes. Clearly a balance needs to be struck in EFL teaching between developing phraseological competence and fostering creative skill (Cowie 1988).

In a chapter which is concerned with the relationship between the academic performance and written proficiency of non-native university students, Peter Howarth asks what contribution is made to the latter by phraseological competence. He shows that imperfect control of idioms and collocations can have an appreciable effect on the effectiveness of student writing, deflecting the attention of the reader from message content to linguistic form, and more generally failing to meet the stylistic expectations of the academic community.

Interestingly, Howarth focuses on the language of the social sciences, noting that this throws up more collocational problems than the languages of the pure sciences. And, in contrast with Altenberg, for instance, he concentrates on restricted *lexical* collocations (specifically verb + noun object), recognizing that these represent the propositional core of the clause. The procedure is to analyse phraseology in a corpus of mature native-speaker (NS) writing in order to establish norms by which the performance of a group of non-native students (NNS) can be assessed.

Once the set of all verb + noun object combinations in the NS and NNS corpora had been drawn up, they were assigned to categories in a continuum embracing free collocations, restricted collocations, and idioms. (The debt to Russian phraseological theory, outlined earlier, will be evident.) As in earlier studies of journalistic writing by Cowie (1991, 1992), Howarth clearly shows that for non-native as well as native writers, idioms form a very small proportion of the items identified, and arguably present less severe problems for the learner. Though they are far from easy to identify, it is restricted collocations which make up the bulk of the phraseological material, and which

constitute the most formidable learning difficulty. Here it is typically not a matter of learning fixed units but of knowing when, and how far, the elements of a collocation are able to recombine with other items.

The combined percentages of restricted collocations and idioms in the two native corpora (over 30 per cent in each case) support the view that there is a high and constant level of collocational use across a broad range of formal registers and genres. As for deviant collocations, very few of the NS items fall into this category.

For non-native writers, the percentage of restricted collocations and idioms is 13 per cent below the average for the two NS corpora. There is, of course, considerable individual variation. But, whatever the level, it is difficult to identify any regular connection between collocation use and any other factor (such as measured linguistic proficiency) or to pinpoint precisely the cause of errors, as there is much less certainty as to the competence students are drawing on.

In certain written genres -- such as the advertisements analysed by Gläser -- the reader's attention is often deflected from content to form by the deliberate manipulation of established collocations. In academic discourse, by contrast, native writers feel impelled to remain detached and allow message content to show through. Detachment, however, calls for the use of ready-made language and especially of restricted collocations. Sensitizing advanced learners to the crucial

importance of this central area of the phraseological spectrum must be a major priority of EFL teachers.

6
PHRASEOLOGY AND THE DICTIONARY

Dictionaries of idioms and collocations for foreign learners have benefited greatly, in the 1980s and 1980s, from the development of large-scale corpora and of sophisticated data-processing tools. One area of research which is potentially of great interest within the computational field, since it is likely not only to affect phraseological dictionaries of the traditional hand-held type but also to improve the design of electronic teaching and learning devices, is the exploitation of machine-readable versions of existing dictionaries, bilingual as well as monolingual. In this expanding field, the work of Thierry Fontenelle is especially noteworthy. First, it has met the challenge of extracting collocational information from the microstructure of a machine-readable dictionary (here the *Collins-Robert French-English English-French Dictionary*). Secondly, it has created a bilingual

database of the collocations extracted which is both highly informative and potentially of great pedagogic value.

Fontenelle contrasts his approach with current research in computational linguistics based on the notion that collocation is a statistical phenomenon. The assumption is that, if two or more words occur within a stated distance of each other more frequently than chance would predict, they are collocations (cf. Sinclair 1991). It should be recognized, though, that a growing body of evidence suggests that, while the *class* of restricted collocation is strongly represented in many types of text, individual restricted collocations are typically infrequent -- whether one considers tokens or types. (For many analysts, as for Fontenelle, collocation is a matter of the co-occurrence of lexemes, not word-forms, collocations being definable as such at the abstract level.)

The Collins-Robert Dictionary in machine-readable form is particularly well suited to the semi-automatic extraction of collocations, because a sophisticated scheme of brackets, parentheses, and elements in italics is used to encode various types of collocation systematically. However, as can be seen from an entry such as **slacken** *vi . . . [gale],* the headword one starts from in the hand-held dictionary is often the 'figurative' element (the so-called collocate or collocator), while the element one moves towards is the 'literal' element (the so-called base). Yet the base is the normal *starting* point when composing in the foreign language, and it is for this reason that collocational dictionaries, such as the BBI Combinatory Dictionary of English (BBI) (Benson et al. 1986), have bases as headwords. One advantage of having access to the machine-readable Collins-Robert is that it becomes possible to access information via any element, including the elements in italics (and thus via the bases).

Identification of a given base, say *enthusiasm,* in all the entries in which it occurs enables the analyst to specify its entire collocational range, but also makes possible the automatic breakdown of that collocate list into grammatical subcategories (e.g. Adj + N, Vtr + Nobj) as a function of the syntactic link between the base and the collocate. The specification of such collocabilities, and subcategories, can be performed automatically, but the output has the limitations of a standard collocational dictionary (such as *BBI*): it fails to differentiate between the meanings of the various collocates (e.g. between *decreasing, intense,* and *unflagging* as collocates of *enthusiasm*).

This limitation can be remedied by describing collocations in terms of Mel'čuk's concept of lexical functions (LFs). Drawing upon the Mel'čuk scheme, Fontenelle has devised a system in which simple and complex LFs are assigned to the 70,000 extracted collocations

and incorporated into the *Collins-Robert* database. This has numerous advantages, of which one is the ability of the user to access unknown collocates via a semantic specification (and to assemble all collocates with the same specification).

One of the great merits of the enriched *Collins-Robert* database is that it enables the user (whether linguist, translator, or languagelearner) to retrieve information via a multiplicity of access points. These include the LF, as has already been shown, but also the base, the collocate, the part of speech, and the translation equivalent. This flexibility of use represents a considerable advance over the traditional collocational dictionary, which normally provides for access via the base alone, and lists its collocates without attempting to formalize the semantic relationship between collocate and base (cf. Cowie, this volume).

Among British linguists, the phraseological research needed to put general and specialized EFL dictionaries on a sound footing began in the late 1920s -- though the specialized works, at any rate, did not appear until over fifty years later. As Anthony Cowie shows in his historical survey, this research was initiated in Japan by Harold E. Palmer -- later to be joined by A. S. Hornby -- and resulted in an elaborate categorization published in 1933. Palmer and Hornby's analysis dealt only with composites (or nominations), and it was firmly grammatical: it broke the multiword units down into a number of finely differentiated syntactic categories -- verbal, nominal, adjectival, and so forth. However, though Palmer and Hornby were certainly aware of the broad gulf -- within each of these grammatical types -- between free combinations and ready-made combinations *in general,* they did not acknowledge any shading-off between idioms and restricted collocations, and hence did not provide the analyst with procedures for bringing it to light.

Subsequent developments in the lexicographic treatment of idioms and collocations have built on the major strength of the PalmerHornby tradition -- its grammatical dimension -- and progressed by adopting a critical stance towards its perspective on phraseology. The first advanced-level phraseological dictionary to be produced in Britain -- the *Oxford Dictionary of Current Idiomatic English*, volume i (*ODCIE I*) (1975) dealt with phrasal verbs, and was underpinned by a syntactic classification. It was, as it were, a development in depth of the relevant parts of the Palmer-Hornby grammatical legacy.

However, the companion volume to that dictionary -- *ODCIE 2* (1983) -- though retaining the same approach to grammatical description, also drew on the Russian, or Russian-inspired, systems of phraseological types described earlier. As regards nominations or

composites, *ODCIE 2* recognized a division into pure idioms (e.g. *spill the beans, a busted flush*), figurative idioms *(do a U-turn, a clean sheet)*, and restricted collocations (e.g. *break one's journey, a safe job*) and the appropriate formal criteria were described in the introduction. But those criteria also helped to determine, during compilation, which of the total number of candidates were actually included. It is perhaps worth noting that from the most fuzzy category -- restricted collocations -- only those items were included which were entirely fixed (e.g. *curry favour*) or which allowed very limited substitution (e.g. *a chequered career/history*).

In limiting coverage of restricted collocations as narrowly as this, *ODCIE 2* was tacitly acknowledging the need for a different kind of dictionary -- one which would deal with those thousands of combinations which, while not quite fixed, were none the less problematical for foreign learners. This much-needed resource was the collocational dictionary. However, collocational dictionaries of English compiled, at least in part, by native speakers have been slow to appear. The first was *Selected English Collocations (SEC)* (1982), followed by the generally more satisfactory *BBI Combinatory Dictionary of English (BBI)* (1986).

Both SEC and BBI have certain features that are characteristic of collocational dictionaries generally. Both have an 'orientation' -- from noun headword to verb or adjective collocate -- which serves their predominant productive or 'encoding' function. Both also follow the practice of arranging collocates, within entries, according to their syntactic function (thus, adjective, verb transitive, verb intransitive, etc.). This feature is of great help to a user attempting to search for a specific item, though, as mentioned earlier, limitations remain in the semantic organization of collocates in these dictionaries.

7
PHRASEOLOGY: PRESENT STANDING AND FUTURE PROSPECTS

As recently as the early 1980s it was still possible to dismiss phraseology as a linguistic activity of only minority interest and with poor prospects of recognition as a level of language or of linguistic description. At the time, the subject was literally peripheral in that the chief focus of theoretical and descriptive work was Eastern Europe, while in Britain (especially) the one activity with a known professional commitment to phraseology was dictionary-making, which, though it

drew increasingly on computer-stored data, was not thought worthy of serious attention by more than a handful of trained linguists. [5]

Phraseology is no longer marginalized, partly as a result of the ending of the political and intellectual isolation of Eastern Europe in general and of Russia in particular, partly because the achievements of specialists working in that region have gradually, since the early 1980s, been revealed to a much wider audience, and not least because of the growing recognition being given to phraseology within the heartland of linguistics -- the USA itself. Phraseology has made measurable progress on at least two levels in America. First, there is the work of those who, like Wong Fillmore (1976) and Peters (1983), have made the linguistic community aware of the crucial role played by the formulaic in first-language acquisition and adult languageprocessing. Secondly, the central edifice of generative linguistics itself has come under increasing attack from a phraseological standpoint. The views of Bolinger, for long almost a lone advocate of the formulaic in American linguistics, received powerful endorsement in the late 1980s from Fillmore, who argued that, far from being able to account for linguistic competence by means of rules of great generality, we need a whole spectrum of mini-grammars to account for the variable speech formulae which exist in such abundance (Fillmore et al. 1988).

Above all, there is the compelling evidence, chiefly from European phraseologists, of an increasing weight of descriptive studies. Analyses of written and spoken data, as reported in this volume, bear eloquent witness to the pervasiveness of a wide range of specific types of wordcombination. In terms of frequency, as Moon has demonstrated, these are not the idioms beloved of traditional lexicographers and amateur wordsmiths, but discourse markers *(now then, of course)* and short and stylistically colourless adverbials *(in short, in time)*. What is perhaps more surprising, however, as Cowie and Howarth make clear in their studies of academic written English, is that those 'restricted' lexical collocations which make up the propositional core of the clause account for over 30 per cent of all combinations of a given structural pattern across a range of discourse types. All this evidence has profound implications for foreign language-learning and teaching.

From the late 1980s on, studies of collocations have pushed the boundary that roughly demarcates the 'phraseological' more and

[5]The majority of whom were EFL lexicologists. One sign of growing interest is that, at the biennial congresses organized by the European Association for Lexicography (EURALEX); a special section is now regularly devoted to 'lexical combinatorics'. A recent bibliography (Cowie and Howarth, with Corpas Pastor, 1996) gives some idea of the range of specialist fields -- including dictionary-making -- in which phraseologists are now active.

more into the zone formerly thought of as free, and it should not surprise us if in future dictionaries consist of a higher proportion than before of collocations, idioms, and formulae. But lexicographers will need to draw on the expertise and research findings of phraseologists, both to identify categories of multiword units that we are only now coming to recognize and to do full justice to their meanings, written and spoken forms, and syntactic and pragmatic functions.

PART 1

Theoretical Perspectives: The Russian Tradition; the Cultural Element

[This page intentionally left blank.]

2

Collocations and Lexical Functions

IGOR MEL'ČUK

1

INTRODUCTION

This'his'chapter discusses collocations from the viewpoint of their theoretical and practical (i.e. lexicographic) description. Although they are, and have long been, a popular topic in linguistics, there is, as far as I know, no universally accepted formal definition of collocations nor a proposal for their uniform and systematic treatment in dictionaries. I hope to fill both these gaps, taking up the following four topics:

- characterization and definition of collocations;
- characterization and definition of Lexical Functions -- the main tool for the description of collocations;
- possible uses of Lexical Functions in linguistics;
- presentation of Lexical Functions in the dictionary.

The literature on collocations is simply overwhelming. Since it is out of the question to present here even a partial survey of it, I will abstain from detailed references to other approaches, limiting myself to an absolute minimum.

2

COLLOCATIONS

Collocations -- no matter how one understands them -- are a subclass of what are known as *set phrases;* [1] they therefore have to be defined in

The first draft of the paper has been read (as always) by L. Iordanskaja; its present form owes much to (sometimes violent) discussions with M. Alonso Ramos, who read the subsequent text, and T. Reuther. Tony Cowie and A. Grosu went through the pre-final version, and Iordanskaja checked the final one, hunting down many remaining inconsistencies. I tender [= Oper1.(*gratitude)]* my heartfelt [= Magn(*gratitude)]* gratitude to all of them, while taking [= Real $_1$.(*responsibility)]* full [= Magn(*responsibility)]* responsibility for all errors and obscurities that survived their scrutiny.

[1] They are also known under a host of other names: fixed (frozen) phrases, wordcombinations, idiomatic expressions, idioms, etc. See the terminological discussion in the Introduction to this volume.

terms of their *differentiae specificae* with respect to set phrases that are not collocations. This establishes my course: first, I define set phrases; then I propose a calculus or typology of set phrases; finally, I point out the place that collocations occupy among set phrases by supplying a formal definition of collocation.

2.1

Set phrases, or phrasemes

People speak in set phrases, rather than in separate words, hence the crucial importance of set phrases. At the same time, set phrases, or PHRASEMES, represent one of the major difficulties in theoretical linguistics as well as in dictionary-making. [2] Therefore, both linguistic theory and lexicography should really concentrate on them (an idea that has been advocated for about twenty-five years by a number of people, including, among others, Becker (1975), Bolinger (1976, 1977), Pawley (1985), Jackendoff (1995), and the present writer). To show what I mean by phrasemes, here are several examples collected from one newspaper column (phrasemes are in bold):

(1) (a) **Of course,** investors **accept the challenge offered** by this region.
(b) Rabin **made** these **remarks** in an interview.
(c) North Queensland is **best known** for its reef.
(d) The rejection by the Bosnian Serbs of the plan **placed** them **on a collision course** with the five powers.
(e) His statement **added fuel to the fire.**
(f) The share **price** of Perilya Mines **collapsed** yesterday **under the weight** of **heavy** share selling **orders.**
(g) **The hardest thing, for instance,** will be **making decisions.**

A good dictionary of language **L** should include *all* the phrasemes of **L,** because the main substantive property of a phraseme is its noncompositionality: it cannot be constructed, for a given Conceptual Representation, from words or simpler phrases according to general rules of **L,** but has to be stored and used as a whole. A phraseme is a lexical unit; and, more crucially, it is the numerically predominant lexical unit: in any language -- i.e. in its lexicon -- phrasemes outnumb er words roughly ten to one. Collocations make up the lion's share of the phraseme inventory, and thus deserve our special attention.

[2] This is so because phrasemes cannot be studied in any one of the traditional divisions of linguistics (thus, not in semantics nor in syntax), precisely because of their non-compositional, 'irregular' semantic and syntactic nature.

2.2
A typology of phrasemes
2.2.1
Preiminary notions
To define a phraseme, I need some preliminary notions: those related to my assumptions concerning the way a speaker produces a text, those related to the linguistic sign, and two auxiliary concepts.
Text production
It should it be emphasized that the following discussion of collocations makes sense only if we look at them from the *speaker*'s viewpoint: in this chapter, phrasemes are considered exclusively in terms of their production or construction (rather than in terms of their interpretation by the addressee).I adopt the following view of text production (for the MeaningText framework, see Mel'čuk 1974: 9-44, 1981, 1988a: 43-101, 1993: 41-79):

- The speaker begins with what I call the Concept(ual) R(epresentation) [or ConceptR] of the situation he wishes to verbalize. The ConceptR is a mental reflection of perceived reality, of the speaker's encyclopaedic knowledge relevant to the situation in question, of his intentions, preferences, wishes, and goals, of his ideas about the addressee, etc. The ConceptR of a given situation contains everything that might be needed in order to say what the speaker wants to say about it.
- Based on the initial ConceptR, the speaker constructs the Sem(antic) R(epresentation) [or SemR] of his intended utterance. He does so according to the Concepts-Meaning Model of his language L [=CMM(L)], which associates with elements and configurations of the ConceptR elements and configurations of the corresponding SemR.
- From a given SemR, the speaker constructs, through a series of steps, the Phon(etic) R(epresentation) of the utterance; he does so according to the Meaning-Text Model of L, or MTM(L), which associates with elements and configurations of the SemR all the actual linguistic elements that make up the corresponding actual utterance.

An utterance is thus produced in two major steps using two models and involving three major representations:

	CMM		MTM	
{ConceptRk}	U=21D4	{SemRi}	U=21D4	{PhonRj}

Linguistic sign

A phraseme (like a word-form, a morph, etc.) is a *linguistic sign.* A linguistic sign is an ordered triple:

$X = \ll$'X' ; /X/ ; $\Sigma_x \gg$,

where 'X' is the signified of the sign **X** (= its meaning), /X/ is its signifier (= its phonetic form), and σ_x, its syntactics (= the set of data on its co-occurrence with other signs) (see Mel'čuk 1992: 40-1, 1993: 117-29). Except for syntactics, which has been added by the author, the concept of sign adopted here is clearly Saussurean. For simplicity of presentation, I leave syntactics out of consideration in my discussion of phrasemes.

Auxiliary concepts

The concepts 'unrestrictedly constructed **E'** and 'regularly constructed **E',** as applied to the signified or the signifier of a multi-unit expression, are crucial to the definition of phraseme. These concepts are to be understood as follows:

> 1 Unrestrictedly constructed **E** = 'an **E** whose components are selected -- for a given starting representation -- according to *arbitrarily chosen* selection (≈on) rules of *L'.*

If the signified/the signifier E of an expression is constructed unrestrictedly, no rules $\{R_E\}$ applied to construct **E** are mandatory: instead of {RE}, the speaker can apply *any* other applicable rules {RE'} to produce an equivalent **E'** . Thus, the signified and the signifier of the phrase No *parking* are not unrestrictedly constructed, because it is not acceptable to express -- on a sign -- any equivalent meaning, for instance 'you should not park here', or the same meaning in a different form, such as *Parking not allowed* or *Do not park,* although lexical and grammatical rules of English allow you to do so. [3] In contrast, the signified and the signifier of the sentence *This dictionary has been compiled by many people* are unrestrictedly constructed, because you can express the same or an equivalent meaning by any other appropriate linguistic means: e.g. *This dictionary is the result of work by many hands,* etc. 'Unrestrictedness' thus means unlimited freedom of choice among (quasi-) equivalent independent meanings and expressions; it has to do with the *selection* of meanings and lexical units and is related to the concept 'selection rules of a language'. However, it should be emphasized that for signifiers an additional proviso is necessary: a complex signifier is not unrestrictedly constructed if one

[3]You do, though, come across signs like Parking prohibited near loading bays.

of its parts is selected contingent on another one. We will see the importance of this condition in Definition 2.

> 2 Regularly constructed **E** = 'an **E** whose components are combined exclusively according to general combination rules (= grammar) of **L'.**

If the signified or the signifier of an expression **E** is constructed regularly, its components are put together, or united, solely by general rules of **L.** Thus, all the expressions mentioned in the previous paragraph are constructed regularly, while the signified of the expression *the chip on N's shoulder* (i.e. 'a grievance which makes **N** permanently discontented and

quarrelsome') is not, because there is no way to construct it out of the signifieds 'chip', 'on', and 'shoulder' by general rules of English.'Regularity' thus means the observance of general rules in the *combination* of meanings and expressions and is related to the concept 'combination rules of a language'. These rules are represented in the formalism of Meaning-Text Theory by the *Operation of Linguistic* Union ⊕:i.e. putting together linguistic items of **L** while constructing expressions of a higher order (Mel'čuk 1982: 41-2, 1993: 137-44). The symbol ⊕ is reminiscent of arithmetical summation, but linguistic union is much more complex than simple addition: it presupposes observing *all* the general combination rules of **L,** and doing this in conformity with the nature of items being united (signifieds are united in a different way from signifiers and syntactics, etc.). Thus, **X** ⊕**Y** denotes the regular union of signs **X** and **Y** (i.e. the expression **X** ⊕**Y** is regularly constructed out of signs **X** and **Y**); 'X'⊕'Y' is the regular union of signifieds 'X' and 'Y' -- and so on.Informally and approximately, a phraseme is a phrase whose signified and signifier *cannot* be constructed both unrestrictedly and regularly.

2.2.2

Free phrases

Definition 1: Free Phrase A FREE PHRASE **A** ⊕**B** in language **L** is a phrase composed of lexemes **A** and **B** and satisfying simultaneously the two following conditions:

1. Its signified 'X' = 'A⊕B' is unrestrictedly and regularly constructed on the basis of the given ConceptR -- out of the signifieds 'A' and 'B' of the lexemes A and B of L;
2. Its signifier /X/ = /A⊕B/ is unrestrictedly and regularly constructed on the basis of the SemR 'A⊕B' -- out of the signifiers /A/ and /B/ of the lexemes **A** and **B.**

Informally, a free phrase **A** ⊕**B** is a phrase such that: (1) its signified 'A⊕B' is freely constructed for the given ConceptR and can be replaced by any other sufficiently close signified 'Y', obtainable from the same ConceptR by rules of **L;** (2) this signified is a regular union of the signifieds of the phrase's components and its signifier is a regular union of their signifiers, such that the phrase **A** ⊕**B** can be produced according to general combination rules of **L:**

A ≪'A; /A/≫⊕**B** ≪'B'; /B/≫ = **A** ⊕**B** ≪'A⊕B'; /A⊕B/≫.

For a phrase to be free means *freedom of selection* (of its signified -- with respect to the given ConceptR, that is, in the ultimate analysis, to the given situation; and of its signifier -- with respect to the corresponding SemR) and *freedom of combination* (of its components: according to their own signifieds and syntactics plus the general combination rules of **L).**

2.2.3

Set phrases, or phrasemes

A SET PHRASE, or PHRASEME, **AB** is a phrase which is *not free.* Being not free can have three sources: either both Conditions 1 and 2 in Definition 1 are violated; or Condition 1 (but not 2) is violated; or Condition 2 (but not 1) is violated:

1. Condition 1 is violated -- such that the signified 'X' = 'A⊕B' is not unrestrictedly constructed on the basis of the given ConceptR (although it is regularly constructed) -- and Condition 2 is violated as well (in the same way). Then, for the given ConceptR, *only* the given signified 'A⊕B' coupled with the given signifier /A⊕B/ is possible: the phrase in question is not unrestrictedly constructed. Not *all* applicable rules of **L** can actually be applied in the construction of **AB** while selecting its components; the choice of an appropriate meaning is reduced to one possibility (or to a few), and so is the choice of the form. As a result, we get PRAGMATIC PHRASEMES, or PRAGMATEMES. For instance, one sees on a restaurant sign *Caesar Salad: All you can eat;* its counterpart in French is *Salade César á volonté,* lit. ('Caesar Salad to [your] wish', i.e. 'as much as you want'). It would be semantically and syntactically correct to say in French #*Salade César: Tout ce que vous pouvez manger;* however, this expression smacks of a calque: this is not the way a Frenchman would put it. [4] (The symbol # indicates pragmatic inappropriateness: #*X* means '*X* should not be used in the given situation'.) Thus, 'X: all you can eat' and *'X à volonté'* are pragmatemes of English and French, respectively.

[4] The converse is true of English: #*Caesar Salad: As much as you like* is fully grammatical and understandable, but it is not what English speakers write on their signs.

Condition 1, but not Condition 2, is violated (as above). Then, for the given ConceptR, still only one given signified 'AEDB' is possible, but it is unrestrictedly expressible, i.e. although you cannot use an equivalent meaning, for 'A⊕B' you can choose any one of several possible (quasi-) synonymous expressions that the rules of **L** allow. Such expressions are PRAGMATEMES as well; for example, signs in a US library meant to prohibit talking say *No talking please, Please do not talk, Please be quiet,* etc. (but not, for instance, #*Don't make any noise please or* #*Keep silent please).*

All ready-made expressions (like greetings, typical phrases used in letters, conversational formulae, technical clichés, proverbs, sayings, etc.), even if they are wholly compositional semantically and syntactically, are pragmatemes: they are non-compositional pragmatically. (In this study, I will not consider pragmatemes: I am concerned solely with collocations.)

Condition 2 is violated, but Condition 1 is not (in the sense that the signified of **AB** is constructed unrestrictedly; yet it is not constructed regularly). Then for the given ConceptR, any signified obtainable by general selection rules is possible, but for a selected signified 'A⊕B', the corresponding signifier /X/ is not unrestrictedly constructed: if 'X' = 'A⊕B', then /X/=/A⊕B/. We thus have SEMANTIC PHRASEMES. [5] (The important distinction between pragmatic and semantic phrasemes was first established in explicit terms in Morgan (1978).) From now on, I shall be concerned with semantic phrasemes only. Let me establish their major types.

Condition 2 of Definition I can only be infringed in the following three ways:

- **AB** = ≪'C'; /A⊕B/≫ | 'C'⊕'A' & 'C'⊕'B' This formula describes FULL PHRASEMES, or IDIOMS ([to] *shoot the breeze, [to] spill the beans, [to] pull* [N's] *leg,* [to] *trip the light fantastic, of course, [to) put* [someone] *up, [a] red herring).* Instead of the regular union 'A⊕B'of the signifieds 'A' and 'B', an idiom **AB** has a different signified, 'C', including neither 'A' nor 'B'.

- **AB** = ≪'A⊕C'; /A⊕B/≫ | 'C' is expressed by **B** such that /A⊕B/ is not constructed unrestrictedly

  These are SEMI-PHRASEMES, or COLLOCATIONS ([to] *land a* JOB; *high* WINDS; [*to] crack a* JOKE, [*to] do* [N] *a FAVOUR, [to] give [N] an ULTIMATUM, [to] launch an ATTACK, [to] stand COMPARISON [with N], strong COFFEE).* The signified of a collocation includes 'intact'

[5]The signifier of a semantic phraseme can be constructed regularly or not; this is immaterial in the present context.

the signified of the one of its two constituent lexemes -- say, of A (shown in the examples in small caps); **A** is freely chosen by the speaker strictly because of its signified. But the other component of its signified, i.e. 'C', is 'problematic': it is expressed by **B,** which is chosen *contingent* on **A** (this means that the signifier of a collocation is not unrestrictedly constructed).

- AB = ≪'A⊕B⊕C'; /A⊕B/≫ | 'C' ⊕ 'A' & 'C' ⊕ 'B'

  These are QUASI-PHRASEMES, or QUASI-IDIOMS ([*to] give the breast [to* N], [*to] start a family, bacon and eggs, shopping centre).* Here the signified of **AB** includes the signifieds of both constituent lexemes, but also contains an unpredictable addition 'C'.

The general classification of phrasemes can be presented schematically as shown in Fig. 2.1.

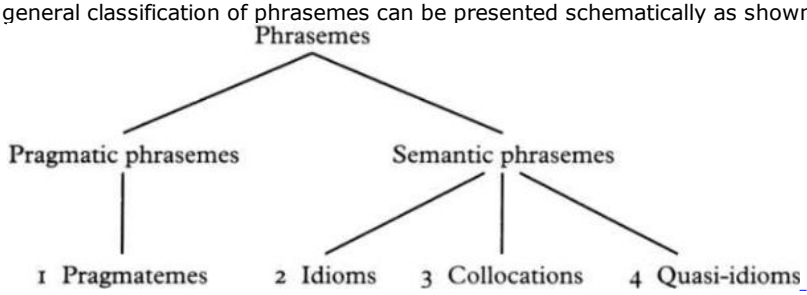


FIG. 2.1. Classification of phrasemes

2.3
The concept of collocation

Definition 2: Collocation (= Semi-Phraseme) A COLLOCATION **AB** of language **L** is a semantic phraseme of **L** such that its signified 'X' is constructed out of the signified of one of its two constituent lexemes -- say, of **A** -- and a signified 'C' ['X' = 'A⊕C'] such that the lexeme **B** expresses 'C' only contingent on **A.** The formulation '**B** expresses "C" only contingent on **A** ' covers four major cases, which correspond to the following four major types of collocations:

1. EITHER 'C' ⊕ 'B', i.e. **B** does not have (in the dictionary) the corresponding signified;

   AND [(a) 'C' is empty, that is, the lexeme **B** is, so to speak, a semi-auxiliary selected by **A** to support it in a particular syntactic configuration;

OR ('b) 'C' is not empty but the lexeme **B** expresses 'C' only in combination with **A** (or with a few other similar lexemes)];
OR 'C' = 'B', i.e. **B** has (in the dictionary) the corresponding signified;

AND [(*a*) 'B' cannot be expressed with **A** by any otherwise possible synonym of **B;**

OR (*b*) 'B' includes (an important part of) the signified 'A', that is, it is utterly specific, and thus **B** is 'bound' by **A].**

Examples (lexeme **A** is in small capitals): Case 1 (*a):* collocations with support (i.e. 'light') verbs, such as [*to] do* [N] *a FAYOUR, [to] give a LOOK, [to] take a* STEP, *[to] launch an APPEAL, [to] lay* SIEGE [*to* N]. Case 1(*b):* collocations such as *black COFFEE, French* WINDOW, Fr. BIéRE *bien frappée (★battue)* 'well chilled (lit. 'beaten') beer'.

Case 2(*a):* collocations with intensifiers, such as *strong (★powerful)* COFFEE, *heavy (★weighty)* SMOKER, *deeply* MOVED, [*to]* ILLUSTRATE *vividly.*

Case 2(*b):* collocations such as *The HORSE neighs, aquiline* NOSE, *rancid BUTTER,* or *artesian WELL.* [6]

Collocations constitute the absolute majority of phrasemes and represent the main challenge for any theory of phraseology. In order to describe collocations in a rigorous, systematic, and exhaustive way, Meaning-Text Theory proposes the apparatus of Lexical Functions.

3
LEXICAL FUNCTIONS
3.1
Introductory remarks

I begin with the general concept of LEXICAL FUNCTION [=LF] and then proceed to a particular one -- the SIMPLE STANDARD LF -- which is of special interest here (Zolkovskij and Mel'čuk 1967; Mel'čuk et al. 1984/1988/1992; Mel'čuk 1996). The term *function* is used in its mathematical sense: f(*x*) = *y,* and the adjective *lexical* indicates that **f** 's

[6]The difference between cases of the type of *black coffee* 1(b) and those of the type of *artesian well 2(b)* is that BLACK does not have in the dictionary the sense 'without milk' among its different senses, because it realizes this sense only with COFFEE, whereas ARTESIAN has -as its only sense -- '[well] such that water in it rises to the surface without pumping'. In other words, the difference between cases 1(*b)* and 2(*b)* depends entirely on the lexicographic treatment

we adopt for phraseologically bound senses. However, the problem of the lexicographic description of lexical units is an independent problem that has to be solved (or presupposed to be solved) prior to any discussion of phraseology.

domain of definition as well as the range of **f'** s values are both sets of lexical expressions.

A LEXICAL FUNCTION **f** is a function that associates with a specific lexical unit [= LU], L, which is the 'argument', or 'keyword', of f, a set $\{L_i\}$ of (more or less) synonymous lexical expressions -- the 'value' of **f** -- that are selected contingent on L to manifest the meaning corresponding to **f:**

f(L) = {Li}.

To put it differently, an LF, particularly a Simple Standard LF, is a very general and abstract meaning, coupled with a D(eep-)Synt (actic) role, which can be lexically expressed in a large variety of ways depending on the lexical unit to which this meaning applies.

About sixty Simple Standard LFs have been recognized so far in natural languages. Let me cite four preliminary examples and then proceed to definitions:

'the one who/which undergoes . . .' [*nomen patientis]*			
S2(to shoot)	= target	S2(to server)	= client
S2(hotel)	= guest	S2(prison)	= prisoner
S2(doctor)	= patient	S2(hairdresser)	= customer
'intense(ly)', 'very' [intensifier]			
Magn(shaveN)	= Close, clean	Magn(naked)	= stark
Magn(easy)	= as pie, as 1-2-3	Magn(thin)	= as a rake
Magn(to condemn)	= strongly	Magn(to rely)	= heavily
'do', 'perform' [support verb]			
Oper1(cryN)	= to let out [ART~]		
Oper1(figureN)	= to cut [ART~] [He cut a miserable figure]		
Oper1(strikeN)	= to be [on~]		
Oper1(supportN)	= to lend [~]		
'realize', 'fulfil [the requirement of]'			
Real2(mineN)	= to strike [ART~] [Their car struck a land mine]		
Real2(testN)	= to withstand [ART~]		
Real2(jokeN)	= to get [ART~]		
Real2(examN)	= to pass [ART~]		

(The symbol ART indicates that an article or a grammatically equivalent determiner should be used, according to grammatical rules.)

3.2
Central concepts: LFs and Simple Standard LFs

Definition 3: Lexical Function

A function **f** associating with a lexical unit L a set **f** (L) of lexical expressions is called a LEXICAL FUNCTION if and only if one of the following two conditions is satisfied:

A. Either **f** is applicable to several Ls; in that case, for any two different L_1 and L_2, if **f** (L1) and **f** (L2) both exist, then:
   1. Any elements of **f** (L_1) and of **f** (L_2) bear an (almost) identical relationship to L_1 and L_2, respectively, as far as their meaning and the DSynt-role are concerned; i.e. for any $L_{f(L1)} \in$ f(L_1) and any $L_{f(L2)} \in$ **f** (L_2), it is true that

      'Lf(L1)' : 'L1' ≈ 'Lf(L2)' : 'L2'.

   2. At least in some cases, **f** (L_1) = **f** (L_2).

B. Or **f** is applicable to one L only (maybe to two or three semantically related Ls).

LFs of type A are called 'normal' LFs; those of type B, 'degenerate' LFs. In **f** (L), L is the 'keyword' [7] of **f,** and **f** (L) is the 'value'.**Definition 4: Standard Lexical Function** A normal LF **f** is called a STANDARD LEXICAL FUNCTION if and only if the following two (additional) conditions are simultaneously met:

3. **f** is defined for a relatively large number of arguments. (To put it differently, the meaning 'f' is sufficiently abstract and general to be applicable to many other meanings.)
4. **f** has a relatively large number of lexical expressions as its value -- such that these expressions are more or less equitably distributed between different keywords.

Normal LFs that do not satisfy both Conditions 3 and 4, on the one hand, and degenerate LFs, on the other, are called NON-STANDARD. (Thus the difference between Standard and Non-Standard LFs is purely quantitative: it concerns the number of possible keywords and value elements.)

Among Standard LFs, a subset of about sixty basic LFs is singled Out: SIMPLE STANDARD LFs. Simple Standard LFs constitute the foundations for the description of irregular derivation and restricted lexical co-occurrence (that is, of collocations).

3.3
The system of Simple Standard LFs

Since full lists of Simple Standard LFs are found in previously mentioned publications as well as in Mel'čuk and Zholkovsky (1988), Mel'čuk *et al.* (1984/ 1988/ 1992), and Melč'uk (1996), I will not supply such a list here. Instead, I will offer a brief substantive characterization

[7]When speaking of LFs, I avoid using the term argument because of its multiple ambiguity.

of LFs, sketch their classification and quote four groups, which include the most current LFs.

3.3.1
Informal Characterization of LFs

Lexical Functions were first introduced by žolkovskij and Mel'čuk (1967). They are used to describe two types of lexical phenomena that turn out to be of the same logical nature, that is, both are readily amenable to a description via the concept of function in the mathematical sense.The first type involves PARADIGMATIC lexical correlates $\{L_{i'}\}_{par}$ of a given lexical unit L; they can be loosely described as (quasi-) synonymous with L. An L' can designate a situation or an object close to or identical with 'L', a generic notion for 'L', a situation implied by 'L', or a participant in the situation (implied by) 'L'. Thus, where L = *school,* $\{L_{i'}\}_{par}$ = *teacher, student, subject, exam, lesson, mark, class, [to] teach, [to] learn,* etc. Where L = [*to] escape,* $\{L_{i'}\}_{par}$ = [*to] flee, [to] break away, escape* $_{(Noun)}$, *escapee, place of confinement,* etc. Such lexical correlates show kinds of derivational relationship with L.The second type involves SYNTAGMATIC lexical correlates $\{L_{i'}\}_{synt}$ of L that form with L collocations like some of those in bold in the examples at (1): *offer/accept the challenge, make a remark, best known, place on a collision course, heavy [selling] orders.* Thus, where L = *school,* $\{L_{i'}\}_{synt}$ = *teach (school), go (to school), graduate (from high school),* etc.; and where L = *escape*($_{Noun}$), $\{L_{i'}\}_{synt}$ = *daring.*LFs represent both types of lexical correlates of L.

3.3.2
Classification of LFs

LFs can be classified from different viewpoints; without having a scientific impact on the issue, such classifications facilitate the task of the user and thus possess pedagogical value. (The present classification and description of LFs follows some suggestions in Alonso Ramos and Tutin 1994.)

- Paradigmatic vs. syntagmatic LFs have been already characterized. Paradigmatic LFs deal with selection; they are aimed at answering questions of the type 'What do you call an object, situation, etc. X, related to Y?' -- while speaking of X rather than of Y. Syntagmatic LFs deal with combination; they are aimed at answering questions of the type 'What do you call the action, characteristics, attributes, etc. X of Y?' -- while speaking of Y rather than of X.
- Standard vs. non-standard LFs are different, first of all, with respect to the number of their possible keywords and value

elements. Another important difference is that standard LFs participate in synonymic paraphrasing while non-standard ones do not. (This distinction will not be fully explained here; for details, see e.g. Mel'čuk 1992b).

- Ten semantic/syntactic groups of Simple Standard LFs can be distinguished, based on the meaning and the DSynt-role associated with the given LF. These are as follows:
  1. BASIC LFs: **Syn** (onym), **Anti** [= antonym], and **Conv** (ersive) $_{ij}$. These embody the main semantic relations that play a special role in MT-Theory -- synonymy, negation, and converseness (*X precedes Y ~ Y follows X).* Since they are relatively well known, I will not discuss them here, except to say that **Syn, Anti,** and **Conv** $_{ij}$ can be semantically exact or approximate, i.e. they can have a richer ($\supset$), poorer ($\subset$), or intersecting ($\supset$) meaning, in which case they are quasi-synonyms, quasi-antonyms, and quasi-conversives. The same subscript symbols are also used for other LFs.
  2. DERIVATIVES are of two subtypes: SYNTACTIC derivatives represent nominalization S_0 . (*rejection* from REJECT), adjectivalization A_0 (*urban* from CITY), verbalization V_0 . (*to attack* from [*the*] ATTACK), and adverbialization Adv_0 (*well* from GOOD); **Pred** is a combination of a meaning with the copula; thus **PredMagn** (*animosity) = runs rampant.*

     SEMANTIC derivatives are, roughly speaking, agent noun S_1 , patient noun S_2 , active adjectival A1 (in *search of* from [*to]* LOOK FOR), passive adjectival A_2 (*under construction* from [*to]* BUILD), place noun S_{loc} , instrument noun S_{instr} , active potential adjective **Able** $_1$ (*inquisitive* from [*to]* ASK), passive potential adjective **Able** $_2$ (*reliable* from [*to]* RELY), etc.
  3. GENERICS: hyperonym **Gener** and metaphoric denotation **Figur** (*curtain* of RAIN).
  4. QUANTIFIERS: singulative **Sing** (*speck* of DUST) and collective **Mult** (*pride of* LIONS).
  5. MODIFIERS: **Magn, Plus/Minus, Ver** (*restful SLEEP*), **Bon** (*valuable CONTRIBUTION, delicious MEAL).*
  6. PHASALS: verbs denoting the three phases of an event -- the beginning (**Incep),** the end (**Fin**), and the continuation (**Cont**). These LFs are often used in combination with other verbal LFs.
  7. CAUSATIVES: verbs denoting the three possible types of causation: causation of existence (**Caus**), causation of non-existence (**Liqu**), and non-causation of non-existence (**Perm**).

It should be noted that the phasals stand in antonymous relation to each other; the same holds true of causatives: **Incep = AntiFin, Liqu = AntiCaus,** etc. Furthermore, causatives and phasals are related, since one can cause the beginning, the end, or the continuation of an event; however, this issue cannot be further discussed here.

AUXILIARIES (= support, or 'light', verbs): these are semantically empty verbs linking a DSynt-actant [= A] of L to L; **Oper $_{1,2}$** takes L as its DSyntA **II, Func $_{0,1,2'}$** as its DSyntA **I,** and **Labor $_{12,21}$ '** as its DSyntA **III** (for more details, see the following subsection).
REALIZATIONS: **Real $_{1,2'}$ Fact $_{0,1,2'}$ Labreal $_{12,21}$.**
VARIA: **Involv, Son, Imper, Degrad, Manif, Sympt.** Simple Standard LFs can form combinations, to produce Complex Standard LFs: such as **AntiMagn, IncepOper $_{1,2'}$ CausFunco $_0$ ' CausPredPlus,** etc.

3.3.3
Illustrative list of LFs

As indicated above, I will explain and illustrate four groups of Simple Standard LFs.

1 Semantic derivatives: actantial and circumstantial nouns S $_i$ is a standard name of the i-th DSyntA of L; it is thus an ACTANTIAL noun: primarily, these are nouns denoting the agent ('the one who Ls') and the patient ('the one whom someone or something Ls').

Syntactically, there are **S $_i$** s of two types. An **S $_i$** (L) of the first type is used in the text, as a rule, instead of L, especially if this L is a verb. An Si(L) of the second type is used in the text together with L, taking it as its own DSyntA **II: S $_I$** →+ΠL, etc.

S1(to teach) =	teacher	S1(letter) =	author; sender [of the letter]
S2 (to teach) =	(subject) matter, subject [in high school]	2(letter) =	addressee [of the letter]
S3 (to teach) =	pupil	S3(letter) =	contents [of the letter]

S $_{instr'}$ S $_{med'}$ S $_{mod'}$ S $_{loc'}$ S $_{res}$ are standard names of instrument, means, mode, location, and result of the situation denoted by L (as a rule, L is a noun or a verb); **S $_n$** s are thus CIRCUMSTANTIAL nouns. Like actantial nouns, **S $_n$** s normally are used instead of their keyword L; if they are not, they also take it as their DSyntA **II: S $_{instr}$** →+ΠL, etc.

Sinstr⊃(to shoot) =	firearm	Sloc(to fight [as of two armies]) =	battlefield

S_{instr}(murderV,N) =	[murder] weapon	Sloc(war) =	theatre (of war)
S_{med}⊃(to Shoot)=	ammunition	S_{res}⊃(to learn) =	knowledge, skills
S_{mod}(to write) =	style [He writes elegantly ~ The style of his Writing is elegant]		

2 Intensifiers

Magn(agree) =	wholeheartedly	Magn(committed) =	deeply
Magn(analysis) =	trenchant	Magn(deserve)	= richly
Magn(bore $_N$) =	crashing	Magn(work $_V$) =	like a Trojan, one's guts out,

3 Semi-auxiliary verbs

The LFs **Operi, Funci,** and **Labor** $_{ij}$ are SUPPORT (or 'LIGHT') VERBS (Gross 1981; Catell 1984); they are semantically empty (or emptied) in the context of the keyword LU.This LU is necessarily a noun whose meaning is or includes a predicate (in the logical sense of the term), thus presupposing actants. In other words, the keyword of these LFs is, as a general rule, the name of an action, an activity, a state, a property, a relation, etc. (It can also be the name of a concrete object, which is defined by its role in a situation. Such is the case with a body part or organ, for instance: they represent what is called 'inalienable possession' and have as the value of the LF **Oper $_1$** the verb [*to*] HAVE or its equivalent.)

Support verbs serve to link, on the DSynt-level, (the name of) a DSynt-actant of L to L itself; they thus play an important semanticsyntactic role and cÃan be loosely called semi-auxiliaries.

1. **Oper $_i$** [Lat. operari '[to] do, carry out']: the DSyntA **I** of this verb (and its SSynt-subject) is the phrase that is described in the Government Pattern [= GP] of L as the *i*-th DSyntA of L, and **Oper $_i$ '** s DSyntA **II** (= its mains [8] S(urface)Synt-object) is L itself. (Further DSyntAs of **Oper $_i$,** if any, are the phrases described in the GP of L as further DSyntAs of L.)

Oper $_1$ (blow $_N$)	= [to] deal [ART ~ toN]	Oper $_2$ (blow $_N$)	= [to] receive [ART ~ fromN]
Oper $_1$ (support $_N$)	= [to] lend [~ to N]	Oper $_2$ (support $_N$)	= [to] receive [~ from N]
Oper $_1$ (order $_N$)	= [to] give [ART ~ toN]	Oper $_3$ (order $_N$)	= [to] receive [ART ~ fromN]

[8] A main Surface-Syntactic Object of a lexical unit L is either its D(irect) O(bject) (if L can have a DO), or its I(ndirect)O (if L cannot have a DO), or the strongest Prep (ositional) O (in the absence of both DO and 10).

Oper $_1$ (resistance)	=[to] put up [ART ~], [to] offer [ART/ ~]	Oper $_2$ (resistance)	= [to] meet [ART ~], [to] run [into ART ~]
Oper $_1$ (control $_N$)	= [to] have [~ over N]	Oper $_2$ (control $_N$)	= [to] be [under N's ~]

The expression in brackets following each element of the value of the LF illustrated is its REDUCED GOVERNMENT PATTERN -- its lexical subentry.2. **Func** $_i$ [Lat. ★*functionare* '[to] function']: the DSyntA **I** of this verb (and its SSynt-subject) is L itself, and its DSyntA **II** (= its main SSynt-object) is the *i*-th DSyntA of L.

Func $_1$ (*blow $_N$*) = *comes [from* N] **Func $_2$** (*blow $_N$*) = *falls [upon* N] **Func $_1$** (*proposal) = comes, stems***Func $_2$** (*proposal) = concerns* [N] [*from* N]

In cases where there is no object at all, i.e. where **Func $_i$** is an absolutely intransitive verb, the subscript $_0$ is used:

Func $_0$ (*snow $_N$*) = *falls [At night, the snow started to fall]* **Func $_0$** (*war) = is raging* **Func $_0$** (*silence) = reigns*

3. **Labor $_{ijk}$** [Lat. *laborare* '[to] work, toil']: the DSyntA **I** of this verb (and its SSynt-subject) is the i-th DSynt-actant of L, its DSyntA **II** (= its main SSynt-object) is j-th DSyntA of L, its DSyntA **III** (= its second SSynt-object) is *k*-th DSyntA of L, and its further DSyntA (= its third SSynt-object) is L itself.

Labor $_{12}$ (*interrogation) = [to] subject* [N *to an interrogation,* where the keyword INTERROGATION is DSyntA **III** of the verb *subject]*

Labor $_{321}$ (*lease $_N$) = [to] grant* [N *to* N *on lease,* where the keyword LEASE $_N$ is DSyntA **IV** of the verb *grant]*

The LFs **Oper $_{o/i}$ ' Funco $_{o/i}$ '** and **Labor $_{ijk}$** can be paired in converse relations:

Oper $_1$ = **Conv $_{21}$ (Func $_1$)** ; **Labor $_{12}$** =**Con $_{132}$ (Oper $_1$);** etc.

These relations may be represented diagrammatically -- for a twoactant LU -- as shown in Figure 2.2.

In Fig. 2.2 , a two-actant lexeme L (= ANALYSIS, with two DSyntAs: **I** -- JOHN, and **II** -- PHENOMENON) is presented; the whole means 'John analyses the phenomenon'. The arrows represent

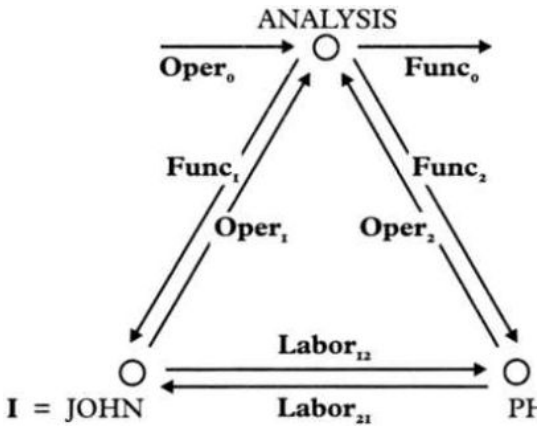


FIG. 2.2. *Support verbs and their DSynt-relationships with their keyword*

the LFs values, i.e. the support verbs in question; the arrow's tail indicates DSyntA **I** of the support verb (= Grammatical Subject), while the head points to its DSyntA **II** (= Main Object). Thus:

Oper $_1$ (*analysis) = [to] carry out [John carries out the analysis of the phenomenon];*
Oper $_2$ (*analysis) = [to] undergo [The phenomenon underwent (careful) analysis (by John)];*
Func $_1$ (*analysis) = is due [The analysis of this phenomenon is due to John];*
Func $_2$ (*analysis) = covers, concerns [John's analysis concerns this phenomenon];*
Labor $_{12}$ (*analysis) = [to] submit [John submits this phenomenon to a (careful) analysis];*
Labor $_{21}$ (*analysis) = leads [The phenomenon leads John to a (specific) analysis];*
Func $_0$ (*analysis) = is in progress [John's analysis of the phenomenon is in progress];*
Oper $_0$ (*analysis) = [one] sees [One sees an analysis of the phenomenon by John].*

This description can be represented by Fig. 2.3.

From Figs. 2.2 and 2.3 it is easy to see why the support verbs are presented as *three* LFs: these LFs are distinguished according to their syntactic behaviour with respect to the major sentence SSyntelements, and there are three such elements, namely -- Grammatical Subject, Main (roughly, Direct) Object, and Second (roughly, Indirect or prepositional) Object.

	DSynt-role of L and of its DSynt-actants with respect to VLF

FIG. 2.3. *Definitions of support verbs*

		DSynt-actant I ofVLF is:	DSynt-actant II ofVLF is:	Dsynt-actant III/IV ofVLF is:
Support Verb V^{LF}	Oper1/2	Ist/IInd DSyntA of L	L	-
	Func0/1/2	L	non/Ist/IInd L DSyntA of L	-
	Labor12/21	Ist/IInd DSyntA of L	IInd/Ist DSyntA of L	L

FIG. 2.3. *Definitions of support verbs*

4 Realizations

Real $_{o/i}$, Fact $_{o/i}$, and **Labrea $_{lij}$,** or FULFILMENT VERBS, mean, roughly, '[to] fulfil the requirement of L' [= '[to] do with L what you are supposed to do with L'] or 'L fulfils its requirement'. The 'requirements' differ with respect to different Ls: thus the 'requirement' of a hypothesis is its confirmation, and the 'requirement' of a disease is the malfunctioning/death of the person affected, while the 'requirement' of an artefact is that it be used according to its intended function. **Real $_{i}$** [Lat. *realis* 'real'], **Fact $_{o/i}$** [Lat. *factum* 'fact'], and **Labreal $_{ij}$** [a hybrid of **Labor** and **Real]** are (more or less) synonymous full verbs, differing with respect to their syntax only; their keywords are nouns whose meaning includes the component corresponding to a 'requirement': 'supposed to . . .', 'designed to . . .', etc.

In sharp contrast to support verbs, which accept as their keywords basically abstract nouns, fulfilment verbs can have both abstract and concrete keywords, provided the latter have actants and imply a 'requirement'. Such concrete nouns are necessarily the names of artefacts or organs, which are by definition 'designed to . . .'.

Syntactically, **Real $_{i}$, Fact $_{o/i}$,** and **Labreal $_{ij}$** are fully analogous to the LFs **Oper $_{i}$, Func $_{o/i}$** and **Labor $_{ij}$,** respectively. This means that the keyword L and its DSyntAs fulfil with respect to **Real $_{i}$** the same syntactic roles as they do with respect to **Oper $_{i}$,** etc. Therefore, they are linked to their keywords in the following way:

Realo/i→+ΠL, Fact $_{o/i}$ ^{I}→+L, and Labrealij→+IIIL.

Reat $_{1}$ (accusation)	= [to] prove [ART~]	Real $_{2}$ (law)	= [to] abide [by ART~]
Real $_{1}$ (car)	= [to] drive [ART~]	Real $_{2}$ (hint)	= [to] take [ART~]
Real $_{1}$ (illness)	= [to] succumb (to ART~]	Real $_{2}$ (demand)	= [to] meet [ART~]
Real $_{1}$ (bus)	= [to] drive [ART~]	Real $_{2}$ (bus)	= [to] ride [on [ART~]
Compare:			
Oper $_{1}$ (promiseN)	= [to] make [ART~],	but Real $_{1}$ (promiseN) = // [to] keep [ART~]	
Oper $_{2}$ (attackN)	= [to] be [under an~ ↦ of N],	but Real $_{2}$ (attackN)	= [to] fall [to ART~ ↦ of N]
Oper $_{2}$ (exam)	= [to] take [ART~],	but Real $_{2}$ (exam)	= [to] pass [ART~]
Fact $_{O}$ (hopeN)	= comes true	Fact $_{O}$ (filmN)	= is playing, is on
ContFact $_{O}$ (luck)	= holds Russ.	ContFact $_{1}$ (udača 'luck')	ne pokidaet [Nacc], lit. 'does not abandon'
Labrea $_{12}$ (gallows)	= [to] string up [N on ART~]		
Labreal $_{12}$ (saw)	= [to] cut [N with ART~]		

3.4
LFs and collocations

LFs cover *all* collocations with the sole exception of those covered by the Government Pattern [= GP] of L. Consider Fr. *assurance vie,* 'life insurance', where life is what you insure, vs. *assurance maladie,* lit. 'illness insurance', where illness is what you insure against (cf. *health insurance);* similarly, *assurance auto,* 'car insurance', vs. *assurance incendie,* lit. 'fire insurance', etc. The restricted co-occurrents in these collocations are Sem-actants of the keyword. Further examples include *un condamné à mort,* lit. 'a person-sentenced to death', vs. *un condamné à vie,* lit. 'a person-sentenced to life (i.e. in prison)'; Fr. *auto-école* vs. Eng. *driving school;* also *sick leave~maternity leave~study leave; hit list~shopping list; life sentence,* etc. All these collocations are described not by LFs of the keyword L, but by the L's GP.

On the other hand, *not all* LFs describe collocations: only the syntagmatic LFs do (whereas the paradigmatic LFs represent the derivatives of the keyword). Thus the set of all collocations and that of

all expressions described by LFs overlap: they have an important intersection.

3.5
The degree of fixedness of LF expressions

An important property of LF expressions (and of course of the collocations they represent) is *fixedness:* the quantity of similar phrasemes that exist for the phraseme under consideration. The phrase *pay attention* is very fixed: **Oper 1** is expressed as PAY only with ATTENTION (*pay heed,* though possible, is formal and rare), VISIT/CALL and COMPLIMENT (cf. ◊*pay a greeting*). [9] On the other hand, the phrase *give [him] a look* is much less fixed: **Oper 1** is expressed as GIVE with scores of nouns (*give* [N] *a pull, a punch, a smile, a tug, a push, a kick, a stroke, a kiss, a try, a greeting,* etc.). Moreover, it is possible to characterize semantically the resulting construction: it means '[affect an object or communicate with a being] voluntarily, performing one unit of the activity involved' (Wierzbicka 1982; Dixon 1991: 348-51; Stein 1991; Stein and Quirk 1991). With body parts, **Oper 1** is fully predictable: it is always HAVE. However, since in very many cases **Oper 1** is expressed by phraseologically bound LUs, *all* the expressions with **Oper 1** are considered phrasemes. The same is true of all LF expressions: once an LF, always an LF. As a result, we can have LF expressions (i.e. collocations) with a very low degree of fixedness: the elements of the LF value may be semantically transparent and their co-occurrence, sufficiently predictable -- yet the expressions in question remain phrasemes by analogy with more restricted cases. The degree of fixedness is thus an independent parameter of phrasemes, cutting across their defining properties (restrictedness of selection and irregularity of combination).

As a result, in spite of the basically idiosyncratic character of LFS, in many cases a given LF has the same values for quite a few different keywords, the reason often being semantic proximity: semantically related LUs can possess the same values for a given LF. This fact can be accounted for by following the general principle of lexical inheritance (Mel'čuk and Wanner 1996):

PRINCIPLE OF LEXICAL INHERITANCE

> All lexicographic data shared by a family of semantically related LUs should be stored just once -- under one LU of the corresponding vocable or under the generic LU of the corresponding semantic field, from where these data are 'inherited' in each particular case.

[9] Compare the discussion by Howarth, this volume.

This principle covers, first of all, LFs; however, I cannot explain here all the techniques of generalizing over common values of LFs.

3.6
LFs vs. semantic restrictions

Not all cases of restricted co-occurrence of LUs are cases of restricted *lexical* co-occurrence. Consider, for example, the Russian verb OŠIBIT'SJA + N_{instr}, which roughly means, 'use or try to use the wrong N':

> (2) (*a*) *On ošibsja dver+ju* 'He passed or tried to pass through the wrong door.' vs. ◊*On ošibsja ključom* 'He used or tried to use the wrong key.' (*b*) On *ošibsja adresom* 'He came to a wrong address.' vs. ◊*On ošibsia avtobusom* 'He boarded or tried to board the wrong bus.' (*c*) *On ošibsja nomerom* 'He called a wrong (telephone) number.' vs. ◊*On ošibsja knigoj* 'He took or tried to take a wrong book.'

The co-occurrence of OŠBIT'SJA in this construction (which, astonishingly, is recorded by no Russian dictionary I have consulted) is extremely limited and looks very capricious. However, the expressions *ošibit=śja dver+0301ju (adresom, nomerom),* and a few others that are possible, are not collocations, but free phrases, the restrictions observed being purely semantic. The meaning of the verb OŠIBIT'SJA here is 'mistakenly try to establish contact with somebody or something at a location identified byY^1 while believing that$Y1$ is$Y2$; therefore, Y can be only something that might be interpreted as being or identifying a location. Stretching things a little, one might say:

> (*d*) *On ošibsja čemodanom* 'He opened or tried to open the wrong suitcase.'

This sentence, however, cannot mean ◊'He took/bought/brought the wrong suitcase'. As we can see, one has to distinguish between, on the one hand, a very specific and therefore highly restrictive meaning -that is, *semantic* constraints in lexicographic definitions -- and, on the other, genuine *lexically* restricted co-occurrence of LUs. Only the latter comes under the jurisdiction of LFs.

4
LFs IN LINGUISTIC APPLICATIONS

To illustrate the role of LFs in linguistics, I will say a few words about their possible uses in the area known as Computational Linguistics. More specifically, I will touch upon LFs in Automatic Translation

and Text Generation. Four aspects are of particular interest: lexical choices, paraphrasing, communicative structure, and text cohesion.

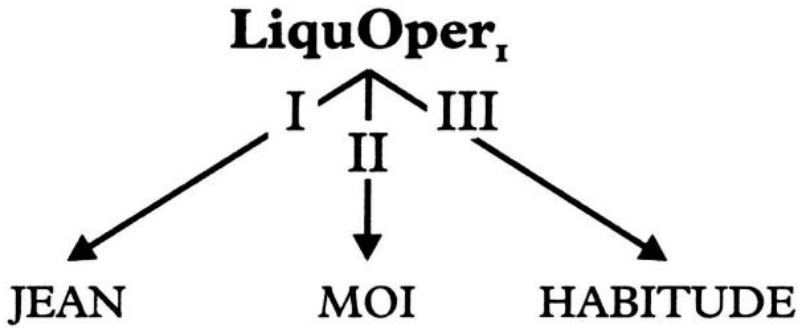
4.1
LFs and lexical choices (collocational aspects)

Suppose that we have a system of Automatic Translation in which the transfer (from the source language into the target language) is done at the level of DSynt-Structure [= DSyntS]; suppose furthermore that we are interested in the translation of collocations. In such a case, it suffices to reduce the source-language collocation to its LFrepresentation, then translate the keyword only, and, finally, to select the value of the LF for the equivalent of the keyword in the target language. Consider, for instance, the French sentence at (3):

(3) *Jean m'a détourné de cette habitude* 'John broke me of this habit'.

Analysis

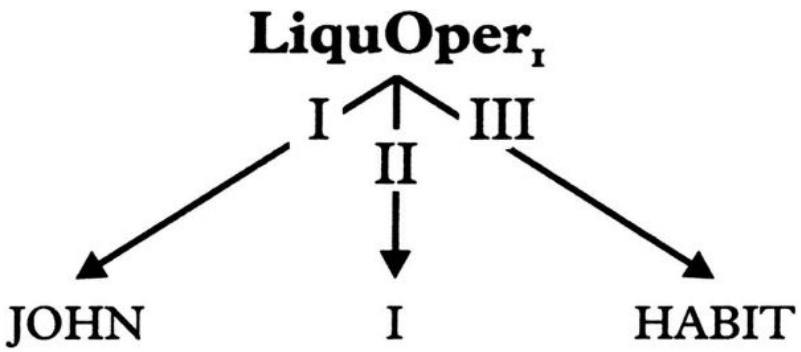
Using a monolingual French dictionary which lists the values of all LFs for all head LUs (plus of course all syntactic mechanisms needed), (3) becomes (3ʻ):



(3ʻ = DSyntS($_3$)

Transfer

Using a bilingual (or multilingual) network of lexical correspondences, the French tree (3ʻ is replaced with the English tree (4ʻ):



(3ʻ = DSyntS($_4$)

Synthesis

Using a monolingual English dictionary, again with the values of all LFs specified for all head LUs (and all corresponding syntactic mechanisms), the tree of (4ʻ is turned into the English sentence (4):

(4) *John broke me of this habit.*

As can be gathered from this simplified example, in a collocation, only the keyword needs actual transfer, i.e. the looking up of its equivalent (Fr. HABITUDE = Eng. HABIT).The search for the 'bizarre' correspondence DÉTOURNER = [*to*] BREAK in the context of HABIT is avoided altogether: [to] BREAK [N *of*~] is computed as an element of the value of the LF **LiquOper $_1$** (HABIT) in an English dictionary -- quite independently of the source language. LFs thus play the role of a transfer interlingua. In this way, multilingual translation does not require many pairwise-arranged transfer dictionaries of collocations. It is enough to have monolingual dictionaries with LFs specified plus indexes of multilingual translation equivalents for keywords only.

The same type of procedure can be used by any system of Text Generation that produces the output text passing by a DSyntStructure. One such system is described in Iordanskaja, Kim, and Polguère (1994). It makes extensive use of LFs for lexical choices and, in particular, does so with an eye to paraphrasing and the Communicative Structure of the sentence to be generated (see sections 4.2 and 4-3).

To make things clearer, let me cite (Table 2.1) a series of seemingly 'bizarre' correspondences that can be easily and naturally expressed in terms of LFs.

TABLE 2.1.
Correspondences between LFs and lexical choices

LFs Eng.	HABIT	⇔	Fr. MABITUDE
IncepOper $_1$	acquire, develop, form [ART~], get [into ART~], take [to ART~]		contracter, prendre [ART~]

LFs Eng.	HABIT	⇔	Fr. MABITUDE
FinOper $_1$	drop [ART~], get out [of ART~], get rid [of ART~], . . .		abandonner, perdre [ART~]
LiquOper $_1$	break [N of ART~], wean away [N from ART~]		détacher, détourner [N de ART~]

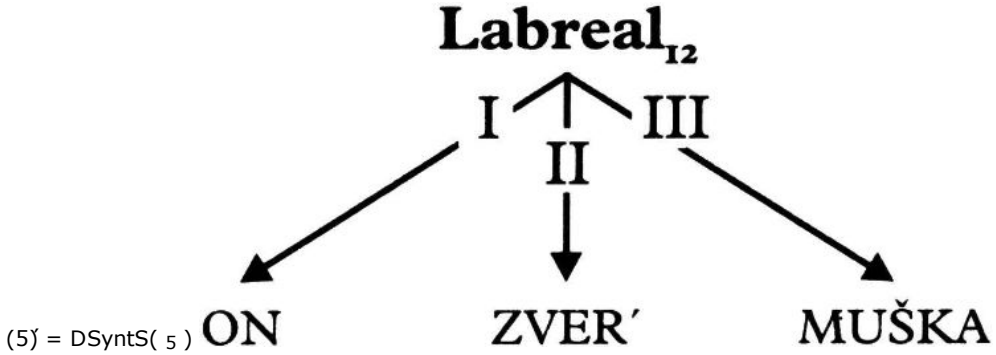
4.2
LFs and paraphrasing (syntactic aspect)

A well-known thorny problem of text generation is the widespread incompatibility of a given LU and the syntactic constructions in

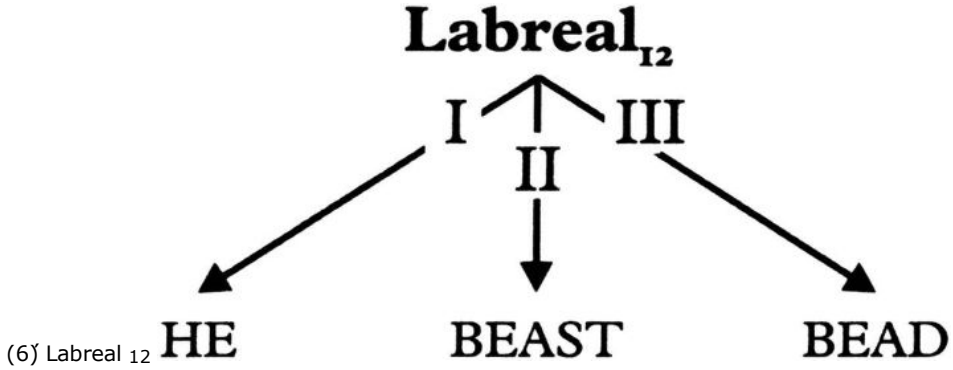
which it must appear; in many cases, the lexical choices that are made entail syntactic restructuring. LFs turn out to be helpful in this respect as well. The fact is that the equations relating LFs [10] allow for a number of important syntactic transformations. Thus, consider the Russian sentence (5), which has to be translated into English:

(5) On vzjal zverja na mušku, lit. 'He took the-beast on bead', i.e. 'He took aim at the beast.'

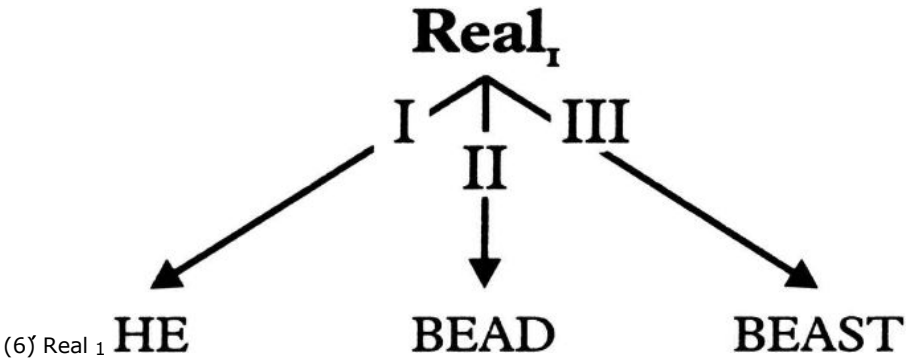
At the analysis stage, sentence (5) is reduced to the DSyntS (5'):



At the *transfer* stage, the Russian nominal lexemes are replaced with their equivalents: HE, BEAST, and BEAD. But the resulting English DSyntS (6ʹ):



cannot be implemented directly, because the English equivalent of MUŠKA, i.e. BEAD, does not have a **Labreal $_{12}$** : **Labreal $_{12}$** (*bead*) = ? Yet BEAD has a **Real1** : [*to*] *draw.* Replacing **Labreal $_{12}$** by **Real $_1$** and performing, at the same time, the standard transformation associated with **Labreal $_{12}$** ⇔ **Real $_1$** substitution, we obtain the correct English tree (6"):



It should be emphasized that standard syntactic transformations (= tree restructuring) associated with a given LF substitution are univer-

[10]For these equations and for a detailed description of the paraphrasing system based on LFs, see Mel'čuk (1992*b*); an example of paraphrasing equations is given in section 3.3.3 (p. 38).

sal: they do not depend on the keyword or on the language involved. At the synthesis stage, the above DSyntS is realized as (6):

(6) *He drew a bead on the beast.*

which is the optimal translation of (5). In this way, LFs take upon themselves the syntactic adjustments needed to carry out the transfer between languages -- in cases where the LFs are involved.

Another telling illustration of the process is the translation of the English sentence (7) in Russian:

(7) *He was stabbed three times, once fatally.*

Its translation appears at (8):

(8) *Emu bylo naneseno tri naẑevyx rany, odna iz kotoryx okazalaś smerteĺ noj,* lit. 'To-him was dealt three knife wounds, one of which turned out mortal'.

Again, if a translation system tries to make the transfer at the level of DSyntS, it can use, to obtain the result shown, the following LFs (the description is fragmentary and approximate, giving only a rough idea of how such transfers can occur):

ANALYSIS (English)	
[to] stab [N]	= Labreal $_{12}$ (*knife*) + CausFunc $_{1}$ (*wound*)
≪≪fatal	= Magn(wound) [the symbol '≪≪' = the extreme value of Magn]
TRANSFER	
Eng. KNIFE	= Rus. NOZ
Eng. WOUND	= Rus. RANA
SYNTHESIS (Russian)	
CausFunc $_{1}$ (*rana*)	= *nanesti* [N $_{dat}$ ~*u* $_{acc}$]
Magn (*rana*)	= *≪≪smertel 'naja*
caused with knife (*rana*)	= *noẑevaja*

If, however, a translation system proceeds via a SemR, then its task (in regard to restricted lexical co-occurrence) is to establish the relevant LF starting from the initial SemR and then to 'compute' its value for the given L, based on a monolingual dictionary of the ECD type. Of course the same procedure is needed for text generation, whatever its underlying representation.

4.3 LFs and the Communicative
Structure of sentences (communicative aspect)

Let us suppose that a text-generation system has to verbalize the meaning of sentence (9) (this example is adapted from Wanner and

Bateman (1990), where the use of LFs in connection with the Communicative Structure of the sentence is discussed):

(9) *The adjective***'electronic'** *indicates to the reader that the dictionaries are dedicated to computers.*

If in the Semantic Structure of (9) the meaning of the phrase *the adjective 'electronic'* is specified as the theme, then the ordering of elements in (9) is appropriate. But if the meaning of the phrase to the reader is specified as the theme, a different syntactic structure is needed, such as in (9ў):

(9ў) The reader gets some indication that the dictionaries are dedicated to computers from the adjective 'electronic'.

To replace *indicate* with *get* an indication, one needs paraphrasing equations of the type

V ⇔ **S $_{o}$** (V) + **Oper $_{2}$** (**S $_{o}$** (V)) [*X analyses Y ⇔ Y undergoes an analysis by X, X resists Y ⇔ Y runs into resistance from X, X orders Y to do Z ⇔ Y receives from X an order to do Z,* etc.],

and most important of all, a dictionary which specifies, for each L, the values of LFs (cf. Melčuk 1992*b*).

4.4
LFs and text cohesion (cohesion aspect)

LFs prove equally useful in selecting the referring expressions in anaphoric links in such a way as to avoid tedious repetitions and guarantee, at the same time, maximum cohesion of the resulting text -- see Tutin (1992) and Alonso Ramos *et al.* (1995). Thus, speaking of an *ambush,* you can refer back to it by calling its participants *attackers*:

(10) *An Indonesian***patrol** *was caught in an***ambush.** *The***attackers** *fired three rockets at the***soldiers** *and sprayed them with automatic fire.*

Here, *attacker* = **S** $_1$ (*ambush*), and *soldier* = **S** $_1$ (*patrol*). This lexical knowledge is used to construct the sentence sequence (10) in an obvious way. [11]

Another example:

(11) *Sales***increased** *slightly in Quebec and Ontario. Modest***gains** *were also reported in British Columbia.*

11Note that *be caught in an ambush, fire rockets,* and *spray with automatic fire* are collocations and can be described in terms of LFs: **Real** $_2$ (*ambush*) = *be caught* [*in* ART~], **Real** $_1$ (*rocket*) *fire* [ART~], and **Labor** $_{12}$ (*automatic fire*) = *spray* [N *with*~].

Instead of simply repeating the same phrase and saying *Sales also increased slightly in British Columbia,* the speaker chooses to use **S** $_2$ (*increase*) = *gain* $_N$ (the amount by which X increased), which allows him to produce a more varied and elegant text. (Note also that *modest* = AntiMagn(*gain*), so that this adjective corresponds semantically to *slightly* = AntiMagn(*increase* $_V$).)

5 LFs IN THE LEXICON

LFs are specified -- for each LU -- in the dictionary, so that they are essentially a lexicographic problem. The MTT presupposes the existence of a special type of lexicon in whose entries LFs occupy an important place and which constitutes one of the central modules of the MT model of natural language. This lexicon is the EXPLANATORY COMBINATORIAL DICTIONARY [ECD]. For a working understanding of LFs a brief description of their representation in an ECD is indispensable; this, in its turn, requires a cursory sketch of the ECD. Since, however, the publications on *ECDs* are numerous (Z+ŭolkovskij and Melčuk 1967; Melčuk et al. 1984/1988/1992; Melčuk and Zholkovsky 1984, 1988; Melčuk and Polguère 1987; Melčuk 1988b, 1989, 1992b; Ilson and Melčuk 1989), I will limit myself here to a very short characterization.

5.1

Main properties of an ECD

The ECD is semantics-and paraphrase-based: (quasi-) synonymous paraphrases constitute the main target as well as the main research tool for an ECD. Its entries are intended to supply *all* the lexical information which might be needed for the two tasks that any linguistic model has to tackle:

- the transition from a Semantic Representation (formally, a network composed of semantic units) to a DSyntR(epresentation) (formally, a dependency tree composed of actual LUs);
- the construction, for a given DSyntR, of all the DSyntRs which are (up to the communicative organization) synonymous with it; this is paraphrasing.

The main *substantive* property of an ECD is that it is a PHRASAL DICTIONARY. It contains set phrases, i.e. phrasemes, (1) as headwords of numerous entries (idioms and quasi-idioms) d (2) as important data within the entries (semi-idioms, i.e. collocations, represented as LF-expressions).

The six main *formal* properties of an ECD are as follows:

1. An ECD is a theoretical dictionary: it is elaborated within a coherent linguistic theory, featuring developed semantic, syntactic and morphological components, or modules, and putting a strong emphasis on the lexicon.
2. An ECD is an active dictionary: it is consistently geared to production, or synthesis.
3. An ECD is a semantic dictionary: it is based on semantic representations of all the expressions it contains, the definition being the central part of a lexical entry.
4. An ECD is a combinatorial dictionary: it is centred around restricted co-occurrence (syntactic and lexical).
5. An ECD is a formalized dictionary: it can be considered as a lexical database.
6. An ECD aims to be exhaustive with respect to individual LUs (lexemes and phrasemes): a lexical entry includes whatever a native speaker knows about the LU in question.

5.2

The structure of an ECD article

All LUs stored in an ECD have dictionary articles of the same structure. An ECD article is divided into three major zones:

1. The SEMANTIC zone: the DEFINITION (= a SemR of the head lexical unit L), which (in the case of LUs with predicative meaning) is based on a propositional form with variables for semantic actants and constitutes a strict decomposition of the meaning of L. For instance, the verb [*to*] HELP (in one of several senses):

X helps Y to Z with, W= 'Y trying to do or doing Z, || X uses X's resources W, adding W to Y's efforts with the goal that W facilitates for Y doing Z'.

(The part to the left of the '||' symbol is a presupposition: it remains asserted even when the entire meaning of HELP is negated: *John didn't help Mary to prepare the dinner* still implies that Mary prepared the dinner.)

The LFs of L are semantically related to some particular semantic components of L's definition. Thus, **Magn** (*help*) = *a lot* intensifies 'facilitate'; the same is true for all LFs.

2. The SYNTACTIC zone: the GOVERNMENT PATTERN (= a subcategorization frame), which specifies, for each Sem-actant, the corresponding DSyntA and lists all surface means of expressing it in the text. Consider the Government Pattern for the verb [*to*] HELP (C stands for 'column', so that C $_{III.1}$ means Column III, line 1).

TABLE 2.2. Correspondences between Sem-actants and DSyntAs

X = I	Y = II	Z = III	W = IV
I. N	1. N	I. V $_{inf}$	1. *with* N
		2. *to* V $_{inf}$	2. *by* V $_{ger}$
		3. *with* N	
		4. *in* V $_{ing}$	
		5. PREP $_{dir}$ N[a]	

[a] 'PREP $_{dir}$ ' stands for 'directional preposition' (*into, up* (*the ladder*), *over* (*the fence*), . . .).

(1) C $_{III.1}$	:'X being directly involved in Z' [= 'X doing Z himself']
(2) C $_{III.2}$	:'X not being directly involved in Z' [= 'X not doing Z himself, but providing some external resources to Y'] [12]
(3) C $_{III.5}$	:**if** Z = 'travel/move [something] in the direction α', **then** [III = L('α') **and** C $_{III}$ = C $_{III.5}$] **is possible**
(4) C $_{III.3}$ + C $_{IV.1}$	:**undesirable**
Impossible	: ★ *Kathleen was helped move the furniture (by Arthur, and not by Jane*) [correct expression: . . . ***to*** *move the furniture*] (General rule of English syntax: no bare infinitive with the passive).
Undesirabl	: *? Kathleen helped Arthur with his work with her advice* [correct expression: either . . . ***in*** *his work* ***with*** *her advice or . . .* ***with*** *his work* ***by*** *advising him*] (Rule 4.).

Kathleen helped the old gentleman finish his preparations ≪helped the boy to finish his studies with her generous financial assistance, helped me in buying my last car with her advice, helped Jack out of his coat, helped Jack up the stairs with a kick in the bottom/by giving him a firm shove/push≫.

LFs of L are related to L's GP in an obvious way: thus, **Oper** $_1$ is different from Labor1 in so far as the former takes L as its DSyntA **II** and the latter, as its DSyntA **III; AntiBon** $_1$ **Involv** (*car*) = *smash* [*into* N] (it is the DSyntA **I** , i.e. the car, that suffers), while **AntiBon** $_2$ **Involv** (*car*) = *run over* [N] (it is the DSyntA **II** , i.e. the person run over, that suffers); etc. Moreover, the values of many LFs have GPs of their own; however, I cannot develop this point here.

12This constraint (stipulating that using TO with the infinitive dependent on HELP implies rather indirect help than help by participation) is not strict and is often violated; many speakers use the *to*-infinitive and the bare infinitive after HELP indiscriminately.

3. The LEXICAL CO-OCCURRENCE zone: LEXICAL FUNCTIONS, which present the RESTRICTED LEXICAL CO-OCCURRENCE of the headword L. The description of restricted lexical co-occurrence of L is fully adjusted to L's definition and to its Government Pattern.

5.3
A sample lexical entry, ECD-style

By quoting a fully-fledged lexical entry I hope to show LFs in their natural habitat -- that is to say, in a dictionary. Among other things, it can be seen how the LFs in the entry are related to the definition and to the Government Pattern.

REVULSION

X's revulsion for Y = X's (strong) negative emotion about Y similar to what people normally experience when they are in contact with something that makes them sick and such that it causes that X wants to avoid any contact with Y [13]

Government Pattern

X = I	Y = II
1. N's	1. *against* N
2. A $_{poss}$	2. *at* N
	3. *for* N
	4. *toward* N

(1) C $_{II.2}$: N denotes something that happens and can be seen or felt (2) C $_{II.4}$: N denotes people *John's (his) revulsion against racism (against greed/the dismal results of his endeavor); John's (his) revulsion at such behavior (at the sight of sea food); John's (his) revulsion for work (for all those killings); John's (his) revulsion for (toward) these scoundres/toward the government; John's (his) revulsion ★at these shouts [correct: . . . for these shouts]*

Lexical Functions

Syn⊂	: distaste
Syn∩	: repugnance; repulsion; disgust; loathing
Anti∩	: attraction

Conv $_{21}$ Anti∩	: appeal
A $_1$	: revulsed

[13]The construction 'Y causes that X wants/does/sees . . .' (instead of the grammatically correct 'Y causes X to want/to do/to see . . .') is used in the semantic metalanguage of the ECD for greater precision and explicitness: it allows for an explicit expression of the subject of the fact that is caused.

Able $_2$	: revulsive
Magn + Able $_2$	:of utmost ~ | G = SCENE, SIGHT [G stands for the syntactic governor of the LF value]
Magn	: deep ≪ extreme ≪≪ utmost
AntiMagn	:slight
Adv $_1$	: in [~]
Propt	: from [~]
Oper1	: experience, feel [~]
Magn + Oper $_1$	: be filled [with~] Magn + Labor $_{21}$ fill [N with~]
Conv $_{21}$ Caus $_2$ Oper $_1$	: be driven [to~]
Adv $_1$ Manif	: with [~]

Examples

He did it from deep revulsion against the bitterness of the sectarian strife. Any revulsion they might feel from fat-ass bastards they ran up against professionally was *ad hominem* and not *ad genus* [Alison Lurie]. Kathleen turned her head away in revulsion. I felt no revulsion for her maternal phantasies, only a practical concern. She met his advances with revulsion. It was a scene of utmost revulsion. Pam was driven to revulsion (by the sight of the dead animal). ≪★The sight of the dead animal drove Pam to revulsion.≫ Revulsion at slaughter cut war short [newspaper heading].

Notations and Abbreviations

-A	: actant
ConceptR	: Conceptual Representation
DSynt-	: deep-syntactic
ECD	: Explanatory Combinatorial Dictionary
GP	: Government Pattern
L	: a particular lexical unit
L	: a given natural language
LF	: Lexical Function
LU	: lexical unit
MTT	: Meaning-Text Theory
-R	: representation
-S	: structure
SemR	: Semantic Representation
SSynt-	: surface-syntactic

[This page intentionally left blank.]

3
Phraseology as a Language of Culture: Its Role in the Representation of a Collective Mentality

VERONIKA TELIYA, NATALYA BRAGINA, ELENA OPARINA, and TRINA SANDOMIRSKAYA

1
INTRODUCTION

Phraseology is a domain of linguistic study which to a high degree illustrates the correlation between language and culture. In a typological approach, it is necessary to define and classify the types of cultural information which are illuminated by lexical collocations. Clearly, the recent development of cognitive linguistics and conceptual analysis in the 1980s and 1990s can provide a reliable foundation for this kind of investigation. A further important reason why cultural information should be included in an account of linguistic meanings concerns the needs of lexicography today. For the practical purposes of compiling a dictionary of lexical collocations, a number of theoretical points must be elucidated concerning the general problem of cultural markedness.

This study suggests a new direction for phraseological research -that is, linguo-cultural studies, or the analysis of

phraseological units for cultural data as represented in linguistic meanings. For language theory and for practical applications of linguistics, cultural data relevant to the form and meaning of linguistic units -- either words or word-combinations -- have to be gathered and formalized. We shall show below how the lexicon can be quarried for cultural information. We shall also show that linguo-cultural analysis -- conducted on a systematic basis -- is best suited for phraseology, and especially for

The research reported here has been partly financed by OSI/RSS. We gratefully acknowledge this generous support.

restricted lexical collocations. The latter abound in cultural information and can hardly be described at all as a class of denominations if their cultural meanings are not taken into account. Indeed, phraseology can be regarded as a testing ground for the anthropomorphic paradigm in linguistics, whose fundamental assumption is that the linguistic world-picture is commensurable with the mental attitudes and culture of a speech community.

The aim of this chapter is to propose a conceptual framework (a) for the description of cultural data stored in the lexicon, and (b) for the lexicographic description of those phraseological units (idioms and restricted collocations) that are acknowledged to show the most pronounced cultural colouring.

2
THE LEXICON AS THE STOREHOUSE OF CULTURAL DATA

The anthropocentric approach in linguistics is directed towards the elucidation of the everyday language world-picture. It is assumed that every language, especially with regard to its figurative meanings, is concerned with the reflection and extension of what Humboldt and Weisgerber called the *Weltansicht*, or 'world-view' (Weisgerber 1929). The world-view shared by all members of a linguo-cultural community makes possible the generation and comprehension, in a subconscious process of insight, of metaphorical linguistic meanings.

Edward Sapir (1964) was the first to postulate explicitly that language represents and conceptualizes reality in a culturally specific manner, so that individual native languages stand in a relation of complementarity to each other. This idea of linguistic relativity was further developed by Whorf (1956). However, for a long time linguistic relativity was viewed as a linguistic-philosophical concept rather than a purely linguistic one. The latest developments in cognitive linguistics seem to be offering fresh scope for practical linguistic application.

In the anthropocentric paradigm, the notion of linguistic relativity can be reformulated as linguistic-cultural relativity: language is the means of representing and reproducing culture. In other words, culture is assumed to be implemented, one way or another, on the content plane of linguistic expressions, reproduced in an act of denomination and transmitted from generation to generation through linguistic and cultural norms of usage. Thus, language can be looked upon as a crucial mechanism contributing to the formation of a collective cultural identity. Culture being thus implemented through

language, cultural norms are not only reproduced in language but are made mandatory for speakers of that language through the linguistic structures they use.The above postulates require that three points be clarified:

- what we understand by culture;
- what we understand by implementation through language structures;
- how we can be sure that cultural norms are made mandatory by language.

By culture, we understand the ability of members of a speech community to orientate themselves with respect to social, moral, political, and so on values in their empirical and mental experience. Cultural categories (such as Time and Space, Good and Evil, etc.) are conceptualized in the subconscious knowledge of standards, stereotypes, mythologies, rituals, general habits, and other cultural patterns. (For further details, see Bartminski 1993 and Teliya 1993.) A set of patterns can be looked upon as an alphabet of culture. When these patterns enter the lexicon, they may act as 'direct' cultural signs (e.g. as proverbs and sayings, with their immediate descriptive and prescriptive functions, and invariable epithets and comparisons, such as Eng. *as happy as a lark, as cunning as a fox,* etc.). On the other hand, when linguistic symbols interpret cultural patterns and categories, then these symbols serve as bodies for those cultural patterns. In that case, language units acquire the status of quasi-standards, quasi-stereotypes, and so on. For example, the idiom *nesti krest*, lit. 'to carry one's cross', interprets the biblical story of the Crucifixion and in its everyday, non-biblical, usage becomes the quasi-stereotype of torment and self-sacrifice. In a similar fashion, Russ. *u cherta na kulichkakh,* lit. 'in the devil's mires', or 'very far away', acts as a quasi-standard of remoteness through its allusion to the Other Space, a dwelling-place of evil spirits. Similarly, in *chernaya zavis",* lit. 'black envy', the collocator bears an allusion to the idea of evil (in general, symbolically represented by the colour black) and through this becomes a quasisymbol of this evil feeling.

Such instances seem to confirm our suggestion that native speakers' capacity for linguistic introspection and cultural reflection derives from their knowledge of cultural-linguistic codes -- that is, from their linguo-cultural competence. Linguo-cultural competence is assumed to be acquired (together with, and in close connection with, knowledge of one's mother tongue) in the process of internalizing collective cultural experience.

3
CULTURAL DATA: WORDS, LEXICAL COLLOCATIONS, AND IDIOMS

That language forms part of culture seems to be beyond question, but whether (and in what way) cultural data are incorporated in lexical meanings remains uncertain. Such uncertainty is very much due to the fact that lexical units vary as to how, and how far, they take up cultural data. Besides, individual words, idioms, and collocations often combine in their semantics more than one type of cultural information. In this chapter we suggest five channels through which language is penetrated by culture: cultural semes, cultural concepts, cultural connotations, cultural background, and discourse

stereotypes.

3.1

Cultural semes

Brought together under this heading are words and wordcombinations that denote idioethnic realia. The cultural component forms a cultural seme -- that is, forms part of the lexical meaning. The cultural seme reflects general knowledge about the realia. Such entries are normally found in encyclopaedic dictionaries with appropriate etymological and cultural commentaries (e.g. Vereshchagin and Kostomarov 1983). Examples include:

1. Material realia: *lapti*, 'footwear made of hast traditionally worn by Russian peasants', *rozhon*, 'a double-edged blade mounted on a Y-shaped spear traditionally used in bear hunting in old Russia'; *chernaya izba*, lit. 'a black cottage', i.e. a traditional peasant timber cottage heated by a stove with no chimney.
2. Social and historical realia: *gorodovoy*, 'a policeman in Tsarist Russia', *kolkhoz*, 'a Soviet collective farm'; *krasnokorichnevuie*, lit. 'the red-browns', i.e. followers of the Russian chauvinist/communist political movement in post-*perestroika*Russia.

3.2

Cultural concepts

These are abstract notions that map and construct the world-picture in a culturally specific way (Arutyunova 1991). Significantly, their specificity is implemented at the cognitive, not the semantic level. In this group, we differentiate between concepts proper (holistic *Gestalten* of meaning), on the one hand, and subconcepts on the other.

1. Concepts proper largely coincide in all European languages but, for some aspects of meaning, corresponding words and phraseologisms show a high degree of cultural specificity, as in the case of Russ. *pravda* and *sovest*'. Thus, pravda denotes truth as an ethical phenomenon with no direct counterpart in English. This is a case of linguistic/cultural lacuna. Similarly, sorest", 'conscience', is a case of partial overlap: the Russian word implies 'the presence of God in one's soul' and only partially coincides with the English meaning of *conscience* as 'knowledge of good and evil'. It would be only natural to expect different cultural implications from different conceptualizations verbalized in different languages.
2. Subconcepts are fragments of concepts proper, when abstract notions are verbalized in concrete nouns. Thus, *baba*, 'a human (lower-class) female' (colloq. or derog.), is a concrete noun, originally the name for a peasant woman. Later, the word came to designate a complex of properties assumed to pertain to women in general. Thus, *baba* is a subconcept of the the concept of 'femininity' which has no translation equivalent in English. (The notion will be described in greater detail below.)

3.3

Cultural connotations

By cultural connotation, we mean the interpretative relation between linguistic signs and symbols of any other cultural non-verbal code (stereotypes, prototypes, myths, and other entities termed cultural patterns above). According to Teliya (1993), cultural connotation arises from an associative relation between the image contained in the inner form of a language sign (Potebnya 1905) and the content of a cultural pattern. Cultural connotations are especially vivid in idioms and restricted lexical collocations. In restricted collocations, for instance, the activation of cultural connotation is connected with the type of cultural information contained in the keyword (the base of the collocation) and the nature of the semantic specialization in the meaning of the collocator. [1] In general, cultural connotations can accompany any culturally marked words when they combine

1Lexical collocations are defined as word-combinations in which one member -- the base according to Hausmann (1985) -- is used in its non-figurative meaning (the denotation of the whole), while the semantically specialized collocator is bound to the nominal base (Teliya 1981) and denotes its features and conceptual parameters. Idioms are defined as completely opaque combinations of two or more lexical (or lexical plus grammatical) words (cf. Benson 1985; Cowie 1994).

in phraseologisms. Cultural connotations are manifested in wordcombinations which activate culturally relevant parameters of the base.

1. Some collocations derive their connotative meaning from allusion to cultural realia (cf. Sect. 3.1, paras. 1 and 2). Consider: *treskuchiy (moroz),* lit. 'snapping (frost)', i.e. very hard frost when trees give out a snapping sound; *korob (novostey),* lit. 'a basketful (of news)', i.e. as much news as if there were a whole *korob* -- a huge basket made of bast and without handles; *zhandarm Evropui,* lit. 'the gendarme of Europe', used with reference to Russia's reactionary foreign policy in the 1840s (the connotation coming from the historical persona of *zhandarm* -- an agent of the repressive secret police in Tsarist Russia). Though *gendarme* is an international word, in Russian it is associated with political power that is cruel, aggressive, and stupid. The same word in English -- also borrowed from Frenchmerely signifies 'a French policeman' (according to the *Longman Dictionary of Contemporary English*)and evokes no negative connotations. Compare also lexical collocations denoting the political leaders of *perestroika: arkhitektorui/kapitanui perestroyki,* lit. 'architects/captains of *perestroika'*. Here, cultural connotation derives from the metaphorical interpretation of political reforms as constructing or seafaring.

In a similar way, cultural connotations invoked by cultural semes can also be found in idioms, as, for example, *lezt na rozhon,* lit. 'to thrust oneself against the Y-shaped spear', i.e. to provoke a dangerous situation; or *zavarivat kashu,* lit. 'to cook porridge', i.e. to stir up trouble. (*Kasha* is a ritual food traditionally cooked by in-laws for a wedding party.) Consider also *mamaevo poboishche,* lit. 'Mamai's slaughter', i.e. a bloody fight between many people -- an allusion to a medieval battle between Russians and Tartars.

2. Cultural connotations can also arise from cultural concepts. *Sovest zazrila* (obs.), lit. 'one's conscience began to see' or 'woke up'; *bol naya sovest,* lit. 'a sore conscience', i.e. an uneasy conscience. Cultural connotations are generated by

metaphor, when moral constraints (conscience) are conceptualized as if they are a living being that can see and feel pain. Such metaphorical conceptualization can easily be traced to the religious notion of the living soul. The interaction between the denomination and the background religious belief is what provides the content of cultural connotation.

Cultural connotations can also derive from the interpretation of subconcepts (cf. Sect. 3.2, para. 2). For instance, *muzhskaya (druzhba),* lit. 'men's (male) friendship', connotes that males are ideal bearers of genuine friendship, while in *bazarnaya (baba),* lit. 'a

market-place (lower-class) woman', the collocator creates the connotation 'ill-bred and loud-mouthed'. These connotations originate from stereotypical assumptions about men being paragons of human virtues, and women being the embodiment of evil. Similarly, cultural connotation in *bazarnaya baba* is related to the behaviour of a lowerclass woman in the market place, and *bazar* ('market place') is a stereotype of noisy, impudent conduct. A quite different stereotype is that of mother as caregiver: *materinskaya nezhnost',lyubot',zabota, laska,* lit. 'maternal tenderness, love, care, caress'.

Idioms also manifest cultural connotations associated with subconcepts. Consider the subconcept of 'loose woman' in the following idioms: *trepat' yubki,* lit. 'to swing one's skirts', i.e. (of women) to be promiscuous; *khodit' po rukam,* lit. 'to pass from hand to hand', i.e. (of women) to change sexual partners; *shlyukha podzabornaya,* lit. 'a stroller from under a fence' or 'a loose woman' (derived from *shlyat'sya pod zaborom,* lit. 'to stroll under the fence' -- to be loose). Cultural connotations in such cases may only be revealed if considerable ideographic fields are studied, including idioms, restricted collocations, proverbs, sayings, and so on. Their cultural specificity only shows through the common kernel metaphor (Lakoff and Johnson 1980), which acts as a hyperonym with respect to the images conveyed by such idioms (i. e. 'strolling' = 'promiscuity'). In the above examples, the image derives from one and the same prototype 'to move about aimlessly' and is closely associated with the epithet *gulyashchaya baba,* lit. 'a strolling woman', i.e. a loose woman. According to folk tradition, also reflected in proverbs, a decent woman cannot walk about by herself; she should stay at home: *Muzhik da sobaka vsegda na dvore, a baba da koshka vsegda v izbe,* lit. 'The man and the dog are always outside, the wife and the cat are always inside'. The same norm, incidentally, was explicitly prescribed in the Russian sixteenth-century private law code *Domostroy.* Thus, such idioms acquire a culturally marked connotation from interaction with three cultural codes: (i) the subconcept of *gulyashchaya;* (ii) a cultural norm that insists that women should stay at home; (iii) norms as reflected in proverbs.

3.4
Cultural background

Cultural background refers to information that is most difficult to formalize, as it is connected with semantics in a very indirect and still unexplored way. We say that a word or a word-combination has 'cultural background' when it possesses a clearly discernible ideological aura associated with a historical situation, a political movement, a

fashionable trend, and so on. Such entities could be compared to visual symbols and emblems. Thus, *russkaya berezka,* lit. 'a Russian birch tree' is unmistakably identified by native speakers as a symbol of the motherland. Such conceptualization could easily be developed into coherent ideological discourse; in fact, cultural background appears to be a reduction of such discourse. Similarly, cultural background can be discerned in such lexical collocations as *serp i molot,* lit. 'hammer and sickle', *britanskiy lev,* lit. 'the British lion', and others.

3.5
Discourse stereotypes

The development of culture involves the repetition, reinterpretation, and multiplication of texts, as well as the creation of new ones. These texts, pertaining to different discourse types (which are described below, with respect to the process of generation of restricted collocation), exert a powerful influence over culture, while cultural change causes the reproduction and reinterpretation of discourse stereotypes in speech practices. Such mutual exchange cannot but affect language. The lexicon, the repository of forms, 'freezes' and organizes certain expressions which regularly occur in different discourse types. Among the forms stored in the lexicon are lexical collocations that could be interpreted as 'open' word-combinations, if certain cultural stereotypes did not exist which restricted their use.

Thus, in a number of cases, cultural data contained in a language unit can be elicited from knowledge of a text, or of a corpus of texts, in which the concept was first described. For instance, the notion of *pervaya lyubov',* lit. 'first love', is clearly associated in the minds of native speakers of Russian with the story by Ivan Turgenev. The expression *pervaya Iyubov'* connotes pure, delicate, refined, and hopeless passion between a sexually inexperienced girl and youth, a love that cannot be consummated, an extremely lyrical combination of desire and innocence. Indeed, *pervaya lyubov'* is a restricted lexical collocation because it is a cultural reflection of the (textual) situation within a specific historical and social context (compare, for instance, the English expression *calf love* and its entirely different connotations). Interestingly, no such discourse stereotype can be found for the combination *poslednyaya lyubov',* lit. 'last love', though one can easily imagine a story about such a relationship. In fact, *poslednyaya lyubov'* seems to be a free rather than a restricted collocation.

Another example, *sal'ericheskaya zavist',* lit. 'Salieri's envy', can be described as a collocation whose meaning is associated by native speakers of Russian with Aleksandr Pushkin's *Mozart and Salieri*: the mediocre Salieri envies Mozart his talent and murders him out of

jealousy. In contemporary Russian, the expression denotes destructive feelings of jealousy towards a gifted person.

To a certain extent, restricted collocations such as *evreyskiy vopros,* lit. 'the Jewish problem', *zhenskiy vopros,* lit. 'the women's problem', *russkaya dusha,* lit. 'the Russian soul', can also be looked upon as discourse stereotypes, though they

were not originally borrowed from the writing of a specific author but rather allude to a whole collection of anonymous texts of the same genre and with a common ideological basis. Such borrowings are described in greater detail below.

As has already been stated, one of the aims of this chapter is to answer the question whether cultural *Weltansicht* is imposed by language, or, in other words, whether native speakers' cultural patterns really are locked up in language. In our view, the analysis of phraseology in terms of cultural components of meaning provides the basis for a positive answer. We provide below, as one set of examples, some Russian phraseologisms describing women. They illustrate the general idea of how the cultural concept of gender is encoded in Russian.

In the Russian mentality, *glupaya baba,* lit. 'a silly common female', serves as a stereotype of the low intellectual capacity of women (compare also sayings like *U babui volos dolog, um korotok,* lit. 'Women have long hair and short intellect'). Phraseologisms such as *devich'ya pamyat',* lit. 'a maiden's memory', i.e. a short memory, *zhenskaya logika,* lit. 'feminine logic', i.e. illogical logic, also refer to the same stereotype. Restricted lexical collocations play the part of signs for such stereotypes and thus become cultural symbols. Compare also a derogatory collocation often used in political debate today, *bab'ya politika,* lit. 'women's politics', i.e. unreasonable, contradictory, and absurd politics, the slang musical term *babiy ritm,* lit. 'women's beat', i.e. no beat at all, and many other collocations with the word *baba* whose connotation is 'incompetence'. In fact, the cultural norm that allows speakers to represent the idea of incapacity, incompetence, and so on through the image of women means the perpetuation of a patriarchal cultural attitude towards women as inferior human beings. Similar explanations could be given for idiomatic expressions referring to the stereotype of *gulyashchaya,* lit. 'strolling' or 'loose'. Also, a cultural norm that views woman as an object is reflected in such collocations as *appetitnaya,* lit. 'appetizing', *puishnaya,* lit. 'plump', and *sdobnaya,* lit. 'shortening', *baba*. The woman is literally an object of consumption for males (an item of food). At the same time, female sexuality is defined in other restricted collocations in an utterly repressive way. *Devich'ya/zhenskaya gordost',* lit. 'maidenly/female pride', and *devichiy/zhenskiy stuid/grekh,* lit. 'maidenly/female shame/sin', imply a very severe moral standard for women to conform to in sexual

relations. Patriarchal gender as represented in these (and many other) expressions still preserves its power -- in spite of the fact that presentday native speakers of Russian were brought up in the Soviet state, which claimed to have emancipated women.To conclude, we would suggest that a number of research procedures could be used in the identification and description of the cultural component of meaning:

- deciphering the metaphor (or other kind of figurative meaning);
- establishing its prototypic relation to other tropes that derive from the same kernel metaphor;
- connecting the perception thus elicited from its associative image-based motivation with a cultural category through specific standards, prototypes, stereotypes, myths, archetypes, etc.;
- representing a phraseologism as a cultural sign expressing speakers' collective mentality.

The point of departure in this chain of cultural reflection would be the image (or any other isomorphic entity) which serves as a source of semantic transposition of a free lexeme into a collocator, or of a collocation into an idiom. The central point would be cultural interpretation (in the sense of insight) that forms the basis of a culturally marked connotation. Clearly, the role of metaphor is crucial, and it will be outlined in the next section.

4
METAPHOR AND THE CULTURAL MARKEDNESS OF LEXICAL COLLOCATIONS

The study of restricted lexical collocations shows the important role of metaphor in the meaning of the collocator describing parameters of non-material objects. Thus, the concept of *voobrazhenie* ('imagination' or 'the human ability to create mental images') has a parameter 'a high degree of activity', which in English collocations is conveyed metaphorically as *vivid/fertile/lively.* These adjectives refer to colour, soil, and living organisms, respectively. In Russian, the same conceptual parameter is described metaphorically by adjectives whose associative motifs correspond only partially to those in English. English *vivid* in collocation with *imagination* can be translated into Russian as *burnoe,* lit. 'turbulent', *zhivoe*, lit. 'lively', or *puilkoe,* lit. 'ardent', 'fervent'.

Another example illustrates the metaphorical use of a verb as a collocator. In Russian, the process of forming a person's character is signified by the collocation *vuikovuivat' kharakter,* lit. 'to forge some-

one's character'. The association is with a blacksmith hammering at a metal object to give it firmness and hardness. In English, the collocation *to mould someone's character* is used, also associated with a firm object but emphasizing, at the same time, the idea of giving shape to an originally shapeless mass. So, each language chooses its own mode of metaphorical conceptualization, the 'nerve' of such specificity being the association underlying the figurative meaning. One can perceive that for the Russian mentality the standard of character is represented in the image of armour, hard and firm (cf. Lakoff and Johnson's kernel metaphor *'Life is War'*), while for the English language such a standard can be associated with the idea of a clear-cut shape as the ideal creation (it should be noted that the verb *to mould* in English is used in its transposed meaning 'to create' not only in collocation with character -- cf. *to mould something upon something, to mould someone into something)*. These examples show that metaphorical collocators denoting conceptual parameters can be used as clues to different approaches to conceptualization and verbalization in different languages -- that is, to cultural data contained in, or associated with, the meaning of restricted collocations. Many more Russian examples are given in the following section.

Now, one might ask: What is the driving force behind the activity of the metaphor as a linguistically creative mechanism? The cognitive foundations of the phenomenon were formulated by Kant as the Als Ob principle (Kant 1994: 400-20). Kant postulated, as inherent in the human mind, the tendency to interpret abstract categories in terms of sensuously perceptible objects, relating the abstract to the concrete through analogy. As a result, abstract categories appear to the mind and in the language as if they were concrete objects, material processes, or properties. The Als Ob principle has been recognized in linguistics as the cognitive mechanism of linguistic metaphor (Black 1979; Apter 1982). Therefore, everyday language and the everyday (naïve) concepts it contains tend to associate abstract notions with physical phenomena.

Consider, for instance, 'talent', a non-material concept. The Russian language conceptualizes its manifestations as *iskra talanta,* lit. 'a sparkle of talent' (as if talent were fire). Similarly, morals are a social code of behaviour, but parameters of the concept are figuratively verbalized as if it were a material object. Thus, the Russian collocation for the deterioration of morals is *padenie nravov,* lit. 'morals fall', with the non-figurative meaning of the collocator referring to the downward movement of a physical object in space. Thus, metaphor is crucial to the process of conceptualization, when concept parameters are ascribed to a base.

Though conceptualization through figurative meaning does not guarantee that a lexical collocation is culturally marked (as not all metaphors necessarily require a culturally oriented interpretation), there are a number of metaphorical collocators whose non-figurative meanings involve concepts associated, one way or another, with categories of culture and cultural patterns (standards, stereotypes, mythologies, ideologies, etc.) These patterns become the focus of metaphorical associations and thus motivate the process of semantic transposition. Similarly, culturally marked metaphorical collocators derived from cultural data give cultural markedness to the whole of the lexical collocation.

The difficulty with analysing linguistic metaphors for the cultural information they might contain lies in the fact that culturally marked collocations differ in that their collocators may refer to different cultural patterns, calling for different types of cultural interpretation. Thus, the collocator in *zerno istinui,* lit. 'a grain of truth', is loaded with cultural associations: grain is part of the rite of sowing connected with the archaic concept of Birth and Rebirth. This cultural connotation is activated in *zerno istinui* in its meaning 'a single and small but true and important idea, generating true and useful knowledge or conclusions'. But these cultural data cannot be derived directly from either the literal or the metaphorical meaning of the collocator. They represent cultural reflection activated through a metaphorical linguistic image. On the other hand, in the collocation *angel'skiy kharakter,* lit. 'angelic character', also associated with the religious tradition, the cultural seme constitutes part of the literal meaning of the collocator. So, metaphorical transposition changes and complicates the relation between the contents of the collocator and the patterns of culture. In metaphor, consequently in the whole of the collocation, the original (biblical) angel serves as a Christian cultural standard of virtue. 'Angelic' features are singled out to become the focus of association and the core of the new metaphorical content. Thus, *angel'skiy kharakter* relates to a person as if he or she were an angel (i.e. conforming to the highest standard of virtue in the Christian world-picture). To interpret the concept, an associative motif needs to be correlated with the cultural pattern.

Metaphors deriving from mythological archetypes often require thorough investigation before the cultural patterns underlying the connotation can be identified. For instance, the Russian collocation *temnaya lichnost',* lit. 'a dark personality', denotes an unknown, suspicious, dangerous person (cf. Eng. a *shady character*), while *svedaya lichnost',* lit. 'a light personality', refers to an individual who is highly respected because of his or her excellent qualities, especially high

moral standards. Both collocations can be traced to the mythological opposition between light and darkness. Genetically related to the same opposition is the concept of *enlightenment* (Fr. le Siècle des Lumières, Russ. *Prosveshchenie*), in which the concept of light is associated with a benign and cultivated world of Reason as opposed to the hostile darkness of reactionary superstition (cf. *obscurantism*). Other examples of mythological archetypes in cultural connotations will be given below.

Cultural markedness in lexical collocations can also depend on specificity of metaphorical conceptualization. Thus, from the Russian standpoint, *a donkey* can be assumed to be a stereotype (a subconcept) of foolish stubbornness, as witness: *oslinoe upryamstvo,* lit. 'donkey obstinacy', *oslinaya glypost',* lit. 'donkey folly'. For native speakers of English, stubbornness seems to find its stereotype in a mule, hence *mulish stubbornness*. Such collocations form part of the linguisticcultural thesaurus and usually stand out against the background of proverbs, sayings, and other verbal folklore. Compare *glup kak osel,* lit. 'as silly as a donkey', and *Osel na osle, durak na durake,* lit. 'a donkey over a donkey, a fool over a fool' (i.e. 'there are too many fools around'; cf. Eng. as *stubborn as a mule*).

Cultural patterns are a productive source of incessant linguistic creativity, a fund of metaphorical quasi-stereotypes that is constantly replenished. Thus, Westerns produced the stereotypical image of the cowboy. This Hollywood image eventually penetrated the Russian collective mentality, resulting in utterances like *Reagan is a political cowboy* and in the collocations *kovboyskoe nakhal'stvo,* lit. 'cowboy insolence', and *kovboyskaya bestseremonnost',* lit. 'cowboy cheek'.

To sum up, interpretation of metaphorical collocations often requires reference to language-specific cultural patterns. These make a linguo-cultural community perceive concepts as if through the prism of culturally associated images. Since these cultural patterns parametrize concepts, they can be discussed not only in terms of the creativity of metaphor -- according to Black (1979) and Apter (1982) -- but also in terms of its linguistic-cultural creativity.

5
LIFE AND DEATH: ELICITING CULTURAL CONNOTATIONS FROM LEXICAL COLLOCATIONS

As we said earlier, cultural connotations result from the interaction between linguistic meanings and other symbolic cultural codes and form a common domain relevant for both verbal and non-verbal meanings. This point could be illustrated through an analysis of such

basic cultural constructs as *Life and Death*, which, in terms of an anthropocentric approach, are assumed to lie at the very centre of the linguistic and cultural world-picture.

Let us now consider some illustrations selected from a corpus of collocations containing the nouns *smert',* 'death', and *zhizn',* 'life'. Conceptual parameters of life are represented through both temporal and spatial metaphors. Represented in terms of space, for instance, are aspectual parameters marking the beginning or end of individual existence: *vstupat' v*

zhizn', lit. 'to step into life', *priyti v zhizn',* lit. 'to come into life' (cf. Eng. *to come to life/to one's senses*), *uyti iz zhizni,* lit. 'to walk away from life'.

Also, the beginning of a new stage in one's life is often subject to metaphorical interpretation in terms of space. Consider: *vstupat' v trudovuyu zhizn',* lit. 'to step into one's working life', *stoyat' na poroge novoy zhizni,* lit. 'to stand at the threshold of a new life', perevalit" za *seredinu zhizni,* lit. 'to pass over the middle of one's life' (as if it were a mountain pass), *pered (kem) raspakhnulis' beskraynie prostorui zhizni,* lit. 'the boundless spaces of life opened wide before one', *(komu) otkruilis' novuie zhiznennuie gorizontui,* lit. 'new horizons of life opened before one'.

Death, it would seem, is also seen as departure into a new space, the space of the Other. Consider: *stoyat' na poroge/u vrat smerti,* lit. 'to stand at the threshold/the gate of death', *uyti v mir inoy,* lit. 'to leave for the other world', *uyti/otpravit' sya k praottsam,* lit. 'to go to one's forefathers'.

Again, a descriptive-meaning parameter (here, periods of life) forms an association with a kernel metaphor of time as space, while the selectivity of the cultural subconscious determines what kind of space will be activated in regularly occurring linguistic metaphors. For instance, quite a number of collocations rely on the metaphor 'life as a road or path'. These metaphors can also be found in archaic and present-day cultural practices. (Compare, for instance, the linguistic metaphor 'life as a road' with the medieval religious practice of going on pilgrimage and today's tourism.)

Further expansion of the phraseological corpus would be governed by the logic of the archetypal concept underlying the metaphor. For instance, the image of the road is associated with a trajectory, as in *proyti dolgiy zhiznennuiy put',* lit. 'to have passed a long path of life', with the vector of movement understood as *tsel' zhizni* and *smuisl zhizni,* lit. 'the goal of life' and 'the sense of life', respectively. Along the path of life *(na zhiznennom puti),* at the crossroads of everyday life *(na zhiteyskikh perekrestkakh),* one is accompanied by good luck *((komu) soputstvuet udacha)* or pursued by ill luck *((kogo) presleduyut*

neudachi). Death dogs one's footsteps *(smert' gonitsya (za kem) po pyatam,* lit. 'chases one on the heels') and sometimes overtakes one *(smert' nastigla (kogo))* or passes by *(smert' proshla mimo)*. People meet and part on this road -- consider: *vstretit' svoyu sud'bu,* lit. 'to meet one's fate', and *ikh sud'bui/zhiznennuie puti perekrestilis'/pereseklis'/ soshlis'/ razoshlis',* lit. 'their destinies/life ways crossed/converged/diverged'. In addition, the logic of the archetype dictates an ethical norm encoded by means of the same linguistic metaphor, e.g.: *zhizn' bez tseli,* lit. 'life without a goal', *besputnaya zhizn',* lit. 'wayless', 'directionless' (i.e. 'dissipated') life, *rasputnaya zhizn',* lit. 'wayward' (i.e. 'profligate') life, *ostupit'sya/poskol'znut'sya v zhizni,* lit. 'to take a wrong step/to slip in life', *otklonit'sya ot puti istinnogo,* lit. 'to diverge or stray from the path of righteousness'.

Besides describing life as a spatial-temporal form (here, a path), language also makes use of another spatial metaphor, the kernel metaphor of 'life as a receptacle'. Fullness (vs. emptiness) is alluded to in describing a satisfying, harmonious life. *Nastoyashchaya zhizn',* lit. 'a true life', is *zhizn' vo vsey polnote ee proyavleniy*, lit. 'life in all the fullness of its manifestations'. To live a good life is *zhit' polnov/napolnennoy zhizn'yu,* lit. 'to live a full life'; *pustaya zhizn',* lit. 'an empty life', is 'no life'. Consider also: *zhizn', napolnennaya sobuitiyami,* lit. 'a life full of events', *soderzhatel'naya zhizn',* lit. 'a life with content', 'a complete life', *zhizn', b'yushchaya/pleshchushchaya/bruizzhushchaya cherez kray,* lit. 'a life that brims over', *zhizn' v ee polnokrovnom techenii,* lit. 'the full flow of life', *zhizn' -- polnaya chasha,* lit. '(having) the full chalice of life' (i.e. '(being) wealthy'). All these are kinds of life that give satisfaction, that make one *duishat' polnoy grud'yu,* lit. 'fill one's lungs with air', and *ispuituivat' vsyu polnotu zhizni,* lit. 'feel the full completeness of life' (compare *ispit' chashu zhizni,* lit. 'drink the chalice of life', *zhazhda zhizni,* lit. 'the thirst of life', *upivat'sya zhizn'yu,* lit. 'to drink in life').

To sum up, the evaluative parameter of the concept (roughly speaking, 'a happy life') alludes to the kernel metaphor 'life as a receptacle', and the cultural factor contributes information on precisely what type of receptacle this is -- a chalice, an archetypal symbol of sacrifice and communion with God.

An important parameter of life is, of course, death. The way death is linguistically conceptualized -- either as a result of deliberate choice by the individual or as the result of an accident in which he is the victim of circumstances beyond his control -- is transformed into meaning through different tropes: through the metaphor of economic exchange, on the one hand, or through an allegorical personification, on the other. The 'time-is-money' kernel metaphor outlined by Lakoff

and Johnson (1980) is represented in our material with its more elaborate version 'life as an economic value, an object of possession and exchange'. Hence we find: *rasschitat'sya s zhizn'yu/svesti schetui s zhizn'yu,* lit. 'to settle accounts with life' (i.e. 'to (be ready to) commit suicide'), *podarit' zhizn', darovat' zhizn' (komu),* lit. 'to make a gift of life to someone', *lishit' zhizni (kogo),* lit. 'to deprive someone of life' (i.e. 'to kill'), *otnyat' zhizn',* lit. 'to take life away from someone', *zaplatit' zhizn'yu (za chto),* lit. 'to pay the price of one's life for something', *otdat' zhizn' (za kogo/chto),* lit. 'to give one's life for someone/ something', *pozhertvovat' zhizn'yu (radi chego/kogo),* lit. 'to sacrifice one's life for someone/something', *prinesti svoyu zhizn' v zhertvu chemu/na altar' chego,* lit. 'to lay one's life on the altar of something', *rastratit' zhizn' (na chto),* lit. 'to waste one's life on something', *ne shchadit' zhizni (radi chego),* lit. 'not to spare one's life'. *dorogo tsenit' svoyu zhizn',* lit. 'to give a high price for one's life', *dorogo prodat', svoyu zhizn',* lit. 'to sell one's life dear'. The 'political economy' of life and death establishes a cultural norm according to which one's life is exchanged for something more valuable, for instance, fame. Note, for example: *otdat' zhizn' za blagoe delo,* lit. 'to give one's life for a good deed', *umeret' dostoynoy (geroicheskoy, pochetnoy) smert'yu,* lit. 'to die a worthy/heroic/honourable death'. In contrast, it appears wretched to throw away one's life on trifles *(rastrachivat' zhizn' na pustyaki)* and to die a shameful death *(umeret' pozornoy smert'yu)*. Note especially: *umeret' pod zaborom,* lit. 'to die under a hedge', *umeret' sobachey smert'yu,* lit. 'to die a dog's death', or *umeret' glupoy/nelepoy/bessmuislennoy smert'yu,* lit. 'to die a foolish/an absurd/a meaningless death'. Though economic exchange cannot be classified as an archetype, it seems to be related, in an indirect way, to the mysterious archaic practices of sacrifice -- as is money in general.

As for the second trope -- the allegorical personification of deathit can be shown to be activated in the conceptualization of the parameter -- 'threat of death'. Here, death poses as an archetypal counteragent, a gargoyle: *popast'sya smerti v*

lapui/zubui/kogti, lit. 'to be caught in death's paws/teeth/nails', *smotret' smerti v glaza/v litso,* lit. 'to look death in the eye/face', *smert' glyadit pustuimi glaznitsami,* lit.'death stares from empty eye sockets', *lik smerti,* lit. 'the visage of death', *smert' osenila kruilami (kogo),* lit. 'death spread its wings over someone'. The subject is conceptualized as waging a mysterious fight with death, as in *obmanut'/perekhitrit' smert',* lit. 'to cheat/outwit death', *brosit' vuizov smerti,* lit. 'to challenge death', *borot'sya so smert'yu,* lit. 'to fight death', shutit"/igrat" so smert"yu, lit. 'to joke/to play with death', *smert' medlit,* lit. 'death is lingering', *smert' stuchitsya u vorot,* lit. 'death is knocking at the gate', *smert' podzhidaet (kogo),* lit. 'death is waiting

for someone', *smert' podkaraulila, podsteregla,* lit. 'death is waiting in ambush', *smert' otstypila (ot kogo),* lit. 'death has retreated from one', *smert' poshchadila (kogo),* lit. 'death had mercy on someone', etc.It can easily be seen from these illustrations that the metaphorical conceptualization of life and death in Russian phraseology is heavily loaded with cultural connotations but hardly shows any signs of ethno-cultural -- i.e. purely Russian -- colouring. Rather, the images and symbols of death could be attributed to European culture in general. Whether this is the result of cultural and linguistic borrowing or a proof that Russian and other European cultures have much in common remains a matter of debate and is outside the scope of the present study. Of much greater importance is the conclusion that cultural connotations give us a better opportunity of finding systematic correlations between everyday language and other -- non-verbalcultural discourses. The cultural analysis of lexical collocations makes clear that information about specific archetypes and other patterns would be of value in a phraseological dictionary, as they are responsible for the conceptualization, in terms of metaphors, of vast and important ideographic areas.

6

A CASE STUDY: CULTURAL DATA IN COLLOCATIONS THAT NAME EMOTIONS

In our analysis, emotional phenomena are clearly connected with images of Nature: natural elements serve as the kernel metaphor motivating the phraseological denotation of feelings. Consider: *burya/ vikhr'/volna/stikhiya/vsplesk chuvstv, chuvstva nakhluinuli, razveyat' tosku,* etc., lit. 'a storm/a whirlwind/a wave/the element/a surge of feelings'; 'a feeling swept through someone', 'to blow away anguish'. Natural phenomena in figuratively transposed collocations occur as a second-order reality, as a linguistic-cultural construct from which lexical collocations acquire their cultural markedness.A number of lexical collocations imply that feelings can move. Language conceptualizes emotions as being capable of transfer from one person or place to another. Feelings can 'enter' a person and 'come out' of him or her, and move from one person to another. The way feelings enter the subject seems to be elaborated in greater detail:

- they can penetrate in a non-violent way: *chuvstvo voshlo, (pronikio) v dushu/v serdtse; chuvstvo posefflos' v dushe/v serdtse,* lit. 'a feeling entered, (penetrated) one's soul/heart', 'a feeling settled in one's soul/heart';

- they can penetrate in a fraudulent, furtive way: *chuvstvo zakralos'/ zatailos'/zapalo v dushu,* lit. 'a feeling stole into/hid in/fell into one's soul';
- they can penetrate in an aggressive way: chuvstvo *(toska, gore) napalo/navalilos'/porazilo/presleduet,* lit. 'a feeling (anguish, grief) attacked/fell upon/hit/pursued someone'; the departure of an emotion is described with less precision: *chuvstvo ushlo, ischezlo, isparilos'/uletuchilos',* lit. 'the feeling went away/disappeared/ evaporated'.

The transfer of emotion from one person to another is also depicted through restricted lexical collocations: *zarazit'sya unuiniem (ot kogo),* lit. 'to be infected with despondency', *radost' peredalas' (komu),* lit. 'joy was communicated to someone'. *(kto) naveyal na (kogo) pechal',* lit. 'someone blew sorrow on someone' -- examples which seem to suggest that language depicts emotions as independent of the subject. However, the 'space' available for transfer is limited and corresponds to the individual space of a person -- that is, his or her spiritual dimension. Language seems to hush up the actual location of feelings when they are external to the subject. It is clear that feelings can exist apart from the subject. We say: *chuvstvo ushlo kuda-to,* lit. 'the feeling has gone away somewhere'. But the 'somewhere' is not defined; one cannot say, for example, ★*chuvstvo ushlo daleko,* lit. 'the feeling has gone far away'.

Emotions are the Other for an individual in the same way that Nature is the Other of Culture (Lotman 1992). Emotions as represented in phraseology have much in common with Chaos (feelings are represented as shapeless, disordered, and uncoordinated), with the Pandora's Box into which Prometheus, as the myth relates, put all people's troubles -- including passions. Hence, in brief, some of the lexical collocations mentioned act as quasi-standards of the Other Space (compare the idiom *u cherta na kulichkakh* above). Some refer to Nature as a kernel metaphor, and all of them are culturally marked.

7

PHRASEOLOGISMS AND DISCOURSE STEREOTYPES: CULTURAL MARKEDNESS THROUGH ASSOCIATION WITH DIFFERENT DISCOURSE TYPES

As has been pointed out, in some cases the material for the formation of a lexical collocation is borrowed from a specific discourse. All collocations of this type are culturally marked through their association

with a particular body of discourse. We have been able to identify at least four discourse types that are very productive in generating the culturally relevant phraseologisms, mostly restricted collocations.

7.1

Religious and philosophical discourse

Lexical collocations such as *demonicheskaya strast',* lit. 'demoniac passion', *sataninskaya gordost',* lit. 'satanic pride', *pravednuiy gnev,* lit. 'righteous wrath' all contain cultural semes in their collocators. Other lexical collocations, for instance *nebesnaya/nezemnaya lyubov',* lit. 'celestial/unearthly love', also represent a cultural concept. Still others, like *ostolbenet' ot strakha/ot uzhasa, okamenet' ot strakha/ot uzhasa,* lit. to turn into a pillar with fear/terror', 'to petrify with fear/terror',

all contain cultural connotations (by allusion to the biblical myth about Lot's wife or to the ancient Greek myth about Medusa the Gorgon). Significantly, all of them are biblical or classical allusions and are more or less easily identified as such by language-users. Hence their high cultural potential.

7.2
Literary discourse

Collocations such as *liricheskaya/romanticheskaya lyubov',* lit. 'lyrical/ romantic love', *poeticheskoe/sentimental' noe chuvstvo,* lit. 'poetic/sentimental feeling', *elegicheskaya grust',* lit. 'elegiac sadness', have a cultural background associated with different literary movements (Romanticism, Sentimentalism). Various 'cultural heroes', and their associated behaviour, feelings and ideas, come to be stereotyped by a linguistic-cultural community and verbalized by means of lexical collocations.

Thus, the notion of *romantic love* (Russ. *romanticheskaya lyubov'*) can be historically associated with the first third of the nineteenth century -- the Romantic period. What contemporary everyday linguistic meaning has absorbed from Romanticism in terms of love obviously amounts to the connotation of romantic love as 'an elevated, idealistic emotion associated with the young'. Just like *pervaya lyubov"* (see above), the lexical collocations *toska po rodine,* lit. 'anguish for the homeland', i.e. 'nostalgia', *dorozhnaya toska,* lit.'road misery', i.e. the acute discomfort one feels when making a long and tedious journey along a Russian road, are phraseological clichés from literary texts that describe an emotional state experienced not only as a fact of reality but also as a fact of culture. Literary discourse gives emotions a kind of cultural dimension. For instance, *toska po rodine* denotes not simply 'acute homesickness' but an emotion that has

been thoroughly elaborated in the Russian literary tradition, especially when one recalls that many generations of Russians have been forced into exile and that many still live outside their homeland (cf. Polish *tesknic* and English *homesick,* analysed by Wierzbicka 1992).Of special interest are collocations referring to the prototypical personal name of a historical character -- e.g. *sal'ericheskaya zavist',* lit. 'Salieri's envy', *shekspirovskie strasti,* lit. 'Shakespearian passions', *vol'terova ironiya,* lit. 'Voltairian irony', *platonicheskaya lyubov',* lit. 'Platonic love'.Lexical collocations such as *gore/sud'ba stuchitsya v dom/v dver',* lit. 'grief/fate knocks at the house/door', reflect the personification and individualization of human destiny typical of Russian literature of the sixteenth and seventeenth centuries. Here, Grief and Fate act as mythological personae, while the house, the door, the knock at the door are cultural signs. The collocation proper could be regarded as a quasi-standard of the Path (see similar examples with life and death above).

7.3
Poetic folklore discourse

Lexical collocations which originate in such a discourse type represent, *en masse,* a closed and well-regulated world-picture that took shape in the past and is not subject to alterations at present. This world-picture has absorbed typical features of conceptualizing reality, a poetic folk view. The following types of semantic relations between the base and the collocator can be identified:

- intensification which is often based on reduplication (the base and the collocator being derived from the same root). Consider: *gore goremuichnoe,* lit. 'grievous grief', *beda bedovaya,* lit. 'troublesome trouble', *skuka skuchnaya,* lit. 'boring boredom', etc.;
- evaluation (bad vs. good): *lyubov' zlaya,* lit. 'vicious love', *pechal' chernaya,* lit. 'black melancholy'. Motivating the collocator is a metaphor associated, again, with the symbolic oppositions of light vs. darkness.

7.4
Political discourse

Quite a number of lexical collocations are associated with texts reflecting a political ideology. Russian official phraseology of the totalitarian period is a very rich field of research for those interested in such discourse stereotypes. For instance, *chuvstvo glubokogo udovletvoreniya,* lit. 'a feeling of profound satisfaction', and *chuvstvo zakonnoy*

gordosti, lit. 'a feeling of justified pride', are notorious clichés associated with Soviet totalitarian texts. They were intended to describe the enthusiasm a Soviet citizen felt for the political achievements of the Soviet state. The use of these collocations at the time was explicitly required in official speeches and newspaper articles; implicitly, they signalled the speaker or writer's loyalty to the political system and the political leadership. In the post-Soviet period, they are often used ironically in critiques of totalitarianism.

8
CONCLUSION

This chapter provides a formulation of a problem rather than its solution. It is a matter of theoretical as well as practical interest to find out to what extent and in what ways cultural information manifests itself in language, especially in lexical collocations, and how this information can be identified, isolated, and lexicographically presented. As we have seen, the cultural component of meaning as represented in phraseology is not homogeneous. Special lexicographic techniques are therefore needed for its representation.

First, the lexicographic definition of phraseologisms in terms of their cultural value should start from the concept. Conceptual analysis is instrumental in bringing the metaphor in the 'inner form' to the forefront: elicitation of the metaphor and the description of the frame should yield an understanding of the principles of conceptualization.

Secondly, and following from the first, the most suitable lexicographic form should be that of an ideographic dictionary. This approach would provide some understanding of how the world-picture contained in phraseology fits into the conceptual map of reality. Such an approach would be especially useful in comparative lexicology, as it would reveal the linguistic lacunae which exist among different modes of conceptualization in different languages.

Thirdly, the description of the pragmatic aspect would have to include a much wider coverage of cultural knowledge responsible for the generation of a speech act with an idiom or a lexical collocation. The speech subject must be conceived

as a subject of culture, a bearer of ethno- or sociocultural patterns. Phraseology, then, can be seen as the language of culture.

Systematic research into phraseology following a linguistic-cultural approach would ultimately help to overcome what might be called 'cross-cultural deafness', a barrier that now hinders the development of a postmodern, multicultural world community.

[This page intentionally left blank.]

PART 2
Phraseology in Written and Spoken Corpora

[This page intentionally left blank.]

4
Frequencies and Foms of Phrasal Lexemes in English

ROSAMUND MOON

1
INTRODUCTION

This chapter reports some of the findings of research into English phrasal lexemes and their behaviour as evidenced in corpus text. This research was carried out in 1992-3, while I was working for Oxford University Press on a research collaboration, the Hector Project, with Digital Equipment Corporation's Systems Research Center in Palo Alto, California.By phrasal lexemes, I mean the whole range of fixed and semi-fixed complex items which dictionaries in the Anglo-American tradition classify and treat as 'phrases' or 'idioms' (according to their metalinguistic terminology): the sorts of item that for reasons of semantics, lexico-grammar, or pragmatics are regarded as holistic units rather than compositional strings. Such items include pure idioms, proverbs, similes, institutionalized metaphors, formulae, sayings, and various other kinds of institutionalized collocation. For practical reasons, I am specifically excluding phrasal verbs such as *give up* and *stick out;* compound nouns, adjectives, and verbs such as *civil servant, selfraising,* and *freeze-dry;* and foreign phrases such as en *passant* and *caveat emptor.*As a corpus lexicographer, I am continually confronted with evidence for certain points when dealing with these items:

- pure (or classical) idioms, such as *spill the beans* or *bite the dust,* are infrequent;
- the expressions which occur most frequently (e.g. *at least, of*

I should like to thank Oxford University Press for allowing me to make continued reference to the corpus data of the Hector Project, even though I am no longer working for them. The Hector Project is described by Atkins (1992) and Glassman *et al.* (1992). A detailed account of the phraseological research and its findings is given in Moon (1998).

course) tend to be functional rather than fully lexical, and to be frozen collocations rather than metaphorical in nature;
- the more colourful, stylistically marked, and metaphorical expressions (e.g. *have your cake and eat it, call a spade a spade*) are often exploited or manipulated in the various contexts in which they appear.

Overall, there seem, intuitively and impressionistically, to be strong correlations between frequency, form, idiomaticity type, and discourse function.

My principal research interest was the discourse functions and pragmatic behaviour of phrasal lexemes. However, I felt it would be useful and interesting to gather some hard facts concerning their distribution and so on in order to test hypotheses about correlations between frequency, form, type, and function, and hence markedness. I therefore assembled a database, specifically oriented towards this end, and drawing on the results of the interrogation of an eighteenmillion word corpus of British English, the Oxford Hector Pilot Corpus, which was being used in Palo Alto for the Hector Project.

The corpus, it has to be said -- and as those involved with it would freely admit -- is not perfect. Journalism constitutes 66 per cent of the whole corpus -- an excessive amount, especially as the largest components were from the Guardian and the Independent, rather than more demotic and stylistically less mannered newspapers. Of the rest, 18 per cent consists of non-fiction texts; 11 per cent of fiction; 2 per cent of ephemera, and only 3 per cent of transcribed spoken interaction. Such imbalance is clearly undesirable, but it does not necessarily invalidate the findings, and I shall, in the final part of this chapter, consider some corpus comparisons and also some genre distributions.

Frequency information derived from this corpus for the rarer items may not be conclusive, but the overall tendencies and the figures for higher-frequency items are likely to be replicable in other corpora. In general, corpus comparisons indeed

show that such phenomena as the frequency distribution of certain lemmas, the phraseological patterns associated with those lemmas, and the relatively high frequencies of delexicalized meanings recur across corpora. In contrast, rarer uses, specialist terminology, and genre- or topic-specific words and meanings show greater variation from corpus to corpus. Such comparisons as I produce at the end will, I believe, support my contention that the broad patterns I observed in the Hector corpus concerning phrasal lexemes will recur elsewhere in other bodies of evidence.

-80-

It is clear that the availability of large text corpora has revolutionized lexicology and lexicography. The literature of corpus linguistics shows repeatedly how corpus data has affected thinking about lexis and grammar (see e.g. Aarts 1991; Halliday and James 1993; Sinclair 1994). As far as phrasal lexemes are concerned, corpus evidence shows that their forms are by no means as fixed as some dictionary inventories appear to suggest, and that the division between multiword and single-word items is blurred, to say the least. Yet, while it is inconceivable to me now that description of vocabulary can proceed without reference to a corpus or substantial body of text, it will be clear that the eighteen million words I had access to are not enough to describe adequately and in detail many of the phrasal lexemes I was looking at -- in particular, pure idioms and proverbs. Even a corpus such as the Bank of English at Birmingham, now more than ten times larger than the Hector corpus, fails to provide evidence of a number of items that it is nevertheless reasonable to consider part of the current English lexicon. This is important; however, it is also important to recognize that, as a corpus lexicographer with ready access to large quantities of data, I am in an unusual position. Many researchers can work only with much smaller corpora. Furthermore, it is not only the size and composition of a corpus that are important. The software tools for accessing and analysing the corpus are crucial: speed, flexibility, delicacy are all needed. It is only fair for me to say that my task of studying a substantial number of expressions in the Hector corpus within a limited time frame was made much easier, perhaps only made possible at all, because the tools developed by Digital in Palo Alto were designed to expedite retrieval of collocational matches and sets, as much as single items. (For accounts of these tools, see Atkins 1992 and Glassman et al. 1992.) I am aware that some colleagues working elsewhere on multiword items, both in Britain and abroad, have encountered many more problems than I did in actually locating and retrieving instances of expressions and their variants.

The database that I assembled was technically a series of structured text files, and it recorded information for some 6,700 expressions of the types already mentioned. The information included frequencies in the Hector corpus, variations observed, some transformations, idiomaticity type, syntactic class or structure, metaphorical character, discourse function, and so on. The set of 6,700 expressions was loosely based on those treated as 'phrases' in the first edition of the *Collins Cobuild English Language Dictionary* (1987), of which I was one of the editors, with a number of additions and omissions: there was in the end an overlap of about 85 per cent between the two sets. I wanted

-81-

to look at a reasonably large number of expressions which would be representative of the set of phrasal lexemes in current English -- the kinds of item and the spread of items which a general dictionary might be expected to cover. As *Cobuild* had itself been based on firsthand observations of lexis in a seventeen-million-word corpus (the *Birmingham Collection of English Text*), I expected it to contain a reasonable and sensible selection. The set of complex items I looked at is clearly neither exhaustive nor exclusive. To put its size into perspective, one should recall that the two-volume *Oxford Dictionary of Current Idiomatic English* (1975-83) contains some 15,000 items, including phrasal verbs: perhaps in the region of 9,000 items of the kind I am discussing here. In contrast, *The Longman Dictionary of Idioms* (1979) contains around 4,500 expressions of various kinds, and Makkai's revised edition of *A Dictionary of American Idioms* (1987) contains 5,000 items, again varying in type.

2
FREQUENCY, TYPE, AND SYNTACTIC FORM

Before setting out some results from my study, I should enter two caveats. First, I counted polysemous -- genuinely ambiguous -- expressions individually as two or more homographs. Minor variations in meaning, however, such as the different usages of *beg the question,* or of phatics such as *you know,* were not distinguished in this way. Secondly, my counts are crude, and I did not attempt to compute mathematically robust statistical analyses of frequencies and significances -- the significances, that is, of the component elements co-occurring within a span. In reporting that an expression was found with a frequency of one in a million, I mean simply that I found eighteen discrete realizations of it in the Hector corpus of eighteen million words, including any inflected forms.

The overall distributions of the expressions, arranged in bands, and represented as rounded percentages, are shown in Table 4.1. It can be seen that the bulk of phrasal lexemes (70 per cent) occur less frequently than one in a million. Moreover, 40 per cent fail to occur at all, or occur with frequencies that must be considered no better than chance, according to Dunning (1992), who discusses the (lack of) significance of low-frequency events in corpora. The important point is that so many phrasal lexemes occur at this frequency level. With frequencies of five or more, it is possible to be slightly more confident of observing them again in other broadly comparable corpora.

-82-

TABLE 4.1. Overall frequencies of phrasal lexemes in the Hector corpus

Frequency range	Percentage of PLs
0	8
1-4	32
5-17	32
1-2/million	12
2-5/million	9
5-10/million	4
10-50/million	3

50-100/million	< 1
> 100/million	< 1

Some examples of expressions with zero frequencies included:

(1) bag and baggage
hang fire
kick the bucket
lose your rag
one man's meat is another man's poison
out of practice
speak for yourself!
when the cat's away, the mice will play

Only sixteen items were found in the Hector corpus with frequencies of over 100 per million, and these included:

(2) at all
at least
in fact
of course
rather than
sort of
take place
you know

together with some multiword grammatical items such as *according to, because of,* and *so that/so as.*

It is interesting to observe what happens when frequency is correlated with idiomaticity type. There are a number of different typologies described in the literature (those based on Russian phraseological theory are discussed by Cowie in this volume). The typology I used in this research was developed in the light of other typologies, and can be related to them. It is very much a lexicographer's typology, based

on an assessment of the idiomaticity (or non-compositionality) of a string, and the reason why a lexicographer might choose to treat it in a dictionary as a complex item. I established three groupings or macro-categories:

- anomalous collocations (involving problems of lexico-grammar);
- formulae (involving problems of pragmatics and discoursal or situational interpretation);
- metaphors (involving semantic problems).

These can be broken down further, and are set out in Table 4.2. Most of the categories are self-explanatory. 'Cranberry' collocations (e.g. *put the kibosh on*) are those which contain lexical items which are unique to the combination (Makkai 1972: 43 and elsewhere). De-

TABLE 4.2. The typology in use

Nature of problem	Major categories	Subcategories
Problems of lexico-grammar	Anomalous collocations	Ill-formed collocations *by and large, of course* Cranberry collocations *put the kibosh on, to and fro* Defective collocations *beg the question, in time* Phraseological collocations *in action, on show*
Problems of pragmatics	Formulae	Simple formulae *in this day and age, you know* Sayings *an eye for an eye* *that's the way the cookie crumbles* Proverbs (metaphorical/literal) *you can't have your cake and eat it, enough is enough* Similes *as nice as pie, as white as a sheet*
Problems of semantics	Metaphors	Transparent metaphors *alarm bells ring, rock the boat* *Semi-transparent metaphors* grasp the nettle, on an even keel Opaque metaphors *bite the bullet, kick the bucket*

fective collocations are expressions which contain items used with meanings which are unique to the combination (e.g. *beg the question*), or which contain delexicalized or semantically depleted component items (e.g. *in time*). Phraseological collocations are those where there is a limited paradigm operating at one (or more) of the word-slots (cf. *on*

show/display), but the realizations are restricted to a small set that is not fully productive (cf. the unacceptable ★*on demonstration*).

In about 25 per cent of cases, expressions had claims to being assigned to two categories. They might, for example, contain unique items and also have developed diachronically as metaphors, as in the case *of blow the gaff* and at *someone's beck and call.* However, in 75 per cent of cases, a unique assignment was possible. In the following, I shall consider only primary assignments to one of the three macrocategories.

Table 4.3 sets out frequencies correlated with the typology. It shows that the majority of metaphorical expressions have frequencies of less than one per million, and that very common expressions are likely to be anomalous collocations of some kind.

Turning to look at each macro-category in finer detail, we find that the grouping of anomalous collocations is divided up, in terms of primary assignments, as 10 per cent grammatically ill-formed items, 4 per cent cranberries, 62 per cent defective collocations, and 24 per cent phraseological collocations. The commonest type by far is the string which contains an item with a meaning unique to the combination, or where there is no paradigm, or only a partial paradigm, operating within it.

TABLE 4.3. Frequencies and typoloy (%)

Frequency range	Anomalous collocations	Formulae	Metaphorical expressions
0	2	2	3
1-4	11	8	13
5-17	15	5	13
1-2/million	7	2	3
2-5/million	6	2	1
5-10/million	3	≪1	≪1
10-50/million	2	≪1	≪1
50-100/million	≪1	≪0.1	0
≫100/million	≪1	≪0.1	0
TOTAL	46	21	33

TABLE 4.4. Frequencies of anomalous collocations

Frequency range	Ill-formed collocations	Cranberry collocations	Defective collocations	Phraseological collocations
0	≪1	≪1	3	1
1-4	2	1	14	5
5-17	3	1	21	7
1-2/million	≪1	≪1	9	3
2-5/million	1	≪1	8	4
5-10/million	≪1	≪0.1	3	2
10-50/million	≪1	≪1	3	1
50-100/million	≪1	0	≪1	≪1
≫100/million	≪1	0	≪1	≪0.1

Table 4.4 maps types of anomalous collocations. The type with the highest frequencies is clearly the defective collocation. Those expressions which are grammatically ill formed or contain unique lexical items generally have lower frequencies.

The grouping of formulae breaks down as follows: 70 per cent of my set were simple formulae, 2 per cent were sayings, 19 per cent were proverbs, and 9 per cent were similes. The proverb component was inflated because I incorporated around 250 proverbs from an ancillary study of proverbs, undertaken with Pierre Arnaud and reported in Arnaud and Moon (1993). Of these proverbs, 67 failed to show up in the Hector corpus, and another 160 occurred with frequencies no better than random chance. Of the proverbs 59 per cent were metaphorical, and this figure can be compared to Norrick's figure of

TABLE 4.5. Frequencies of formulae (%)

Frequency range	Simple formulae	Sayings	Proverbs	Similes
0	5	≪0.1	5	2
1-4	20	1	11	6
5-17	21	≪1	3	≪1
1-2/million	9	≪0.1	≪0.1	≪0.1
2-5/million	8	0	≪0.1	0
5-10/million	4	0	0	0
10-50/million	3	0	0	0
50-100/million	≪0.1	0	0	0
≫100 million	≪0.1	0	0	0

about 56 per cent in his examination of the proverbs in *The Oxford Dictionary of English Proverbs* (Norrick 1985).

Table 4.5 maps types of formulae. As far as proverbs, sayings, and similes are concerned, their occurrence or non-occurrence in a corpus of this size is almost entirely a matter of chance.

Of the metaphors, I analysed 37 per cent as transparent, 51 per cent as semi-transparent, and 12 per cent as opaque. It is important to note that few metaphorical expressions are actually opaque (i.e. true idioms). The majority can be decoded through real-world knowledge and perhaps some specialist knowledge (consider *touch base, go off the rails, on an even keel*). Because the line between semi-transparent and opaque is arbitrary and subjective (cf. Cowie1981), the two are grouped together as 'non-transparent' in Table 4.6, which maps types of metaphors. It is worth noting that, of the metaphors that I classified as opaque, only four have frequencies greater than one in a million:

(3) bite the bullet
over the moon
a red herring
a tall order

No metaphor occurs more frequently than fifty times per million, and the sorts of high-frequency metaphors that do occur are far from being pure idioms: for example, *play a part in (something),* and *take steps to (do something).*

For reasons of space, I cannot deal in detail with the syntactic structures of phrasal lexemes, though this is clearly a rich field to explore, as Healey (1968) and Makkai (1972) have shown. A rough breakdown by syntactic type of the expressions found in my set is given

TABLE 4.6. Frequencies of metaphors (%)

Frequency range	Transparent metaphors	Non-transparent metaphors
0	2	6
1-4	13	26
5-17	14	24
1-2/million	5	5
2-5/million	2	2
5-10/million	≪1	≪1
10-50/million	≪1	≪0.1
50-100/million	0	0
≫100/million	0	0

TABLE 4.7. Grammatical types of phrasal lexemes (%)

Grammatical types	Percentage of PLs
Predicates	40
Nominal groups	9
Adjectival groups	2
Modifiers, quantifiers	1
Adverbial groups, submodifiers	28
Sentence adverbials	5
Conventions, exclamations, and subordinate clauses	12
Fillers, others	1

in Table 4.7. The commonest type (40 per cent) is the predicate -- a verb and its arguments or complementation; the next commonest (28 per cent) consists of adverbials of different sorts. There are some predictable correlations with idiornaticity types. For example, metaphors cluster as predicates, adverbials, and nominal groups; anomalous collocations as predicates, adverbials, and sentence adverbials; and formulae as conventions or exclamations. There are also predictable correlations between grammatical type and frequency band, and these are set out in Table 4.8. Predicates are generally infrequent. Expressions with grammatical functions, such as disjuncts and conjuncts, are more prominent in the higher frequency ranges, and this is entirely in keeping with the relative frequencies of lexical and grammatical words in the lexicon as a whole.

The assembly of the database and the examination of corpus evidence should have provided an excellent opportunity to gather hard evidence concerning the inflections of component words in expressions and the transformational potential of, in particular, idioms. This is the kind of information that has interested phraseologists working within a generative perspective (Fraser 1970; Newmeyer 1974). I had wanted to see to what extent the relationship between surface lexical meaning and deep idiomatic meaning reflect passivization potential and so on. Newmeyer (1974), for example, observed that *bury the hatchet* and *spill the beans* passivize, whereas *kick the bucket* and *blow your top* do not, and that this has to do with the passivization potential of their meanings, not of the verbs realizing their metaphors. The database records such mismatches, and this provides some neat examples and pleasing confirmation of the traditional concrete-toabstract pattern that characterizes metaphors; however, in the best cases, idioms, there was typically too little evidence for anything definite to be said about passivization. The same is true, by and large,

TABLE 4.8. Grammatical types and frequencies (%)

Grammatical types	Proportion (per million and as % of database)		
	≪1/m. 72% of database	1-10/m. 24% of database	≫10m. 4% of database
Predicates	32	8	≪1

Nominal groups	7	2	≪0.1
Adjectival groups	2	≪1	≪0.1
Modifiers, quantifiers	≪1	≪1	≪0.1
Adverbial groups, submodifiers	17	9	2
Sentence adverbials	2	2	1
Conventions, exclamations, and subordinate clauses	11	1	≪1
Fillers, others	≪1	≪1	≪1

of aspect and tense, and person. Clearly, a much bigger corpus is needed for firmer conclusions to be reached, though, even in the 211million-word Bank of English, all 117 examples of *spill the beans* are active, and only 4 out of 77 examples of *bury the hatchet* are passive. (None of the 18 examples of *kick the bucket* or the 64 examples of *blow your top* is passive.)

What emerges clearly from the Hector corpus -- as from other corpora -- is that many phrasal lexemes are fossilized in certain phraseological structures or lexico-grammatical patterns. Rules cannot be established conclusively, but clear tendencies and patterns can be observed. For example, it was not possible to state unequivocally which expressions passivize, as an absence of passive forms may simply result from there being very few instances anyway; similarly, an isolated passive form may not be significant. On the other hand, some expressions passivize freely:

(4) bear something in mind, something is borne in mind
haul someone over the coals, be hauled over the coals
nip something in the bud, be nipped in the bud
settle a score, a score is settled
arm yourself to the teeth, be armed to the teeth

A few expressions are fossilized in the passive -- the Hector corpus, at any rate, does not provide evidence of active uses:

(5) be cheered to the echo
be laughed out of court

be mentioned in despatches
be hoist by your own petard
the die is cast

Similar points could be made about negatives and other kinds of transformations.

Equally important, however, are the structures and words which co-occur with phrasal lexemes. The status of such structures and words is not always clear: they could, arguably, be regarded as part of the actual expression. For example, *over the moon* is typically associated in the Hector corpus with an adjunct or a when-clause; *pull your weight* is typically associated with a negative; the intensifier *to a fault* typically follows a positively-evaluating adjective (e.g. *generous* or *honest*). *Rock the boat* is nearly always associated with modal verbs or other modalizers, and usually with negatives or negative-oriented words; as in: [she] . . . *attempts damage limitation by telling her boss to keep his mouth shut and not rock the boat* (Hector corpus, journalism).

Some phrasal lexemes themselves collocate with other phrasal lexemes: for example, *as they say* (though not to *coin a phrase,* which collocates with other kinds of cliché). Schiffrin draws attention to the case of *you know,* which signals consensual truths (1987: 267-8):

(6) **Y'know** *they say***an apple a day keeps the doctor away.**

Francis (1993: 145) points out that be a case of typically co-occurs in the Bank of English with phrasal lexemes or other expressions and functions as a preface which compares a situation already established in the discourse with another one familiar to the reader/hearer. Her examples include *It's simply a case of keeping your fingers crossed* and *For it may not just be a case of having egg on your face.* Another similar example of a preface *is not be one to (do something).* The expression here refers generally to the sort of action which is accepted or recognized as untypical of the subject (here, *Mr Baker* and *Robinson*):

(7) *But Mr Baker***is not one to go out on a limb.** (Hector corpus, journalism)

*With £10.8m turnover, Robinson***is not one to rest on laurels** [*sic*]. (Hector corpus, journalism)

3
POLYSEMY AND AMBIGUITY

My analysis suggested that around 5 per cent of my set of phrasal lexemes are polysemous. This can be compared with a figure cited by Klappenbach, who reports research (1968: 183) suggesting that 8-9

per cent of Russian phraseological units of equivalent kinds are polysemous. I found polysemy more strongly associated with anomalous collocations than other idiomaticity types, and more strongly associated with adverbial groups than other grammatical categories. Of phrasal lexemes in my database 40 per cent are predicates, and 28 per cent are adverbial groups; however, of polysemous expressions, 31 per cent are predicates and 52 per cent adverbials. Many polysemous expressions have different structures or collocations associated with their different meanings, and so very few are truly ambiguous in context. The most heavily polysemous items I found were *give way, in line,* and *take care,* and here the

different uses or meanings were indeed associated with different forms and collocations.

Searching the Hector corpus for the metaphorical expressions revealed surprisingly few literal counterparts, and in fact Chafe hypothesized (1968: 111) that the idiomatic meanings of strings would be commoner than their literal counterparts, where such exist. Chafe's view can be compared with the reports of psycholinguistic experiments by Gibbs and others, showing that idiomatic meanings are generally processed before, or faster than, literal ones (Gibbs 1980, 1986). The absence or infrequency in performance of literal counterparts is significant, especially in the light of conventional assumptions that true idioms must have literal analogues. Corpora support the view that the anomalous collocation or metaphor blocks use of the literal equivalent, if indeed there is one. For example, all instances of the lemma bean occurring within fifteen words of the lemma *spill* in the Hector corpus form part of the idiom *spill the beans.* Similarly, the lemmas *storm* and *weather,* or *rock* and *boat,* unconstrained as to word class, might be predicted to co-occur in their literal meanings, simply through coincidence, since they are co-members of lexical sets or common to single semantic fields. However, 22 out of 23 co-occurrences of *storm* and *weather* realize the expression *weather the storm,* and 29 of 32 co-occurrences of rock and boat realize the idiom *rock the boat.* The remainder, the literals, all have very different structures. Arguments concerning the literal counterparts of idioms and so on may well, therefore, have little basis in reality, though plausible in the abstract. This, of course, ignores cases where puns are intended on the literal meanings of idioms, as in:

(8) *A companion volume, Blades Trivia (surely the club's entire history?), asks how Leeds United's hopes of a win***went up in smoke** *when Steve Cammack scored a late equaliser on April Fools' Day 1975. The sponsors' firework display, timed to start on the final whistle, went off early, and during the distraction Cammack scored.* (Hector corpus, journalism)

4 LEXICO-GRAMMATICAL VARIATION

Corpora provide strong evidence of patterning, though only incomplete or inconclusive evidence of transformational potential. Corpora also show overwhelming evidence of the instability of the forms of phrasal lexemes, and the frequent indeterminacy of their lexical elements. In some cases, of course, the lexis is non-negotiable, as in *kick the bucket, steal someone's thunder,* and so on. However, around 40 per cent of the expressions I looked for in the Hector corpus did not have frozen and fixed canonical forms, and I believe this figure will be the same or even higher in larger corpora. This has serious implications for teaching, lexicography, and such matters as the automatic or computational detection of such items in corpora.

While the variations differ in type and degree, certain patterns can be observed. For example, there are quasi-systematic variants, of the kind that Cowie draws attention to in the introduction to the *Oxford Dictionary of Current Idiomatic English*, ii (Cowie et al. 1983: pp. xxxiii-iv). They can be seen as systematic because they show some sort of regularity, or perhaps predictability. These often express notions, loosely, of possession:

(9) get a raw deal
have a raw deal
get your eye in
keep your eye in
have your eye in [1]
have full play
give full play to something
allow no play to something (etc.)

have got cold feet
develop cold feet
have your finger in the pie
get fingers in too many pies (etc.)

have the measure of someone
get the measure of someone
take someone's measure (etc.)

have (no, an) axe to grind
with(out) an axe to grind
have your feet on the ground
keep your feet on the ground
with your feet on the ground

[1] Here, as in other sets, the verb alternation corresponds to the contrast between 'causative', 'durational', and 'stative' meaning.

More commonly, expressions have synonymous variants: different lexicalizations, where typically the substituting lexemes are not necessarily synonymous, although the variant expressions are:

(10)	*at all events*	*at any event*
	burn your boats	*burn your bridges*
	drag your feet	*drag your heels*
	hit the roof	*hit the ceiling*
	lower your guard	*let your guard down*

sweeten the pill	*sugar the pill*
take someone down a peg/notch or two	*knock someone down a peg or two*
take the biscuit	*take the cake*
throw/toss/chuck in the towel	*throw in the sponge*
two a penny	*ten a penny*
up the ante	*raise the ante*

Some of these are restricted in terms of language variety, with, say, British and American English preferring different lexicalizations:

(11)	*blow your own trumpet* (Br.)	*blow your own horn* (Am.)
	cap in hand (Br.)	*hat in hand* (Am.)
	the boot is on the otherfoot (Br.)	*the shoe is on the otherfoot* (Am.)
	wouldn't touch something with a bargepole (Br.)	*wouldn't touch something with a ten-foot pole* (Am.)

It is not always easy for lexicographers to know whether or not these represent independent items or variants. They can be compared with sets of realizations of phraseological frames which have regular meanings:

(12) in bloom/in blossom/in bud/in flower/in leaf
in a fix/in a hole/in a mess/in a paddy/in a spot
in debt/into debt/out of debt
in touch/into touch/out of touch [2]

We can also compare the sorts of groupings to which Ruhl (1975, 1978) and Rose (1978) draw attention:

(13) hit the deck/hit the hay/hit the sack
make a fortune/make a killing/make a mint/make a pile

[2]If verb collocates are added to some sets, alternations appear which are similar to those at

(9): be in debt
get into debt
get someone into debt
be in touch
get in touch
(cf. *be out of touch, get out of touch*)

Some variations may be seen as focusing -- adapting metaphors, say, to their contexts or indicating degrees of completion or intensity:

(14) add fuel to the fire, fuel the fire, throw fuel on the fire
cut your coat according to your cloth, cut your cloth according to something
go into raptures, send someone into raptures, throw someone into raptures
put someone off the scent, throw someone off the scent
at all hours, at all hours of the day and night
at least, at the least, at the very least
in bloom, in full bloom
to advantage, to good advantage, to best advantage

An interesting subset are those where the variations are effectively lexico-grammatical transforms:

(15) another nail in the coffin, a final nail in the coffin, to nail down the coffin, to drive the first nail into the coffin (etc.)

by no stretch of the imagination, not by any stretch of the imagination, a stretch of the imagination, to stretch the imagination

to rap someone on the knuckles, to rap someone over the knuckles, to get your knuckles rapped, to have your knuckles rapped, a rap on the knuckles

to let the cat out of the bag, the cat is out of the bag

to pass the buck, the buck passes somewhere (cf. the buck stops here)

the writing is on the wall, to see the writing on the wall (etc.)

to turn the tables on someone, the tables are turned

to wipe the slate clean, the slate is clean (cf. a clean slate)

The integrity of the metaphors remains, but, as with the previous set, the variants foreground different parts of the metaphor.

Among the more interesting and amusing of the minor types of variation are those where variant spellings or forms have arisen through homophony or erroneous etymology. Many of these are trivial, but the variants encourage different interpretations of the metaphors involved. Some pseudo-variations arise through malapropism or jocular manipulation, and these variations may themselves become institutionalized:

(16)		
	be a shoo-in	*be a shoe-in*
	have another think coming	*have another thing coming*
	put a damper on something	*put a dampener on something* [3]
	rack and ruin	*wrack and ruin*
	have one's knuckles rapped	*have one's knuckles wrapped*
	sell like hot cakes (Br.)	*sell like hotcakes* (mainly Am.)

[3] I am grateful to Andrew Delahunty for pointing out the cases of *have another think/thing coming* and *put the damper/dampener on something.*

straight and narrow	*strait and narrow*
the spitting image of -----	*the spit and image of ----*
toe the line	*tow the line*
to whit	*to wit*
with bated breath	*with baited breath*
muddy the waters	*muddle the waters*
to all intents and purposes	*to all intensive purposes*
cast aspersions	*cast nasturtiums*
in suspense	*in suspenders*

There are also more extreme cases of variation -- ones where the only canonical form of the expression is a syntactic frame, although there may also be restrictions on the kinds of lexis realizing the frame: this type, the so-called formal idiom, is described by Fillmore, Kay, and O'Connor (1988).

(17) INTERROGATIVE or INTERROGATIVE (to indicate certainty): *Am I right or am I right?*

If-CLAUSE, CLAUSE (to indicate acceptance of a situation): *If it rains, it rains.*

INTERROGATIVE (as a response to an unnecessary question): *Is the Pope Catholic? Do bears shit in the woods? Do ducks swim? Does Dolly Parton sleep on her back?* (etc.)

QUANTIFIER + NOUN GROUP 1 + *short/shy of* + NOUN GROUP 2 (to indicate inadequacy, especially mental inadequacy): *two sandwiches short of a picnic, a few bricks shy of a full load, a few pickles short of a jar, one shingle shy of a roof, a flying buttress short of a cathedral, one hot pepper short of an enchilada* (etc.)

These are essentially variable items. In particular, the use of the last of these frames demands creativity in the selection of the realizations. Other cases of extreme variation present more conventional problems:

(18) shake/quake in your shoes
shake/quake/quiver in your boots
quake in your Doc Marten's
go down the chute/drain/pan/plughole/toilet/tube/tubes
view . . . through rose-coloured lenses
look at through rose-coloured glasses
observe. through rose-tinted glasses
her rose-coloured idea
a rose-tinted vision
this rosy view
. . . recollection . . . is less rosy.

The first group could be represented as 'oscillate in your footwear',

though this is not a lexicalization that actually occurs. Francis points out another case in her discussion of variations in the Bank of English on the theme of I haven't the faintest idea (1993: 144). *Lack* and *display* substitute for *have; least, slightest, foggiest,* and *remotest* substitute for *faintest;* and *conception* and *notion* substitute for *idea* (which can also be ellipted in *I haven't the foggiest*). However, not all permutations of realizations of the slots are possible, and complex collocational principles, as well as semantic ones, underlie such sets.

It may be more appropriate to consider such sets of variations as realizing'idiom schemas'. An idiom schema has an underlying, often metaphorical, concept or conceit with two or more preferred lexicalizations. In addition to the cases just mentioned, it can be seen with the loose expression *chicken-and-egg* or *which came first? the chicken or the egg?* and so on, where a single concept is represented but expressed in different ways. Similarly, with *carrot and stick:*

(19) *Well a number of things. I've mentioned a few of them. I would have a* **carrot and stick** *approach. The***carrot** *would be much better Park and Ride, which is what they are talking about. Now what the Labour Group are doing is super, but it doesn't go far enough.* (Hector corpus, transcribed speech)

T*he career opportunity***carrot** *at the end of the exacting training***stick** *is, it has to be said, pretty appetising.* (Hector corpus, journalism)

*Ronald Reagan reached the limits of abuse; the macho bullying,***the stick and the carrot.** (Hector corpus, journalism)

*No***carrot was dangled** *for the SACU as a quid pro quo for any decision it may take to cancel the tour.* (Hector corpus, journalism)

It is here that the notion of the canonical form, or of fixedness, is undermined most strongly. And this has nothing to do with the phenomenon of the deliberate exploitation or infixing of idioms and other expressions for stylistic effect or discoursal reasons. Exploitation is not a negligible phenomenon, and there are many examples in everyday usage, but it is not as significant as simple variation.

5
CORPUSL COMPARISONS

It is important to relate my observations of phrasal lexemes in the Hector corpus to other corpus-based research. There are very few published reports, and it is difficult to compare them. Altenberg (1991, 1993, and this volume) has worked on pre-patterning and recurrent combinations in the London-Lund corpus of spoken English. The five fixed expressions which he observed most commonly (1991)

are *at all, I think, thank you, you know,* and *you see.* These are all amongst the commonest expressions that I observed in the Hector corpus.

Norrick, Strässler, and Sorhus all looked at spoken data. Norrick (1985: 6-8) hunted for proverbs in Svartvik and Quirk's *A Corpus of Spoken English Conversation* -- 170,000 words -- but found only two, of which one was in a non-canonical form. This is far fewer than the 19 (9 canonical, 10 non-canonical) which I found in roughly three times as much transcribed spoken interaction in the Hector corpus: this spoken data, however, included broadcast journalism as well as conversation.

Strässler (1982: 77-84) looked at 106,000 words of spoken English interaction of various kinds, and found ninety-two idioms. He has a fairly narrow view of idiom, but he also identifies some instantial metaphors as idioms. A cautious estimate suggests that metaphors in the Hector corpus overall are around 40 per cent commoner than in Strässler's corpus.

Sorhus (1977) finds a very high density of fixed expressions (her term) in her study of 131,536 words of spontaneous Canadian speech, and calculates that 'the national Canadian average . . . [is] one fixed expression every five words'. [4] She is looking at fillers and formulae, however, and includes single-word items such as *say, right, well,* and *please. At times* and *of course* in her study are significantly commoner than in the Hector corpus, although most of her commonest expressions also have high frequencies in the Hector corpus: for example, *of course, at all,* and *at least,* and the fillers *I think* and *you know. You know* appears to be nine times commoner in her data than Bengt Altenberg found in the London-Lund corpus. In contrast, *in fact* occurs very frequently in the Hector corpus but does not feature in Sorhus's list of commoner expressions, which suggests that *in fact* is more prominent in written texts than spoken.

Finally, Akimoto (1983) quantified some 400 predicate expressions in the Survey of English Usage (385,000 words) and an unspecified amount of 1970s journalism and other writing. His commonest expressions are *take place, take action, take part in, play a part/role in,* and *take steps (to do something*). These all occur with high frequencies in the Hector corpus, although the relative proportions vary.

Equally important and interesting are comparisons between the Hector corpus and other corpora. Ken Church, of AT&T Bell Laboratories, supplied me with some data concerning seven items in

[4] Compare Pawley and Syder (1983) on the close relationship between the clause and the incidence of ready-made sequences.

TABLE 4.9. Corpus comparisons: Church data (permillionages)

Phrasal lexeme	OHPC 18m.	AP 1990 46m.	AP 1991 47m.	WSJ 61m.	CHE 18m.
spill★ the beans	0.39	0.02	0.02	0.06	0.11
beg★ the question	1	0.09	0	0.18	0.78
call★ the shots	0.94	0.71	0.93	0.89	0.94
without exception	2	0.28	0.53	0.47	2.28
ups and downs	1.83	1.47	1.97	1.93	0.78
from afar	0.89	0.39	0.72	0.36	0.22
of course	242	28.98	34.32	67.88	282.17

Note: AP = Associated Press; CHE = Canadian Hansard, English Version; OHPC = Oxford Hector Pilot Corpus; WSJ = *Wall Street Journal*.

four large corpora: three consisting of US journalism -- the AP newswire text from 1990 and 1991, and the *Wall Street Journal* -- and the fourth the English-language version of the Canadian Hansard. [5] He ran strictly linear searches, and the

results can be seen in Table 4.9. Permillionages are the numbers of tokens occurring per million words of corpus text. Distribution is clearly variable. It is noticeable that the differences are not consistent across all the tested expressions. Some, such as *call the shots* and *ups and downs,* bear up fairly well under comparison, whereas *of course* in particular is remarkably divergent.

I checked to see how this related to the evidence in the 211-millionword version of the Bank of English at Cobuild. I looked for the same seven expressions, and the results are given in Table 4.10. In this case, it was possible to run non-linear searches, and to include infixed and transformed instances. The figures show less of a disparity.

These two sets of comparisons suggest the effect of genre distinctions, and this can be investigated more closely. For example, *beg the question* has a frequency of forty-five in the Hector corpus. It occurs particularly strongly in non-fiction, and its presence in journalism is particularly associated with serious newspapers. *Beg the question* is less common overall in the Bank of English, but features even more strongly in its subcorpora of serious journalism than it does in the Hector corpus. It also seems that *beg the question* is significantly less common in American than British English. Genre and register may therefore account for disparities between corpora.

The distribution statistics of four further expressions, as evidenced

[5]I am indebted to Ken Church for his help.

TABLE 4.10. Corpus comparisons: Bank of English data (permillionages)

Phrasal lexerne	OHPC 18m.	Bank of English 211m.
spill ★ *the beans*	0.39	0.56
beg ★ *the question*	2.5	1.24
call ★ *the shots*	0.94	1.21
without exception	2	1.18
ups and downs	1.83	2.38
from afar	0.89	0.95
of course	242	217

Note: OHPC = Oxford Hector Pilot Corpus.

in the Hector corpus and an earlier 120-million-word version of the Bank of English, are set out in Table 4.11 : note that it was not possible to separate the fiction and non-fiction components in the Bank of English. The expressions are generally less common in spoken data than might have been predicted. *Spill the beans, call the shots,* and *ups and downs* are consistently commoner in journalism. In contrast, *from afar is* commoner in fiction/non-fiction: slightly so in the Hector

TABLE 4.11. Corpora, genre, and phrasal lexemes (%)

Corpus and genre	Genre as percentage of corpus	Percentage of tokens per genre			
		spill the beans	***call the shots***	***ups and downs***	***afar***
OHPC journalism	66	75	83	79	62
BofE journalism	33	49	54	49	22
OHPC speech	3	0	0	3	0
BofE speech	28	6	23	19	11
OHPC fiction & non-fiction	29	25	17	18	38
BofE fiction & non-fiction	38	45	23	31	66
OHPC other	2	0	0	0	0
BofE other	1	0	0	≪1	≪1

Note: Bole = Bank of English; OHPC Oxford Hector Pilot Corpus.

corpus, dramatically so in the Bank of English, which contains substantially more fiction than the Hector corpus. This kind of comparison could be the basis for a stylistics of phrasal lexemes.

6
CONCLUSION

I do not wish to overstate my case, since such small amounts of data are involved. However, it seems reasonably certain that most phrasal lexemes of the kind discussed here are indeed infrequent. Huge corpora are needed to provide robust information confirming hypotheses concerning transformations and so on. New models of the variations and canonical forms of phrasal lexemes are also needed, in order to account for the overwhelming evidence of the phenomenon of variability and extreme flexibility. There seem to be decided genre preferences for phrasal lexemes in general, as well as for individual expressions, and again more work is needed to confirm this. All this has, I believe, implications for lexicography and lexicology, as well as other fields.

Clearly, my research was not conclusive, nor could it be, given that I was working with a comparatively small, unbalanced corpus. At best, it provides an overview, with benchmarking statistics, leading to some plausible probabilities. These themselves suggest areas in which further corpus investigations are needed in order to describe the complex set of English phrasal lexemes more robustly.

5
On the Phraseology of Spoken English: The Evidence of Recurrent Word-Combinations

BENGT ALTENBERG

1
INTRODUCTION

Phraseology is a fuzzy part of language. Although most of us would agree that it embraces the conventional rather than the productive or rule-governed side of language, involving various kinds of composite units and 'pre-patterned' expressions such as idioms, fixed phrases, and collocations, we find it difficult to delimit the area and classify the different types involved. Indeed, as Pawley and Syder (1983) and others have pointed out, the existence of a large number of more or less prefabricated expressions in language blurs the distinction between lexicon and grammar and strongly suggests that 'lexicalization and productivity are matters of degree' rather than a clear-cut dichotomy. This state of affairs creates problems of description for both the empirical and the theoretical linguist, at the same time as it provides a challenge to anyone who wants to get a better understanding of language and language use.

In this chapter I will adopt a rather non-committal approach to these problems and examine the phraseology of spoken English on the basis of recurrent word-combinations in the London-LundCorpus of Spoken English (for a description of the corpus, see Greenbaum and Svartvik 1990). By 'recurrent word-combination' I here mean any continuous string of words occurring more than once in identical form (see Altenberg and Eeg-Olofsson 1990). To judge from the London-LundCorpus, spoken English is very rich in such recurrent strings: the corpus (which consists of nearly half a million running

This study is part of a larger investigation of the phraseology of spoken English financed by the Swedish Council for Research in the Humanities and Social Sciences. I gratefully acknowledge this support.

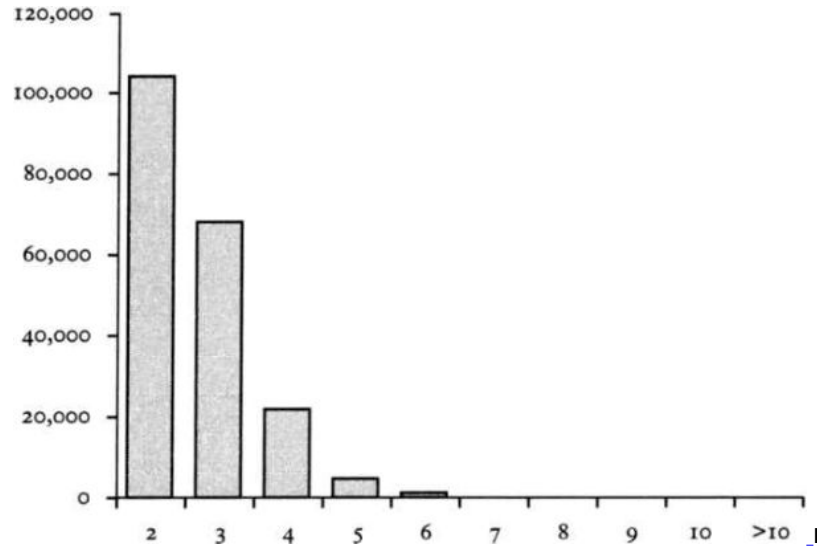


FIG. 5.1. *Length and frequency of recurrent word-combinations in the London-LundCorpus*

words) contains over 201, 000 recurrent word-combinations, representing over 68, 000 different types of varying length and frequency (see Fig. 5. 1). A rough estimation indicates that over 80 per cent of the words in the corpus form part of a recurrent word-combination in one way or another.

Although many of these combinations are of little phraseological interest, since they consist of mere repetitions or fragments of larger structures (e.g. *the the, and the, in a, out of the),* the material is a useful starting point for an examination of the phraseology of spoken English. At the same time, the sheer bulk of the material makes some sort of selection necessary. For practical reasons I will therefore limit my examination to word-combinations consisting of at least three words occurring at least ten times in the corpus. These limitations are to a large extent arbitrary. Neither length nor frequency is a criterion of phraseological status, but the frequency threshold gives at least some guarantee that the selected word-combinations have some currency in spoken discourse and that they are of some interest from that point of view. The length restriction was chosen partly to reduce the number of fragmentary sequences, but mainly to reduce the material to a manageable size. The drawback of excluding two-word combinations, which are by far the most common type of recurrent

word-combination in the corpus (see Fig. 5. 1), is that it also excludes a number of phraseologically interesting idioms and collocations. This restriction should be borne in mind in the following discussion.From this initial selection I eliminated all examples involving unintentional repetition or stuttering (such as *the the the, I was I was, but I I),* since these are of little phraseological interest. The resulting material consisted of 6, 692 examples (tokens), representing 470 different types of word-combinations. It should be emphasized that this is a very small sample of the total number of recurrent wordcombinations in the London-LundCorpus -- 3 per cent of the tokens and a mere 1 per cent of the types. Moreover, despite the length condition, the word-combinations are on the whole not very long: they range from 3 to 5 words, with a mean length of 3.15 words. In other words, continuous recurrent word-combinations in speech tend to be fairly short.

2
GRAMMATICAL TYPES

Grammatically, these word-combinations represent a wide range of structures. It is useful to distinguish the following broad categories:

- Full clauses independent dependent
- Clause constituents multiple single
- Incomplete phrases

As shown in Table 5.1 , most of the examples in the material consist of

TABLE 5.1. Main types of structures

Type of structure	*n*	%
Full clauses	660	10
independent	533	8
dependent	127	2
Clause constituents	5, 101	76
multiple	3, 379	56
single	1, 362	20
Incomplete phrases	931	14
TOTAL	6, 692	100

multiple or single clause constituents (56 per cent and 20 per cent respectively), while those forming full clauses or incomplete phrases are less common (10 per cent and 14 per cent respectively). In the following sections I will describe each of these categories in turn.

3
FULL CLAUSES
3. 1
Independent clauses

The most interesting types of independent clauses are listed in Table 5.2 together with their frequencies. The examples have been divided into major functional types. Forms varying merely in terms of tense or some optional expansion have been conflated. Thus the sequence *yes I do/did,* for example, represents two recurrent forms in the material, *yes I do* (13 examples) and *yes I did* (12), while *(yes) that's right (yes)* represents three recurrent forms, *yes that's right* (22 examples), *that's right yes* (12), and *yes that's right yes* (11). Functionally, the great

TABLE 5.2. Recurrent types of independent clauses

Functional type		Example	*n*
Responses	thanks	*thank you very much (indeed)*	44
		thanks very much	15
	reassuring	*it's all right*	13
		that's all right	10
	acknowledgement	*oh I see*	38
		yes I see	13
	agreement	*(yes) that's right (yes)*	45
	positie (polar)	*(yes) oh yes (yes)*	30
		yes I do/did	25
		yes it it is	22
		yes of course	12
		yes I hae	10
	negatie (polar)	*no I don't/didn't*	24
		no I haven't	18
		oh no no	13
	disclaimer	*well I don't know*	14
		oh I don't know	10
Epistemic tags		*I don't know*	47
		I'm not sure	12
Metaquestions		*what is/was it*	22

majority of these clauses are responses, reflecting the interactive nature of the spoken material. In addition, two clause types function as 'epistemic tags' -- that is, they act as modal comment clauses that are either inserted parenthetically in, or added to the end of, an utterance, as illustrated in the following examples:

(1) C why don't you comm\ute# and sort of meet in N\ottingham or something (-- laughs) B yeah well (-- laughs)

you kn/ow#. I suppose I suppose that's the other . altvemative# *I don't kn\ow#* (LLC 8.2.695)
(2) and they'd had a {d\ishwashing} mach\ine# since -- *I don't know* for y/ears# (LLC 4.3.204)
(3) Dan's doing qu\ite well# he was [?]offered a vVery good job in the United States# which he w/ouldn't t\ake# -- he belongs now to a gr\oup [?ə] of and I think he's chief s/alesman# *I'm not s/ure#* (LLC 1.13.597)

Finally, one type functions as a 'metaquestion', reflecting problems of encoding in spontaneous speech. As shown in (4), this type, too, is typically inserted in the middle of an utterance. Though interrogative in form, it is self-oriented rather than interactive in character.

(4) well Time is very g\ood# and it's ch\eap# -- -- ((at least)) I th\ink it sells {each w\eek# {for *what \is ir#}#* . s/ixpence# to ninepence ((per \issue#)) (LLC 2.1. 1578)

The responses can be divided into functional subtypes reflecting the nature of the response. Most of these functions are fairly transparent and need no comment. The 'reassuring' expressions act as responses to apologies; the 'acknowledgements' typically signal the reception of new information; the 'agreements' serve to support a statement (rather than a question) uttered by the previous speaker.

Formally, the expressions represent three main types: those having a clearly recognizable clause structure (e.g. *it's all right, what is/was it, I don't know);* those having a deficient clause structure (e.g. *thank you very much);* and those forming a combination of several independent elements, either a clause and some reinforcing word (e.g. *yes I do, no I haven't, oh I don't know)* or a cluster of response items (e.g. *oh no no, yes of course, (yes) oh yes (yes)).* Most of the combinations consist of a core expression accompanied by an expansion of some kind (e.g. *thank you [very much] [indeed], [well] I don't know);* in some cases the distinction between core and expansion is blurred (e.g. *[yes] [oh yes] [yes], [yes] [of course])* and the term 'cluster' seems more adequate.

Phraseologically, few, if any, of the examples seem to be entirely 'frozen'. All are fairly transparent semantically; only a few seem to be

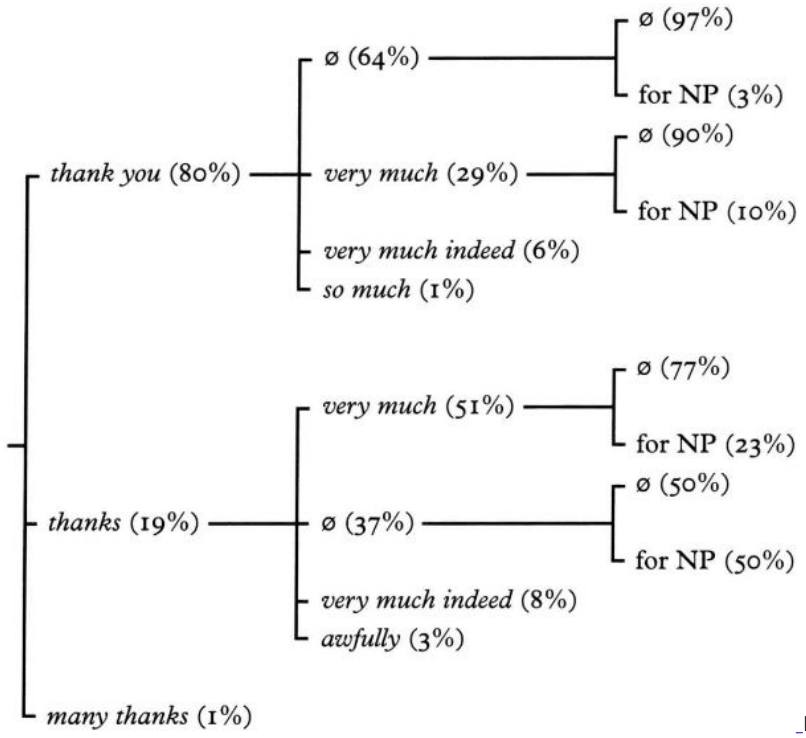


FIG. 5.2. *Variant expressions of thanks in the London-LundCorpus (n* = 308)

syntactically fixed (e.g. *that's right)* and even those expressions which have a fixed core generally permit some expansion *(thank you [very much] [indeed])* or variation (it's/that's all right). In fact, most of the examples appear to be fully rule governed -- that is, formed by productive grammatical rules rather than picked as wholes from the mental lexicon.

However, what makes all these expressions interesting from a phraseological point of view (apart from their frequency) is their pragmatic specialization. Not only are all of them speech-specific, they are also restricted to particular speech situations and, indeed, fully interpretable only through these situations. In other words, even if they are not fully lexicalized expressions, they represent conventional ways of conveying specific pragmatic meanings.

It should be added, however, that -- unlike most idioms -- none of the expressions serves as a unique expression of its particular pragmatic function. Rather, as the grouping suggests, each function seems to be served by a set of expressions; some sets are comparatively large (and represented only by the most common variants in the selected sample), as, for instance, the positive and negative polar responses,

while other sets are relatively small, containing only a few variants. An example of such a restricted paradigm is the expression of thanks in English. The variants occurring in the London-LundCorpus are shown in Fig. 5.2.

An interesting feature that has only been touched on so far is the 'composite' nature of most of the examples. As already mentioned, many of the expressions can be regarded as consisting of a core expression accompanied by an optional expansion, or simply as a 'cluster' of independent items. Even if some of them may have a fixed core (e.g. *that's right),* the variable or composite nature of the examples makes it difficult to describe them as phraseological units. Yet, what is interesting about these composites -- even from a phraseological point of view -- is that the expanded variants are generally more common than the non-expanded 'cores'. This is clearly illustrated by *I see,* which is twice as common with some accompanying element as without one in the corpus (see Table 5.3). Similarly, *that's right* occurs

TABLE 5.3. Recurrent combinations with I see

Combination	*n*
oh I see	66
I see	58
yes I see	31
[m] I see	7
oh yes I see	4
I see OK	4
ah I see	3
I see right	2

together with some reinforcing item in 60 per cent of the cases in the corpus (see Table 5.4). The same applies to *yes* and *no,* which generally co-occur with another response item (see Table 5.5 , and cf. Stenström 1984: 199).

The tendency of speakers to use composite (expanded) responses is further highlighted by the expression *I don't know,* which typically occurs in its expanded form *well/oh I don't* know as a disclaimer, but in its non-expanded form as an epistemic tag (see Table 5.2). Hence, to judge by the frequency of the different variants, the composite or expanded expressions can be regarded as the normal or unmarked form, and the non-expanded expressions as the less common or marked form.

TABLE 5.4. Recurrent combinations with that's right

Combination		*n*
	that's right	28
yes	*that's right*	22
	that's right yes	12
yes	*that's right yes*	11
yeah	*that's right*	9
	that's right yeah	5
oh	*that's right*	4
[m]	*that's right*	3
	that's right yes yes	3
yeah	*that's right so*	3
yeah	*that's right yeah*	3
	that's right that's right yes	2
	that's right yes that's	2
oh yes	*that's right*	2
yeah well	*that's right*	2
yes	*that's right*	2
yes	*that's right but*	2
yes	*that's right and*	2
sure	*that's right*	1
	that's right that's right	1
yes I think	*thats right*	1
well	*that's right*	1
	that's right well	1
...	*that's right . . .*	37

TABLE 5.5. Recurrent combinations with yes

Combination	*n*
yes yes	348
oh yes	224
yes but	143
yes well	132
[m] yes	87
well yes	74
yes [m]	69
ah yes	56
yes quite	37
yes OK	23

TABLE 5.6. Recurrent combinations with oh

Combination	*n*
oh yes	224
oh no	113

oh well	110
oh I see	64
oh dear	47
oh God	32
oh yeah	31
oh good	17
oh I don't know	16
oh really	14

The tendency of responses to take the form of composite clusters makes them reminiscent of collocations. Since their frequency and motivation appear to be discourse determined rather than lexically determined, I hesitate to call them collocations, but it is interesting to note that many composite responses display privileges of cooccurrence in much the same way as collocations. Thus, although *I see* co-occurs with a number of other response items, oh is clearly its favoured 'collocate' (see Table 5.3). *Oh,* in turn, typically co-occurs with *yes* (see Table 5.6), but it is also selected by a number of other items, acquiring different nuances of meaning in the process -- from reinforcing or emphatic (with *yes* and *no)* to 'qualifying' with *well* and 'surprised' with *I see.* Similarly, *yes* also tends to select certain 'collocates' *(yes, oh, [m], well, OK),* although here the choice is wider.

3.2
Dependent clauses

The recurrent dependent clauses in the material can be divided into three functional types (see Table 5.7): comment clauses expressing the speaker's view on either the form or content of an utterance; indirect conditions expressing politeness or a metalinguistic comment; and the reformulatory apposition marker *that is to say,* which is functionally (though not formally) similar to the comment clauses. All are mobile in their matrix clause and function as parenthetic adverbials (cf. Quirk et al. 1985: 1112-18).

Semantically, all except *as it were* are fairly transparent. Syntactically, most of them are relatively fixed, although some can be reduced *(that is [to say])* or expanded (e.g. *as you [probably] know [already])* or varied for tense (as I say/said). But again, what makes them phraseologically interesting is their pragmatic specialization. Each expression has a distinctive conventionalized discourse function, serving to

TABLE 5.7. Recurrent types of dependent clauses

Functional type	**Example**	***n***
Comment clauses	*as it were*	23
	I should think	20
	as you know	18
	as I say	12
Indirect conditions	*if I may*	12
	if you like	11
Apposition marker	*that is to say*	11

express the speaker's view of how an utterance is to be interpreted. Some of them are part of larger paradigms, but even so there is generally some functional difference between related variants: *as you know,* for example, is truth affirmative in contrast to the neutral alternative you know (cf. Quirk et al. 1985: 1116); *I should think* is more tentative than the shorter *I think;* and, although if you like and *if I may* both act as politeness formulae, the former is restricted to offers and the latter to impositions. In addition, both have specialized functions as metalinguistic hedges, as illustrated in (5):

> (5) yeah but you've you and your {st\udents are sh\aring# a great deal alr\eady# [ə?əm] y\ou are in f/act {*if you l/ike#}#* engaged on a a c\mmon# [ə:] explor/ation# (LLC 5.7.177)

4
CLAUSE CONSTITUENTS
4. 1
Multiple clause constituents

The majority of the recurrent word-combinations in the sample do not constitute full clauses but sequences of two or more clause constituents, such as *and you know, and then I, there is a, just sort of, of course is, to do it, it at all.* The phraseological status of these sequences varies depending on the nature of the sequence, in particular its grammatical structure. Generally, closely related constituents (e.g. a verb and its complement) are more likely to be fused into phraseological units than weakly related constituents (e.g. a sentence adverbial and a subject).

On the whole, most of the clause element sequences in the sample do not form phraseological units, partly because so few of them contain lexical (open-class) words. However, most of them represent interesting clusters -- that is, highly frequent sequences of elements that can tell us a great deal about the way speakers compose utterances in discourse.

Since sequences of multiple clause constituents can represent practically any part of a clause, it is fruitful to see them in a textual perspective and divide them into broad functional categories depending on their function and position in the (unmarked) linear organization of the clause. For this purpose I shall use the positional scheme (partly inspired by Halliday 1985: ch. 3) displayed in Fig. 5.3 , which shows the most common functional slots occupied by the multiple constituents in the sample (cf. Altenberg 1990). The positions are indicated by clause-element symbols at the top of the figure and the various types of sequences have been given names suggesting

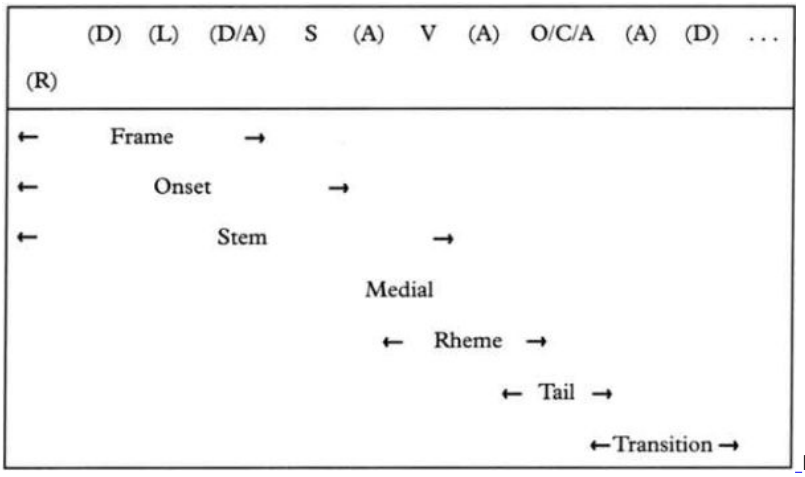


FIG. 5.3. *Linear distribution of multiple clause constituents Abbreviations:* A = adverbial, C = complement, D= discourse item, L = linking word, O = object, R = response, S = subject, V = verb element.

their textual functions. Thus 'frames' consist of elements in presubject ('first theme') position; 'onsets' consist of thematic elements, including a subject, occurring before the finite verb; 'stems' consist of subject and verb (plus any preceding thematic elements) but lack a rhematic post-verbal element, etc. These types of sequences overlap and can be regarded as successive building blocks that speakers use again and again to compose utterances in on-going discourse.

As shown in Table 5.8 , some of these building blocks are more common than others: the great majority are stems (73 per cent) but onsets, frames, and transitions are also fairly common. Together these four categories make up no less than 97 per cent of the recurrent clusters, whereas what can be regarded as the propositional core of utterances -- the 'theme' -- is seldom recurrent in the same form (2 per cent). This division clearly reflects the tendency of utterances to be produced in two parts: a thematic 'springboard' containing given information -- typically a pronominal subject -- and some framesetting element (connective, discourse marker, pragmatic particle, etc.) -- followed by the propositional core conveying new information. The former tends to be composed of items drawn from a restricted store of frequently utilized items, while the latter is composed of variable items drawn from an open set. As a result, the

TABLE 5.8. Types of multiple clause elements

Type of element	*n*	%
Frame	288	7.7
Onset	382	10.2
Stem	2, 738	73.2
Medial	32	0.9
Rheme	63	1.7
Tail	25	0.7
Transition	211	5.6
TOTAL	3, 739	100.0

springboard generally consists of highly recurrent and to some extent conventionalized clusters, while the rest of the utterance is more variable and less stereotyped (cf. Chafe 1987 and Altenberg 1990; on recurrent propositional core expressions in the London-Lund Corpus, see Altenberg 1993).

Since space does not allow a detailed presentation of all these types, I shall confine myself to two of the most frequent types: frames and stems.

4. 1.1
Frames

Frames consist of 'thematic' elements in pre-subject position. Functionally these elements serve various interactive, interpersonal, and textual functions (cf. Halliday 1985). Some common frames are shown in Table 5.9 . They typically consist of combinations of connectors (e.g. *and, but, that, because),* response items (including the initiator

TABLE 5.9. Some common frames

Example	*n*
and you know	24
well I mean	20
well of course	19
well you see	17
but you see	16
that you know	15
but you know	15
I mean you know	15
because I mean	15

well), discourse items *(you know, you see, I mean),* and modal adverbs (e.g. *of course),* which occur in a particular near-fixed order.

The frames in the material do not form phraseological units in a strict sense and they can hardly be called (lexically determined) collocations, but their frequency in the material suggests that they act as routinized sentence or clause openers in interactive speech. Structurally, we can regard them as realizations of a series of linearly ordered options, where each option offers a choice from a restricted set of items. To what extent each choice restricts the choice of the next item is unclear, but the relative frequency of the various combinations suggests the existence of certain pragmatic restrictions. For example, if the initiator *well* is chosen in the first slot (R in Fig. 5.3), it is typically followed by certain clarifying markers in the D slot:

(6) *well I mean*	20
well of course	19
well you see	17
well you know	13

Pragmatically, this seems reasonable: *well* normally signals a reservation of some kind and is most naturally reinforced by a clarifying or explanatory marker. Similarly, *but* (in the L slot) introduces an objection or contrast and is naturally followed up by a clarifying discourse marker or reinforcing contrastive connector (e.g. *on the other hand):* [1]

(7) *but you see*	16
but you know	15
but I mean	13
but on the other hand	10

4.1.2
Stems

Stems form the largest subclass of multiple clause elements. They can be described as extended onsets, involving not only a subject (and other thematic elements) but also a verb. Typically, however, they have an empty slot for the following object or complement: *there is a, this is the, and they were, and I think.* In other words, stems form the springboard of utterances leading up to the communicatively most important -- and lexically most variable -- element. [2]

[1] The association between *but* and *on the other hand* is particularly strong. Although the former cannot be said to 'select' the latter, the reverse is often true: whenever *on the other hand* is preceded by another connector, but is the preferred choice. Thus, the selection of items in these frame combinations is not necessarily a successive left-to-right phenomenon, but a clustering of items that is natural on pragmatic grounds.

[2] In this respect, the 'stems' as defined here differ from Pawley and Syder's (1983) 'sentence stems', which frequently involve full (but variable) clause-like constructions like 'NP be+TENSE

TABLE 5.10. Some common types of stems

Type	Example	*n*
Epistemic	*I think/thought (that)*	323
	I don't think (that)	53
	I suppose (that)	43
	it seems to me (that)	21
	I would have thought	10
Existential	*there*BE Det $_{indef}$	143
Reporting	*and he/I/she said (well)*	81
	so I said	16
Interrogatie	*do you know*	16
	do you think	14
	have you got	14
	do you have	11
	are you going to	10
	how do you	11
	how would you	11
	what about the	12
	what is the	11
Other	*I (don't) want to*	30
	I can't remember	15
	my name is	13
	I went to	13
	I tried to	13
	I don't mind	10
	I'm trying to	10

The most common stems in the material are normally composed of closed-class items, especially in their thematic part. This, together with the fact that they are structurally incomplete and apparently rule-governed constructions, makes it

difficult to describe any of them as lexicalized phraseological units. Yet, again, many of these stems are highly routinized expressions with various important discourse functions. Some categories are shown in Table 5.10 . As the examples indicate, some combinations appear to be more conventionalized than others: for example, the formula *my name is* (the only item with an open-class subject in the sample); the expressions *I don't mind* (always non-assertive) and *I can't remember* (usually non-

---

sorry to keep+TENSE *you waiting', 'Who (the* EXPLETIVE) do+PRES NP $_{i}$ *think* PRO $_{i}$ be+PRES'. The reason why full sentence stems of their kind do not appear in the present sample (except as complete clauses discussed in section 3) is that they are either too rare or too variable in the corpus to form continuous recurrent strings of sufficient frequency.

assertive); the reporting clause *(so/and) I/he/she said (well),* where *well* is a direct speech marker in much the same way as a colon or quotation mark in writing; the epistemic set, offering various alternatives with different nuances of meaning; the existential formula *there* BE Det $_{indef}$, which serves to introduce a new referent in discourse; and the fixed *wh*-interrogative *what about,* which is the only verbless construction in the list.

4.1.3
Speech production as linear composition or 'stitching'

As shown in Figure 5.3 , the recurrent clause element sequences in the material can be regarded as a series of overlapping and interlocking options that are utilized again and again by speakers in ongoing discourse. These sequences do not necessarily constitute phraseological units, but, if we regard them as more or less prefabricated or routinized building blocks that are at the speaker's disposal in the production of discourse, we can use them to illustrate Nattinger's (1988: 76) characterization of speech as 'basically a "compositional" process, one of "stitching together" preassembled phrases into discourse', or Sinclair's (1991: 109-15) description of text production as a continuous alternation between two modes of processing, the open-choice principle (each position offers a choice) and the dominant idiom principle (the use of semi-preconstructed phrases). If we simplify matters and concentrate on the most common clauseelement sequences in the sample (shown in Table 5.11), we can use these interlocking sequences as a simple generative model of speech production and create portions of discourse by stitching together frequently selected realizations of each sequence.

If we go by frequency alone and mechanically choose the most common word sequence available at each 'selection point', the model soon runs out of lexical material (since closed class items and very

TABLE 5.11. The most common clause element sequences

Type	Clause element sequences							*n*
Frame	L	D						108
Onset		D	S					155
Stem			S	V				845
Rheme				V	0			20
Tail					0	A		13
Transition						A	L	13

Note: A = adverbial; D = discourse item; L = linking word; O = object; S = subject; = verb element.

general lexical words predominate in the sample), producing ludicrous or impossible strings:

(8) *and you know you know I I have a to do it it at all at all and*

But if we relax the frequency condition and select more freely among the recurrent sequences, it is possible to generate more plausible strings, even with the lexically restricted material represented in the sample:

(9) *but I mean are you going to do it at all because you see I don't want to see you at the moment and you know we've got to do so you know really*

It is obvious that the combinatorial possibilities of highly recurrent word-combinations are restricted by a number of factors, lexical, grammatical, semantic, and pragmatic (cf. Altenberg 1990), but what this generative exercise suggests is that a 'stitching' model of discourse production is in fact a quite plausible, and perhaps frequently utilized, alternative to a more rule-governed mode of production.

What I hope this section has demonstrated is that there exists a large number of interlocking routinized clause-element sequences with varying discourse functions (frames, onsets, stems, transitions, etc.) that speakers make use of again and again in discourse. But such sequences are not the only building blocks available to the speaker, and to fill out the picture we must now turn to recurrent single clause constituents.

4.2
Single clause constituents

The recurrent single clause elements in the sample typically consist of a complete phrase (plus any accompanying grammatical word) functioning as a clause constituent. The categories that are of particular phraseological interest are noun phrases, prepositional phrases, and adverb phrases. In contrast, verb phrases are less common as threeword combinations, and adjective phrases are almost totally absent in the material. I will here draw attention to certain functional types, which are shown in Table 5.12 .

A lot can be said about these categories, but I will confine myself to

TABLE 5. 12. Some common types of single clause elements
TABLE 5. 12. Some common types of single clause elements

Type of element	Example	*n*
Vagueness tags	*and so on*	47
	or something like that	16
	and all that	14
	and things like that	14
	something like that	13
	sort of thing	11
	that sort of thing	10
Qualifying expressions	*more or less*	28
	in a way	15
	in a sense	11
	on the whole	10
Intensifiers/quantifiers	*the whole thing*	24
	a bit more	17
	a little bit	16
	a lot more	11
	a little more	10
	the whole lot	10
Connectors	*first of all*	17
	at any rate	15
	in other words	15
	on the other hand	14
	at the same time	10
	as a matter of fact	10
Temporal expressions	*at the moment*	43
	all the time	26
	in the past	22
	the other day	16
	in the morning	15
	at that time	14
	for the first time	12
	in the afternoon	11
	in the end	11
	at this stage	11
	for a long time	10
	for a moment	10
	in the future	10
	at the time	10
	at the same time	10
Spatial expressions	*in this country*	22
	at the back	11
	in the country	11
	in the field	11
	in the house	10
	in the world	10

three observations. First, most of these examples have adverbial functions -- that is, they serve to add various kinds of circumstantial information to the propositional core of utterances. Secondly, the fact that these categories turn up in the selected sample suggests that they serve important needs in spoken discourse. For example, real-time production leaves no room for precision, which creates a demand for various vagueness tags that can be appended to lists and imprecise descriptions; the need of speakers to be polite and express positive solidarity with their interlocutors requires various qualifying (hedging) and intensifying (downtoning or amplifying) expressions (cf. Holmes 1984); the same cooperative spirit may also explain the need for connectors to signal cohesion in discourse; finally, a host of temporal and spatial expressions emphasizes the need of speakers to anchor their utterances in time and space.

Thirdly, to serve these needs the language offers a number of more or less conventionalized expressions. Phraseologically, those listed in Table 5.12 illustrate very clearly the fuzzy boundary between fully lexicalized units and free expressions. While some are completely fixed, semantically and/or grammatically (e.g. *and so on, more or less, on the whole, as a matter of fact, in the end),* others appear to be fully rule governed (e.g. *a bit more, at this stage, at that time, in the house).* Between these extremes there are a number of expressions that are more or less flexible, permitting some reduction, expansion, or substitution (e.g. *or something [like that], and all that [stuff/nonsense], at the back [of* X], *the other day/night/week);* expressions that appear to be semantically transparent and grammatically flexible but are part of very restricted paradigms (e.g. *in the past/future, in the morning/afternoon);* and others again that seem to be flexible *(all the/this/that time, at the/ this/that time, at the/this/that moment),* but where the variant with the definite article has lost its specific (anaphoric) reference and acquired a function of its own. Hence, despite this comparative transparency and flexibility, most of the expressions in Table 5.12 represent conventionalized and preferred ways of encoding various

meanings. If we add expressions of this kind to the 'stitching' model outlined in the previous section, the productive power of the model increases enormously.

5
INCOMPLETE PHRASES

A fair number of recurrent examples in the sample consist of incomplete (simple or complex) phrases of various kinds (see Table 5.13). What is missing in these examples is generally a lexical word acting as

either the head or postmodifier of the phrase. Although incomplete phrases are less likely to form phraseological units, there is in fact a large number of expressions which must be regarded as more or less fixed. Apart from certain multiword compounds such as *point of view*

TABLE 5.13. Some frequent incomplete phrases

Example	*n*
out of the	42
a sort of	34
the sort of	30
a lot of	25
because of the	24
a couple of	20
what sort of	19
part of the	19
one of the	19
a kind of	19

and *more or less,* there is one recurrent phenomenon in particular that deserves attention -- namely, what Renouf and Sinclair (1991) have called 'collocational frameworks'. These consist of discontinuous grammatical frames such as as . . . *as, a(n) . . . of, in . . . of* enclosing lexically and grammatically related sets of words (see also EegOlofsson and Altenberg 1994). Many sequences of this kind are very frequent in discourse and some of them have fused into more or less fixed units. Some examples of this kind are listed in Table 5.14 . These sequences illustrate particularly well the scalar nature of phraseological 'fixedness' (on complex prepositions, see Quirk and Mulholland 1964). Some sequences are completely fused into multiword units of various kinds: prepositions *(as well as, in terms of),* conjunctions *(as far as),* premodifying quantifiers *(a couple of, a great deal of);* others are comparatively fixed but permit some modification (e.g. the quantifiers *a lot of, a bit of,* a number of, and the intensifier a bit of a), permutation (in favour of), or separation *(in order to).* Others are located further down the scale of fixedness, behaving more like free combinations (e.g. *as much as, the whole of, at the beginning/end of, in the middle of),* but even these have an unmistakable conventional character, especially in comparison with less common alternatives (cf. *as frequent as, the sum of, at the outset/conclusion of).*

TABLE 5.14. Some common collocational frameworks

Sequence	Example	*n*
as+ Ad + *as*	*as far as*	16
	as well as	14
	as much as	11
a(n)+ N + *of*	*an (awful) lot of*	49
	a (little) bit of	24
	a couple of	20
	a number of	13
	a great deal of	12
	a bit of a	12
the+ N + *of*	*the whole of the*	23
Prep + *the*+N+*of*	*at the end of the*	13
	at the beginning of the	11
	in the middle of	10
in+ N + *of*	*in terms of*	14
	in favour of	10
Other	*in order to*	23

6
CONCLUSION

This examination of the phraseological properties of spoken English has been based on a small sample of recurrent word-combinations from the London-LundCorpus. The material has been restricted to examples of a certain minimal length and frequency, and it goes without saying that the picture that emerges is both biased and incomplete. For example, idioms and (lexical) collocations, which tend to be either short or infrequent (or both), are very sparsely represented in the sample.

Yet, despite these limitations, it is obvious that even a small sample of recurrent word-combinations can tell us a great deal about the phraseology of speech. What is perhaps the most striking impression that emerges from the material is the pervasive and varied character of conventionalized language in spoken discourse. The use of routinized and more or less prefabricated expressions is evident at all levels of linguistic organization and affects all kinds of structures, from entire utterances operating at discourse level to smaller units acting as single words and phrases.

Another striking observation is that, at each of these levels, there

are comparatively few examples that are completely 'frozen', semantically or grammatically. Rather, the great majority of the examples occupy a position along the cline between fully lexicalized units and free constructions. In other words, they can best be described as more or less' conventionalized expressions, and they illustrate very clearly the difficulty -- or impossibility -- of making a sharp distinction between lexicon and grammar.

The material also clearly demonstrates that spoken language is exposed to various types of phraseological pressure which differ somewhat from one level to another. Conventionalization of complete sentences or clauses is mainly a pragmatic phenomenon: certain expressions are needed to convey various recurrent speech acts and discourse strategies and many of these are conventionalized by frequent use. This pragmatic pressure for routine expressions also operates at lower levels, leading to the creation of fixed 'stems' (e.g. the introduction formula *my name is . . .)* and a range of discourseoriented phrases and complex words (e.g. vagueness tags and connectors). But at the lower levels we also find a large number of conventional expressions serving various 'propositional' (semantic and grammatical) functions, ranging from temporal adverbials (e.g. *at the moment, the other day)* to compound nouns (e.g. point of view), prepositions *(as well as, in terms of)* and conjunctions (e.g. *as far as, in order to).* Thus, depending on the forces at work and the function of the conventionalized expressions, we may talk of 'pragmaticalization', 'lexicalization', and 'grammaticalization', even if these processes are not always easy to distinguish in individual cases.

At clause level, sequences of clause elements also tend to appear in recurrent 'clusters' reflecting routinized ways of unfolding and presenting information in continuous discourse. Different types of clusters can be seen as serving certain thematic needs in the linear organization of clauses, acting as interactive, textual, interpersonal, and topical frames or onsets, or as more extended stems or springboards for the theme. These clusters can be seen as interlocking building blocks of varying size and function, and, although their combinatorial possibilities are constrained by various factors, pragmatic as well as semantic and grammatical, they represent an important phraseological resource in speech production.

The picture that emerges from this small sample of recurrent wordcombinations emphasizes rather than clarifies the fuzzy character of phraseology. Speakers engaged in spontaneous interaction are in constant need of easily retrieved expressions to convey their intentions and reactions in discourse. At their disposal they have a large stock of recurrent word-combinations that are seldom completely fixed but

can be described as 'preferred' ways of saying things -- more or less conventionalized building blocks that are used as convenient routines in language production. These building blocks come in all forms and sizes, from complete utterances to short snatches of words, and they display varying degrees of flexibility. To explore this complexity and explain it in theoretical as well as practical terms is a great challenge for phraseological research in the future.

PART 3

Phraseology in Special-Purpose Languages and Foreign-Learner Language

[This page intentionally left blank.]

6

The Stylistic Potential of Phraseological Units in the Light of Genre Analysis

ROSEMARIE GLÄSER

1 INTRODUCTION

Phraseology as a subdiscipline of the linguistic system is an expanding field of research and has attracted interest from many sides. Scholarly attention has been focused on the semantic and syntactic properties of phraseological units, on various approaches to their synchronic and diachronic description, on their pragmatic function in discourse, and, quite recently, on cultural peculiarities of idioms and phrases in the light of a cross-cultural and contrastive approach. Several of these aspects are reflected in the range of topics represented in the present volume.

This chapter will focus on stylistic properties of phraseological units and their textual environment -- in short, their stylistic potential. This approach entails a number of subordinate aspects: the modification of the phraseological unit in certain contexts and its relation to a particular genre, punning with idioms in the light of intertextuality, and so on. At the outset, however, some key terms need to be explained and their related concepts defined.

2 DEFINITIONS

A 'phraseological unit' is a lexicalized, reproducible bilexemic or polylexemic word group in common use, which has relative

syntactic and semantic stability, may be idiomatized, may carry connotations, and may have an emphatic or intensifying function in a text. As the dominant subtype within this all-embracing category, an idiom is a lexicalized, reproducible word group in common use, which has syntactic and semantic stability, and may carry connotations, but whose meaning cannot be derived from the meanings of its constituents. Thus, an idiom is characterized by a specific choice and combination

of semantic components (or semantic markers) carried by its constituents, as in the following examples:

(1) a lame duck, a dog in a manger
every Tom, Dick and Harry
to eat one's words, to bark up the wrong tree
born with a silver spoon in one's mouth
shipshape and Bristol-fashion
hook, line and sinker
before you can say Jack Robinson

3 THE PHRASEOLOGICAL SYSTEM

Phraseological units constitute the 'phrasicon' of a language -- that is, the whole inventory of idioms and phrases, both word-like and sentence-like set expressions. Word-like phraseological units are 'nominations' and designate a phenomenon, an object, an action, a process or state, a property or a relationship in the outside world. They are manifest in the traditional parts of speech which are related to these conceptual entities: nouns, verbs, adjectives, and adverbs. They represent the centre of the phraseological system (in the model of centre and periphery applied to the phrasicon) and embrace idioms and non-idioms (i.e. non-idiomatic restricted collocations). Idioms form the majority and may be regarded as the prototype of the phraseological unit. Non-idioms have transparent meanings and include technical terms (terminological word groups), onymic entities (i.e. phrases which are proper names), clichés, paraphrasal verbs, and other set expressions. Examples of the latter category include:

(2) unconditional surrender, the benefit of the doubt
the Black Sea, the Golden Twenties
an eloquent silence
of paramount importance, gainfully employed
wet/drenched to the skin
beyond compare, beneath contempt

Sentence-like phraseological units are 'propositions' and designate a whole state of affairs in the outside world. Their logical structure consists of a nomination and a predication. The finite verb as part of the predicate may be absent in the case of reduction or ellipsis. Propositions form the periphery of the phraseological system. The transition area between nominations and propositions is occupied by phraseological units which thus have a dual character. These comprise: irreversible binomials, stereotyped comparisons, proverbial

sayings, fragments of proverbs, and allusions and fragments of quotations:

(3) kith and kin, wait and see
as blind as a bat, to swear like a trooper
to put the cart before the horse
a new broom, to be or not to be
a thing of beauty

Sentence-like phraseological units (or sentence idioms) include:

- PROVERBS: *Make hay while the sun shines. One swallow does not make a summer.* (All proverbs are idiomatic because in their figurative meaning they refer to a different state of affairs; most of them have an educative function.)
- COMMONPLACES: *Boys will be boys. We live and learn. It's a small world.* (Commonplaces may be trite formulae and truisms. They do not have an educative function, but rather serve as conversational fillers. As a rule, they are not idiomatic.)
- ROUTINE FORMULAE: *Come again? Mind the step. Many happy returns (of the day). Hold your horses.* (These phrases may also include idioms.)
- SLOGANS: *Value for money. Safety first.* (As a rule, slogans are selfexplanatory and therefore not idioms.)
- COMMANDMENTS AND MAXIMS: *Thou shalt not kill. Thou shalt not steal. Know thyself. Do it yourself. Be relevant. Be brief* (cf. the Gricean conversational maxims).
- QUOTATIONS AND WINGED WORDS: *Where ignorance is bliss, 'tis folly to be wise. A Jekyll and Hyde. Catch 22.*

The structure of the phraseological system and the composition of the phrasicon are summarized in Fig. 6.1.

4 *THE STATUS OF CONNOTATIONS*

As indicated above, the characteristic features of a phraseological unit are the following:

+ lexicalization
+ common usage
+ reproducibility
+ syntactic and semantic stability
+/- idiomaticity
+/- connotations
+/- expressive, emphatic, or intensifying functions in a text.

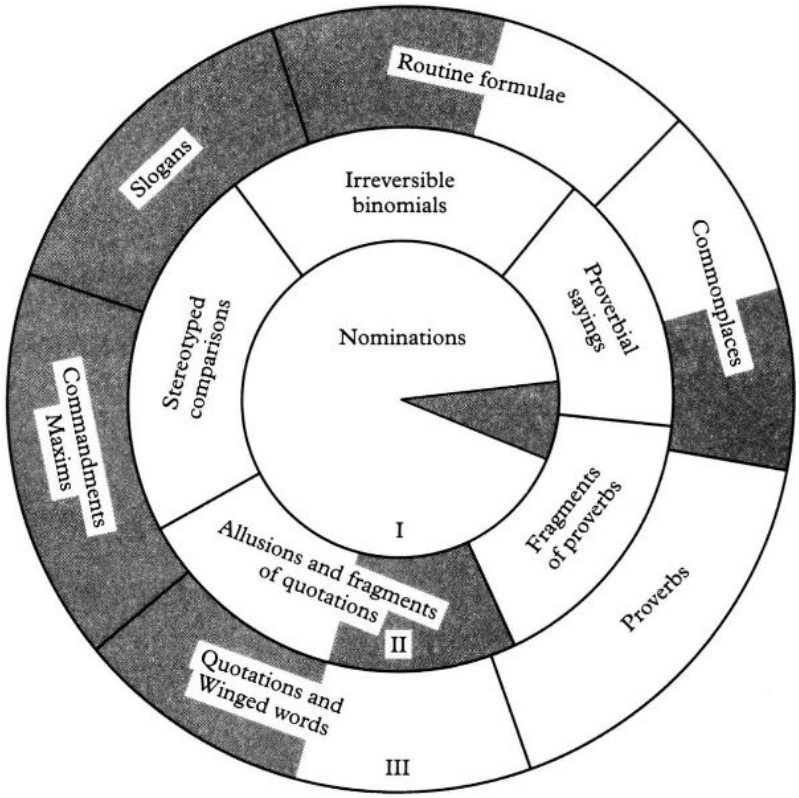


FIG. 6.1. *The phraseology system of Modern English*

The term 'connotation' calls for further comment, because it is a key to the stylistic properties of set expressions. Connotations are additional semantic markers which are associated with the value judgements of a speech community (i.e. a class or social group) or of an individual speaker or writer. Connotations are supplementary to the denotation of a word or phrase; in other words, they enrich their cognitive content by means of emotive and/or attitudinal semantic markers. With reference to phraseological units as word-group lexemes, connotations can be subdivided according to the criteria used for describing simple and complex lexemes.

Phraseological units may carry connotations which in lexicology and lexicography are described as 'usage labels' or 'style markers'. A

distinction is commonly drawn between 'expressive' and 'stylistic' connotations, and between both of these and 'register markers'.Expressive connotations include the lexicological markers 'derogatory', 'taboo', 'euphemistic', and 'jocular/humorous' or 'facetious'. Stylistic connotations cover the well-known indicators of different stylistic levels. Thus, the markers 'colloquial' and 'slang' are characteristic of stylistically lowered phrases, whereas the elevated stylistic level is represented by the markers 'formal', 'literary', 'archaic', and 'foreign' (i.e. of foreign origin). Register markers appear in the dictionary (as, for example, in the *Longman Dictionary of Contemporary English* (1995) and the *Oxford Advanced Learner's Dictionary* (1995)) as references to a particular field or province of discourse, e.g. 'med(ical)', 'phys(ics)', 'leg(al)'.These major types, with examples, are set out below:

- EXPRESSIVE CONNOTATIONS
  - derogatory: *mutton dressed as lamb; to breed like rabbits*
  - taboo: *get stuffed; son of a bitch*
  - euphemistic: *the great divide; to live in sin; of a certain age*
  - jocular/humorous: *Darby and Joan; to have a bun in the oven*
- STYLISTIC CONNOTATIONS
  - colloquial/inforrnal: *green fingers; every man Jack; full of beans; fine and dandy; before you can say lack Robinson; clear off!*
  - slang: *reach-me-downs; to kip down; on the never-never*
  - formal: *the compliments of the season; a bone of contention; gainfully employed; to be the question; under the aegis*
  - literary: *the alpha and omega; hermetically sealed; irretrievably lost; between Scylla and Charybdis*
  - archaic: *in days of yore; as it came to pass; thou shalt not kill*
  - foreign: *in casu belli; sine qua non; carte blanche; comme il faut*
- REGISTER MARKERS
  - astronomy: *black hole; red giant*
  - economics: *a high flier, idle funds; intermittent dumping*
  - judicial: *burden of proof, minister without portfolio; persona non grata*
  - medical: *corpus luteum; benign tumour, Caesarian section; pepper-and-salt fundus* (= fundus oculi)

5

VARIATIONS AND MODIFICATIONS OF PHRASEOLOGICAL UNITS

Phraseological units may vary in their stability in that they encompass constituents which allow variations within the constraints of the lexicological/phraseological system. In other words: they are 'systemic variations' of idioms and phrases. The substitutes of particular

constituents may count as 'contextual synonyms' as they have a common semantic marker (cf. Moon, this volume):

(4) to tremble/quiver like an aspen leaf
to be in accord/harmony/tune with sb./sth.
to have a flair/gift for
to tear to pieces/to shreds
as black as coal/ink/jet/midnight/pitch/soot
every now and then/again

These systemic variations of phraseological units, however, must be distinguished from 'creative modifications' of phraseological units made by the individual speaker or writer who intends a particular stylistic effect. These changes to the basic form of a phraseological unit and the associated playing with its literal and transferred meaning are always bound to a particular textual environment. Here, they can deploy their stylistic potential.Phraseological units are unevenly distributed in texts, and their occurrence depends on the text type and the given genre. Text types are understood as basic cognitive operations which are manifest in text segments and speech acts -- for example, description, narration, exposition, argumentation, and instruction (cf. Werlich 1976). These text types constitute traditional text forms, known as genres and subgenres, and stretch from fiction to non-fiction. Thus, genres cover both literary discourse and technical or scientific discourse.As a literary term, 'genre' subsumes more literary types or classes of texts than those subsumed under the classical terms epic, tragedy, lyric, comedy, and satire, but also text forms which display an elaborated literary style, such as novels and short stories, radio plays, essays, reportage, and verse tales.Another concept of genre is that relating to scientific and technical discourse. Swales (1990: 58) has suggested the following working definition: 'A genre comprises a class of communicative events, the members of which share some set of communicative purposes. These purposes are recognized by the expert members of the parent discourse community and thereby constitute the rationale for the genre. This rationale shapes the schematic structure of the discourse and influences and constrains the choice of content and style.' The genres which will be studied in the following sections with special reference to the use of idioms and phrases, include:

- popular-scientific articles as specimens of expert-to-non-expert communication (LSP discourse on the fringe of journalism);
- academic-scientific monographs and research articles as specimens of expert-to-expert communication;

- student textbooks and correspondence texts for adult learners (the Open University of Great Britain);
- texts in commercial advertising and journalism;
- novels as other genres of prose fiction (including aspects of literary translation).

These genres, which, in a different dimension, may be allocated to functional styles (or functional varieties of usage), cannot be described by an exhaustive in-depth analysis, but will be dealt with by way of sample analysis. Their stylistic properties will be discussed in pragmatic and functional communicative terms. 'Style', in this context of discourse, is understood as the choice of linguistic means made by the individual speaker or writer in accordance with the requirements of the communicative situation. Style depends on the following parameters: the author's personality and communicative intention; the time and place of discourse; the medium and province of discourse. Of no lesser importance are the textual norms which constrain the choice of linguistic means (vocabulary, sentence types, text patterns, and macrostructures).

6

METHOD OF ANALYSIS

Effective procedures for identifying the stylistic effect of an (idiomatic) phraseological unit are:

- the substitution test
- the paraphrase test
- the deletion test.

The substitution test is based on the replacement of a phraseological unit by a more or less synonymous simple or complex lexeme (i.e. a simple, compound, or derived word). The resulting version will convey the same information, but may have a different stylistic effect.

The paraphrase test is applied when the lexicon does not provide a substitute for a phraseological unit so that its denotational meaning must be restated in other words. Here again, the stylistic effect of the original version is hardly ever matched.

The deletion test may be applied to check whether a particular phraseological unit functions as a 'padding element' and is therefore redundant -- for example, a routine formula or commonplace in conversation.

Text-related methods of analysis like these, however, presuppose a representative number of informants and, to a no lesser degree, the

raising of the stylistic awareness of the test subjects. But in many cases the native speaker's linguistic competence and stylistic experience and intuition may be a clue to the stylistic potential of a phraseological unit in its communicative setting. The following observations are not derived from, and underpinned by, quantitative data, but were made during classroom work based on English texts, and similar analyses of phraseological units in German texts. In addition, several English native speakers were consulted.

7
PHRASEOLOGICAL UNITS IN TEXT GENRES
7.1
Popular scientific articles

Popular-scientific articles appear in semi-specialist journals which aim at disseminating specialist knowledge to a general audience of interested lay-persons. In contrast to academic research articles, popular-scientific articles prefer a wide variety of stylistic devices aimed at attracting and retaining the reader's attention and interest. Popular-scientific writers tend to apply linguistic and stylistic means which are well tested in journalism -- for example, colourful headings, and openings consisting of an anecdote, an episode, a recent event in a particular field of discourse, a bold personal assertion, a proverb or saying, or an allusion to the common cultural heritage. Furthermore, authors of popular-scientific articles prefer figures of speech (striking similes, metaphors, metonymies, parallelisms, rhetorical questions, antitheses, inversions, and other means of emphasis). Analogy from the reader's background experience plays an important part and provides the inductive basis for elaborating a complex problem.

The following text has been taken from the international journal *New Scientist* (8 January 1994: 3) and illustrates the writer's use of phraseological units.

> (5) *How to give science a bad name* A 59-year-old British business woman gives birth to twins on Christmas Day following a fertility treatment in an Italian clinic. A 37-year-old black woman undergoing fertility treatment in Rome opts for a white baby, allegedly to spare her child from racism. Are we *on the brink of a brave new world* of 'designer babies' and 'unnatural' post-menopausal mothers?
>
> The alarmist responses of many doctors and politicians wrongly suggest we might be. Last week European newspapers were rife with lurid references to *'Pandora's box'* and Frankenstein-style biologists playing God with motherhood . . . Swap these ages around to produce a 59-

---

> year-old father and what was previously a hot news story becomes a dead donkey.

The heading itself contains a phraseological unit which has the stylistic connotation 'colloquial', 'not formal' (*to give someone/something a bad name* is defined in the *Longman Dictionary of English Idioms (LDEI)* (1979) as: 'to harm the reputation of (a person or thing) with which one is connected because of one's bad character, behaviour, appearance, etc.'). The given heading has a stronger appeal than its non-idiomatic wording -- e.g. 'How to disparage scientific advancement'.

The phrase *brave new world* has become a catchphrase and designates-according to the *Oxford Dictionary of English Idioms (ODEI)* (1983) -- 'a new era brought about by revolutionary changes, reforms, etc.'. It has negative expressive connotations derived from the utopian novel by Aldous Huxley, *Brave New World*. In terms of intertextuality, the book's title is an allusion to Miranda's words in Shakespeare's *The Tempest*. In the present article, the phrase rouses a whole complex of associations and has more expressive power than its possible nonidiomatic substitute *an alarming/threatening future.*

The phrase *Pandora's box* is an allusion to antiquity, the Greek cultural heritage, and has literary connotations. It is absent from idiomatic dictionaries, but the *Concise Oxford Dictionary (COD)* (1995) gives the following explanation: 'the box from which the ills of mankind were released by Pandora, the first mortal woman, only Hope remaining'. Here again, we could replace the figurative phrase by a stylistically neutral expression -- e.g. *the root of all evil, the consequences of biological manipulation,* and the expressive value of the sentence would be considerably weakened. Moreover, the allusion to Frankenstein, the humanoid, blood-sucking monster-figure, would be isolated from the previously mentioned mythological context, and the whole sentence structure would be unbalanced. The phrases *hot news* and *dead donkey* are reminiscent of journalist jargon, the latter usually meaning a human-interest story.

Another example comes from a popular scientific article in *The New Scientist* (20 March 1993: 31) which deals with laser beams and their application in medicine:

> (6) The *Achilles heel* of the X-ray laser turned out to lie in how tightly the beam can be focused. Chemical, free-electron and most other lasers, the light is forced to travel back and forth between a pair of mirrors so that it is amplified and forms a tightly directed beam.

The idiomatic phrase *Achilles heel,* another allusion to ancient history and Greek antiquity, has literary connotations, and is suitable in the

---

present context. It could be substituted by a wordy circumlocution like *the weak* point (or *problematic nature) of the X-ray laser.* The alternative, however, is less pithy, and the loss of stylistic expressiveness can be felt.

7.2
Academic-scientific monographs as specimens of expert-to-expert communication

Authors of monographs, research articles, academic essays, conference papers, and scholarly book reviews in specialist journals tend to employ a wide variety of stylistic devices to give their own elaboration of a topic more prominence. Thus, in the text types of description, narration, exposition, and argumentation we may come across different figures of speech, among them metaphors and allusions to proverbs and quotations, and striking modifications of phraseological units. These may occur in genres of both the natural and social sciences but vary in the individual text. In the last instance, authors have their personal preferences regarding stylistic means.

The following extract from the monograph *Forms of Talk* by the sociolinguist Erving Goffman (1981) may serve as an

example. Goffman clearly enjoys playing with idiomatic phrases and quotations, as in the Introduction:

> (7) Thus, in talk about how individuals acted or will act, we can get by with a small repertoire of alludings and simulations. Fiction writers and stage performers extend these everyday capacities, carrying the ability to reinvoke beyond that possessed by the rest of us. But even here only sketching is found. So it remains to *microanalysts of interaction to lumber in where the self-respecting decline to tread.* A question of pinning with our ten thumbs what ought to be secured with a needle. With my own thumbs, in this volume I want to hold up three matters for consideration . . . (Goffman 1981: 2)

Goffman's dual allusion can only be decoded if the reader is familiar with English literature, but the discovery of the original source which the author has modified may give the reader intellectual pleasure. Goffman implicitly refers to the famous quotation from Alexander Pope's *Essay on Criticism* (1711) which reads: 'Fools rush in where Angels fear to tread.' This quotation has' become a so-called winged word and is to be found in most of the current dictionaries of English quotations. It means: 'Foolish people act hastily and do or say things that wiser people would avoid. Angels here represent people of wisdom' (Ridout and Witting 1967: 69). Goffman resumes the idea of rashness and clumsiness by his allusion to the proverbial saying ten

thumbs (meaning 'unskilled hands': one can say of an awkward person: 'His fingers are all thumbs').

As shown in the following example from the same source, Goffman seems fond of modifying quotations from English literature. The formulation 'social life is but a stage' refers to Shakespeare's play As You Like It (where Jacques says meditatively: 'All the world's a stage, And all the men and women merely players . . .'). Goffman writes:

> (8) So three themes, ritualization, participation framework, and embedding. It is their interplay that will be at issue. . . . In what follows, then, I make no large literary claim that *social life is but a stage,* only a small technical one: that deeply incorporated into the nature of talk are the fundamental requirements of theatricality. (Goffman 1981: 4)

In a similar way, the monograph *English for Specific Purposes* (1987), written by Tom Hutchinson and Alan Waters and intended for applied linguists and practising foreign language teachers, abounds in literary allusions. The authors have chosen a motto from English literature and world literature (including the *Bible* and Chinese philosophy) to introduce a new chapter, and they have used the structure, plot, and vocabulary of a fairy tale to illustrate the diversification of technical discourse by means of an allegory, entitled *The City of ELT* (i.e. English Language Teaching). In addition to these stylistic devices, they use idioms and phrases to make their description livelier and more convincing:

> (9) (*a*) The target situation analysis stage marked a certain *'coming of age'* for ESP. What had previously been done very much in a *piecemeal way,* was now systematized and learner's need was apparently placed at the centre of the course design process. It proved in the event to be *a false dawn.* As we see in the following chapters the concept of needs that it was based on was far too simple. (Hutchinson and Waters 1987: 12)(*b*) Conclusion: Other options . . . We have noted that there is much *common* ground between learners of apparently very different subject specialisms . . . Students and sponsors might feel that they are only getting their money's worth if they get a *tailor-made* course. Even so, a few strategic *cosmetic changes* may solve this problem. . . .For those who, in the end, feel they have to write new materials, here are a few hints:
> A. Don't *re-invent the wheel.* Use existing materials as a source for ideas.
> B. It's better to work in a team, if only to *retain your sanity.*
> C. Don't *set out* to write the perfect materials on the first draft.

> Materials can always be improved. Do what you can and *try it out.* Use what you learn from this experience and revise and expand the materials.
> E. Don't underestimate the time needed for materials writing. It can be a very time-consuming business.
> F. *Pay careful attention* to the appearance of your materials. If they look boring and scruffy, they will be treated as such.
> G. *Good luck!* (Hutchinson and Waters 1987: 125)

The authors use the metaphor *to come of age* (a phraseological unit meaning 'to become old enough to be responsible in law'). The phrase *in a piecemeal way* means 'gradually', 'piece by piece', 'one thing after the other'. A *false dawn is* a metaphor (not a proper idiom) designating 'a deceptive new beginning'. The expression *tailor-made* stands for the idea of 'appropriate (ness)'. *Cosmetic changes* is a cliché, an idiomatic phrase which refers to superficial changes which have no lasting effect and do not touch the essence of a problem. The hints for the teacher who has embarked on writing teaching material contain a proverbial saying to *re-invent the wheel* ('to do something superfluous; to spend or waste one's energy on the wrong object'). The colloquial, familiar note of address is evident in the use of the phrasal verbs *(set out, try out)* and in the routine formula *Good luck!*

7.3
Phraseological units in student textbooks

Textbook authors who write for schoolchildren, students, and adult learners (including those who attend courses at the Open University) tend to enliven their presentation of a specialist subject by a variety of stylistic devices (in particular, figures of speech) which enhance the intelligibility and memorability of a textbook unit. The following examples have been drawn from student textbooks on history, psychology, and physical geography. The history textbook *The Early Modern Age* (Snellgrove 1979) has been chiefly designed for 'pupils aged about thirteen or fourteen'. The author stresses that 'great care has been taken to keep the language clear and simple' (p. vii).

> (10) (*a*) London was in a violent and angry mood that Christmas. Armed bands roamed the streets, often with drawn swords. Mobs ran wild and rumours *spread like a forest fire [sic].* For the first time two abusive

nicknames were heard. Royalists called their opponents 'roundheads' because so many were short-haired apprentices. Parliamentary mobs shouted 'cavalier', comparing their opponents to the brutal Spanish caballeros who had slaughtered the Protestant Dutch in the days of Alva. Charles felt *the time was ripe to hit back.* (Snellgrove 1979: 150)

(*b*) Lilburne was a Leveller. This group demanded something quite extraordinary for those days -- a vote for most males over twentyone years of age. . . . Cromwell did not *see eye to* eye with them. (Snellgrove 1979: 162)

The idioms occurring in this text are everyday usage, and can be assumed to exist in the pupil's vocabulary: *to spread like a forest fire* -'to spread rapidly'; to hit back -- *'to retaliate';* to *see eye to eye* -- 'to agree'.

The textbook *Psychology -- an Outline for the Intending Student* (J. Cohen 1972) sets out 'to introduce the subject of Psychology to pupils in the upper forms of secondary schools interested in taking it up at university or in becoming familiar with its problems, methods and goals as part of what has come to be called General Studies'. The author uses occasional idioms like the following: 'We can see even *with the naked eye* that it is not uniform in structure' (i.e. 'quite easily') (J. Cohen 1972: 23). Or again: 'The actual relations of *Tom, Dick and Harry* or the behaviour of *Jack and Jill* may go down in our field notebooks . . .' (J. Cohen 1972: 147-8). Here, the phrase *Jack and Jill* refers to the well-known nursery rhyme; *Tom, Dick and Harry* means 'everyone'. A further example is:

(11) In general, friendship patterns tend to develop among people with similar personality characteristics and attitudes, a fact already enshrined in proverbial wisdom -- *' Birds of a feather flock together'.* (J. Cohen 1972: 167)

Here, the author aptly uses a proverb meaning 'people of the same character associate together', but certainly does not imply the negative connotation of 'unscrupulous attitudes'.

The textbook Processes and Patterns in Physical Geography (Keith 1983) deals with 'recent changes in the scope or nature of physical geography and caters for the changing needs of the student population in the 16-19 range . . . it covers concepts, facts and skills necessary for modern A-level syllabuses'. Examples of phraseological units include: '*Solid as a rock* is a well-used phrase, particularly by advertising copy writers, but is it the earth that is solid?' (Keith 1983: 4). Here the author has recourse to the young person's general knowledge. Or again: 'What causes its rebirth and transformation into Plate Tectonics? In a nutshell, new ideas and new facts' (Keith 1983: 7). Here, *in a nutshell* means 'in short', 'in the fewest possible words' (from the Latin phrase: *in nuce*).

Correspondence texts of the Open University -- the following examples come from teaching materials on English literature -- use phraseological units at the colloquial stylistic level, chiefly phrasal

verbs, proverbial sayings, and routine formulae, a range which indicates the author's explicit dialogue with the adult addressee:

(12) PHRASAL VERBS: to *back up* . . . his thought; to *come up* with some thoughts about . . . ; *jot down* answers; *to round off* this Introduction; *to rule out* any points that do not have a direct bearing on . . .; *'skim' through* the relevant chapters.

OTHER VERBAL PHRASES: it would be hard *to put one's finger* on any specifically socialist ideas in it; the first question *which springs to mind; tackle your work* for this course.

IRREVERSIBLE BINOMIALS: not to expect everything to be *cut-and-dried;* I am not wishing to propose a *hard and fast* dividing line; and persuasion is *part and parcel* of their enterprise; the issue of *winning and losing.*

PROVERBIAL SAYINGS: Just to *get the ball rolling;* I am going to suggest . . .; we have to take more care to *strike the right note;* Wyatt is *kicking against the pricks* of the courtly conventions of aristocratic love-making. (Gläser 1986c: 343)

A special characteristic of correspondence course materials are 'routine formulae', which authors use in their written dialogue with the adult student as an expression of personal opinion: *I feel . . ., I should think . . ., I would say It seems to me that . . ., my own conclusion is . . .*

7.4
Phraseological units in commercial advertising

Advertisements for consumer goods, technical devices, investments, and services tend to make use of idioms and phrases. The headings of such texts may have the function of advertising slogans, as is borne out by the following excerpt from a student dissertation:

(13) Big is beautiful; Blended to your taste; Britain at its best (England having the top position in a certain field); Safety first (Safety devices at work); Time is money; Too good to be true; The bottom line is excellence (the firm's basic principle is excellent quality). (Polzer 1989: 53)

Examples from a monograph by Bürli-Storz (1980) highlight deliberate ambiguity in advertising. Oil's well that ends up in an Avis car (an advertisement for an oil company) involves modification of the proverb All's Well that Ends Well; while *Familiarity breeds content* (an advertisement for Swan Vesta Matches) involves modification and semantic conversion of the proverb *Familiarity breeds contempt.* In the given context, a match, as an indispensible aid to daily living in the household, is associated with familiarity in a favourable sense. Thus, the well-known product gives the user satisfaction. In

another

example, *straight out of the flying plane into the foyer* (an advertisement for an airline company), the wording is a modification of the proverbial saying *out of the frying pan into the fire,* whose meaning ('from a bad situation to one that is worse') has been completely reversed. Emphasis is placed on the easy passage from the plane into the reception area of the airport, with air-conditioning and all modern conveniences. The phonetic pattern is of particular interest (Bürli-Storz 1980: 29).

The following example (from the *Sunday Times Magazine* of 25 September 1983: 124-5) is the entire text of an advertisement containing several idioms:

(14) You've heard of the BAKER'S DOZEN Well, Tesco Didn't Just Stop There.

Many bakers would have *called it a day* after coming up with all our white and brown breads: crusty, uncut, sliced, wheatgerm, bran and wholemeal.

Most would certainly have *rested on their laurels* after baking all of our buns (burger, bath, fruit, iced, Belgian, Chelsea), and all of our rolls (snack, finger, morning, bran and muesli).

But at Tesco we then *went on (and on)* to include crumpets, muffins, fruited batch, floured baps, sultana scones, syrup pancakes, pitta bread and croissants.

In all we've introduced well over sixty different bakery lines, and we're not *at the end of the line* yet.

Why not *give some a try?*

Though we feel we should warn you, once you've started you'll probably find it very difficult to stop.

We couldn't.

Today's TESCO

The stylistic effect of this advertisement for the bakery line originates not only from the cumulative enumeration of the whole assortment (which in fact constitutes a commercial nomenclature), but also from the skilful application of idioms. The idiom *a baker's dozen* has historical associations: 'Formerly bakers were punished if they sold loaves of bread below a lawful weight. To each dozen (12) loaves that were sold, therefore, an extra loaf was added free, to keep the weight above the lawful standard' (*LDEI*).

The verbal idioms *to call it a day* (*ODEI* (informal), 'decide or agree to stop (doing sth.), either temporarily or for good') and *to rest on one's laurels* (*LDEI*, coll., 'to be content with successes already gained and not attempt to increase them') are contextual synonyms and form a logical antithesis to the restless endeavours of the TESCO bakery line

(note *went on and on* -- a phrasal verb). The last idiom in this text is the paraphrasal verb *to give (something) a try,* another indicator of the colloquial level of the whole advertising text. The substitution test -replacing idioms by non-idiomatic vocabulary -- would result in a substantial loss of stylistic colour.

An advertisement for an insurance company, the National Provident Institution (NPI), contains a modification of the proverbial saying *to keep up with the Joneses:*

(15) *The Joneses kept up with each other until they retired.*

(The accompanying illustration shows two different couples at different garden gates with different facial expressions: No. 50 dejected vs. No. 52 amused and happy).

Numerical data: £2,732 p.a. state pension vs., £8,320 p.a. NPI pension.

The text reads:

They were earning the same salaries. With identical houses. Comparable cars. And matching lifestyles. But the similarities stopped when they stopped work.

Mr. and Mrs. Jones at number 50 found themselves struggling to scrape by.

While *the Joneses* at number 52 carried on as before, *living happier ever after.*

Not surprisingly, the difference between a retirement dream and a pensioner's nightmare is:

money.

Or, more accurately, the lack of it. . . .

Ask your broker, bank manager or other adviser about NIP.

Or clip the coupon.

After all, *which of the Joneses would you rather keep up with?*

The idiom *to keep up with the Joneses* (*LDEI*, coll., 'to compete with other people for a better social position, e.g. to buy

more or better material things than one's neighbours') and the phrase which finishes English fairy-tales, and *they lived happily ever after* (here in the modified form of the comparative), add to the stylistic effect of this advertisement.

7.5
Phraseology in prose fiction

A literary author, depicting a reality of his or her own imagination, has recourse to the whole wealth of the national standard language with its expressive means on different stylistic levels. The fictional world may be presented to the reader in a quite realistic way, when the author uses direct speech to characterize a real situation and real figures (*dramatis personae*).

(16)(*a*) *'Hello, old man,'* he said. *'Long time no see.'*

'We've been playing *Box and Cox*.' I said. 'When we came back from our holidays you went on yours.' (Braine 1957: 212)

(b)'It's *lovely to see you both again,*' she said. 'We've kept missing each other . . .'

' *Box and Cox*', Sybil said. '*Many happy returns,* Susan.' (Braine 1957: 214)

This stretch of discourse contains typical conversational formulae at a colloquial level of discourse: formulae of salutation *(hello, old man; long time no see)* and a congratulatory formula on the occasion of someone's birthday and the idiom Box and Cox (ODEI, 'from a story -- also the theme of a short Gilbert and Sullivan opera -- of two lodgers named Box and Cox who shared the same room unknown to each other, one occupying it by day, the other by night').

As an example of a very elaborated technique of the interior monologue (represented speech) which forms the artistic principle of a whole novel, I have chosen the German author Christa Wolf and her book *Kein Ort. Nirgends* (1979). Christa Wolf experiments with language and also taps the potential of idioms and phrases, as the following passages exemplify. I refer to the English translation from the German by Jan van Heurck (1982). The translator has carefully observed Christa Wolf's habits of style and has been able to offer a functional equivalent of the German phrase. The leading figure of the quoted passage is the German poet and playwright Heinrich von Kleist, who committed suicide.

(17) (*a*) Gedanken nutzen sich ab wie Münzen, die *von Hand zu Hand gehen,* oder wie Vorstellungen, die man *sich immer wieder vors innere Auge ruft.* (Wolf 1979: 11)

Thoughts *get worn out,* like coins which are *passed from hand to hand,* or like images that one *calls to mind over and over.* (Wolf 1982: 7)

(b)Er *hat die Wahl* -- falls das eine Wahl ist -- das verzehrende Ungenügen, sein bestes Teil, planvoll in sich abzutöten oder *ihm freien Laufzu lassen* und am irdischen Elend zugrunde zu gehn. Sich Zeit und Ort nach eigner Notwendigkeit zu schaffen oder *nach gewdhnlichem Zuschnitt zu* vegetieren. *Recht hübsch das.* Die Mächte, die ihn *in ihren Klauen haben* -- durch Geringschätzigkeit beleidigen sie ihn nicht. Das wird die einzige Genugtuung sein, die er in seinem Leben erf?hrt. Und er wird *sich ebenbärtig zeigen.* Kein anderer wird das *Urteil* an ihm *vollstrecken* als er selbst. Die Hand, die *schuldig werden* mußte, *vollzieht die Strafe.* Ein Schicksal*nach seinem Geschmack.* (Wolf 1979: 43-4)

He *has the choice* -- assuming that it can justly be called a choice -either to systematically annihilate in himself that consuming

dissatisfaction which is the best thing in him, or to *give free rein* to it and be destroyed by his temporary misery. To create time and space in accordance with the necessity of his own being, or simply to vegetate in a *run-of-the-mill existence.* Really, *a nice touch, that.* The Powers who *have him in their clutches* do not demean him by esteeming him lightly. This is the only compensation he will know in his life. And he is determined *to show himself worthy of it.* No other than himself will *execute judgement upon him.* The hand which was fated *to commit the crime will carry out the sentence.* A destiny *after his own heart.* (Wolf 1982: 29)

Christa Wolf's text of narrative and reflection has a remarkable density of idioms, phrases, and collocations which underpin the rich imagery of this passage. The phraseological units can be grouped into various thematic fields. One of them is characterized by dynamism (relating to Kleist's personal state of unrest, anxiety, and despair): *von Hand zu Hand gehen(to be passed from hand to hand);sich immer wieder vors innere Auge rufen(to call to mind over and over);freien Lauf lassen (to give free rein).* The second thematic field includes phrases referring to social status (Kleist could not adapt himself to the life style of Prussian feudalism): *nach gewöhnlichem Zuschnitt(a run-of-the-mill existence);recht hübsch das(a nice touch, that)* -- an old-fashioned routine formula used in polite conversation; *sich ebenbürtig erweisen(to show oneself worthy of it).* The third thematic field covers collocations of legal language. Kleist, through his administrative and clerical work in the Prussian state, was familiar with them; Christa Wolf deliberately uses these phrases to describe Kleist's stream of consciousness in interior monologue: *schuldig werden(to commit a/the crime);das Urteil vollstrecken(to execute judgement);die Strafe vollziehen(to carry out the sentence).* The latter two are contextual synonyms and support the effect of stylistic variation. The English translator has convincingly proved his understanding of the stylistic effect of the phraseological units in the source language text and succeeded in conveying a similar impressive value in the target language text. His translation amply reflects the mood of Christa Wolf's narrative.

8
CONCLUSION

The stylistic potential of phraseological units is richer and more versatile in literary and everyday genres than could be shown in the text analyses provided here. On the whole, the use of these stylistic devices depends on the individual author's intention and personal taste and thus cannot be added to the linguistic features of a particular genre or

subgenre. In general, idioms, being metaphors or metonymies, may add to the imagery of any text, ranging from advertising and instruction to academic discussion and prose fiction. Authors of scientific writing are prone to modify idioms, proverbs, and quotations for intellectual punning and sophisticated allusions. Journalists will exploit idioms and phrases in headlines and commentaries. The authors of textbooks intended for pupils and students may use phraseological units to enhance the intelligibility and memorability of a text.

The stylistic potential of the phrasicon is unchallengeable. Since phraseological units may be relevant stylistic devices, they are strong evidence of a level inside the stylistic system which has been termed 'phraseo-stylistics' (Gläser 1986b) and which combines the systemic and communicative aspects of linguo-stylistic analysis.

[This page intentionally left blank.]

7
Prefabricated Patterns in Advanced EFL Writing: Collocations and Formulae

SYLVIANE GRANGER

1
INTRODUCTION

Since the mid- 1980s, the use of prefabricated language ('prefabs') has become a major focus of interest in EFL, arguably for three main reasons. First, the emergence of the concept of lexico-grammar, inspired by Halliday and Sinclair, has promoted the syntagmatic investigation of lexis. The traditional association between syntagmatics and grammar, on the one hand, and paradigmatics and lexis, on the other, which was challenged by Michael Halliday as early as 1966, is now clearly a thing of the past. Secondly, corpus linguistics has played an important role, giving linguists the computational means to uncover and analyse lexical patterns. Abundant information about wordcombinations can now be obtained with ease using text-retrieval software. Finally, pragmatics has become a major field of study in its own right, in linguistics, and now in EFL. Pragmatic competence has come to be viewed as an essential part of learners' competence. The formulaic nature of many pragmalinguistic rules has necessarily contributed to bringing the study of prefabs to the fore.

2
WORD-COMBINATIONS IN LEARNER WRITING

Work at Louvain on word-combinations was inspired by the International Corpus of Learner English (ICLE) project, the aim of which was to gather and computerize a corpus of EFL writing from learners

I gratefully acknowledge the support given to this research by the Belgian National Scientific Research Council and the University of Louvain Research Fund. I also wish to thank Carol Edgington for assisting me in collecting and analysing the data and Stephanie Tyson for helping me to clarify my vision of the final shape of this chapter.

of various mother-tongue backgrounds (Granger 1993). Although the corpus is not yet complete, research already undertaken in the fields of lexis and discourse has demonstrated its potential to uncover factors of non-nativeness in advanced learners' writing.

The methodology employed for most of this research I have termed Contrastive Interlanguage Analysis (CIA) (Granger 1996). CIA may involve two types of comparison: a comparison of native and nonnative varieties of one and the same language: L1 vs. L2, or a comparison of several non-native varieties: L2 vs. L2. The investigations presented in this chapter are based on the former type of comparison. The initial hypothesis was that learners would make less use of prefabs, or conventionalized language, in their writing than their native-speaker counterparts, given that the use of such language is universally presented as typically native-like. I hypothesized that learners would make much greater use of what Sinclair (1987: 319) calls the 'open-choice' principle than native speakers, who have been found to operate primarily according to the 'idiom' principle. To use Kjellmer's metaphor (1991: 124), I expected the learners' building material to be individual bricks rather than prefabricated sections The data compared came from a corpus of native English writing and a similar corpus of writing by advanced French-speaking learners of English. The learner, or non-native speaker (NNS), corpus is a subcorpus of the ICLE database. The native-speaker (NS) corpus is made up of three main parts: the Louvain essay corpus, the student essay component of the International Corpus of English (ICE), and the Belles Lettres category of the Lancaster-Oslo-Bergen (LOB) corpus. [1]

TWO
TYPES OF WORD-COMBINATIONS: COLLOCATIONS AND FORMULAE

For the purposes of the investigation, I focus on two types of wordcombination: collocations and formulae. The term collocation is used to refer to 'the linguistic phenomenon whereby a given vocabulary item prefers the company of another

item rather than its "synonyms" because of constraints which are not on the level of syntax or

[1]The breakdown of the two corpora and their total number of words are given in the Appendix. Clearly, the size of the corpora used raises some questions for the study of prefabs. However, my research so far has demonstrated that, to quote Johansson (1991: 305-6), 'something may still be said . . . for smaller, carefully constructed sample corpora', and this is, in my view, especially true for learner language, which is an extremely heterogeneous variety of English.

conceptual meaning but on that of usage' (Van Roey 1990: 46). This phenomenon is illustrated by combinations such as *commit suicide, sound asleep,* or *pitched battle,* which Benson et al. (1986) call 'lexical collocations' and Aisenstadt (1979) 'restricted collocations'. [2] I contrast this type of combination with pragmatic phrases, such as *be that as it may* or *it seems* (*to me*) (*that*) **X,** which Nattinger and DeCarrico (1992) refer to as 'lexical phrases', but which are elsewhere called 'pragmatic idioms' or 'formulae' (cf. Gläser 1986a, 1988a).

3.1
Collocations
3.1.1
Collocational study of amplifiers

For the collocational study, one category of intensifying adverbs was selected: amplifiers ending in *-ly* and functioning as modifiers, such as those in examples 1-3 below:

1. *although this feeling is* ***perfectly*** *natural.*
2. *themes in Les Mouches which are very* ***closely*** *linked with . . .*
3. *a young man who is* ***deeply*** *in love.*

These constitute a particularly rich category of collocation, involving as they do a complex interplay of semantic, lexical, and stylistic restrictions and covering the whole collocational spectrum, ranging from restricted collocability--as in *bitterly* cold--to more open collocability--as in *completely different/new/free,* etc. In including adverbs such as *bitterly* in *bitterly cold* or *unbearably* in *unbearably ugly,* I have adopted a much wider notion of amplifier than linguists such as Bäcklund (1973), who rejects adverbs such as these, which express both degree and manner.

Using the text-retrieval software TACT, all the words ending in *-ly* were automatically retrieved from the NS and NNS corpora and then manually sorted according to pre-defined semantic and syntactic criteria. As a first step, the number of types and tokens in the two corpora were compared, revealing a statistically very significant underuse of amplifiers in the NNS corpus, both in the numbers of types and in the numbers of tokens (see Table 7.1).

The next step was to establish whether this underuse was general or due to underuse of particular amplifiers or categories of amplifiers. Of the individual amplifiers, only three demonstrated statistically significant differences, as shown in Table 7.2 . As can be seen, *completely*

[2]This category can be traced back to the 1940s and to the work of the Russian phraseologist Vinogradov (cf. Cowie, this volume).

TABLE7.1. Raw frequencies of amplifiers based on the NS and NNS corpora

Types/tokens	NS	NNS
Types	75	41-★★
Tokens	313	230-★★

Notes: The NS corpus is 234,514 words.
The NNS corpus is 251,318 words.
Very significant levels of overuse or underuse on the part of learners are indicated by a plus or a minus sign followed by a double asterisk.

TABLE 7.2. Raw frequencies of completely, totally, and highly in the NS and NNS corpora

Amplifier	NS	NNS
completely	15	42+★★
totally	18	46+★★
highly	31	11-★★

and *totally* were overused by the learners and *highly* was underused. On the whole, however, the frequencies of the individual amplifiers were often too low for meaningful conclusions to be drawn.

The wide variety of words with which the learners combined *completely* and *totally*--thirty-six different collocates for *completely* and thirty-four for *totally*--suggested that they are used as all-round amplifiers or 'safe bets'. Indeed, practically none of the combinations produced was felt to be unacceptable or even awkward by native speakers. One possible explanation for their overuse may well be that they have direct translation equivalents which are very frequent in French--*complètement* and *totalement*--and which display similarly few collocational restrictions. There may be an equally feasible interlingual explanation for learners' underuse of *highly,* whose literal equivalent, *hautement,* is only used in formal language and is relatively much less frequent. It is striking that the few combinations that the learners actually used--such as *highly developed/civilized/specialized/ probable*--translate very nicely into French.

When it came to examining the amplifiers by category, I chose to apply the classification of amplifiers into 'maximizers' and 'boosters' provided by Quirk et al. (1985: 590). Maximizers are amplifiers such as *absolutely, entirely, totally,* which express the highest degree; while boosters, such as deeply, strongly, highly, merely express a high degree.

TABLE 7.3. Raw frequencies of maximizers and boosters in the NS and NNS corpora

Amplifiers	Types		Tokens	
	NS	NNS	NS	NNS
Maximizers	10	10	106	150
Boosters	65	31-★★	207	80-★★
TOTAL	75	41-★★	313	230-★★

As shown in Table 7.3 , learners use the same number of types and a slightly higher number of tokens (mainly due to overuse of *completely* and *totally*) in the maximizer category, but the overall figures are not statistically significant. However, the categorization revealed an underuse of boosters by the learners significant enough to explain the general underuse of amplifiers attested to earlier.

The category of boosters represents 66 per cent of the amplifiers in the NS corpus as compared with only 35 per cent in the NNS corpus and the number of types is much higher than in the category of maximizers, understandably, given that boosters represent an openended set. Quoting examples such as *admirably fair* and *dazzlingly clear,* Bolinger (1972: 25) has pointed out that 'virtually any adverb modifying an adjective tends to have or to develop an intensifying meaning'.

By further subdividing the boosters into three categories--those that are exclusively used by the native speakers, those that are exclusively used by the foreign learners and those that are common to both groups (see Table 7.4)--I was able to reveal further differences in the use of boosters by natives and learners.

It will be noted that the majority of the NNS boosters (77.5 per cent) were used by native speakers too, while the majority of the NS boosters (63 per cent) were used exclusively by natives. Broadly

TABLE 7.4. Boosters: Types exclusive to natives or learners and types common to both

Corpora	Native speakers only	Native speakers and learners	Learners only
NS corpus	41 (63%)	24 (37%)	
NNS corpus		24 (77.5%)	7 (22.5%)

TABLE 7.5. NS and NNS collocations with closely, deeply, and severely

Boosters	NS	NNS
closely	*linked*(4)	*linked*(3)
	integrated	*involved*
	attached	*related*
deeply	*moved*	*moved*
	convinced	*convinced*
	affected	*rooted*(8)
		hurt
		in love
		changed
		divided
severely	*punished*	*punished*
	restricted	
	shaken	
	attacked	
	depleted	
	complicated	
	felt	
	flogged	

speaking, the native-exclusive combinations fell into two categories: stereotyped combinations such as *acutely aware, keenly felt, painfully clear, readily available, vitally important,* and creative combinations such as *ludicrously ineffective, monotonously uneventful, ruthlessly callous, astonishingly short.* [3] Both types of combination were significantly underused by the learners. The learner corpus, it is true, contained some rare examples of creative combinations, such as *ferociously menacing, shamelessly exploited,* but these were not always very successful: *dangerously threatened* and irretrievably *different* might seem odd to a native speaker.

Interestingly, the few stereotyped combinations used by the learners typically have a direct translation equivalent in French, or are 'lexically congruent', to use Bahns's (1993) terminology. For

[3]The following is a selection of the booster combinations used exclusively by native speakers: *acutely aware, astonishingly short, bitterly disillusioned, blatantly clear, blindingly obvious, brilliantly clever, devastatingly shocking, extensively excavated, extraordinarily painful, gravely disorganised, horribly disfigured, intensely aware, intimately bound up, irredeemably tied, irrevocably affected, keenly felt, ludicrously ineffective, mercilessly hard, monotonously uneventful, painfully clear, powerfully represented, profoundly shocked, readily available, ruthlessly callous, singularly stupid, steeply dipping, unbearably ugly, unusually small, vitally important.*

example, *closely* and its French equivalent étroitement have very similar collocational ranges, as do *deeply* and profondément. Several of the combinations used by the learners are typical combinations both in English and in French: *closely linked, closely related, deeply moved, deeply convinced, deeply rooted, deeply hurt* (see Table 7.5). The collocation *deeply rooted,* for instance, which occurs eight times in the NNS corpus, corresponds to *profonégment enraciné,* which is included as a typical combination in most French dictionaries. The case of *severely* is particularly striking. Of all the combinations used by the natives, the only one that translates into French is precisely that used by the learners--*severely punished,* which corresponds to *sévèrement puni.* All the other combinations used by the natives would be impossible in French: *sé?vèrement restreint/ébranlé/attaqué/diminué,* etc. There is also evidence that learners use non-congruent combinations, albeit comparatively few of them. In fact, there are only three obvious ones, and these are *badly injured, finely detailed,* and *widely held.*

So far, then, the investigation has supported the initial hypothesis that learners use fewer prefabs than their native-speaker counterparts. Further, there is evidence that the collocations used by the learners are for the most part congruent and may thus result from transfer from LI. But the general picture is one of learners who seem to use amplifiers more as building bricks than as parts of prefabricated sections. They tend to use some amplifiers as 'general-purpose' items, a tendency confirmed by their use of the amplifier *very,* which, although not part of the present investigation, was analysed independently. The analysis showed a highly significant overuse of *very,* the all-round amplifier *par excellence.* [4] From the figures in Table 7.6 one could postulate that the learners' underuse of *-ly* amplifiers is compensated for by their overuse of very.

7.6. Relative frequencies of -ly amplifiers and very based on 200,000 words per variety

Amplifiers	NS	NNS
-ly amplifiers	267	183-★★
very	190	329+★★

[4] In saying this, I do not disagree with Igor Mel'čk, who has pointed out (personal communication) that the use of *very* is not totally unrestricted and cites *very tired* and ★*very rested* to demonstrate this. Nevertheless, *very* combines with more adjectives than any other amplifier and can, I think, still be termed the 'all-round amplifier *par excellence'.*

3.1.2
Sigificant collocation

So far, we have established that learners are using collocations, but that they underuse native-like collocations and use atypical wordcombinations. The results of an independent study I carried out suggest that this is probably due to an underdeveloped sense of salience and of what constitutes a significant collocation. The aim of this study was to extract introspective data on collocations and involved submitting a word-combination test to 112 informants, 56 French learners of English and 56 native-speakers of English. [5] Informants were asked to choose, from a list of 15 adjectives in each case, the acceptable collocates of II amplifiers, by circling all the adjectives which in their opinion collocated with the amplifier. If they were unsure about a particular adjective, they were instructed to underline it and if they felt that one adjective was more frequently associated with the amplifier than all the others, they were requested to mark it with an asterisk.

It was the comparison of the forms marked with an asterisk by the learners and the natives, and which therefore indicated those combinations which were particularly salient in the subjects' minds, that yielded particularly interesting results. All in all, the learners marked with an asterisk over 100 fewer combinations than the natives (280 vs. 384). Table 7.7 gives clear evidence of the learners' weak sense of salience. *Readily available,* for instance, was asterisked by 43 native speakers but by a mere 8 learners. *Bitterly cold* was selected by 40 native speakers but only 7 learners. For *blissfully,* the native speaker selections were evenly distributed between *blissfully happy* and *blissfully ignorant,* asterisked by 19 and 20 informants respectively, while not one single learner marked the latter combination and only 4 selected the former.

On balance, the learners marked a greater number of types of combinations than the natives, indicating that the learners' sense of salience is not only weak, but also partly misguided. Although there was evidence of a good sense of salience among a significant number of learners for some combinations, such as *fully aware,* and *fully reliable,* the learners also considered four other combinations to be

[5] The eleven amplifiers presented were: *highly, seriously, readily, blissfully, vitally, fully, perfectly, heavily, bitterly, absolutely, utterly.* The format of the test was as follows:
readily significant reliable ill different essential aware miserable available clear happy difficult ignorant impossible cold important
bitterly significant reliable ill different essential aware miserable available clear happy difficult ignorant impossible cold important

TABLE 7.7. Native-speaker and learner responses to word-combining test

Amplifiers	Native-speaker responses	Learner responses
readily	*readily available(43)*	*readily available(8)*
bitterly	*bitterly cold(40)*	*bitterly cold(7)*
		bitterly aware(3)
		bitterly miserable(2)
blissfully	*blissfully happy(19)*	*blissfully happy(4)*

Amplifiers	Native-speaker responses	Learner responses
	blissfully ignorant(20)	
fully	*fully aware(33)*	*fully aware(21)*
	fully reliable(3)	*fully reliable(15)*
		fully different(6)
		fully significant(5)
		fully impossible(3)
		fully available(2)
highly	*highly significant(33)*	*highly significant(15)*
	highly reliable(3)	*highly reliable(7)*
	highly important(2)	*highly important(6)*
	highly aware(3)	*highly impossible(6)*
		highly difficult(5)
		highly essential(4)
		highly different(2)

significant collocations, none of which was selected by the native speakers, thus: *fully different/significant/impossible /available.* Besides selecting *highly significant,* learners also marked six other combinations with *highly,* four of which were not marked by native speakers. In fact, *highly impossible/difficult/essential/different* were together selected more often than *highly significant.* This is somewhat paradoxical when considered in the light of evidence that learners underuse *highly* in their writing, but this could perhaps be put down to the production/ reception distinction.

Apart from demonstrating that introspective data can play a role in revealing features of learner language, the study also suggests that this type of test could be valuable in providing a clearer notion of what constitutes a significant collocation. Certainly, there is a problem with using corpus data. As Clear points out: 'By far the majority of lexical items have a relative frequency in current English of less than 20 per million. The chance probability of such items occurring adjacent to

each other diminishes to less than 1 in 2,500,000,000! Reliable evidence of patterning between such items can be obtained *only from very substantial text corpora* . . .' (Clear 1993: 274; emphasis added). This statement is supported by the fact that Kjellmer's (1994) dictionary of collocations, based on the one-million-word Brown Corpus, does not contain such familiar combinations as *blissfully happy, highly significant,* or *seriously ill.*

3.2

Formulae: Sentence-builders

Within the wide spectrum of word-combinations are a number of categories which have a pragmatic rather than a syntactic function. Among these are 'routine formulae' (Gläser 1986a, 1988a), which include fixed expressions (e.g. *good morning, how are you?*) that function in discourse as greetings, apologies, etc., and more or less frozen patterns used to regulate conversation (e.g. *come again? you were saying?*). As research at Louvain focuses on learner writing, I chose to investigate formulae in writing, examining in particular the category of 'sentence-builders', phrases which function as macro-organizers in the text. This study fits in well with wider research being conducted at Louvain into coherence in learners' writing.The study is based on two discourse frames--one passive, the other active--which are used to state the discourse purpose. Both frames are instances of what Pawley and Syder (1983) call 'productive speech formulas'--that is, constructions whose lexical content is only partly specified. A precise description of the two frames is given below:

- Passive frame

  it + (modal) + passive verb (of saying/thinking) + *that-clause*

  Examples: *it is said/thought that . . . ; it can be claimed/assumed that . . .*
- Active frame

  I or *we/one/you* (generalized pronoun) + (modal) + active verb (of saying/thinking) + *that*-clause.

  Examples: *I maintain/claim that . . . ; we can seelone could say that . . .*

Every instance of the pronouns *it/I/we/you/one* followed by *that* within a span of 1-5 words was taken from the two corpora used in the collocational study and the relevant active and passive structures selected. The results are presented in Table 7.8 .

The results were most striking. While the learners made a similar use of the passive structure to the native writers--both quantitatively

TABLE 7.8. Relative firequencies of passive and active frames based on 200,000 words per variety

Frame	NS	NNS
Passive frame	77	52
Active frame		
we/you/one	56	269+★★
I	53	130+★★
TOTAL	109	399+★★

and qualitatively--they massively overused the active structure (c.400 vs. *c*.100). Some of the frequently recurring sequences in the learner corpus are listed at (4). Two of the most striking examples of overuse were sequences with *say*--used 75 times by the learners but only 4 times by the native speakers--and sequences with *think*--72 in the learner

corpus compared with only 3 in the native-speaker corpus. [6] *Notice* and *not forget* were not used at all by the native speakers. Here again the reason for the overuse may be partly interlingual. French uses many more phatic introductory phrases than English. Phrases such as *we can say* that fulfil exactly the same function as *actually* or *as a matter of fact,* which have also been found to be overused by French learners in a study of connector usage in native and non-native writing (Granger and Tyson 1996).

(4) Active frame: some recurring phrases in the learner corpus:
a. *we/one/you can/cannot/may/could/might* ***say*** *that . . .* (75 occurrences vs. 4 in NS corpus)
b. *I* ***think*** *that . . .* (72 occurrences vs. 3 in NS corpus)
c. *we/one can/could/should/may/must* ***notice*** *that . . .* (16 occurrences vs. no occurrences in NS corpus)
d. *we/one may/should/must* ***not forget*** *that . . .* (13 occurrences vs. no occurrences in NS corpus)

Clearly then, while the foreign-soundingness of learners' productions has generally been related to the *lack* of prefabs, it can also be due to an excessive use of them. Examples (5) to (7) below give evidence of the kind of verbosity this causes. In the three examples, two other lexical phrases, *the fact that* and *as far as***X** *is concerned,* are also

6Bengt Altenberg (this volume) refers to the high frequency of the epistemic stem I think that in spoken English. Arguably then, learners' overuse of this phrase may be a register-related problem.

highlighted. [7] These also have been found to be overused by learners and these too increase the impression of verbosity.
5. *Opinions are divided on this question, but AS FAR AS I AM CONCERNED* ***I truly believe that*** *this task can only be performed by each student individually.*
6. *I said unfortunately because* ***I think that*** *THE FACT THAT TV has too much importance for some has many bad consequences.*
7. *As a conclusion,* ***I would say that*** *we cannot deny THE FACT THAT university degrees are more theoretical than practical but* ***I think that*** *it is too easy to deduce that degrees are of little value.*

The use of all these phrases and frames could be viewed as instances of what Dechert (1984: 227) calls 'islands of reliability' or 'fixed anchorage points, i.e. prefabricated formulaic stretches of verbal behaviour whose linguistic and paralinguistic form and function need not be "worked upon" '. In other words, learners' repertoires for introducing arguments and points of view are very restricted and they therefore 'cling on' to certain fixed phrases and expressions which they feel confident in using.

4 *PEDAGOGICAL IMPLICATIONS*

Conscious of the importance of prefabricated patterns in language, several EFL specialists have advocated a teaching method based on the pattern of L1 acquisition which Nattinger and DeCarrico (1992: 12) represent by means of Fig. 7.1. According to this view of L1 acquisition, the child first acquires chunks and then progressively analyses the underlying patterns and generalizes them into regular syntactic rules. Willis suggests following the same pattern for secondlanguage acquisition--that is, exposing learners to the commonest patterns and then relying on 'the innate ability of learners to recreate for themselves the grammar on the basis of the language to which they are exposed' (1990: p. iii). A word of caution is necessary here. It

7In his investigation of the German subcorpus of ICLE, Lindner (1994) finds a similar overuse of *the fact that* in the English of advanced German learners. His explanation for this is that, 'Apart from interference of German *die Tatsache, das,* a missing flexibility on the part of the learners may also play a role. Their syntactic/phrasal repertoire when giving evidence for an observation is limited. Also, they may feel that expository-argumentative texts need a high degree of verbal factualness to be convincing' (1994: 37). This suggests that the overuse of this phrase and others may be partly due to transfer but also be partly a common feature of learner writing.

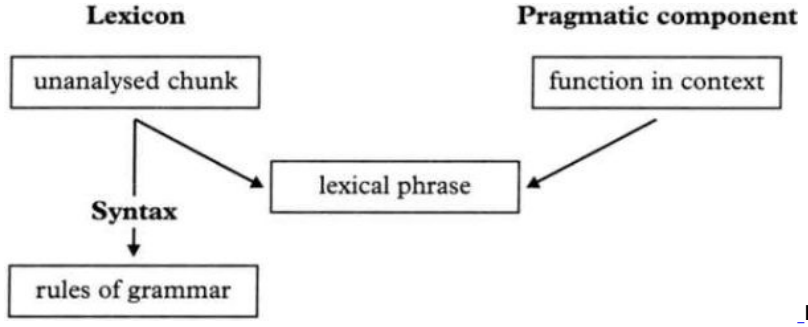


FIG. 7.1. First-language acquisition

is undoubtedly important to lay greater emphasis on prefabricated language in ELT, especially in the case of EFL learners who have very little exposure to the L2, but it seems dangerous to overemphasize the role of prefabs in SLA, as research in this area is very much in its infancy. Krashen and Scarcella (1978) and Weinert (1995) have surveyed several investigations into the part played by routine patterns in the development of syntactic structures, in both first- and secondlanguage acquisition, and it is quite clear from these surveys that the results are very inconclusive. If anything, the studies seem to indicate that the two strategies--routines and creative constructions--develop independently of each other and this view is supported by neurolinguistic evidence--that is, automatic speech has been proved to be neurologically different from creative language. Within the context of L1 acquisition, Peters (1977) has demonstrated that children use

two learning strategies: 'analytic' ("from the parts to the whole") and 'gestalt' ("from the whole to the parts") and suggests that domination of one strategy or the other will depend on individual personality and context of use.

There is very little data relating to prefabs in adult L2 acquisition. The only investigation reported by Krashen and Scarcella (1978: 295)--namely, that by Hanania and Gradman (1977)--shows that the routines used by adult L2 learners resist segmentation. In other words, at this level gestalt language fails to develop into analytic language. A more recent study by Yorio points in the same direction: 'Unlike children, they [i.e. adult L2 learners] do not appear to make extensive early use of prefabricated, formulaic language, and when they do, they do not appear to be able to use it to further their grammatical development' (1989: 68). In other words, there does not seem to be a direct line from prefabs to creative language, or, to use

Sinclair's (1987) terms, from the idiom principle to the open choice principle. It would thus be a foolhardy gamble to believe that it is enough to expose L2 learners to prefabs and the grammar will take care of itself. While research into the role of prefabs in L2 acquisition remains inconclusive, it seems wise to advise course designers not to overstress phraseological knowledge at the expense of creative skills.

Nevertheless, prefabs certainly need to play a greater role in EFL than they have in the past. The investigations presented in this chapter demonstrate that learners' phraseological skills are severely limited: they use too few native-like prefabs and too many foreignsounding ones. But if we are to devise the 'ideal' pedagogical tools, a great deal more empirical data on prefabs is required. Richards (1983: 115) considers that 'many of the conventionalized aspects of language are amenable to teaching' but he adds that 'applied linguistic effort is needed to gather fuller data on such forms (through discourse analysis and frequency counts, for example) with a view to obtaining useful information for teachers, textbook writers, and syllabus designers'. I would suggest that we need three major categories of data:

1. Detailed descriptions of English prefabricated language. The existence of computer corpora makes the compilation of collocational dictionaries possible. Kjellmer's three-volume work (1994) is the first major dictionary of this kind and makes a valuable contribution to the description of English prefabs. However, more work of this type using the new gigantic corpora is essential, if we wish to draw up lists of statistically significant collocations. As for lexical phrases, Nattinger and DeCarrico (1992: 174) stress the need for additional empirical fieldwork and Lewis (1993: 132) is equally adamant that 'A resource book of lexical phrases, including sentence heads and institutionalised utterances, should be an important priority for one of the major publishing houses'.

However, this type of data alone does not suffice. Learners clearly cannot be regarded as 'phraseologically virgin territory': they have a whole stock of prefabs in their mother tongue which will inevitably play a role--both positive and negative--in the acquisition of prefabs in the L2. The influence of L1 routines has been brought out by psycholinguistically oriented investigations of L2 speech production. In his description of learners' communication strategies, Raupach (1983: 208) notes that 'many factors that constitute a learner's fluency in his L1 are liable to occur, in one form or another, in the learner's L2 performance', while Mohle and Raupach (1989: 213)

stress the complexity of L2 processing 'where the learner's L2 procedural knowledge is activated in combination with parts of transferred L1 procedural knowledge'. It is thus necessary to have access to two other types of data: contrastive data and learner data, which will allow us to select the most useful prefabs for teaching purposes.

2. Good descriptions of prefabricated language in the learneps' mother tongues. These are necessary to assess the potential influence of the mother tongue and consequently to produce the appropriate pedagogical aids for specific mother-tongue groups. Comparisons between the different mother tongues and English will be made easier thanks to the bilingual computer corpora which are being collected today.
3. Good descriptions of learner use of prefabs. We need these descriptions as well as contrastive descriptions because not all learner problems are transfer-related. Computer-learner corpora such as ICLE which cover different language backgrounds will make it possible to distinguish the phraseological features common to several categories of learners from the L1dependent features.

5
CONCLUSION

Prefab-oriented approaches to teaching are currently, in the late 1990s, in vogue, with EFL specialists suggesting that teaching procedures be based solidly on them. Yet when we consider how little we know about them, how they are acquired, what production difficulties they cause, and how L1 and L2 prefabs interact, this is quite alarming. We possess insufficient knowledge to decide what role they should play in L2 teaching: we do not know what to teach, how much to teach, and least of all how to teach, hence the urgent need for empirical work. This should be greatly facilitated by the wide variety of large computer corpora currently being assembled. However, the value of introspective tests in this field should not be underestimated.

My own results indicate that the L1 plays an important role in the acquisition and use of prefabs in the L2. For obvious commercial reasons, most EFL material is aimed at all learners, irrespective of their mother tongue. Given the essentially language-specific nature of prefabs, this is a major issue that must be addressed if we are serious about giving learners the most efficient learning aids. Developing EFL materials from the types of data outlined above would go a long way towards solving this problem.

APPENDIX

The breakdown of the two corpora is as follows:

Native-speaker corpus (NS)

Louvain essay corpus 16,686 words: untimed argumentative essays 72,839 words: timed literature exam papers
International Corpus of English (ICE) 50,202 words: timed and untimed student essays
LOB 94,787 words: Belles Lettres and essays (categories G36-G77)

TOTAL: 234,514 words

Learner corpus (NNS)

ICLE subcorpus: French-speaking learners

164,190 words: untimed argumentative essays

24,174 words: timed argumentative essays

62,954 words: timed literature exam papers

TOTAL: 251,318 words

8
The Phraseology of Learners' Academic Writing

PETER HOWARTH

1
INTRODUCTION

This chapter describes research into the phraseological performance of non-native writers of English and, in particular, presents some of the findings of a large-scale study of collocations in academic writing (Howarth 1996). The specific university setting in which the study has been conducted has had a considerable influence on the aims of the research and the approach adopted. The focus of attention is the written production of foreign students, from a variety of language backgrounds, studying at postgraduate level at British universities. My particular interest in these students is twofold. First, I am involved in teaching on postgraduate courses in applied linguistics, evaluating native and non-native students' written performance from an academic point of view, and, secondly, as a teacher of English for Academic Purposes (EAP) I am involved in the teaching of academic writing to non-native students from a wide range of subject areas.

From this dual perspective it is apparent that, in attempting to fulfil the very demanding requirements of academic assessment, even advanced non-native writers may fail to communicate effectively their understanding of the subject matter for reasons of incomplete linguistic competence rather than academic weakness. This lack of proficiency takes many forms and can be seen at all levels of linguistic description, from word formation to the discourse structure of texts. Many difficulties correspond to traditional fields of language study (including grammar and vocabulary), which teachers, lecturers and the writers themselves recognize and understand, and which are extensively treated in grammars and dictionaries. However, there is one type of error, resulting from a lack of phraseological competence, that has received relatively little close attention from linguists. These are errors that most learners are only dimly aware of, that many

teachers can identify as a problem though cannot describe, and which linguists do not fully understand. Although they are not all major errors in themselves, and while the degree to which intelligibility is affected varies, they can have an appreciable impact on the effectiveness of a piece of writing, and their cumulative effect can be a serious loss of precision. A single example from a non-native student's essay should illustrate the point:

(1) *How much and to what extent can one accept the* ***findings reached*** *by Gardner and Lambert?*

The writer has successfully manipulated a fairly demanding structure on the grammatical level, but uses an inappropriate verb with the noun *findings.* We REACH a *conclusion* but ARRIVE AT or PRODUCE *findings.* The problem in this instance is the selection of the appropriate co-occurring items, and it is part of the much more general phenomenon of arbitrary lexical and/or grammatical restriction. Knowledge of arbitrary restriction on collocation is required in order to conform to the expectations of the academic community. The difficulty for teachers wishing to explain such errors as *reach findings* is that their cause is not simply a matter of breaking general rules of the language (though there may in addition be grammatical errors) nor of the mis-selection of individual vocabulary items.

The difficulty faced by learners has been described in this way: 'So often the patient language-learner is told by the native speaker that a particular sentence is perfectly good English . . . but that native speakers would never use it. How are we to explain such a state of affairs?' (Allerton 1984: 39) The answer is that learners do not know 'how to form the vast number of arbitrary collocations that are essential to spoken and written communication' (Benson 1985: 189).

Pawley and Syder (1983: 215, 193) place the difficulties facing learners in a wider context:

> It is a characteristic error of the language learner to assume that an element in [a linguistic] expression may be varied according to a phrase structure or transformational rule of some generality, when in fact the variation (if any) is much more restricted. . . . Only a small proportion of the total set of grammatical sentences are nativelike in form.

Recognition of the phraseological deficit experienced by learners can be traced back at least to the 1930s. Palmer (1933: 8) lists a selection of verb + object combinations (e.g. *to ask a question, to do a favour, to give trouble, or to have*

patience) and points out:

> While these are fairly regular they show the learner (what sooner or later in the course of his study he must come to know) that this particular verb may

> be followed by this particular object . . . Without such information the learner tends to form such combinations by guess work or on the analogy of his mother tongue, and we can imagine him coining such unusual expressions as
>
> To make a question
> To perform a favour
> To do trouble
> To keep patience . . .

It has been my experience that when this phenomenon is brought to their attention, EAP teachers react with immediate recognition, but are generally unaware of any descriptive framework for analysing their students' writing from this perspective. They are unable to specify what it is about the combination *do* (*someone*) *a favour* that leads learners into error (here, both a syntactic and a lexical matter) and are therefore ill equipped to provide assistance. It is part of the aim of this chapter to provide such a descriptive framework.

The objective of the research project was an empirical study of non-native academic writing aimed at identifying and analysing nonstandard phraseology. However, in order to investigate and describe non-native phraseological competence, it was first necessary to establish native-speaker norms in academic writing. The scope of the research was narrowed down in three ways. First, it focused on the language of the social sciences, since this is the broad academic area that I operate in as a subject teacher; it is the register that is most frequently drawn on by teachers of EAP; and it possibly throws up more problematic lexical and collocational features than does the language of physics or biology. Secondly, within the whole range of phraseological categories the study was limited to lexical collocations, [1] since for advanced learners these present a much more serious challenge than grammatical collocations; and, thirdly, it concentrated on verb + noun complement lexical collocations, because from a writer's point of view these represent the propositional core of the fully formed clause. Additionally, they provide the closest point of contact with other published studies.

Empirical studies of verbal collocations used by native writers and speakers have been carried out on written English by Cowie (1991, 1992), and on spoken English by Altenberg (1993). Research on nonnative phraseology includes comparative and separate studies of the

[1] 'Lexical collocations' are distinguished by Benson (1985) from 'grammatical collocations'. The former consist of two 'equal' open-class words, such as verb + noun or adjective + noun (*adopt a policy, an aquiline nose*), while the latter are combinations of noun/verb/adjective + a closed-class word, often a preposition (e.g. *an argument about*).

language of German and Polish learners of English, performing translation tasks focused on Li and L2 collocations (Gabrys-Biskup 1992; Bahns 1993; Bahns and Eldaw 1993). This second group comprises direct studies of L2 competence, in that they use as experimental data lists of specific collocations embedded in specially constructed L1 sentences to discover learners' knowledge of the L2. My research differs from these interlingual studies in two main ways. First, I am necessarily concerned with multilingual groups of learners, seldom having the opportunity to examine groups of learners with a single mother tongue, so the approach must be intralingual. Secondly, the present research examines complete texts and focuses on their phraseological features, not being concerned with a predetermined set of collocations. Any conclusions about competence arise from the empirical study of natural language performance.

2
THEORETICAL APPROACH

The framework for the collocational analysis of natural language texts is adapted from the continuum model developed by Soviet phraseologists (see Arnold 1986), and discussed by, among others, Aisenstadt (1981), Cowie (1981), and Gläser (1986*a*). It is examined in a historical perspective by Cowie in this volume. The continuum is distributed across three major categories: free combinations, restricted collocations, and idioms. The last may additionally be divided into figurative idioms and pure idioms:

(2)	*blow a trumpet*	(free combination)
	blow a fuse	(restricted collocation)
	blow your own trumpet	(figurative idiom)
	blow the gaff	(pure idiom)

The term 'conventional' is used by several writers on phraseology in language-learning (e.g. Yorio 1980; Richards 1983; Alexander 1987; Ter-Minasova 1992) and is adopted here to refer to that part of the spectrum encompassing restricted collocations and idioms.

To date, findings have been reported (in Cowie 1991, 1992) on the application of this model to verb + direct object collocations and idioms in native-speaker journalistic prose, giving quantitative results in the following ranges:

(3)	Page 1 news stories:	37.5%, 44%
	Feature article:	46%

which means that of all the verb + direct object constructions in the

texts under investigation between 37.5 per cent and 46 per cent can be classified as restricted collocations or idioms.

Broadly speaking, the present study aims to extend the application of this approach first to native-speaker academic writing, which, as far as I know, has not been analysed in this way before, and, secondly, to non-native writing in the same register. It might be thought that academic writing is produced under circumstances which would result in a much smaller proportion of conventional forms, since, as we know, academics, unlike journalists, compose their texts in tranquillity, with plenty of time for selection of *le mot juste* and opportunities to rewrite and fine-tune before publication. They are also cliché-phobes, most likely to avoid deliberately the common or familiar expression. A second objection, from applied linguists, might be that to make a study of learners' use of familiar collocations in order to compare it with native-speaker norms implies reducing language-learning to phrase-book memorization and repetition. Both these objections are countered by a crucial finding of phraseological studies: the expression of fresh or individual ideas does not entail the generation of novel word-combinations from scratch; originality of thought is not incompatible with familiarity of expression. I believe that as a by-product of this research these claims will find firm support.

3
NATIVE-SPEAKER DATA

The data for the native-speaker analysis consists of two corpora. The first comprises the social-science texts extracted from the LancasterOslo-Bergen (LOB) Corpus. These amount to twenty-nine texts, of 2,000 words each, totalling therefore about 58,000 words. The great advantage of this source is the availability of a parallel grammatically tagged version of the texts, allowing a certain amount of grammatical information to be used in the searches. However, it was not regarded as a sufficiently large quantity of material, so a second corpus was collected from texts donated on disk by staff of Leeds University. These consisted of papers on law, chapters from a book on language studies, and a complete book on social policy. These miscellaneous texts were combined into a single corpus totalling 180,000 words. The total of the two corpora therefore approaches a quarter of a million words, though they were kept separate during the stage of computer searches and analysis. The corpora were prepared as databases (named LOBSS and LUSS) for use with the text retrieval software TACT.

4
NON-NATIVE DATA

The learner data consists of ten essays (totalling about 25,000 words) written by students towards the end of the first term of a one-year, full-time masters course in Linguistics and English Language Teaching. Nine are teachers of English as a second or foreign language, while one teaches German to English-speaking children. They come from seven countries: Botswana, Germany, Greece, Hong Kong, Japan, Taiwan, and Thailand. They are, therefore, linguistically very heterogeneous, both in their first languages and in their sociolinguistic backgrounds.

An important point of difference between the analyses of the two sets of data is that, while searches of the native-speaker data were conducted on the corpora as wholes, without distinguishing between individual texts, the learner essays were analysed as separate texts (though they were additionally assembled into a single computer corpus). The justification for these different treatments was that the attempt to describe certain features of native-writer phraseological competence was based on an assumption that the forty or more authors of those texts were all mature, fully competent writers, and that shared phraseological norms could be identified from their overall combined performance. On the other hand, in order to draw conclusions about learner language, it was necessary to take into account the considerable variability between learners in their use of conventional language forms (resulting from their first language, level of proficiency, and linguistic experience). An approach that produced only an average performance among a group of non-native writers would have limited pedagogical value. For this reason, the texts were prepared in two forms: as word-processed texts, printed out for manual analysis, and as a TACT database.

5
PROCEDURE

For the reasons given, different procedures were followed when analysing the two sets of data. A brief outline is given here of stages in the two processes of analysis.

5.1
Native-speaker corpora

These were the procedures followed in analysing the native-speaker data:

- A list of verb forms was produced from the tagged LOB corpus and the verb forms lemmatized manually.
- The list of verb lexemes was reduced to those whose forms had a frequency of occurrence of 10 or more (136 lexemes).
- The same list of lexemes was then used for searches of the larger LUSS corpus.
- The verb pattern of each occurrence was noted, using the scheme in the *Oxford Advanced Learner's Dictionary* (Cowie 1989b: 155670), and those verb lexemes whose forms occurred more than 10 times in one of five transitive (T) or di-transitive (D) verb + object patterns in both corpora were selected for further analysis: Tn, Tn.p, Tn.pr, Dn.n, Dn.pr. This reduction resulted in 5,379 occurrences of 63 verb lexemes.
- The collocation identified in each occurrence was placed in one of three categories:
  free collocation
  restricted collocation

idiom (either figurative or pure)

5.2

Non-native texts

Each essay was of sufficiently manageable size for all the analysis to be done on paper. The machine-readable corpus was, however, useful for rapid checking of the original context when the final list of collocations was being analysed.

All verb + noun collocations were recorded in a word-processed file.
Verb forms were lemmatized.
Each lexemic collocation was assigned to one of the five verb patterns.
Each was then classified according to its collocational category.
Additionally, deviant forms were noted.

The analysis was conducted manually on each printed text in full (i.e. there was no sampling from the total number of verb lexemes), thus the resulting data used for collocational categorization was not intended to match directly the native-writer data. The method did not depend on there being occurrences of the same collocations in both sets of data.

The total number of collocations extracted from the three corpora for analysis exceeded 6,500: over 5,000 from the native corpora and more than 1,000 from the learner material.

6

ANALYSIS

6.1

Native-speaker phraseology

The final stage of the procedure outlined in section 5.1 involves assigning each collocation to one of the categories in the phraseological continuum from freely combining combinations to the most frozen idioms, and presents the greatest challenge to the approach. The main criteria used are:

- the degree of semantic unity of the whole expression (e.g. *take steps* = 'act'); or, where such unity is lacking:
- the degree of semantic specialization of either element in the collocation (e.g. the figurative sense of the verb *adopt* in *adopt a policy*); and, in such cases:
- restrictions on the substitutability of each element (e.g. limits on the choice of nouns that could be used appropriately with adopt in that sense).

These criteria can relatively easily identify those combinations that fall into the categories at either end of the continuum. Free collocations, for example, consist of two elements, both of which are used in their literal senses, and which permit the substitution of either element without affecting the sense of the other. In examples (4) (*a*) and (*b*) both the verbs *publish* and *put* are used in their primary senses, and the substitution of other noun objects (as in *publish a book/an article/a prospectus* or *put a comment/an address/a full stop*) makes no difference to the sense of the verb.

(4)(*a*) *the* ***programme*** *of ' Investing in Coal' which they* ***published*** *in 1956 (b) The relevant* ***questions*** *were* ***put*** *at the end of the questionnaire*

Notice, however, that there is a quite different combination *put a question,* in the sense of 'ask a question', in which the verb is used in a more specialized sense and which would be classified as a restricted collocation. This example highlights the need to study collocation on the lexemic level, rather than as the co-occurrence of surface forms.

Similarly, idioms can be quite clearly recognized by means of the well-established criterion of semantic unity. *Put a premium on sth* in (5) is regarded as having a figurative sense as a whole ('set a special value on sth'):

(5) *economic needs at the present time* ***put a premium on*** *the emergence of ability*

Of all categories of conventional lexical combinations, idioms have received the most attention in linguistic theory and description (e.g. Hockett 1958; Katz and Postal 1963; Chafe 1968; Weinreich 1969; Fraser 1970; and Makkai 1972 in the USA; Cowie and Mackin 1975; Cowie 1981; Fernando and Flavell 1981; and Cowie et al. 1983 in Britain). Idioms have also been almost the only phraseological category to be recognized in ELT materials (though consider Rudzka et al. 1981, 1985 and McCarthy 1990). However, they are the least frequent category in the type of texts under discussion here [2] (see results below), and arguably present less severe problems to learners. Far more significant is the central area of the spectrum -- restricted collocations. Furthermore, within this group of collocations the greatest descriptive challenge arises not from distinguishing the most restricted collocations from idioms (a distinction well defined in Russian phraseology), but from separating collocations that are to some degree restricted from those that are quite free. Though the borderline must remain fuzzy, it is important to establish the distinction in principle, since it is an essential aspect of linguistic competence to know when the elements of a collocation are free to recombine with other items, and when such extension is blocked.

In order to refine the analysis at this place in the continuum, a more detailed specification was devised of the criteria given earlier, based on a preliminary study of the data. In the case of verb + noun combinations, an essential criterion is that the verb is used in a specialized sense, the focus of the further subcategorization being the degree of substitution permitted of one or both lexical elements. Application of these criteria results in the subdivision of collocations at the following levels, Level 1 being the most free and Level 5 the most restricted:

Level 1 Freedom of substitution of the noun; some restriction on the choice of verb:
- an open set of nouns
- a small number of synonymous verbs
- *adopt/accept/agree to a proposal/suggestion/recommendation/convention/plan, etc.*

Level 2 Some substitution of both elements:
- a small range of nouns can be used with the verb in the given sense
- a small number of synonymous verbs
- *introduce/table/bring forward a bill/an amendment*

[2]Idiomatic grammatical collocations (*at first sight, on the other hand*) may be much more prominent in written texts (see Moon 1992a).

Level 3 Some substitution of the verb; complete restriction on the choice of noun:
- no other noun can be used with the verb in the given sense
- a small number of synonymous verbs
- *pay/take heed*

Level 4 Complete restriction on the choice of verb; some substitution of the noun:
- a small range of nouns can be used with the verb in the given sense
- there are no synonymous verbs
- *give the appearance/impression*

Level 5 Complete restriction on the choice of both elements:
- no other noun can be used with the verb in the given sense
- there are no synonymous verbs
- curry favour

A decision can then be taken as to where the cut-off point is -- for particular descriptive purposes -- between restricted and free collocations. For the present analysis, Levels 2-5 constitute the category of restricted collocations.The semantic specialization of the verb is often regarded (e.g. by Aisenstadt 1979; Cowie 1991) as having three possible manifestations:
- a figurative sense
- a delexical sense
- a technical sense.

There are clearly important issues to be addressed in defining these

TABLE 8.1. Restricted collocations

Level	Figurative	Delexical	Technical
1	assume importance	get satisfaction	
	require qualifications	give evidence of	
2	assume a role	give emphasis to	carry a motion
	follow a procedure	have a chance	consider a bill
3	bring up children	have access to	bring an action
	reach a conclusion	make an application	receive Royal Assent
4	pay attention	do one's best	obtain a warrant
	put sth.to use	take precautions	publish a bill
5		make an investment	
		have a bearing on	

three types of semantic specialization, but a few examples must suffice here to indicate the distinction. The three types are found at almost all levels of restrictedness, as Table 8.1 shows.

7

NATIVE-SPEAKER RESULTS

Table 8.2 gives the percentage figures for the two corpora separately as well as the total combined figures. The first thing to note is the broad similarity between the overall collocational complexions of the two native-writer corpora: the combined percentages of restricted collocations and idioms are 31 per cent in the LOB subcorpus and 40 per cent in the

Leeds (LUSS) corpus. Considering their very dissimilar size and composition, this similarity could be seen to add weight to the view that there is in native writing an identifiable core of collocational conventionality.Comparison with the results reported in Cowie (1991 and 1992), with figures for restrictedness in newspaper language of between 37.5 per cent and 46 per cent, suggests that the academic writing studied here displays on average less conventionality than journalistic prose, though not by a very great degree. The higher levels found in news stories might be due to the pressures of time under which they are composed and to the requirements that the writer maintain anonymity through a detached conventional style. The level of restrictedness in the academic data represents an appreciable proportion of conventional lexical combinations and the result supports the view that there is a norm of native-speaker collocational use across a range of formal registers.

7.1

Native-speaker deviation

There are very many interesting features of native-speaker academic phraseology that could be discussed, but the most significant aspect for purposes of comparison with non-native writing is the extent to which native speakers deviate from standard collocational forms, either deliberately or unintentionally. In fact the number of deviant forms is relatively very small: approximately thirty forms out of more than 5,000. Interest lies in the ways in which L1 writers deviate. Two main types of deviation have been identified in the data:

- grammatical modification
- lexical substitution.

TABLE 8.2. Native-speaker results

Verb	Free	Restricted collocation					Idiom	Total
		collocation						
		2	3	4	5	Total		
LOB (n)	1,010	174	140	70	28	412	45	1,467
1(%)	69					28	3	100
2(%)		42	34	17	7	100		
LUSS (n)	2,354	550	515	179	113	1,357	201	3,912
1(%)	60					35	5	100
2(%)		41	38	13	8	100		
TOTAL (n)	3,364	724	665	249	141	1,769	246	5,379
1(%)	62					33	5	100
2(%)		41	37	14	8	100		

Note: in each case row 1 gives percentages for the major categories, while row 2 gives percentages for the levels of restricted collocation.

7.2

Grammatical modification

Overall, there is very little deviation arising from the structural modification of standard collocations; as one would expect, native speakers seem generally aware of the grammatical restrictions associated with collocations. It is not always clear whether the modifications that are made are structural or involve the simple addition or omission of grammatical items. The distinction is not regarded as crucial here.Most of the modifications are very minor omissions or additions, which cause little disruption to comprehension and may even pass unnoticed by the non-linguist. First, here are two examples of the insertion of the indefinite article:

(6) ★*prohibiting a person to* ***take an advantage*** *of a mistake*

(7) ★*we will never* ***reach a justice*** *whose decisions are always conform to substantive truth.*

In both these cases there is a complex of irregular forms. In (6) the complementation of *prohibit* (to-infinitive rather than *from* + *-ing*) is unusual, and could be regarded as a deviant grammatical collocation (according to the classification of collocations in Benson et al. 1986). Furthermore, the presence of the article is not sanctioned by any reference work and is the only occurrence in five instances of *take* + *advantage* in both corpora. In (7) the conjunction of the article with a noun which is unequivocally uncountable is quite perverse, and this is the only case in 325 occurrences of *justice.* (There is, of course, an alternative reading, whereby *a justice* is interpreted as *a kind of justice,* which illustrates the difficulty of making judgements at this linguistic level.) Additionally, the verb *reach* strikes a discordant note which will be discussed below. [3]

Secondly, there is a grammatical collocation in which the wrong preposition is used:

> 8) ★*Such decisions as are made are likely to* ***have*** *real and lasting* ***consequences on*** *the lives of those persons involved.*

This could be seen as confusion between two prepositions or as a blend of *have consequences* for and *have effects on.*

7.3

Lexical substitution

This category spans quite a wide range of occurrences, some clearly deviant, the great majority arising from the substitution of one verb

[3]Although this is not a matter of phraseology, (7) displays another peculiarity in the use of conform as an adjective. That this is not just a slip and probably an instance of technical terminology is suggested by there being four occurrences in one (legal) text.

for the expected or conventional collocate -- as was shown in this earlier example:

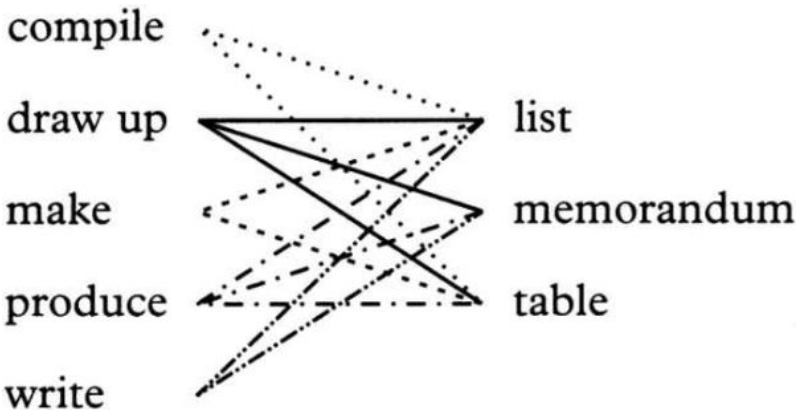
(7) ★*we will never **reach a justice** whose decisions are always conform to substantive truth.*

This combination of verb and noun appears discordant: achieve might be a more conventional choice. Other examples are:

(9) ★*however good my reasons may be for **suggesting** a **proposal** to you*
(10) ★*concerning **proposals done** by historian*
(11) ★*Our **description** of civil service English is therefore **done** by analysing letters*
(12) ★*the continued and persistent **questioning** of Pc 2, such as one other defendant **makes***

Suggest a proposal (9) is repeated three times within a few lines by the same author, which might indicate that the combination is firmly fixed in his lexicon. That the use of *suggest* is, however, unconventional is confirmed by the semantic anomaly of the combination (the tentative sense of *suggest* is contained in *proposal*). The result is to divert the reader's attention from content to form, which would not have occurred had the verb been *make, put,* or *put forward.* In (10) and (11), the problem arises from the choice of delexical verb (in both cases *make* would be acceptable). It is not surprising that such errors of selection occur, given the somewhat arbitrary way in which one delexical verb rather than another forms a conventional combination with a noun. (Consider ★*make an assurance* compared with give an *assurance.*) Example (12) is not so easy to explain, since no alternative verb suggests itself: what does one do with a *questioning?Put,* by analogy with *put a question,* does not seem appropriate. *Engage in* or *resort to* are possibilities, though they may not convey the sense required by the context.

In classifying such lexical mis-collocations, it is possible to distinguish two main categories: collocational overlaps and blends. The first is related to the type of collocational restriction classified at Level 2, above, in which a set of nouns have partially shared collocability (called 'overlapping collocability' in Cowie 1986). The following diagram shows the kind of collocational 'clusters' discovered in the data, in which certain combinations of verb and noun are acceptable (e.g. *draw up a list/memorandum/table, compile a list/table*) but other combinations are non-standard or collocationally 'blocked' (e.g. ★*compile a memorandum*):



(13)

In (14), for example, the error can be seen as the result of filling in a collocational gap within a partially overlapping cluster:

(14) ★*The **contrast** is **drawn** again a few pages later . . .*
★DRAW a contrast
DRAW a distinction
MAKE a contrast
MAKE a distinction

In blends, on the other hand, there exist two distinct collocations with no lexical overlap. While the nouns are semantically similar, neither verb collocates with both:

(15) ★*and we can **pay** particular **care** . . . to look at the fortunes of United Kingdom trade*
PAY attention
TAKE care
(16) ★*appropriate **policy** to be **taken** with regard to such inspections*
TAKE steps
ADOPT a policy

The distinction between the two categories cannot be rigidly maintained, as there are cases where either analysis is possible, depending on acceptability judgements of potential combinations such as ?ATTACH *emphasis to sth:*

(17) ★(4.1) ***places** overmuch **weight on** the influence of average employment*
PLACE weight on
PLACE emphasis on
ATTACH weight to
?ATTACH emphasis to

The alternatives provided here are not intended to represent the only ones possible, but to offer plausible explanations of how the nonstandard forms came about. The variants are all attested in either *SEC* (Kozlowska and Dzierżanowska 1988) or *BBI* (Benson et al. 1986), two specialist collocational dictionaries. In a few of the deviant collocations involving verbs

and nouns in transitive verb + object relationships it is hard to say which has been substituted for which. It

is generally considered that the noun is the base of a verb + noun collocation (Hausmann 1979) and that the direction of determination is from the noun to the verb, which seems to hold good in (15) and (16), where the 'wrong' verb has been produced to collocate with the noun previously selected. In (14) and (17), however, the verb might be the dominant element in the lexical selection and it is the noun that is wrongly chosen. There must always be some doubt about the order in which items are selected, especially when there is a close semantic connection between the two nouns (e.g. *contrast* and *distinction*).

The significance of blends for native speakers is that intelligibility is generally maintained. Though PAY *care* is lexically deviant, there is no problem of comprehension, at least for native speakers, since they have the 'originals' to refer to. That blends can be understood is evidence of shared phraseological competence, and it may be that many pass unnoticed as a result of the way in which semantically similar lexical complexes are stored nearby in the mental lexicon. Blends can also perhaps be seen as the native-speaker intralingual equivalent of non-native interlingual interference, with two target forms competing for selection.

Intelligibility may also be maintained in the case of more radically deviant forms, which depend on the lexical co-occurrence to overcome the structural oddities. Consider:

(18) ★*the guides take a hardline criterion against the birth parents* TAKE a hard line against APPLY a criterion to

The analysis forced on the reader results in the collocations ★TAKE *a criterion against sb* and ★*hardline criterion,* for which few native speakers, I imagine, would have entries in their phrasicon. It is not a simple blend of *take a hard line against sb* and *apply a criterion,* because hardline functions as a modifier, not as a head noun. Nevertheless, most readers can derive sense indirectly from the whole by means of the co-location of the three (or four) lexemes: *take, hard(line),* and criterion. This could be regarded as justification for ignoring the structural relations between the collocates and treating them as cooccurring word-forms. However, it would appear that this is itself a mark of the deviance: the surface lexical information is sufficient in this case to overcome the breakdown in the expected collocational structure, but this is an abnormal processing effort. This phenomenon recalls Aitchison's (1987: 3) answer to the question: 'Are humans primarily good at remembering, or at working things out?' She concludes (1987: 14) that 'they start by using memory, and routine possibilities. If this proves inadequate, they turn to computation.'

7.4
Analysis

The chief effect of many instances of collocational variation, whether intentional or not, is to force the decomposition or analysis of a normally unanalysed complex. Analysis results in one element or more in a composite being encountered by the listener or reader as an item to be retrieved from the lexicon and comprehended separately, as if it were a free combination with an entirely compositional meaning. The analysed meaning of the detached element clashes with the meaning of the target collocate, causing disruption in the flow of language. For example, in ★*we can* ***pay*** *particular care,* the presence of *pay* rather than the expected, though not necessarily the intended, *take* momentarily compels the reader to 'look up' the meaning of *pay* in this lexical context, finding it incompatible with *care. Pay attention* would set off no such process.

The conclusion to be drawn from the data above is that, with fewer than 1 per cent of collocations producing any disruptive effect, the academic register represented in these corpora is highly conventional, designed to be processed with the minimum of attention drawn to any individual style of the writer. What deviation there is appears to be the result chiefly of unintended blends between two similar collocations.

8
NON-NATIVE-SPEAKER RESULTS

Having established the phraseological norms that we assume learners are aiming at, we now turn to the evidence of non-native performance. Looking first at the overall distribution of learner writing across the categories of collocational type (as shown in Table 8.3 , where the results from the ten texts are added together), we find that the equivalent percentage of conventional collocations is 25 per cent, compared with 38 per cent for native-speakers. (X is the category of deviant collocations.) It could therefore be said that native speakers employ about 50 per cent more restricted collocations and idioms (of a particular structural pattern) than learners do, on average.

These learner figures, of course, disguise the extent to which the distribution of collocational categories varies between individuals: Table 8.4 shows the range for each of the categories. There may be ways in which individual learners differ systematically in their collocational performance, revealing underlying levels of competence. However, in practice it is very hard to see any pattern that connects collocational use with any other factor, such as linguistic proficiency (as measured by the Leeds University English Language Test, which

TABLE 8.3. Non-native-speaker results

Note. In each case, row 1 gives percentages for the restricted collocations, while row 2 gives percentages for the levels of collocation. 'X' denotes collocational errors.

Verb	Free collocation	Restricted collocations				Idiom	X	Total
		2	3	4	5	Total		
NNS (n)	857	185	63	26	22	296	12	79	1,244
1(%)	69					24	1	6	100
2(%)		63	21	9	7	100			
NS (n)	3,364	724	655	249	141	1,769	246		5,379
1(%)	63					33	5	≪1	100
2(%)		41	37	14	8	100			

TABLE 8.4. Range of use of collocational categories (%)

Categories	Range
Free collocations	58-74
Restricted collocations	16-33
Level 1	0-19
Level 2	44-70
Level 3	9-38
Level 4	0-21
Level 5	3-26
Idioms	0-4
Deviant	0-13

all foreign students are required to take on admission), academic grade at the end of the course (which correlates very highly with language proficiency), or whether English is spoken as a second or as a foreign language. Statistical tests of correlation by rank order revealed no significant result. Even the percentage of collocations in which there is some grammatical inaccuracy has little correlation with levels of proficiency.It is clear that one needs to look more closely at particular instances of non-standard forms to gain an understanding of underlying processes. The following analysis uses the same categories of NNS phraseological deviation as were uncovered in the NS data. (The examples below include some free combinations as well as conventional collocations.)

8.1

Grammatical modification

The importance of a grammatical perspective on the study of nonnative phraseology is in enabling us to distinguish grammatical error from phraseological deviation, and errors within grammatical collocations from those in lexical collocations. As would be expected even at an advanced level of proficiency, a variety of grammatical errors are found within the verb patterns under investigation:

(19) ★*decline the effectiveness*
(20) ★*respond students' need*
(21) ★achieve to gain

Since this study does not take a comprehensive approach to error analysis, grammatical errors such as these are ignored in the description.

Although (19) seems to reflect confusion between a transitive and an intransitive construction (cf. *effectiveness declined*), the other examples involve errors in forming grammatical collocations -- namely, the omission of a preposition (*respond **to** students' need*), and the wrong choice of verb complementation (cf. *achieve **a** gain*). Cases of this kind are not analysed further (unless they illustrate an additional lexical problem), and their deviance is ignored in the numerical analysis. It should be noted, incidentally, that incorrect grammatical collocations account for little more than 1 per cent of the total number of errors.There are, of course, cases in which, within a verb + direct object structure, the morpho-syntactic form is fixed (for example, number in the noun phrase), and where learner writers seem unaware of the restriction. In (22) one cannot separate the grammatical and phraseological characteristics of the combination:

(22) ★*losing his **tie** with his own culture group (**ties**)*

An additional feature of the grammar of learners' writing that is relevant to its phraseology is the extension of a (possibly incorrect) collocation to include a word that is morphologically related to a word that should be present, or to include a modifier to such a word:

(23) ★*DRAW a conclusive comment*
(24) ★*reach a high achievement*

In (23) the familiar relationship between DRAW and *conclusion* has been transferred to the modifying adjective *conclusive,* suggesting that the collocation has been learnt without sufficient knowledge of how its collocability is constrained by structural limitations. In (24) there is an echo of a *high achiever* or *a high level of achievement.*

8.2

Lexical substitution

This category represents about 6 per cent of the total number of nonnative collocations analysed, or approximately eighty forms. While many of these also include grammatical inaccuracies, the focus is on the lexical material. One of the most telling findings (as was indicated in section 8) is the lack of correlation between general proficiency and the number of deviant collocations. In an earlier small-scale pilot study (Howarth 1993) it was suggested that the more proficient learners made more collocational errors than the less proficient, since they had the confidence to experiment and take risks with combinations. In the larger study this observation does not seem to apply and collocational dysfunction appears as a highly individual variable.

There are a number of instances of combinations which are clearly nonce forms and which additionally fail to communicate a clear meaning as a result of the unnaturalness of the lexical co-occurrence. They can be regarded as not only non-institutionalized but also not semantically rule governed. The anomaly can generally be accounted for as semantic deviance, rather than the result of substituting an unconventional collocate for a conventional one. In these cases a greater amount of context is required to confirm the anomaly:

(25) ★NEED a factor *We know if we want to become an expert at learning language or else, we **need a** strong powerful **factor** to support us to achieve this goal.*
(26) ★ACCOMPLISH interest *Yet, he strongly believes that it is through the positive interaction between learners and instructors that a genuine **interest** in learning a second language can be **accomplished.***

The great majority of deviant combinations can be understood in terms of the semantic specialization of the verb and the

restrictions on the potential for the verb in that sense to combine with a range of nouns. Of the three categories of specialization, technical senses pose the least problem (though they are also by far the least frequent), suggesting, perhaps, the greater degree of lexicalization and therefore familiarity of such collocations as *acquire a language.* Delexical senses are the second most common category, though they are found much less frequently than in the NS data (13 per cent against 21 per cent of all collocations of the patterns studied). This may be the result of the avoidance by some learners of such combinations due to uncertainty over appropriate collocability. This is not surprising given that even native speakers are capable of confusing delexical verbs (for example, *make an assurance for give an assurance*). Although there is occasional doubt over the 'correct' form (views might differ on the acceptability of *do a* study, for example), there are some clear cases of direct confusion between two delexical verbs:

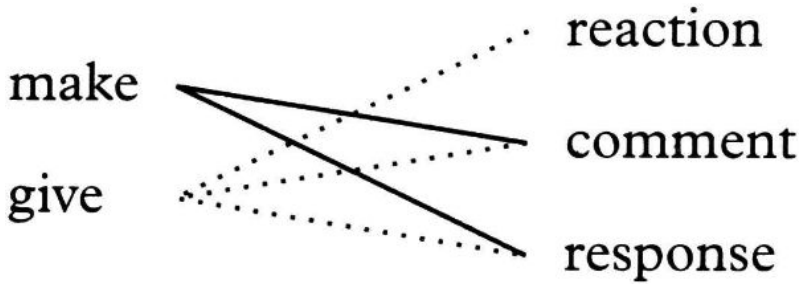
(27) ★*Do attempts* (MAKE)
(28) ★*Do a measurement* (MAKE)
(29) ★*GET contact with* (*either* MAKE contact with *or* GET IN contact with)
(30) ★*MAKE a reaction* (GIVE)

The third and largest category involves the figurative use of verbs, and, as with the NS data, the interrelationship between figurative sense and restricted commutability can be analysed in terms of collocational overlap and blends.

It was shown above that an important category of non-standard

collocations produced by native speakers were identifiable as blends (e.g. *pay care*), which could be regarded as slips of the pen or performance errors, since underlying competence could be assumed. In analysing the production of non-native writers, there is much less certainty concerning the competence that they are drawing on -whether they 'really know' the assumed target collocation or whether the deviant form is the result of incomplete knowledge of an associated set of collocations. The following example could be analysed either as a blend of two collocations that have a semantic similarity in their noun element or as the result of gaps in an overlapping cluster of collocations:

(31) ★*MAKE a reaction*
(*a*) a blend of
GIVE a reaction and
MAKE a response
(*b*) partial overlap within the following cluster



There can be no definitive choice between the two interpretations. It is not sufficient in the case of non-native writers to identify a pair of collocations assumed to be known that result in a blend, since there is no way of knowing what the state of each writer's competence is. At the other extreme, a list of all the potential collocates in an associated cluster might be endless in view of the much looser way in which lexical items are seen to collocate in the examples in (25) and (26) above. The distinction between overlaps and blends must therefore be seen as a continuum, and the following sets of examples are the result of a pragmatic attempt to draw the line for descriptive purposes. The criterion employed is the degree of semantic similarity between the nouns in the underlying collocations. If they are close in meaning, it suggests that there is a short diversion from the normal process of selecting a collocate, caused by the proximity of the two nouns in the mental lexicon.In the following examples the nouns that link the erroneous verb and the target verb are considered to be semantically too dissimilar to be easily confused in the lexicon and there exists a verb which collocates with two of the nouns in the hypothetical set.

(32) ***Attaching*** *such a key* ***role*** *to English Language not only gives pressure to students but also . . .*
★ATTACH a role to

ATTACH a value to
ASSIGN a role/a value to
GIVE a role/a value to

(33) *. . . a means that is cast aside when the* ***reward*** *is* ***attained***
★ATTAIN a reward
ATTAIN a goal/an objective/a target
GAIN an objective/a prize
RECEIVE a prize/a reward

(34) *. . .* ***attempts*** *and researches have been* ***done*** *by psychologist [sic] to find . . .*
★DO an attempt
DO a study
MAKE an attempt/a study

(35) *Those learners usually* ***pay*** *more* ***efforts*** *in adopting a new language by getting more contact . . .*
★PAY effort
PAY attention/a call
MAKE a call/an effort

(36) *. . . the teacher forced students to memorize the **dialogue** inorder [sic] to make them **play it** in front of the other students*

★PLAY a dialogue
PLAY a part
ACT out a dialogue/a part

(37) *The learners may suffer difficulty in communicating in the target language*

★SUFFER difficulty
SUFFER pain
EXPERIENCE difficulty/pain

(38) *. . . he would not take extra efforts to overcome the unfavourable . . . conditions*

★TAKE efforts
TAKE a decision
MAKE a decision/an effort

In some cases the origins of the deviance are much harder to trace. Although the figurative sense of the verb is familiar in an academic context, there does not seem to be a clear semantic link with established collocations. These examples can be assigned to the category of semantic anomalies discussed above:

(39) *Finally, a conclusive **comment** will be **drawn** on motivation and language learning . . .*

★DRAW a comment
DRAW a conclusion
MAKE a comment

(40) *The role of examinations . . . have normally **exerted** a strong **motive** for students*

★EXERT a motive
EXERT pressure
ASSIGN a motive

Blends are proportionately fewer in the NNS data. In the following examples the nouns in the collocations offered as originals of the blend are taken to be semantically close and there is no verb that collocates with both nouns.

(41) *If the students can **achieve** all the **tasks** with no difficulty . . .*

★ACHIEVE tasks
ACHIEVE goals
PERFORM tasks

(42) *We have drawn a favourable correlation between motivation and proficiency in language . . .*

★DRAW a correlation
DRAW a comparison
MAKE a correlation

(43) *. . . **emphasis** will be **given** on the individual learners in relation to second language*

★GIVE emphasis on
GIVE weight to
PLACE emphasis on

(44) *Spaulding **places** more **importance** on the latter kind of motivation . . .*

★PLACE importance on
PLACE emphasis on
ATTACH importance to

A final example in this section illustrates the complexity of lexical cooccurrence relations:

(45) *The concept of expectancy and value refers to **the level of task difficulty** a learner chooses to **perform** after having experienced success or failure in a previous task*

A syntactic analysis would identify the collocation *perform the level* (*of task difficulty*) -- which appears unnatural. Three possible interpretations suggest themselves:

- The deviance could be resolved by altering the verb to *perform at* (*the level*)
- It could be regarded as a blend of PERFORM *a task* REACH *a level*
- Or, more likely, the presence of task in the post-modifying phrase

has had a strong, possibly subconscious, determining influence on the choice of verb.

It is not a coincidence that this extract is taken from the writing of a student who has most nearly achieved native-speaker competence, though not brought up as a bilingual. The example most clearly resembles the instance of native-speaker deviation discussed earlier:

(18) *the guides **take** a hardline **criterion against** the birth parents*

which was taken as a blend of *take a hard line against* and *apply a criterion to,* with the verb influenced by the pre-modifying adjective.

In their detailed study of blends in learner essay-writing, Dechert and Lennon (1989) place such phenomena in the context of Tarone's Continuum Paradigm of variable competence (1983). This suggests that learners' interlanguage performance will depend on the kind of linguistic task assigned, and a range of linguistic tasks performed by learners will reveal the variety of styles in the learners' interlanguage repertoire, from 'vernacular' to 'careful'. The more careful the style, the greater the amount of variability in the (conscious) application of rules and the greater the proportion of both target L2 and Li forms; the more vernacular style, found in automatic unattended speech, is more stable and represents unmonitored competence. Essay elicitation tasks of the kind administered by Dechert and Lennon result in a careful style, in which learners' attention is focused on linguistic form, hence the much greater prevalence of blends in their written data than in casual speech. They conclude by asking 'why was it so full of collocational blends? . . . why, in spite of their many

years of learning the language plus, in some cases, months of exposure to native speakers in England, had their apprehension of collocational affinities between lexical items not reached a state of proceduralized automaticity?' (1989: 165-6). It is proposed here that blends are a normal part of native-speaker performance (rather than examples of aphasia, as Dechert and Lennon hint) and a sign of competence rather than incompetence. One can only produce blends of collocations that are known. If this view is accepted, blending in learners' production is not so mysterious, and, as in the native-speaker data, blends do not necessarily impair comprehension.

9
CONCLUSION

In certain registers, the reader's attention can be deflected from message content to linguistic form by the conscious manipulation of standard collocations: newspaper headlines, feature articles, and

advertisements are examples of discourse types where such changes are commonly made for deliberate effect. In academic discourse, however, the native writer feels under pressure to conform to an impersonal style and to keep the individual personality suppressed. Adherence to the conventional allows message content to be clearly perceived; linguistic divergence permits the personality to intrude.

A much greater incidence of non-standard phraseology is, of course, found in non-native writing, reflecting the learner's general lack of awareness of the phenomenon. It must, however, be recognized that one of the main reasons why the great majority of learners do not reach this state of awareness is that teachers of EFL themselves, both native-speaker and non-native, have little understanding of the phraseological mechanisms of the language. Teaching for the most part leads learners to believe that, in producing stretches of English, they combine items of vocabulary according to their knowledge of syntactic structure. As proficiency develops, those learners additionally come into contact with and memorize an increasing number of idioms, which are learned as fully lexicalized fixed expressions to be used when appropriate. The lexical resources of the language thus appear to them to be divided into two categories: free combinations and idioms, and at an advanced level of proficiency learners are able to control the two extreme ends of the spectrum. From the evidence of NNS collocational deviation, however, it would seem that many learners fail to understand the existence of the central area of the phraseological spectrum between free combinations and idioms. It is in handling restricted collocations that errors of both a lexical and grammatical nature constantly occur.

Moreover, learners need to understand that restricted collocations make up a significant part of a typical native speaker's production in both speech and writing. Conforming to the native stylistic norms for a particular register entails not only making appropriate grammatical and lexical choices but also selecting conventional collocations to an appropriate extent. Such combinations are not optional stylistic adornments on the surface of text; they are essential for effective communication, and their use by non-native writers is a clear sign that these learners have made an essential adjustment to the academic culture they are entering.

PART 4
Phraseology and the Dictionary

[This page intentionally left blank.]

9
Discovering Significant Lexical Functions in Dictionary Entries

THIERRY FONTENELLE

1
INTRODUCTION

The generative tradition of the 1960s and 1970s in linguistics tended always to neglect the role of the lexicon in the description of language. The advent of computer technology, however, enabled linguists to test their hypotheses and intuitions and it soon became apparent that the development of a whole range of natural language processing (NLP) systems required a great deal more information than was needed to parse a lexically simple sentence such as *John loves Mary*. Developers were quickly faced by what has come to be known as the 'lexical acquisition bottleneck' (Wilks et al. 1989). The problem was very simple: in order to develop a large-scale NLP system -- for example, a machine translation system or an information-retrieval system-one needed to feed the lexical component of the system with the description of tens of thousands of lexical items. The source of these descriptions, however, was rather problematic. Should the developer hire a team of highly specialized lexicographers to code the whole lexicon from scratch or should other resources be tapped to reduce the costs? A number of researchers in computational linguistics were and are convinced that the former solution is impracticable because it is too time-consuming and because lexicographical skills for computational applications cannot readily be found. The latter solution -- that is, exploiting existing resources to build the lexicon -then seems a more promising approach. A distinction is generally

The work reported here was carried out at the University of Liège in part within the framework of the DECIDE

project (*Designing and Evaluating Extraction Tools for Collocations in Dictionaries and Corpora* -- MLAP 93/19), which was partly financed by the European Commission (Multilingual Action Plan). I wish to thank Jacques Jansen, computer engineer at the University of Liège, who was responsible for the computing aspects of the research reported here. Nothing would have been possible without his invaluable contribution. I also wish to thank Luc Alexandre, who joined the project at a later stage, and wrote some of the retrieval programs used in this work.

drawn between two main types of resources -- namely, existing dictionaries and lexicons, on the one hand, and corpora, or huge collections of machine-readable texts, on the other. Here again, the research community is divided and not everyone is convinced that traditional commercial dictionaries can be considered a reliable source of information for NLP. (Atkins 1991, and Atkins and Levin 1991, point to a number of shortcomings in an approach based exclusively on dictionary analysis.) Other researchers have tried to exploit the machine-readable versions of commercial dictionaries, hoping to find in them a substantial part of the lexical information required by NLP systems, and extracting automatically or semiautomatically the semantic, syntactic, and pragmatic information discovered in the microstructure of dictionary entries. Work by Michiels (1982), Byrd (1989), Byrd et al. (1987), Boguraev and Briscoe (1989), and Wilks *et al.* (1989) is particularly relevant here, since all these researchers have also shown that the lexical acquisition process requires a great deal of preparatory work and demands a thorough study of the coding practices and lexicographic devices used by the dictionary compilers.

Corpora are also commonly used as a source of information for populating the lexicon of an NLP system. Some researchers would even argue that they should be the sole type of resource employed, on the grounds that naturally occurring texts, which are now available in unprecedented quantities, can yield a deep linguistic description based on objective facts, unlike many existing dictionaries, which are still based on the intuitions of the compilers. (Note, however, that, in the UK, monolingual dictionaries for the foreign learner have, since the 1980s, drawn on corpus material, a lead more recently followed by bilingual dictionaries.)

This chapter refers to work carried out mainly on machine-readable dictionaries, and specifically on a bilingual dictionary, namely the Collins-Robert English-French French-English Dictionary (Atkins and Duval 1978, 2nd edn. 1987), henceforth CR. I wish to show that the machine-readable version of this dictionary can be used as a starting point for the construction of a bilingual collocational database. In the following sections, I will illustrate the extraction procedure with various examples. I will also address various problems that are posed by the construction of this database. Before tackling these problems, however, it is essential to define the concept of 'collocation' as used in our project and to give a short overview of current research in the acquisition of co-occurrence knowledge. I will also briefly describe the approach chosen for the Liège project, based on Mel'čuk's concept of lexical function.

2
DEFINING COLLOCATIONS

There does not seem to be any clear-cut, non-controversial definition of the term 'collocation'. A vague definition would be to say that collocations are groups of words which frequently occur in combination with each other. This definition is totally unsatisfactory, since it says nothing about the number of elements involved, the degree of frequency of occurrence, or the classes of words which can combine. In the literature, it is generally agreed that collocations are different from idioms, although the latter could also be defined as 'groups of words which frequently occur in combination with each other' (for comparisons of these categories, see Cowie 1981, 1988, 1994; Gläser 1986a; Hausmann 1989). Idioms are part of the larger class of relatively fixed multiword units. They are frequently described as semantically opaque word-combinations -- that is, combinations whose global meaning is different from the sum of the individual meanings of the constituent parts (although Nunberg et al. 1994 provide strong arguments against non-compositionality and argue that idioms are not opaque but can be accounted for in terms of situational metaphors). Other authors (Fraser 1970; Michiels 1975) lay stress, in defining idiomaticity, on an expression's resistance to a number of syntactic manipulations (passivization, pronominalization, fronting, clefting, insertion of material, etc.). Carter (1987) gives various examples of idioms, including *it's raining cats and dogs,* which is immutable in so far as it cannot be passivized, does not allow insertion, and has a fixed order (cf. ★*it's raining dogs and cats*). *To drop a brick* (meaning 'to commit a blunder') is an idiom of another structural type, which is resistant to other classes of manipulations. Insertions of material and passivization are permitted but pronominalization is not (★ *John dropped a brick yesterday and Tom dropped one too* is unacceptable). All this shows that there are various structural categories of idioms and that one can certainly talk of degrees of frozenness (Fraser 1970). What is important here is that idioms are to be found at one extreme of a continuum ranging from totally free combinations of words to completely frozen, fixed multiword units. Collocations will be found in the fuzzy area half-way between free combinations and idioms (cf. Howarth, this volume).

Cowie (1986) distinguishes between 'free' collocations and 'restricted'collocations. The former are combinations in which one word is open to partnership with a wide range of lexical items. The verb *hire,* for example, can collocate with an impressively large collection of items such as *staff, clerk, secretary, worker,* etc. In such cases, one

might be tempted to consider the constraints on co-occurrence as selection restrictions rather than collocational restrictions. Restricted collocations, on the other hand, are 'word-combinations in which one element (usually the verb) has a technical sense, or a long-established figurative sense which has since lost most of its analogical force' (Cowie 1991: 102). Cowie illustrates this definition with examples such as *run a deficit, abandon a principle,* or *champion a cause,* in which the object noun limits the choice of verb to one or two. It should be noted that the class of restricted collocations also includes what Gross (1981) calls 'support verb constructions' -- that is, constructions in which the role of the verb is limited to 'supporting' the direct object with which it co-occurs, thereby establishing a link between the object and the subject and conveying information on number, tense, and aspect. In *have a drink, have a bath, or take a decision,* the

delexical pattern is composed of a grammaticalized verb (*often have, make, take, give,* or *do*) and a deverbal noun. Incidentally, such constructions often prove to be a stumbling block for foreign-language students, who frequently have difficulty in selecting the appropriate support verb.

Benson (1989a) insists that collocations are 'arbitrary' recurrent word-combinations: the term 'idiosyncratic' is also frequently used to point to the unpredictable nature of collocations. No rule whatsoever can account for the acceptability of *pay attention* and the illformedness of do attention. A cross-linguistic comparison with French testifies to this, since *payer attention* is inadmissible in French (*faire attention* being the only acceptable combination).

It should not be inferred from the above examples that collocations are exclusively made up of a verb and a noun. In fact, any word class may be involved: so-called 'grammatical' collocations usually consist of a verb, an adjective or a noun, and an item belonging to a closed class (typically, though not exclusively, a preposition, e.g. *tantamount to, different from/than/to*). 'Lexical' collocations involve open-class items such as the verb-noun combinations examined above. Other categories frequently found in combination with one another are adjectives and nouns (A + N), adverbs and verbs (Adv + V), or even nouns and nouns (N + N) (*a confirmed bachelor, to appreciate deeply, a school of fish* being cases in point). It should be noted that in lexical collocations (of various types) a distinction is often drawn between the 'base' (the noun in the case of a V + N or A + N collocation) and the 'collocate' or 'collocator' (Hausmann 1979; Benson 1985; Cop 1988).

3
COLLOCATIONS AND CORPORA

The availability of huge corpora of texts in machine-readable form has made it possible to shed new light on the concept of collocation. Statistical tools are now being developed to extract groups of words that co-occur significantly (see below). The basic assumption is that words do not occur at random in a text. Sinclair recognizes an 'idiom principle', stating that 'a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable as segments' (1991: 110). This is equivalent to Cowie's notion of 'prefabricated chunks of language' stored in a speaker's lexicon.

Much current research in computational lexicography stems from the basic idea that collocations are a statistical phenomenon. Starting from a 'node' (the focal item in collocational analysis), it is possible to identify potential collocates by counting the 'span' of words which appear to the left or to the right of the node. If some words appear more frequently together than chance would predict, they are considered collocations. Smadja (1993) applies this hypothesis in the Xtract program: this software package extracts significant collocations from corpora which have been pre-processed by a part-ofspeech tagger. Two words are considered to co-occur if they are in a single sentence and if they are separated by fewer than five words.

It should be noted that different extraction tools do not necessarily retrieve the same types of combinations. In other words, there may be disagreement as to whether a given pair of lexical items should be granted collocation status. Church and Hanks (1990), for example, have developed a system which retrieves, among many other types of combinations, pairs of words such as *doctor-nurse* or *telephonetelevision.* Smadja (1993) has implemented a set of filters to avoid retrieving just these combinations on the grounds that they do not form a single syntactic construction (as *a school of fish does*) but are associated merely for semantic reasons (they belong to the same domain).

The computational tos mentioned above have obvious applications in lexicography: as Church and Hanks (1990) put it, they help lexicographers decide what to look for, and provide a quick listing of the main collocates of a given item. They make it possible to answer such questions as 'What can one do to a telephone?', which yields the verbs *answer, disconnect, tap, hang up, pick up,* among others.

Grefenstette (1994) also shows how corpus analysis makes it possible not only to extract significant collocations (i.e. those words

which are found recurrently in the immediate vicinity of a given item), but also to group similar words along axes of meaning. His technique enables him to derive what he calls third-order affinities from a corpus. He defines first-order affinities as relations 'describing collocates of words; second-order affinities show similarly used words and third-order affinities create semantic groupings among similar words' (1994: 288). Pairs of closely related words are used to define a semantic axis and similar words can subsequently be placed along these axes.

4
COLLOCATIONS IN THE COLLINS-ROBERT DICTIONARY

As I have already pointed out, machine-readable dictionaries have been the subject of much research in natural language processing but researchers have tended to concentrate almost exclusively on monolingual learners' dictionaries such as the *Oxford Advanced Learner's Dictionary* (Cowie 1989b), the *Collins Cobuild English Language Dictionary* (Sinclair*et al.* 1987), or the *Longman Dictionary of Contemporary English (LDOCE)* (Procter 1978). Some attempts have been made to extract co-occurrence knowledge from *LDOCE* (Wilks et al. 1989; Guthrie et al. 1991) but bilingual dictionaries tend to have been somewhat neglected by the research community, perhaps because publishers have been more reluctant to make the machine-readable versions of these dictionaries available to universities. Another, perhaps more fundamental reason may be the less fully structured format in which these bilingual dictionaries are published. However, it has to be admitted, as Atkins and Levin (1991: 255) have pointed out, that 'the explicit treatment often accorded to [restrictions on subjects or objects of verbs) in dictionaries for the foreign learner (i.e. bilingual dictionaries) renders such works a valuable source of material for the semi-automatic construction of a lexical database'. In the remainder of this section, I wish to show that the system used by the Collins-Robert lexicographers to capture selection restrictions and collocations is explicit enough to enable automatic, or at least semiautomatic, extraction of co-occurrence knowledge.

The *CR* dictionary owes its reputation to the wealth of collocational information it contains. A sophisticated and consistent

system of brackets, parentheses, and elements in italics is used to code various types of collocations systematically. The base of the collocation frequently appears in italics and its relationship to the collocate (the headword) is indicated as follows:

- a typical subject of a verb headword appears in square brackets;
- a typical object of a verb or the typical noun modified by an adjective headword appears unbracketed next to the part-of-speech indicator;
- a noun which usually complements another noun headword appears in square brackets;
- a synonym or hyperonyrn appears in parentheses.

The following examples illustrate this system:

> (1) Sample entries from the *Collins-Robert Dictionary* **devouring** *adj hunger, passiondévorant;zeal, enthusiasmardent* **unflagging** *adj person, devotion, patience* infatigable, inlassable; *enthusiasminépuisable;interestsoutenu jusqu'au bout* **chill** *vt (fig) enthusiasmrefroidir* **spark** *vt rebellion, complaints, quarrelprovoquer, déclencher;interest, enthusiasmsusciter, éveiller* (*in sbchez qn*) **slacken** *vi . . . [gale]diminuer de force;[speed]diminuer,[activity, business, trade]ralentir, diminuer;[effort, enthusiasm, pressure]diminuer, se relâcher* **wear out** *vi [clothes, material, machinery]s'user;[patience, enthusiasm] s'épuis)er* **burst 1** *n [shell etc]* explosion, écatement; *[anger, indignation]* explosion; *[anger, laughter]* éclat; *[affection, eloquence]* élan, transport; *[activity]* vague; *[enthusiasm]aceès, montée;[thunder]coup; [applause]salve; [flames]jaillissement, jet*

The examples above show that the base of the collocation appears in italics while the collocator itself is the headword. In an encoding perspective, this way of presenting collocations may not prove very useful if one wants to start from the base in order to discover a collocate which expresses a given meaning. The availability of the machine-reable version, however, makes it possible for the linguist to access information via any element of the dictionary, including (but not exclusively, of course) the elements in italics. Querying any base in italics yields a list of items (the headwords with which it is associated) which can be considered as potential collocates. Not all pairs of words are collocations, however. Consider the occurrence of the word *person* under **unflagging.** *Person* is actually a metalinguistic indicator which stands for any [+HUMAN] noun. We are then faced with a selection restriction and not with a collocational constraint. It is, therefore, necessary to filter out these elements before embarking on the actual construction of a collocational database.

The design of the database needs to take into account the typographical conventions used by the lexicographer. As is pointed out above, the presence (or absence) of parentheses or brackets is of

utmost importance since it signals a surface syntactic link between the base and the collocate. The following example shows, however, that the identification of the link is not trivial:

> (2) **slacken** *vi . . . [effort, enthusiasm, pressure]diminuer, se relacher*

The square brackets are used to signal a typical subject for a verb. For reasons of economy, though, only one pair of square brackets is used, which means that *enthusiasm* is not directly bracketed. The coding thus needs to be 'decompacted' and the syntactic information redistributed to make sure that *enthusiasm* inherits the properties indicated by the brackets. The decompacted entry looks like the following:

> (3) **slacken** *vi [effort], [enthusiasm], [pressure]diminuer, se relâcher*

This procedure will make sure that **slacken** is included in the list of verbs that can have *enthusiasm* as subject.

The acquisition of lexical knowledge can be performed on a large scale, since the dictionary contains approximately 70,000 items in italics, discounting subject field codes, which can easily be filtered out because they are capitalized. At an intermediate stage, it is possible to extract the list of all headwords which contain some reference to *enthusiasm* in italics. The global list can then be broken down automatically into subcategories as a function of the syntactic link which unites the base and the collocate. Applying this to the whole dictionary (and not just the seven entries given above) yields the following set of items for *enthusiasm:*

> (4) ENTHUSIASM Adj + N collocations catching, decreasing, devouring, infectious, intense, undamped, unequalled, unflagging, wild N1 + N2 collocations burst, decrease, diminution, excess, flame, spurt, surge, tidal wave, wave, wildness, withering Vtr + Nobj collocations chill, communicate, damp, diminish, drum up, fire, infuse, lose, quench, recapture, rekindle, simulate, snuff out, spark, summon up, whip up Nsubj + V collocations decrease, ebb, fall off, fizzle out, flag, ooze away, slacken, be spent, wear out, wilt, wither

Acquiring such co-occurrence knowledge can be performed automatically but the output resembles what can be found in a standard

collocational dictionary such as *The BBI Combinatory Dictionary of English (BBI)* (Bensone*t al.* 1986). As I have argued elsewhere (Fontenelle 1992b, 1994, 1997), however, such compilations do not provide any systematic semantic interpretation: this means that the user has to work out the exact nature of the link which unites the components of the collocation. In the above examples, *drum up, chill, quench,* and *spark* all appear in the same subcategory of Vtr + Nobj collocations; yet, it hardly needs to be pointed out that these verbs do not express the same (or even a similar) meaning and that they are far from being interchangeable. A similar criticism against standard dictionaries and their failure to differentiate collocations has been levelled by Smadja (1991), who admits that the type of semantic interpretation he has

in mind cannot be performed automatically. In the Liège database, we have adopted Mel'čuk's approach based on the concept of lexical function. In the following section, I would like to briefly outline Mel'čuk's approach and show how it can be used to enrich our collocational database with lexico-semantic relationships.

5
LEXICAL FUNCTIONS AND THE EXPLANATORY COMBINATORY DICTIONARY

Mel'čuk's Meaning-Text Theory (Mel'čuk et al. 1984/1988/1992; Apresyan et al. 1969), and more specifically the lexical component of MTT -- namely, the Explanatory Combinatory Dictionary (or ECD), offer a systematic description of what Mel'čuk calls 'restricted lexical co-occurrence'. The concept of lexical function is the basic device used in that theory to account for collocational phenomena. Each ECD entry includes a Lexical Co-occurrence Zone which contains information about the arbitrary collocations that can be formed with the headword. Lexical functions of the form **f** (*X*)=*Y* are used to indicate that a lexical-semantic relation **f** holds between the headword (*X,* or the base of the collocation) and the value *Y* (the collocate). Mel'čuk has identified around sixty standard lexical functions. The **Magn** LF, for example, indicates that the keyword is associated with a word meaning 'much, very, intense, or to a large extent', as in:

(5) **Magn** (*bachelor*) = *confirmed* **Magn** (*pain*) *excruciating* **Magn** (*fear*) = *mortal* **Magn** (*contrast*) = *shar, vivid*

The **Oper** $_n$ lexical function provides for the keyword (*X*) a (nearly)

semantically empty verb (*Y*) that takes *X* as its direct object (the subscript refers to the nth actant, which is taken as subject):

(6) **Oper** $_1$ (*complaint*) = *lodge* **Oper** $_1$ (*attention*) = *pay* **Oper** $_2$ (*attention*) = *draw* **Oper** $_1$ (*support*) = *lend*

The **Func** $_n$ lexical function also represents a support verb which is semantically empty. In this case, the keyword is the subject (the nth participant, if any, is the object):

(7) **Func** $_0$ (*idea*) = *come to*

It is important to realize that lexical functions, which can be conceived of as lexical-semantic relations, can appear either in isolation or in combination. Some of them, like **Plus** (more) or **Minus** (less), hardly ever occur in isolation: they are usually associated with **Pred,** which means'to be an *X*'and with another LF expressing the aspectual component of a collocate. Aspect is conveyed through the use of Incep (denoting the beginning), **Fin** (the end), or **Cont** (the continuation). Causative operators (**Caus** -- to cause -- and **Perm** -- to permit/allow) may also come into play in order to build fairly complex LFs:

(8) **CausPredPlus** (*price*) = *raise, increase* **IncepPredPlus** (*price*) = *rise, increase, rocket, surge* **CausPredMinus** (*anger*) = *assuage, soften, soothe* **FinFunc** $_0$ (*anger*) = *subside, wear off* **FintOper** $_1$ (*influence*) = *lose* **ContOper** $_1$ (*habit*) = *keep*

The interested reader will find more detailed descriptions of lexical functions in the first three volumes of the French *ECD* (Mel'čuk et al. 1984/1988/1992) and in various other articles on the topic (Mel'čuk and Zholkovsky 1988; Mel'čuk in the present volume). What is important to stress here is that MTT and the theory built on the concept of lexical function represent a formal apparatus which can be used to express the relationship between a base and a collocate in a systematic way.

6
ENRICHING THE DATABASE

As noted earlier, the automatic extraction process makes it possible to construct a database of collocations, but the subcategories are rather superficial, since they capre only basic surface-syntactic relations

(Adj + N; N1 + N2;Vtr + Nobj; Nsubj + V). It was, therefore, decided to enrich the database by assigning a lexical function to each of the 70,000 pairs of collocations extracted from the dictionary. Since it proved to be impossible to automate this labelling process, the assignment of this 'semantic tag 'had to be performed, for the most part, manually. A special field was then created in the database and, whenever possible, filled with the appropriate lexical function. The following examples illustrate the type of data that are contained in the database for the noun *enthusiasm.* It should be realized that only the LF field required manual assignment. The translations were also extracted automatically, which required a good deal of preliminary work to ensure proper identification of the translation field.

(9) **ENTHUSIASM** -- enthousiasme

Adj + N collocations		
CATCHING	(*contagieux, communicatif*)	LF: **Qual** $_1$ **CausFunc** $_1$
DECREASING	(*décroissant, diminué*)	LF: **A** $_0$ **IncepPredMinus**
DEVOURING	(*ardent*)	LF: **Magn**
INFECTIOUS	(*communicatif*)	LF: **Qual1CausFunc** $_1$
INTENSE	(*vif, énorme*)	LF: **Magn**
UNDAMPED	(*non reftoidi*)	LF: **A** $_0$ **AntiPredMinus**
UNEQUALLED	(*inéegalé*)	LF: **Magn**
UNFLAGGING	(*inépuisable*)	LF: **Magn**

WILD	(*débordant, dérant*)	LF: **Magn**
N1 + N2 collocations		
BURST	(*accès, montée*)	LF: **Sing**
DECREASE	(*baisse, refroidissement*)	LF: **S $_0$ IncepPredMinus**
DIMINUTION	(*diminution, affaiblissement*)	LF: **S $_0$ IncepPredMinus**
EXCESS	(*excès*)	LF: **S $_0$ Excess**
FLAME	(*flamme, ardeur, feu*)	LF: **Magn+Figur**
SPURT	(*sursaut, regain*)	LF: **Sing again**
SURGE	(*vague, montée*)	LF: **Magn+Figur**
TIDAL WAVE	(*immense vague, flot*)	LF: **Magn+Figur**
WAVE	(*vague*)	LF: **Magn+Figur**
WILDNESS	(*ferveur*)	LF: **Magn+Figur**
WITHERING	(*évanouissement*)	LF: **S $_0$ FinFunc $_0$**
Vtr + Nobj collocations		
CHILL	(*reftoidir*)	LF: **CausPredMinus**
COMMUNICATE	(*communiquer, faire partager*)	LF: **CausFunc $_1$**
DAMP	(*reftoidir*)	LF: **CausPredMinus**
DIMINISH	(*diminuer, amoindrir*)	LF: **CausPredMinus**
DRUM UP	(*susciter*)	LF: **Caus $_2$ Func $_0$**
FIRE	(*enflammer, échauffer, exciter*)	LF: **Caus $_2$ Func $_0$**

INFUSE	(*inspirer, insuffler*)	LF: **Caus $_2$ Func $_0$**
LOSE	(*perdre*)	LF: **FinOper $_1$**
QUENCH	(*refroidir*)	LF: **CausPredMinus**
RECAPTURE	(*retrouver*)	LF: **Oper $_1$ again**
REKINDLE	(*ranimer, raviver*)	LF: **Caus $_2$ Func $_0$ again**
SIMULATE	(*simuler, feindre, affecter*)	LF: **AntiVerOper $_1$**
SNUFF OUT	(*éeindre*)	LF: **Liqu**
SPARK	(*susciter, éveiller*)	LF: **Caus $_2$ Func $_0$**
SUMMON UP	(*faire appel*)	LF: --
WHIP UP	(*donner un coup de fouet à*)	LF: **CausPredPlus**
Nsubj + V collocations		
DECREASE	(*se calmer, se refroidir*)	LF: **IncepPredMinus**
EBB	(*décliner, baisser, être sur le dîlin*)	LF: **IncepPredMinus**
FALL OFF	(*baisser, tomber*)	LF: **IncepPredMinus**
FIZZLE OUT	(*tomber*)	LF: **FinFunc $_0$**
FLAG	(*tomber*)	LF: **FinFunc $_0$**
OOZE AWAY	(*disparaitre, se dérober*)	LF: **FinFunc $_0$**
SLACKEN	(*diminuer, se relâcher*)	LF: **IncepPredMinus**
BE SPENT	(*être tombé*)	LF: **AntiFact $_0$**
WEAR OUT	(*s'épuiser*)	LF: **IncepPredMinus**
WILT	(*diminuer*)	LF: **IncepPredMinus**
WITHER	(*s'évanouir*)	LF: **FinFunc $_0$** .

A few remarks ought to be made at this juncture. First, it should be pointed out that the method for building the lexical co-occurrence zone of an entry differs slightly from the approach adopted in Mel'čuk's *ECD*. Mel'čuk starts from an entry (a lexeme) and a predetermined list of LFs and then tries to find out which values (the collocates) are yielded when one applies the LF to the lexeme. In our database, we work the other way round, since we start from potentially interesting collocations and try to discover the LF -- that is, the lexical-semantic relationship which holds between the lexeme and its collocate. This poses a number of problems to which I shall return below.

Now that the database is complete (Fontenelle 1997), it is possible to access information from various angles and not only via the collocate, as was formerly the case with the printed version of the dictionary. The user is now able to select and extract entries which meet a specific criterion or a combination of criteria. Typical queries are the following:

- Q Which adjectives can express an intense form of enthusiasm? A **Magn** (*enthusiasm*) = *devouring, intense, unequalled, unflagging, wild*

• Q Which nouns can denote a 'portion' or 'single unit' of enthusiasm? A **Sing** (enthusiasm) = burst, spurt • Q Which transitive verbs can take enthusiasm as direct object and mean'decrease'? A **CausPredMinus** (**enthusiasm**) = **chill, damp, diminish, quench** • Q Which intransitive verbs can take enthusiasm as subject and refer to the end of somebody's enthusiasm? A **FinFunc $_0$** (*enthusiasm*) =*fizzle out, flag, ooze away, wither.*

The two-part Appendix, below, illustrates collocational material for the words *anger* and *hatred* respectively. Together with enthusiasm, these nouns belong to the field of emotion lexemes, a semantic set which has attracted much attention. Mel'čuk and Wanner (1996) try to show that, for German emotion nouns, useful generalizations about LF values along semantic lines are possible. Their contention is that it should be possible to avoid redundant repetitions of lexical information by specifying default LF values which would be valid for any emotion lexeme. To some extent, they show that this is partly feasible in German, but they remain aware that 'it is often impossible to find correlations between lexical co-occurrence of the key lexemes and their semantic features'. They add: 'after. all, language is notoriously capricious and unpredictable' (Mel'čuk and Wanner 1996: 337).

The data extracted from the CR dictionary seems to confirm this view of the unpredictable character of language. Indeed, it seems to be very hard to detect generalizations in the values of the lexical functions for the three emotion nouns in question, which, phraseologically speaking, all behave in rather idiosyncratic ways. However, it should be noticed that other types of generalizations can be made as to the *nature* of the lexical functions associated with emotion nouns. The verbs which co-occur with emotion lexemes can be classified as a function of the various stages in a cycle: they can all be described in terms of verbs expressing a beginning (**Incep**), a growth (**Plus**), a decline (**Minus**), or an end (**Fin**). When the emotion reaches a peak, the **CulM** (= culmination) lexical function has to be used (the noun *paroxysm* seems to serve as a default value for this function). The **Caus** LF also comes into play when an external factor (a stimulus) brings about the emotion or is responsible for its appearance or decline. It is interesting to draw a parallel with the approach adopted by Cohen in her Lexique des co-occurrents -- *Bourse et conjoncture économique* (B. Cohen 1993: 505-6). Her contention is that it is possible to describe the behaviour of stocks, bonds, interest rates, and share prices in terms of the various stages in an economic cycle. Interestingly,

emotion lexemes seem to pattern in very much the same way and the study of their phraseological environment provides ample evidence that these basically cyclical phenomena need to be described in terms of very specific lexical functions.

7
A FEW PROBLEMS

In some cases, the lexicographer who has to assign a lexical function to a pair of restricted collocations is faced with the problem of choosing among several alternatives. Apresyan (personal communication) asserts that no battery of tests is available to ensure consistency in the assignment of lexical functions. In the field of nouns expressing emotions and feelings, for example, a choice has to be made between complex LFs such as **IncepPredMinus** and **HnFunc $_0$.** As we saw earlier, the former LF means that the referent of the keyword decreases in intensity while the latter LF refers to its disappearance. For verbs such as *slacken* or *decrease,* the meaning is fairly obvious:

(10) **IncepPredMinus** (*enthusiasm*) = *slacken, decrease* . . .

For verbs such as *flag* (tomber), *wither* (s'évanouir), *wear out* (s'épuiser) or *ooze away* (disparaitre, se dîrober) which can take the noun *enthusiasm* as subject, the lexicographer might hesitate and be tempted to assign the **IncepPredMinus** function as well. In doubtful cases, however, it is possible to apply a fairly simple test -- namely, to use the verb in the present perfect and verify whether the concept denoted by the keyword is still present or not:

(11) *His enthusiasm has withered* (i.e. it no longer exists) à **FinFunc $_0$** *His enthusiasm has slackened* (i.e. it still exists, albeit in a small quantity) → **IncepPredMinus**

The same holds for the French verbs which can be submitted to a test using the *passé composé*:

(12) *Son enthousiasme s'est évanoui* (i.e. it no longer exists) → **HnFunc $_0$** *Son enthousiasme s'est épuisé* (i.e. it no longer exists) → **FinFunc $_0$** *Son enthousiasme a baissé/déché* (i.e. it still exists) → **IncepPredMinus**

It seems that some verbs are vague in so far as the test does not enable the lexicographer to judge precisely whether the emotion is still present or not. Mel'čuk himself is aware of this vagueness. In the *ECD* (Mel'čuk et al. 1984: i. 77), we find:

(13) **FinFunc $_0$ ou IncepPredMinus** (*colère*) = *s'apaiser, se calmer, se refroidir*

We may find cases of vagueness in English as well, as the *LDOCE* entry for *wear off* testifies:

(14) **wear off** *v adv* [IØ] to be reduced until it disappears: *the pain is wearing off*

Reduced clearly refers to the **IncepPredMinus** function, while *disappears* definitely implies **FinFunc $_0$** . Both LFs should therefore be assigned to capture this vagueness.

8
DISCOVERING NEW FUNCTIONS

In this section, I would like to show that other types of relationships can be discovered in dictionary entries. The preceding sections concentrated mainly on abstract nouns and I wish to show that *CR* also contains significant material for artefacts. The entry for *brake* below makes it abundantly clear that the collocates associated with a concrete object are of a different nature altogether and that we have to resort to a different set of lexical functions to model its phraseological environment.

(15) **BRAKE** -- frein

Adj + N collocations		
OVERHEATED	(*qui chauffe*)	LF: **A1ExcessT^0**
ON	(*serré, mis*)	LF: **A $_1$ Real1**
OFF	(*desserré*)	LF: **A $_1$ AntiReal1**
N1 + N2 collocations		
DRAG	(*sabot, patin*)	LF: **Part**
DRUM	(*tambour*)	LF: **Part**
LINING	(*garniture*)	LF: **Part**
RELEASE	(*dégagement, desserrage*)	LF: **S $_0$ AntiReal1**
SCREECH	(*grincement*)	LF: **S $_0$ Son**
SHOE	(*sabot*)	LF: **Part**
SOUND	(*bruit*)	LF: **S $_0$ Son**
SQUEAL	(*grincement*)	LF: **S $_0$ Son**
Vtr + Nobj collocations		
JAM	(*bloquer, coincer*)	LF: **CausObstr**
LINE	(*garnir*)	LF: **PreparFact $_0$**
OPERATE	(*faire marcher, faire fonctionner*)	LF: **CausFact $_0$**
RELINE	(*changer la garniture de*)	LF: **PreparFact $_0$**
Nsubj + V collocations		
DRAG	(*frotter, se gripper*)	LF: **Degrad**
FAIL	(*lâcher*)	LF: **FinFact $_0$**
GRIP	(*mordre*)	LF: **MagnFact $_0$**

JAM	(*se bloquer*)	LF: **Obstr**
SCREAM	(*hurler*)	LF: **Son**
SCREECH	(*grincer*)	LF: **Son**
SQUEAL	(*grincer, hurler*)	LF: **Son**

The first point to note is that we have introduced the **Part** lexical function to account for part-whole relations (for instance, to indicate that brakes are made up of the elements drag, drum, lining, shoe, etc.). Interestingly, Mel'čuk does not recognize the existence of this function, such information being considered part of the encyclopaedic description of a lexeme, while Apresyan *et al.* argue that it belongs to the 'lexical universe' of the keyword, which they define as 'an informal description of a sufficiently "broad piece" of reality including the given situation as a constituent element' (1969: 19). Although I basically agree with Mel'čuk that part-whole relations are purely semantic and not lexical, I tend to share Grimes's opinion (1990: 358) that 'making lexical functions of them [i.e. of part-whole relations] allows them to be accessed by the function mechanism, which may be an advantage in information management'. One cannot deny that part-whole relations are a subclass of collocational relations, and making them accessible in a collocational database is of paramount importance if one wants to use that database in applications such as language-teaching or information retrieval.

The dictionary also includes other types of information, such as, for example, the typical sound made by the referent of the keyword (denoted by the LF **Son**). In most textbooks and articles on MTT, **Son** is typically associated with animal noises (e.g. **Son** (*dog*) = *bark;* **Son** (*frog*) = *croak*). It hardly needs to be pointed out, however, that artefacts (such as doors, hinges, and bells) can also make specific sounds, which can be of practical relevance in a translation perspective.

The Fact function is also an important element in the description of the lexical co-occurrence zone of concrete objects. Unlike **Func,** which means 'to exist' and which is described by Mel'čuk as referring to (almost) semantically empty verbs, **Fact** means 'to function, to work, to be realized, fulfilled, implemented'. It can also be used in conjunction with aspectual or causative operators to form complex lexical functions and having access to such information enables us to answer the following questions:

- Q Which verb(s) can mean that brakes cease to function? A **FinFact $_0$** (*brake*) = *fail* • Q Which verb(s) can mean that one causes brakes to function? A **CausFact $_0$** (*brake*) = *operate*

9
CONCLUSION

In this chapter, I have tried to show that a bilingual machine-readable dictionary such as the *Collins-Robert English-French Dictionary* can be used as a starting point for the construction of a collocational database. The vocabulary items printed in italics can be considered as the bases of potentially interesting collocations and the availability of the dictionary in machine-readable format makes it possible to extract this collocational material readily. The database is then manually enriched with lexical functions *à la* Mel'čuk, which enable the lexicographer to capture significant lexical-semantic relations holding between the base and the collocate. The database then contains a bilingual description of the lexical co-occurrence potential of a given entry. Multiple access points enable the linguist to access information via the base (the keyword, or element in italics), the collocate (the dictionary headword), the lexical function, the translation, or the part of speech, all of which means that this type of collocational dictionary can fill a gap in lexicography: indeed, traditional collocational

dictionaries usually provide one type of access key only or simply list potential collocates without any attempt at formalizing the meaning relationship holding between the constituents of the collocation. The collocational description of a lexeme can then be viewed as a semantic network in which a node (the base) is linked to its collocates through arrows (the lexical functions) which are labels for specific lexical-semantic relationships. Potential applications of such a database range from language teaching to information retrieval and machine translation and, more specifically, language generation, where selecting the appropriate word in context is of crucial importance.

APPENDIX

(16) **ANGER** -- colère

Adj + N collocations		
JUST	(*juste, légitime*)	LF: **Ver**
MOCK	(*simulé, feint*)	LF: **A $_0$ AntiFact $_0$**
RIGHTEOUS	(*juste, justifié*)	LF: **Ver+Bon**
RISING	(*croissant*)	LF: **A $_1$ IncepPredPlus**
SPEECHLESS	(*muet -- de colère*)	LF: **A $_1$ Degrad(speech)+ Sympt $_{23}$**
UNCHECKED	(*non maîtrisé, non réprimé*)	LF: **A $_2$ Perm1Fact $_0$**
WILD	(*fou*)	LF: **Magn**

N1 + N2 collocations		
ACCESS	(*accès*)	LF: **Sing**
BURST	(*explosion, éclat*)	LF: **Sing+Manif**
ERUPTION	(*explosion*)	LF: **Sing**
INTENSITY	(*intensité, force, violence*)	LF: **S $_0$ Magn**
OUTBREAK	(*explosion*)	LF: **Sing+Manif**
OUTBURST	(*explosion, bouffée, accès*)	LF: **Sing+Manif**
PAROXYSM	(*accès*)	LF: **Culm**
SPURT	(*sursaut, regain*)	LF: **ContSing**
Vtr + Nobj collocations		
AIR	(*exhaler*)	LF: **Caus $_1$ Manif**
APPEASE	(*apaiser, calmer*)	LF: **CausPredMinus**
AROUSE	(*exciter, provoquer*)	LF: **CausFunc $_0$**
ASSUAGE	(*soulager, apaiser, calmer*)	LF: **CausPredMinus**
BRAVE	(*braver*)	LF: **CausAntiAbleFact $_0$**
CONTAIN	(*contenir, refréner, maîtriser*)	LF: **nonPerm $_1$ Fact $_0$**
EXCITE	(*exciter, aviver*)	LF: **CausPredPlus**
INCUR	(*s'attirer, encourir*)	LF: **Oper $_2$,**
INFLAME	(*attiser, allumer*)	LF: **CausFunc $_0$**
KEEP DOWN	(*réprimer, contenir*)	LF: **nonPerm $_1$ Fact $_0$**
KEEP IN	(*contenir, réprimer*)	LF: **nonPerm $_1$ Fact $_0$**
KEEP UNDER	(*contenir, maétriser*)	LF: **nonPerm $_1$ Fact $_0$**
MOLLIFY	(*apaiser, calmer*)	LF: **CausPredMinus**
POUR OUT	(*donner libre cours à*)	LF: **Perm $_1$ Fact $_0$**
RESTRAIN	(*contenir, réprimer, maîtriser*)	LF: **nonPerm $_1$ Fact $_0$**
GIVE FREE REIN TO	(*lâcher la bride*)	LF: **Perm $_1$ Fact $_0$**
SMOTHER	(*contenir, réprimer*)	LF: **nonPerm $_1$ Fact $_0$**
SOFTEN	(*adoucir, atténuer*)	LF: **CausPredMinus**
SOOTHE	(*apaiser*)	LF: **CausPredMinus**
STIFLE	(*réprimer*)	LF: **nonPerm $_1$ Fact $_0$**
STILL	(*calmer*)	LF: **CausPredMinus**
STIR UP	(*exciter*)	LF: **CausFunc $_0$**
SUFFOCATE	(*suffoquer de*)	LF: **Degrad (breathing)+ Sympt $_{23}$**
SWALLOW	(*ravaler*)	LF: **nonPermManif**
VENT	(*décharger*)	LF: **Real $_1$**
WREAK	(*assouvir*)	LF: **Real $_1$**

Nsubj + V collocations		
BLAZE	(*éclater*)	LF: **MagnFact $_0$**
BLAZE FORTH	(*éclater*)	LF: **MagnFact $_0$**
BLAZE OUT	(*éclater*)	LF: **MagnFact $_0$**

BLAZE UP	(*éclater*)	LF: **MagnFact0**
COOL DOWN	(*se calmer, s'apaiser*)	LF: **IncepPredMinus**
ERUPT	(*exploser*)	LF: **MagnFact0**
EXPLODE	(*éclater*)	LF: **MagnFact0**
FESTER	(*couver*)	LF: **AntiMagn+Fact0**
FLAME UP	(*exploser*)	LF: **MagnFact0**
FLARE UP	(*éclater*)	LF: **MagnFact0**
MELT	(*tomber*)	LF: **FinFunc0**
MELT AWAY	(*se dissiper, tomber*)	LF: **FinFunc0**
RISE	(*croître, grandir*)	LF: **IncepPredPlus**
SIMMER	(*couver, monter*)	LF: **IncepPredPlus**
SOFTEN	(*s'atténuer*)	LF: **IncepPredMinus**
SUBSIDE	(*tomber, se calmer*)	LF: **FinFunc0**
SUFFOCATE	(*suffoquer de --*)	LF: **Degrad(*breathing*)+** Sympt32
TO LET RIP	(*éclater, exploser*)	LF: **MagnFact0**
WEAR OFF	(*s'apaiser, passer*)	LF: **FinFunc0**
(17) **HATRED** -- Haine		
Adj + N collocations		
BITTER	(*acharné, profond*)	LF: **Magn**
DEADLY	(*mortel, implacable*)	LF: **Magn**
FIERCE	(*implacable*)	LF: **Magn**
GROWING	(*grandissant*)	LF: **A0IncepPredPlus**
INTENSE	(*intense, violent, profond*)	LF: **Magn**
IRRECONCILABLE	(*implacable*)	LF: **Magn**
MORTAL	(*mortelle*)	LF: **Magn**
UNREMITTING	(*opinicitre, constant*)	LF: **A0ContFact0**
N1 + N2 collocations		
GENERATION	(*engendrement*)	LF: **S0CausFact0**
INTENSITY	(*intensité, force*)	LF: **S0Magn**
SHOW	(*manifestation, démonstration*)	LF: **S0Manif**
Vtr + Nobj collocations		
INFLAME	(*attiser, allumer*)	LF: **CausPredPlus**
PROFESS	(*professer*)	LF: **Caus1Manif**
STIR UP	(*attiser*)	LF: **CausPredPlus**
Nsubj + V collocations		
BE SPENT	(*être tombé*)	LF: **FinFunc0**
BLAZE OUT	(*éclater*)	LF: **Fact0**
CENTRE	(*se concentrer*)	LF: **Fact1**
GROW UP	(*naître, se développer*)	LF: **IncepPredPlus**

[This page intentionally left blank.]

10
Phraseological Dictionaries: Some East-West Comparisons

A. P. COWIE

1
INTRODUCTION

This chapter provides a critical survey of some major developments in British and Russian phraseology over the period 1930 to 1960 and traces their influence on dictionaries of collocations and idioms intended chiefly for foreign learners of

English and published between 1975 and 1995. [1] British EFL dictionaries, whether general or specialized, have from their beginnings in the 1930s drawn on research into lexical and grammatical structure to an extent seldom approached, then or since, by works compiled for the native speaker of English (Cowie 1989a). Parallel Russian developments have been more extensive, as well as more centrally directed, and they have been sustained by impressive achievements in research (e.g. Melcuk 1974, 1981). Official policy within the former Soviet Union towards national and leading international languages favoured ambitious programmes of dictionary development, and led to the compilation of a wide range of general and specialized pedagogical dictionaries, bilingual as well as monolingual, aimed at learners of Russian and various other major languages, including of course English. Collocational dictionaries form an impressive part of this output, reflecting the stress laid by Russian lexicographers on meeting the needs of students wishing to write in the foreign language (Tomaszczyk 1981; Morozenko 1986; Benson 1988).

There is little sign to date of British phraseological research having influenced Russian dictionary-making. The flow of ideas in phraseology since the late 1960s has been almost entirely from East to West (Klappenbach 1968; Weinreich 1969; Lipka 1974; Aisenstadt 1979; Cowie 1981; Gläser 1986a). In Western dictionaries, however, the two

[1] I am grateful to Veronika Teliya for her helpful comments on an earlier version of this chapter. Imperfections which remain are, of course, my own.

traditions are sometimes interestingly combined, as will be shown by reference to two dictionaries in current use, one of idioms, the other of collocations. These will be compared with examples of the same two types produced entirely within the Soviet dictionary sphere.

Categorization is notoriously difficult in phraseology because of the bristling array of variables -- syntactic, pragmatic, stylistic, semantic-which the material is constantly throwing up. For analysts, the crucial problems are those of determining where the domain of phraseology actually begins, and, within that expanse, of dividing the more invariable and opaque items from the more recombinable and transparent ones. Solving these problems calls for the selection of appropriate criteria and the elaboration of a framework of phraseological categories. One category which is attracting particular attention in current work on the analysis of written texts (Cowie 1991, 1992; Howarth 1993, 1996) is the so-called restricted collocation, a category which was recognized, though under different names, by Russian phraseologists from the late 1940s onwards. One aim of this chapter will be to throw further light on this category, enabling me to comment in a more principled way on the selected collocational dictionaries.

Phraseology is a field bedevilled by the proliferation of terms and by conflicting uses of the same term, and I hope to contribute to the understanding of a complex area by comparing the terms and categories used in a number of superficially different -- though fundamentally related -- phraseological systems. At the same time it should be made clear that, even within the limited periods surveyed in this chapter, there are major sources of ideas that are not drawn upon. What might be called 'classical' Russian phraseology later acquired disciples in the then German Democratic Republic and beyond, and these in turn became sources of further theoretical development (see especially Fleischer 1982; Gläser 1986a). While the contribution of such scholars is readily acknowledged, the present choice of major East European figures is selective and determined by their pioneering role in identifying the categories that I am attempting to clarify, extend, and apply.

2
H. E. PALMER AND A. S. HORNBY

It is not widely known that matters of selection and categorization in phraseology were of interest to two of the founding fathers of EFL lexicography, H. E. Palmer and A. S. Hornby, or that in the late 1920s Palmer set up a research project, in which he was later joined by

Hornby, with the aim of collecting and classifying a wide range of multiword units. Palmer's theoretical views were set out in his Introduction to the first major report on the project, the so-called *Second Interim Report on English Collocations* of 1933, in which he drew attention to the widespread use in speech and writing of wordcombinations (called by Palmer 'collocations' or more colourfully 'comings-together-of-words'). [2] He also spelt out the consequences for language-learning of their idiosyncratic forms and form-meaning relations. In fact much of the emphasis in his discussion was placed on the learning difficulty they represented and the approach that must therefore be taken to acquiring them. His definition was as follows: 'A collocation is a succession of two or more words that must be learned as an integral whole, and not pieced together from its component parts' (Palmer 1933: 5).

But Palmer was also interested in categorization, and, though he had not clarified in 1933 the fundamental difference (discussed in the Introduction to this volume) between word-like and sentence-like expressions, he was able to focus fruitfully on the former -- on wordcombinations, that is, which function as elements in the simple sentence. So much is clear from his singling out, in the *Report*, of 'Collocations that are classifiable as Verb-collocations, Nouncollocations, Adverb-collocations, Preposition-collocations' (Palmer 1933: 18). [3]

Palmer also gave thought to the important difference between collocations as a general category and what he referred to as 'free phrases' or 'free combinations'. These were a manifestation of the language system and 'could be put together by dint of the application of the commonest and best-known rules of grammar' (1933: 5).

At the same time, his approach had certain limitations. Very few analysts would now apply the term 'collocation' to the whole range of multiword units surveyed by phraseologists. Compare, for instance, Gläser's umbrella term 'phraseological unit' (Gläser 1986a, 1988a), Alexander's 'fixed expression' (Alexander 1978, 1987), and TerMinasova's 'word-combination' (Ter-Minasova 1992). Many would limit the term collocation to word-like combinations (i.e. nominations or composites) which are not idioms but which lie in the fuzzy zone between free combinations and idioms proper (cf. Hausmann 1989). It is this fuzzy zone that I wish to bring into sharper focus later on. Not surprisingly, since Palmer did

not recognize a gradation of

[2]A tentative First Interim Report was issued in mimeographed form in 1931. No copies are now obtainable.

[3]Palmer returned to the problem of defining sentence-like units ('propositions') in 1942, when he produced a detailed -- though as yet unpublished -- subcategorization (Bongers 1947).

idiornaticity within noun-collocations, verb-collocations, and so on, he did not provide us with procedures for differentiating between the more and the less idiomatic cases.

All the same, we should not ignore the positive qualities of the Interim Report, the chief of which is that Palmer and Hornby provided a meticulous classification of the syntactic patterns in which wordcombinations are found. As I have pointed out, the major categories of the Report were phrase patterns or predicate patterns, introducing substantial lists of collocations. For instance, it included the construction 'preposition x *one's* x noun' (category number 35115.1) and provided such examples as:

(1) *at one's ease at one's feet at one's leisure at one's post at one's time of life*

The analytical approach pioneered by Palmer and developed by Hornby greatly influenced the treatment of multiword units in the general-purpose learners' dictionaries of the 1930s and 1940s (Cowie 1990). One aspect of the approach was Palmer's aversion to idioms of the kind traditionally favoured by lexicographers -- that is, structurally complex and culturally marked expressions such as *put one's shoulders to the wheel, skate on thin ice, or buy a pig in a poke* (Palmer 1938). While such idioms were certainly not excluded from Hornby's *Idiomatic and Syntactic English Dictionary* (1942), as much prominence, if not more, was given to the stylistically and culturally neutral types recorded in the *Interim Report* (e.g. *to ask a favour (of), to call attention to,* and *to catch the eye of),* and which current computer-based studies show to occur with much greater frequency in written texts than the highly evocative combinations spurned by Palmer (Moon 1988 and this volume).

Another part of the Palmer-Hornby legacy is the emphasis now laid in virtually all British dictionaries of idioms on the syntactic categorization and analysis of entries. This derives in part from the elaboration, by both men, of grammatical 'codes' for use in general dictionaries (Cowie 1989a), but it also owes something to the *Interim Report*. Once the practice had been established, in the *Report*, of classifying word-combinations according to form and function, it was a natural next step to provide a more elaborate specification, showing the clause functions of the elements in such idioms as *cut a dash* or *spill the beans* and indicating transformational possibilities and restrictions. Two entries from the *oxford Dictionary of Current IdiomaticEnglish*

English, vol. ii (Cowie et al. 1983), illustrate one approach to denoting the elements '(transitive) verb' and '(direct) object', and signalling the possibility or otherwise of the passive transformation:

(2) cut a dash [V + O] spill the beans [V + O pass]

By about the mid- 1970s, the main focus of interest in British (and American) phraseological theory was shifting elsewhere. Palmer and Hornby, as we have seen, had laid out an elaborate network of grammatical subtypes, but had failed to investigate at all closely another kind of variable, cutting across the grammatical one. This was the scale or cline of idiomaticity, which manifested itself in varying degrees of semantic opacity and/or structural stability. Palmer and Hornby knew that the learning difficulty in the case of *at one's leisure or at one's post* had to do with the arbitrary choice of elements. *(Reply at your leisure,* not ★*Reply at your ease;* He *died at his post,* not ★*He died at his position.)* Here, the choices of noun were quite fixed; but in other equally arbitrary cases *(a chink/crack in one's armour)* they were not. What was needed was a more elaborate scheme, one which took account of degrees of variation, and of the connection which often existed between a limited choice of words at one point and the meaning of the word at the other. In *pick a fight/quarrel/argument,* for instance, the sense of *pick* ('instigate') is somehow contingent on the arbitrarily limited set of nouns. New approaches sensitive to these various factors needed to be found, and by the middle of the 1970s these were increasingly East European.

3
V. V. VINOGRADOV AND N. N. AMOSOVA

From the mid- 1970s, and increasingly throughout the 1980s, one of the strongest influences on British phraseological theory was the work of a group of leading Russian scholars who had been active about thirty years earlier. [4] These included Victor Vladimirovich Vinogradov, widely regarded as the father of Russian phraseology, and Natalya Nikolaevna Amosova, deserving of special mention because of a formulation of 'phraseologically bound' meaning which differed from that proposed by Vinogradov (1947) and which was later widely used by Russian scholars. Both devised a tripartite scheme of categories, and it is probably true to say that the differences between the two,

[4]The first to be aware of the distinctions discussed here was Charles Bally (1909), whose scheme formed the starting point of Vinogradov's categorization.

except in the crucial respect mentioned above, were differences of emphasis and terminology, rather than of substance.

Vinogradov's most inclusive category was the 'phraseological unit', perhaps the most widely used umbrella term in Russian phraseology and defined by Rosaliya Ginzburg and her colleagues in 1979 as follows: 'Phraseological units are

non-motivated word-groups that cannot be freely made up in speech but are reproduced as ready-made units' (Ginzburg et al. 1979: 74). The apparent echo here of Palmer's stress on the learning and reproduction of whole units is striking, though it is unlikely -- given the circumstances -- that Vinogradov knew anything of Palmer's activities in 1930s Japan, or he of Vinogradov's a decade or so later. [5]

Vinogradov went on to draw a distinction within this general category between 'phraseological fusions' (also called 'idioms'), 'phraseological unities', and 'phraseological combinations'. The first subcategory was made up of unmotivated units -- those in which there was no relation whatsoever between the meaning of the whole combination and those of its components. This definition, with its stress on the impossibility of interpreting wholes in terms of the meanings of parts, represents the standard approach to idiomaticity at both the popular and technical levels in Britain and the USA, as a reading of the entry for *idiom* in any conventional English dictionary will confirm (cf. Cowie 1988, 1994). But, unlike many lexicographers, Vinogradov did not stop there. He saw that the degree of semantic fusion was often correlated with the rigidity or otherwise of the expression, or, as Arnold (1986: 170) put it, 'with the possibility of changing the form or the order of components, and of substituting the whole by a single word'. Clearly, shoot the breeze, spill the beans, and trip the light fantastic fit this category, which in other frameworks is referred to as 'pure idiom' (cf. Cowie 1981).

Beyond the fusion, Vinogradov recognized a partially nonmotivated type, the 'phraseological unity', whose sense could be perceived as a metaphorical or metonymic extension of the whole expression (Ginzburg et al. 1979: 75). Examples of this category, called in my own categorization 'figurative idioms', are do a U-turn and blow off steam, both of which have undergone metaphorical development from a (still active) technical sense. It has often been noted that the boundary between 'fusions' and 'unities' is not clear-cut

5According to his daughter, Dorothée Anderson, Palmer visited Moscow during the world tour he embarked on in 1931. The Soviet language-teaching experts he met were already familiar with the ideas on vocabulary control and text simplification developed by himself, Michael West, and others. No mention is made, however, of any contacts with Soviet phraseologists (Anderson 1969: 152).

(Cowie 1981: 229). For some speakers, a given expression (say, *burn one's boats*) has a figurative sense which is not yet completely fossilized; for others the same expression is entirely opaque. Igor Mel'čuk (personal communication) is reluctant to recognize the category phraseological unity precisely because of its indeterminacy: assigning expressions to it depends on the differing linguistic and cultural experience of individual speakers.

Vinogradov's third category, the 'phraseological combination', can be dealt with briefly, but it is a crucial type since it introduces the notion of contextual determination of meaning. It is defined by Irina Arnold (1986: 170) as follows: 'Phraseological combinations are not only motivated but contain one component used in its direct meaning while the other is used figuratively: *meet the demand, meet the necessity, meet the requirements.'* Here, it will be noted, *meet* is the element used in a figurative sense, while the nouns *demand, necessity,* and *requirements* form a variable determining context.

Amosova's tripartite system modified Vinogradov's in one respect. She too recognized a division between unmotivated and motivated idioms (between *spill the beans,* say, and run off the rails), though she had no special name for the second type. However, she identified a distinction within Vinogradov's category of 'phraseological combinations' which called for the recognition of two subclasses (Amosova 1963). As we have seen, Vinogradov had identified combinations in which one element had a figurative sense determined by its context (it was, according to his formulation, 'phraseologically bound'). The binding or determining context was a single word, or a limited set, as in the example given above. But for Amosova, the difference between binding word and binding words was central. For a combination to be phraseological, she argued, the bound sense must have a single determining item. This was obviously true of the binary combination *grind one's teeth,* to which Amosova applied the term 'phraseme'. The same term could also be applied to each of the combinations *small talk, small hours,* and *small change* (Arnold 1986: 171), where the sense of *small* varied with its context ('trivial', 'early', 'of low value'), and where that context was a single item in each case.

However, in cases where there were several determining items for the same figurative sense *(pay one's respects/a compliment/court to someone),* the combinations were separated off as 'phraseoloids', and as outside phraseology proper, 'because the determining minimum on which the meaning of *pay* depends [was] not constant' (Arnold 1973: 159). But, as Arnold (1973: 159) was quick to add, this second category was not free either, 'because the group of nouns possible as the second component [was] not limited by the character of the designata

but by stylistic tradition'. In other words, choice in such expressions was governed not by co-occurrence restrictions but by conventiona strong argument for regarding these too as falling within the domain of phraseology.

At this point it may be helpful to set out the terms and categories recognized by Vinogradov and Amosova in a table, indicating at the same time the terms used in my own 1981 framework (see Table 10.1). (Compare the broader framework set out in Table 1.2, in the Introduction to this volume.)

Two further points must be made before leaving the phraseologically bound category, divided as we have seen by Amosova, but unified and phraseological within the other two schemes. The first is that, in two-word combinations such as those we have been considering, limited choice may operate at either of the two structural places, or at both. Consider again the example *pay one's respects/a compliment/court to someone.* We notice that there are actually three factors involved in delineating the type:

> (3) (a) a verb used in a figurative sense (here 'offer', 'extend'), and (b) contextual determination of that sense by an arbitrarily limited set of nouns -- *one's respects/a compliment/court* -- each in this case with further grammatical idiosyncrasies -- but also (c) determination of no other verb in the same sense by the same limited context: *present,* if possible at all as a synonym, is only collocable with plural *compliments,* as in *present one's*

compliments to someone.

Now clearly in some cases we can reverse the direction of determination indicated at (b). That is, a noun in a specialized sense can have as its determining context a limited set of verbs (Cowie 1981: 228):

(4) *cause/create/make a stir*

This is obviously a restricted collocation too. Or we can relax condition *(c),* recognizing as phraseological any case in which a small number of synonymous verbs is determined by the limited set of nouns. By so doing, we of course add greatly to the number of collocations regarded as restricted. But collocability in such expressions is still strictly limited (at both points) and the relevant cases are in practice often listed in collocational dictionaries. Consider the following example:

(5) *make/achieve progress/headway*

Subtypes of restricted collocations can be systematically ordered, and in a recent large-scale analysis of collocations in the language of the social sciences Howarth (1996) has provided a rigorous categorization along a scale from most to least restricted, taking account of

TABLE. 10. 1. Subcategories of word-like units ('nominations')

Author	General category	Opaque, invariable unit	Partially motivated unit	Phraseologically bound unit
Vinogradov (1947)	Phraseological unit	Phraseological fusion	Phraseological unity	Phraseological combination
Amosova (1963)	Phraseological unit (b) Phraseoloid	Idiom	Idiom (not differentiated)	(a) Phraseme
Cowie (1981)	Composite	Pure idiom	Figurative idiom	Restricted collocation

the factors of place (or places) of substitution and degree of limitation of choice. If one considers the types illustrated above, at (3), (4), and (5), to which should be added Amosova's totally invariable phraseme, one has a progression, which extends the zone of restricted collocability some considerable way towards the area of free collocability. (Cf. Howarth's more detailed treatment of the continuum in this volume.) Such a scheme of course does not simply throw light on the nature of collocational restriction; it also provides a principled basis both for the critical appraisal of existing dictionaries and for the compilation of new ones.

4
A. V. KUNIN, ENGLISH-RUSSIAN PHRASEOLOGICAL DICTIONARY

Though British idiom dictionaries based on linguistic principles of analysis began to appear in the 1970s (Cowie and Mackin 1975; Long and Summers 1979), works reflecting the greater depth and precision of Russian phraseology did not emerge till the 1980s. In the meantime, foreign learners' reference needs were in practice often met by phrase books whose compilers had little idea of the limits of phraseology, and who were inclined to include 'alongside set-phrases [i.e. collocations and idioms] . . . free phrases and even separate words' (Ginzburg et al. 1979: 214). Before its radical restructuring by Seidl in 1978, *McMordie*'s widely used and much reprinted *English Idioms* displayed most of the deficiencies later noted by Ginzburg. Differences between one type of word-combination and another -- differences crucial to the foreign learner -- were not made explicit, and, in entry after entry, free phrases (lose a pencil, turn a wheel), restricted collocations (lose a leg, turn a somersault), and idioms (lose heart, turn one's coat) were presented with little consistency of arrangement and no explanation of the combinatorial problems involved (McMordie 1954).

What the learner needed, especially perhaps at the advanced level, were dictionaries in which theoretical principles for the choice and classification of phraseological units were rigorously worked out and applied. A. V. Kunin's *English-Russian Phraseological Dictionary* (*ERPD*), in its first edition of 1955, was the earliest dictionary of English phraseology to be based on such principles, and in its second (1967) and third (1984) editions it has grown in descriptive precision and authority.

In one respect, the ERPD is a conservative work, since it is both

strongly historical in its orientation and partly dependent for its coverage of present-day colloquialisms on British and American dictionaries. Its illustrative quotations range back to the nineteenth century and even earlier, and the work is predominantly literary in its choice of examples. Kunin's judgement of what is, or is not, acceptable current usage, British or American, is sometimes fallible, so that, while the picturesque expression *like a cat in a strange garret* is listed in *A Dictionary of Americanisms* (1951) as having originated in the USA, the phrase *a cat in gloves catches no mice,* though recorded in the Oxford *Dictionary of English Proverbs* (1970, s.v. *Cat in gloves* (or *gloved, muffled, muzzled cat) catches no mice, A),* has no illustrative example later than 1758. [6]

The true strength and originality of Kunin's dictionary lie elsewhere, in the principled selection and systematic analysis of entries. Items are chosen for inclusion strictly according to their membership of categories recognized by the compiler as phraseological. Of the various items incorporating edge, for instance, we notice that on the edge of smth., over the edge, and take the edge off smth. are figurative idioms ('semi-idioms' according to Kunin), since each has an extant literal as well as a figurative sense (cf. the Introduction to ERPD (1984: 14)). Kunin is careful too to limit coverage of restricted collocations (in his terminology 'phraseomatic units') to those which allow no variation (e.g. the naked truth) or minimal variation (e.g. ask/look for trouble).

Kunin's systematic handling of variation is an impressive feature of the dictionary. Where there is a basic stem which can occur unaltered, but in which substitution is permitted at one or more points, the stem is given in bold print and the substitutes in brackets:

(6) have (keep . . . put) **all one's eggs in one bas tread** (walk) **on** (upon) **egg-shells**

Where, however, variation involves more fundamental structural (and semantic) differences, the variants are regarded as separate though related expressions (members of the same 'phraseological series'), treated in distinct numbered entries and cross-referenced to each other:

(7) 118. **care killed a** (the) **cat** 190. **a cat has nine lives** 199. **a cat with nine lives** 208. **have as many lives as a cat**

Such fine-grained analysis is remarkable and admirable in a non-native

[6]References throughout this section are to the 1984 edition of *ERPD*.

lexicographer who, during the period of compilation of the first edition especially (the late 1940s and early 1950s), had little access to native informants and a severely restricted range of contemporary written data.

The examples quoted above are for the most part verb-phrase (or 'predicate') idioms functioning below the level of the sentence ('nominative' expressions, according to Gläser 1988a). But Kunin also includes various types of 'functional expressions' -- that is, expressions of sentence length (simple as in the case of *every bullet has its billet,* complex as in *it's as broad as it's long),* which function as proverbs, catchphrases, or slogans.

Despite its limitations, which arise chiefly from the difficulties experienced by the compiler in gaining access to up-to-date texts and, in particular, modern non-literary material, the *English-Russian Phraseological Dictionary* is a meticulous work of scholarship and a model of theory-driven lexicography.

5
A. P. COWIE, R. MACKIN, AND I. R. MCCAIG, OXFORD DICTIONARY OF CURRENT IDIOMATIC ENGLISH, VOL. ii

The two-volume *Oxford Dictionary of Current Idiomatic English* (*ODCIE*) (1975/ 1983) was the first large-scale phraseological dictionary of English to be compiled by native speakers. [7] It was also the first completed work of its kind in Britain to be based almost entirely on source materials written (and in some cases spoken) since the end of the Second World War . It was thus more up to date than even the 1984 edition of *ERPD* and more attentive to non-literary varieties of the language.

The second volume of *ODCIE* (much more than the first) was a theoretical hybrid. As we have already seen, its system of grammatical categories and codes belonged to a tradition stemming from the work of Palmer and Hornby. However, the dictionary clearly owes much to Russian influences as well (Cowie 1981). Though *ODCIE 2*'s compilers had no detailed knowledge of Kunin's two earlier editions, it is clear that the dictionaries have a common theoretical basis. *ODCIE 2*, a volume devoted to idioms and restricted collocations of a wide

[7]Vol. i (first published in 1975, with a second, retitled edition appearing in 1993) is devoted to phrasal verbs and related verbal idioms such as *blow off steam, take up the slack*.

range of structural and functional types, drew on Russian phraseological theory in two key respects. First it recognized and applied the framework of categories, derived from the work of Vinogradov and Amosova, that was described earlier in this chapter. This was the tripartite scheme of categories called in my own work pure idioms, figurative idioms, and restricted collocations. These were described and illustrated in some detail in the introduction to the volume, and provided the yardstick against which candidates for inclusion were assessed.

In the case of restricted collocations, only those items were included which were entirely invariable (e.g. *break one's journey, curry favour)* or which displayed limited collocability (e.g. *a chequered career/history, do the necessary/needful).* Thus, the dictionary did not go far beyond the limits for phraseology set by Amosova. A second feature which owed a good deal to Russian precedents was the inclusion of a broad range of 'functional' expressions, including catchphrases *(the buck stops here),* proverbs *(out of sight out of mind),* and expletives *(God damn it, get stuffed).*

Pure and figurative idioms, as Kunin had already shown in the earlier editions of his work, often displayed limited internal variation. Such variation was indicated in ODCIE 2, as earlier in ODCIE 1, by means of an oblique stroke, as in a *race/fight against time* and *the devil/Satan finds/makes work for idle hands.* But idioms can also form collocational ties outside their own strict limits with other sets of words. They may, that is, have their own, more or less restricted, collocational ranges (Mitchell 1971; Cowie 1981). The point can be illustrated by the figurative idiom *hold water,* which can collocate with the nouns *statement, description, account,* and *theory,* among others. These nouns could be listed in the appropriate entry, but it was felt that a simple device was needed in *ODCIE 2* to signal that they functioned as the grammatical subject of the idiom (as in the sentence *The theory doesn't hold water).* The device chosen was a function label (in this case **S**) inserted before the collocate list. The following partial entry shows how the conventions work in practice:

(8) **hold water** . . . be sound, valid . . . **S** : theory, argument; explanation, reason, excuse; belief, need . . .

As in a standard collocational dictionary, collocates such as these serve the needs of production ('encoding') by providing concentrated lexical information that can be used with confidence by a non-native speaker or writer. As far as I am aware,

this design feature has no parallel in any idiom dictionary produced in Russia, where the writer

or translator has often to depend on the uncertain support of example sentences.

6
C. D. KOZŁOWSKA AND H. DZIERÉANOWSKA, SELECTED ENGLISH COLLOCATIONS

Leaving aside Friederich and Canavan's *Dictionary of English Words in Context* (1979), which is not strictly a collocational dictionary, but one containing idioms too, in an often haphazard arrangement, Kozłowska and Dzieréanowska's *Selected English Collocations* (*SEC*) (first edition 1982) is the first dictionary worthy of its title to be compiled -- at least in part -- by an English native speaker. [8]

The dominant use of the collocational dictionary is as a tool for writing in, or translating into, the foreign language, a function which explains its characteristic structure and content. Typically, it lacks the information on meaning and style provided in general learners' dictionaries (and indeed idiom dictionaries); typically, too, it contains a bare sufficiency of grammatical information. Compilers assume that users possess this information already (or can find it elsewhere), and that they turn to the collocational dictionary simply for guidance on the combinatorial properties of words. The bare, skeletal appearance of entries in most dictionaries of the type is illustrated by the following extract from the first edition of *SEC*:

> (9) **DOUBT** V. add to, air, bear out, be shrouded in, cast aside, conceal, confirm, dismiss, dispel, dissipate . . . ~ V. ~ arise, be focused on, centre on, concern, creep in, crop up, disappear, flee, go away, grow, harass . . . Adj. broad, gnawing, grave, growing, intermittent, lingering, niggling, passing, persistent, profound . . . ~

All collocations listed in *SEC* contain a noun, and in every case the noun appears as the headword. This characteristic 'orientation', from noun headword to verb or adjective collocate, as in the entry shown, reflects the dominant 'encoding' function of the work and the recognition that, when writing, the direction of enquiry is from noun to verb or adjective, and not the other way about. It is clear in the case of *lingering doubt,* for example, that *doubt* is unproblematical, since it occurs with unaltered meaning in very many contexts; the problem is

[8]Christian Douglas Kozłowska is of Scottish origin.

one of choosing an acceptable adjective from an arbitrarily limited set (*lingering*, not *loitering; persistent,* not *persisting*) (Hausmann 1979; Cowie 1986; Cop 1988).

To enable the user to focus quickly on the part of the entry where the appropriate collocate is to be found, another design feature is required: classification by part of speech. In the above entry, differences between transitive and intransitive verbs correspond to differences of collocability, and collocates are accordingly arranged in blocks: transitive verbs in the first, where the swung dash marking the position of the noun follows; intransitives in the second, where the swung dash precedes. Adjectives appear in the third block.

The organization of SEC is well attuned to the encoding needs of the user. What of the lexical content? Here it seems that the compilers often provide suitable material despite considerable confusion as to what that material should be. Kozłowska and Dzierζanowska recognize the arbitrariness with which one word often selects its partner, yet claim that their dictionary 'does not give phraseological units (restricted collocations), but "free" or "open" collocations, which means that a range of other words can be added at will' (1988: 8). There is no misunderstanding of terms here: free collocability is what the authors say it is. And yet their dictionary -- fortunately for the user -- provides the opposite of what they claim it provides. On the one hand, it does contain such examples as *pay attention* and *drive a hard bargain* (restricted collocations in the narrowest sense). On the other, it does not provide free collocations as the term was understood by Palmer, Vinogradov, or Amosova. The recurrent free collocates of book include *open,close,read,* yet none of these is listed in SEC. Instead, we find verbs at the outer limits of the restricted range: digest, launch, *promote, review, scan.*

There is a lingering analytical problem which may have contributed to misunderstanding in this case. As we have seen, classical Russian theory stopped short of defining more than two bands of collocational restriction: one in which there was absolute limitation on choice (e.g. *grind one's teeth);* and one in which limited substitution was possible at one point (e.g. *pay one's respects/a compliment/court (to someone)).* This rather narrow view of restriction leaves out of account very many combinations which native speakers perceive as arbitrarily constrained (e.g. *digest a book),* so that it would hardly provide the basis for a viable collocational dictionary, and I have suggested here and elsewhere that collocations in which there is limited choice at two or more points are also restricted (cf. *give/deliver an address/speech/ lecture)* (Cowie 1991). This modification would, of course, have the effect of extending the restricted zone to include many combinations

now thought of as free (Howarth 1996). One benefit of such an extension would be to provide lexicographers with a principled means of deciding whether to include or exclude specific combinations.

In SEC, as in the various technical dictionaries discussed by Tomaszczyk (1981), collocates are not semantically grouped, and in an earlier critical analysis of SEC (Cowie 1986) I suggested that it was desirable to put together, within each grammatical class, words of related meaning, possibly introduced by superordinate terms as keywords. Introducing the adjective collocates of doubt, for instance, we might have PAINFUL followed by *gnawing* and *tormenting.* This device would enable the user armed with the meaning of the desired collocate to home in on the collocate itself. In the second (1988) edition of the dictionary its authors have introduced a convention which can sometimes speed up such retrieval. Noting that the collocability of a noun sometimes varies according to its 'countability', they have for many nouns

introduced two headwords, one marked [U] and the other [C], and divided up verbs and adjectives according to their collocability with uncountable and countable uses. However, collocates cannot always be sharply separated in this way, and the idea often breaks down in practice. Consider the following lists of adjective collocates for *conflict* [U] and [C]:

(10) **CONFLICT** [U] Adj. constant, direct, head-on, internecine, open ~ **CONFLICT** [C] Adj. acute, all-out, armed, bitter . . . ~

Here it is doubtful whether the two sets are entirely separate in collocability. Rather, because of the closeness of meaning of the nouns, there is collocational overlap. Compare the restrictions indicated above with these examples: *We are now heading for all-out, armed conflict, where conflict* is [U] and *You are risking a head-on conflict with his father, where conflict* is [C]. Both seem entirely acceptable.

SEC was influenced by the approach to describing restricted collocations advocated by Mel'čuk and Zholkovsky and applied in the West in the *Explanatory Combinatorial Dictionary of Modern Russian* (1984), as well as in all dictionaries of collocations compiled in Russia after 1970. [9] Key design features include: restriction of content to the purely lexical; an entry 'orientation' that serves the encoding function; and an entry structure designed to facilitate access to collocates. These are all helpfully in place. It has to be said, though, that much of the practical value of the work comes from the exceptional breadth, authenticity, and up to dateness of its example collocates, gathered by methods long established in British lexicography.

[9] I am grateful to Veronika Teliya for this information.

7
M. BENSON, E. BENSON, AND R. ILSON, THE BBI COMBINATORY DICTIONARY

The BBI Combinatory Dictionary of English (1986) is based not on examples collected from original sources, but on made-up material. This may explain the narrower range of examples often found in BBI entries (as in the case of *gift,* where BBI lists *give, present, heap, lavish on* but *SEC gives appreciate, decline, get, give, make, offer, present with, receive, refuse, reject,* and *send).* However, the organization of BBI is much superior to that of its older rival, so that, as well as providing collocations of a greater variety of structural types, it is more successful at directing the user to them.

Morton Benson, the senior editor of BBI, is a distinguished Slavist, and it is not surprising that he, like the SEC compilers, has adopted a number of basic structural features from Russian collocational dictionaries. As in SEC, the orientation of entries is one that favours the needs of the writer of English. So, for instance, the verb + noun collocations *display tact, exercise tact, show tact* can be formed by reference to the entry for the noun *tact.* Similarly, the adjective + noun combination *wild frenzy* is formed by reference to *frenzy.* Again, when an entry deals with transitive and intransitive uses of verbs, these are followed or preceded by a tilde indicating the position of the object or subject noun.

However, these comments give little idea of the greater complexity -- and greater multiplicity of sources -- of the BBI entry structure. BBI supplies information about word-combinations on several linguistic levels: it does not confine itself to lexical collocations, as SEC does. On one level it is a syntactic dictionary, supplying information about the complementation patterns of verbs (they enjoy watching television), nouns (an agreement that she would represent us), and adjectives (ready to leave). The complementation in each of these cases is a finite or non-finite clause -- in other words a grammatical category, not a lexical item -- and to call the pattern of which the complementation is part a 'grammatical' collocation is a misuse of the term, the more so as the dictionary lists a number of genuine grammatical collocations such as stray from (verb + preposition) and happy about (adjective + preposition). It is arguable too that the clause complementation of verbs and nouns should be treated in a valency dictionary, not a collocational dictionary.

Arranging lexical collocations (combinations of major-class words) according to their grammatical patterns is of course another matter, and here the *BBI* team is on firmer ground. Two points are worth

making about its approach to presenting lexical combinations as the exponents of specific constructions. The first is that all collocations are assigned in the introduction to a labelled pattern -- a procedure that clearly harks back to the grammatical codes and tables of Palmer and Hornby. Pattern LI, for instance, is transitive verb + noun (e.g. commit treason) and Pattern L3 is attributive adjective + noun (e.g. strong tea). Examples of the patterns are then consistently arranged in the various entries for which those types are appropriate. Thus, if a transitive and an adjectival collocation are called for in an entry, as they are at sentiment, the transitive one precedes:

(11) **sentiment** *n.* ['feeling'] 1. to express a~2. . . . 3. a growing; lofty; patriotic; shocking; strong~

The second point is that probably nothing has been lost by deciding (in the great majority of cases) not to include pattern codes in the entries themselves. In the entry for *sentiment,* syntax is clearly and economically conveyed through examples.

What of the nature of the examples themselves? How suitable are they for inclusion in a collocational dictionary? Unlike Kozłowska and Dzierzanowska, the *BBI* compilers define with a fair degree of accuracy the types of lexical word-combination that need to be included and excluded (though, like their competitors, not always picking them out in practice). Their definition of 'free combinations', for example, as consisting of 'elements that are joined in accordance with the general rules of English syntax and allow free substitution' (Benson et al. 1986: p. ix), is close to H. E. Palmer's. However, their description of collocations (i.e. restricted collocations) as 'fixed' and 'recurrent' reveals two widely shared misconceptions. As we have seen, not all combinations that show arbitrary restrictions are entirely fixed, while recent studies of formal (including academic) prose have shown that in those written styles at least, fixed collocations form a minority of all restricted collocations (Cowie 1991, 1992; Howarth 1996). As regards recurrence as a measure of restrictedness, there is no clear evidence to date of a close correlation between measured frequency of occurrence and

collocational restriction. Fruitful discussion of the issue is to some extent frustrated at present by the insistence of some scholars involved in the computerized analysis of large corpora that frequency of co-occurrence is the only significant measure of conventionality' in language (e.g. Sinclair 1991).

Uncertainty about the defining properties of restricted collocations may help to explain why BBI includes, alongside many combinations that are unquestionably restricted, a fair number of combinations that are free. So, for instance, at the entry for *book* we find *write* and

publish, which have literal senses and occur in many contexts, alongside *pirate* and *renew* (i.e. at a library), whose specialized meanings are contextually determined. Similarly, in the entry for *letter,* we find the free combination *receive a letter alongside the restricted collocation forward a letter.* Focusing specifically on the example *receive a letter,* Howarth offers an interesting explanation for the presence of such collocations in BBI (and to some extent in SEC). 'What do we do with letters?' he asks. 'Typically we write, read, send and receive them' (Howarth 1996: 182). In other words these are acts which contribute centrally to our understanding of the nature of letters (as losing them and tearing them up do not). This analysis, which draws on Mel'čuk's concept of 'lexical function', explains the inclusion in BBI of receive a letter and most of the other free collocations found in the dictionary.

8
CONCLUSION

This chapter has traced a relationship of growing complexity between two traditions in phraseology and a selection of dictionaries of idioms and collocations published in recent years. The survey has had a deliberately narrow focus, aiming to identify movements in theory which have already had a significant impact and which seem likely to stimulate further developments.

Russian phraseology in the 1950s and 1960s (to which Kunin also made an original contribution, which there has not been space to discuss here) is remarkable for having established systems of categories which enable the analyst, and of course the lexicographer, to separate the domain of phraseology from the field of operation of general rules of word-combination. Equally important has been the role of such schemes in helping us to distinguish idioms (aptly called 'fusions' and 'unities' in Vinogradov's terminology) from collocations (looser combinations in which one element determines the sense of the other). Though neither Vinogradov nor Amosova attempted to extend the subclassification of collocations beyond types in which there was a very limited determining context, they provided the means for others to do so. As we venture further into the fuzzy zone between the fixed and the free, we are in fact charting an area where most items which intuition assigns to the collocational dictionary are to be found.

This then is the range of categories which Russian analysts have set up, and it can be extended. But Russian phraseological theory has also left its mark on Western lexicography. Among the most interesting works to have emerged in recent years are those in which Russian

influences have merged with a British tradition deriving from H. E. Palmer and A. S. Hornby. The coming together of these traditions is most clearly seen in ODCIE 2 and BBI. In ODCIE 2, a Russianinspired system of phraseological types determines the outer limits of the work, keeping out all but the most restricted of collocations, while a Hornby-type scheme of grammatical patterns indicates the functions of entries and their transformational possibilities. In BBI, syntactic classification of items, of a kind already present in the Interim Report, is found within entries whose 'general orientation', following Russian models, is designed to meet the user's productive requirements.

This fruitful merging of traditions was under way even before the political and economic changes of the early 1990s began to facilitate closer and more regular contacts between phraseologists in Russia, Britain, and the USA. Collaborative research and cooperation in dictionary-making are now practical realities, and are already bearing fruit in a new generation of Russian-English phraseological dictionaries (Benson and Benson 1993; Kozyreva et al. 1995).

References

Aarts, J. (1991), "Intuition-Based and Observation-Based Grammars", in Aijmer and Altenberg (1991), 44-62.

----- de P. Haan, and Oostdijk, N. (1993) (eds.), *English Language Corpora: Design, Analysis and Exploration* (Amsterdam: Rodopi).

Aijmer, K., and Altenberg, B. (1991) (eds.), *English Corpus Linguistics: Studies in Honour of Jan Svartvik* (London: Longman).

Aisenstadt, E. (1979), "Collocability Restrictions in Dictionaries", in R. R. K. Hartmann (ed.), *Dictionaries and their Users* (Exeter Linguistic Studies 4; Exeter: University of Exeter), 71-4.

----- (1981), "Restricted Collocations in English Lexicology and Lexicography", ITL*(Instituut voor Toegepaste Linguistiek) Review of Applied Linguistics* 53: 53-61.

Aitchison, J. (1987), "Reproductive Furniture and Extinguished Professors", in Steele and Threadgold (1987), ii. 3-14.

Akhmanova, O. (1974) (ed.), *Word-Combination: Theory and Method* (Moscow: Moscow University Press).

Akimoto, M. (1983), *Idiomaticity* (Tokyo: *Shinozaki Shorin).*

Alexander, R. J. (1978), "Fixed Expressions in English: A Linguistic, Psycholinguistic, Sociolinguistic and Didactic Study (Part 1)", *Anglistik und Englischunterricht,* 6: 171-88.

----- (1987), "Problems in Understanding and Teaching Idiomaticity in English", *Anglistik und Englischunterricht,* 32: 105-22.

Allerton, D. (1984), "Three (or Four) Levels of Word Cooccurrence Restriction", *Lingua,* 63: 17-40.

Alonso M. Ramos, and Tutin, A. (1994), *"Les fonctions lexicales du Dictionnaire explicatif et combinatoire pour l'étude de la cohésion lexicale", Lingvisticae investigationes,* 17: 161-88.

----- ----- and Lapalme, G. (1995), "Lexical Functions of Explanatory Combinatorial Dictionary for Lexicalization in Text Generation", in Saint P. Dizier and E. Viegas (eds.), *Computational Lexical Semantics* (Cambridge: Cambridge University Press), 351-67.

Altenberg, B. (1990), "Speech as Linear Composition", in G. Caie*et al.* (eds.), *Proceedings from the Fourth Nordic Conference for English Studies* (Copenhagen: University of Copenhagen), 133-43.

----- (1991), "The London-Lund Corpus: Research and Applications", in *Using Corpora* (Oxford: UW Centre for the New OED), 7-83.

----- (1993), "Recurrent Verb-Complement Constructions in the LondonLund Corpus", in J. Aarts*et al.* (1993), 227-45.

Altenberg, B., and Eeg-Olofsson, M. (1990), "Phraseology in Spoken English: Presentation of a Project", in J. Aarts and W. Meijs (eds.), *Theory and Practice in Corpus Linguistics* (Amsterdam: Rodopi), 1-26.

Amosova, N. N. (1963), *Osnovui anglijskoy frazeologii* (Leningrad: University Press).

Anderson, D. (1969), "Harold E. Palmer: A Biographical Essay", in H. E. Palmer and H. V. Redman, *This Language-Learning Business* (Language and Language Learning 22; London: Oxford University Press), 133-66.

Apresyan, Y., Mel'uk, I., and Zholkovsky, A. (1969), "Semantics and Lexicography: A New Type of Unilingual Dictionary", in F. Kiefer (ed.), *Studies in Syntax and Semantics* (Dordrecht: Reidel), 1-33.

Apter, M. (1982), "Metaphor and Synergy", in D. S. Miall (ed.), *Metaphor. Problems and Perspectives* (Brighton: Atlantic Highlands), 55-70.

Arnaud, P., and Béjoint, H. (1992) (eds.), *Vocabulary and Applied Linguistics* (London: Macmillan).

----- and Moon, R. E. (1993), *"Fréquence et emploi des proverbes anglais et français",* in C. Plantin (ed.), *Lieux communs: Topoï, stéréotypes, clichés* (Paris: Kime), 323-41.

Arnold, I. V. (1973), *The English Word* (Moscow: Vuisshaya Shkola).

----- (1986), *The English Word* (3rd edn., Moscow: Vuisshaya Shkola).

Arutyunova, N. D. (1991), "Editor's Note", in N. D. Arutyunova (ed.), *Kulturnuie kontseptui: A Collection of Papers* (Moscow: Nauka), 3-4.

Atkins, B. T. S. (1991), "Building a Lexicon: The Contribution of Lexicography", *International Journal of Lexicography*, 4/ 3: 167-204.

----- (1992), "Tools for Computer-Aided Lexicography: The Hector Project", in Kiefer*et al.* (1992), 1-59.

----- and Duval, A. (1978), *Robert-Collins dictionnaire français-anglais, anglais-français* (2nd edn., 1987) (Paris: Le Robert; Glasgow: Collins).

----- and Levin, B. (1991), "Admitting Impediments", in Zernik (1991), 233-62.

Bäcklund, U. (1973), *The Collocations of Adverbs of Degree in English* (Studia Anglistica Upsaliensia 13; Uppsala: Uppsala University Press).

Bahns, J. (1993), "Lexical Collocations: A Contrastive View", *ELTJ* 47/ 1: 56-63.

----- and Eldaw, M. (1993), "Should We Teach EFL Students Collocations?", *System*, 21/ 1: 1-14.

Baker, M., Francis, G., and Tognini-Bonelli, E. (1993) (eds.), *Text and Technology: In Honor of_John Sinclair* (Philadelphia: John Benjamins).

Bally, C. (1909), *Traité de stylistique française* (3rd edn., 1951) (Paris: Leroux).

Bartminski, J. (1993), *"O profilovaniu pojeg w slowniku etnolingwistycznym",* in *Profilowane polec* (Lublin).

Becker, J. D. (1975), "The Phrasal Lexicon", in R. Schank and B. Nash Webber (eds.), *Proceedings of Interdisciplinary Workshop on Theoretical Issues in Natural Language Processing* (Cambridge, Mass.: Bolt, Bernback & Newman), 70-3.

Béjoint, H., and Thoiron, P. (1992), *"Macrostructure et microstructuredans un dictionnaire de collocations en langue de specialité"*

dans un dictionnaire de collocations en langue de specialité', *Terminologie et traduction,* 2/ 3: 512-22.

Benson, M. (1985), "Collocations and Idioms", in R. Ilson (ed.), *Dictionaries, Lexicography and Language Learning* (ELT Documents 120; Oxford: Pergamon), 61-8.

----- (1988), "Soviet Lexicography: A Survey", in Snell-Homby (1988), 217-27.

----- (1989a), "The Structure of the Collocational Dictionary", *International Journal of Lexicography*, 2/ 1: 1-14.

----- (1989b), "A Collocational Dictionary of Russian", *Slavic and East European Journal*, 33/ 4: 593-608.

----- and Benson, E. (1993), *Russian-English Dictionary of Verbal Collocations* (Amsterdam: John Benjamins).

----- ----- and Ilson, R. (1986), *The BBI Combinatory Dictionary of English: A Guide to Word Combinations* (Amsterdam: John Benjamins).

Black, M. (1979), "More about Metaphor", in A. Ortony (ed.), *Metaphor and Thought* (Cambridge: Cambridge University Press), 499-520.

Bloch, H. (1989), *Phraseologische Einheiten, ihre Struktur und Funktion in der amerikanischen Kunstreportage* (student dissertation, University of Leipzig).

Boatner, M. T., Gates, J. E., and Makkai, A. (1966), *A Dictionary of American Idioms* (2nd edn. revised by A. Makkai 1987) (New York: Barron's).

Boguraev, B., and Briscoe, T. (1989) (eds.), *Computational Lexicography for Natural Language Processing* (London: Longman).

Bolinger, D. (1972), *Degree Words* (The Hague: Mouton).

----- (1976), "Meaning and Memory", *Forum linguisticum*, 1/ 1: 1-14.

----- (1977), "Idioms have Relations", *Forum linguisticurn*, 2/ 2: 157-69.

Bongers, H. (1947), *The History and Principles of Vocabulary Control* (2 vols.; Woerden: Wocopi).

Braine, J. (1957). *Room at the Top* (London: Eyre & Spottiswoode).

Bürli-Storz, C. (1980), *Deliberate Ambiguity in Advertising* (Bern: Franke Verlag).

by rd, R. (1989), "Discovering Relationships among Word Senses", in *Dictionaries in the Electronic Age* (Oxford: UW Centre for the New OED), 67-79.

----- Calzolari, N., Chodorow, M., Klavans, J., Neff, M., and Rizk, O. (1987), "Tools and Methods for Computational Lexicology", *Computational Linguistics*, 13/ 3-4: 219-40.

Carter, R. (1987), *Vocabulary: Applied Linguistics Perspectives* (London: Unwin Hyman).

----- and McCarthy, M. J. (1988) (eds.), *Vocabulary and Language Teaching* (London: Longman).

Cattell, R. (1984), *Composite Predicates in English* (Sydney: Academic Press).

Chafe, W. (1968), "Idiomaticity as an Anomaly in the Chomskyan Paradigm", *Foundations of Language*, 4: 109-27.

----- (1987), "Cognitive Constraints on Information Flow", in R. S. Tomlin (ed.), *Coherence and Grounding in Discourse* (Amsterdam: John Benjamins), 21-51.

Chernuisheva, I. I. (1964), *Die Phraseologie der gegenwärtigen deutschen Sprache* (Moscow: Vuisshaya Shkola).

Church, K., and Hanks, P. (1990), "Word Association Norms, Mutual Information and Lexicography", *Computational Linguistics*, 16/ 3: 22-9.

Clear, J. (1993), "From Firth Principles: Computational Tools for the Study of Collocation", in Baker*et al.* (1993), 271-92.

Cohen, B. (1993), *"Méthodes de repérage et de classement des cooccurrents lexicaux", Terminologie et traduction,* 2/ 3: 505-11.

Cohen, J. (1972), *Psychology -- an Outline for the Intending Student* (London: Routledge & Kegan Paul).

Cop, M. (1988), "The Function of Collocations in Dictionaries", in T. Magay and J. Zigány (eds.), *BudaLEX '88 Proceedings: Papers from the Euralex Third International Congress* (Budapest: Akadémiai Kiadó), 35-46.

Corpas G. Pastor (1996), *Manual defraseología española* (Madrid: Gredos).

Cowie, A. P. (1981), "The Treatment of Collocations and Idioms in Learners' Dictionaries", *Applied Linguistics*, 2/ 3: 223-35.

----- (1986), "Collocational Dictionaries -- a Comparative View", in Murphy (1986), 61-9.

----- (1988), "Stable and Creative Aspects of Vocabulary Use", in Carter and McCarthy (1988), 126-37.

----- (1989a), "Information on Syntactic Constructions in the General Monolingual Dictionary", in Hausmann*et al.* (1989), 588-92.

----- (1989b) (ed.), *Oxford Advanced Learner's Dictionary* (4th edn., Oxford: Oxford University Press).

----- (1990), "Language as Words: Lexicography", in N. E. Collinge (ed.), *An Encyclopaedia of Language* (London: Routledge), 671-700.

----- (1991), "Multiword Units in Newspaper Language", in S. Granger (ed.), *Perspectives on the English Lexicon: A Tribute to Jacques van Roey (Louvain-la-Neuve: Cahiers de l'Institut de Linguistique de Louvain),* 101-16.

----- (1992), "Multiword Lexical Units and Communicative Language Teaching", in Arnaud and Béjoint (1992), 1-12.

----- (1994), "Phraseology", in R. E. Asher and J. Simpson (eds.), *The Encyclopedia of Language and Linguistics* (Oxford: Pergamon Press), 3168-71.

----- and Howarth, P. (1996), "Phraseological Competence and Written Proficiency", in G. M. Blue and R. Mitchell (eds.), *Language and Education* (British Studies in Applied Linguistics 11; Clevedon: Multilingual Matters), 80-93.

----- ----- (with Corpas G. Pastor) (1996), "Phraseology -- a Select Bibliography", *International Journal of Lexicography*, 9/ 1: 38-51.

----- and Mackin, R. (1975), *Oxford Dictionary of Current Idiomatic English*, i (2nd edn., retitled Oxford Dictionary of Phrasal Verbs 1993) (London: Oxford University Press).

----- ----- and McCaig, I. (1983), Oxford Dictionary of Current Idiomatic English, ii (retitled *Oxford Dictionary of English Idioms 1993*) (Oxford: Oxford University Press).

Cruse, D. A. (1986), *Lexical Semantics* (Cambridge Textbooks in Linguistics; Cambridge: Cambridge University Press).

Danlos, L., and Samvelian, P. (1992), "Translation of the Predicative Element of a Sentence: Category Switching, Aspect and Diathesis", in *TMI-92* (Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation) (Montréal: Les Presses de l'Université de Montréal), 21-34.

Dechert, H. (1984), "Second Language Production: Six Hypotheses", in H. Dechert , D. Mohle, and M. Raupach (eds.), *Second Language Productions* (Tübingen: Gunter Narr Verlag), 211-30.

----- and Lennon, P. (1989), "Collocational Blends of Advanced Second Language Learners: A Preliminary Analysis", in W. Olesky (ed.), *Contrastive Pragmatics* (Amsterdam: John Benjamins), 131-68.

Dixon, R. M. W. (1991), *A New Approach to English Grammar, on Semantic Principles* (Oxford: Clarendon Press).

Dunning, T. (1992), "Accurate Methods for the Statistics of Surprise and Coincidence", *Report No. MCCS-91-223: Computing Research Laboratory* (Las Cruces, N. Mex.: New Mexico State University).

Eeg-Olofsson, M., and Altenberg, B. (1994), "Discontinuous Recurrent Word-Combinations in the London-Lund Corpus", in U. Fries, G. Tottie , and P. Schneider (eds.), *Creating and Using English Language Corpora* (Amsterdam: Rodopi), 63-77.

Fernando, C., and Flavell, R. (1981), *On Idiom: Critical Views and Perspectives* (Exeter Linguistic Studies 5; Exeter: University of Exeter Press).

Fillmore, C. J., and Atkins, B. T. S. (1994), "Starting Where the Dictionaries Stop: The Challenge of Corpus Lexicography", in B. T. Atkins and A. Zampolli (eds.), *Computational Approaches to the Lexicon* (Oxford: Oxford University Press), 349-93.

----- Kay, P., and O'Connor, M. C. (1988), "Regularity and Idiomaticity in Grammatical Constructions: The Case of Let Alone", *Language*, 64/ 3: 501-38.

Fleischer, W. (1982), *Phraseologie der deutschen Gegenwartssprache* (Leipzig: VEB Verlag Enzyklopädie).

Fontenelle, T. (1992a), *Collocation Acquisition from a Corpus or from a Dictionary: A Comparison*, in Tommola*et al.* (1992), 220-8.

----- (1992b), *Co-occurrence Knowledge, Support Verbs and MachineReadable Dictionaries*, in Kiefer*et al.* (1992), 137-45.

----- (1992c), "Using a Bilingual Computerized Dictionary to Retrieve Support Verbs and Combinatorial Information", *Acta linguistica hungarica,* 41/ 1-4: 109-21.

----- (1994), "Towards the Construction of a Collocational Database for Translation Students", *META* 39/ 1: 47-56.

----- (1997), *Turning a Bilingual Dictionary into a Lexical-Semantic Database* (Lexicographica, Series major 79; Tübingen: Max Niemeyer).

Francis, G. (1993), *A Corpus-Driven Approach to Grammar*, in Baker*et al.* (1993), 137-56.

Fraser, B. (1970), "Idioms within a Transformational Grammar", *Foundations of Language*, 6: 22-42.

Friederich, W., and Canavan, J. (1979), *Dictionary of English Words in Context* (Dortmund: Verlag Lambert Lensing).

Gabrys-Biskup, D. (1992), "L1 Influence on Learners Renderings of English Collocations: A Polish/German Empirical Study", in *Arnaud and Béjoint* (1992), 85-93.

Galperin, I. R. (1977), *Stylistics* (2nd edn., Moscow: Vysshaya Shkola).

Gibbs, R. W. (1980), "Spilling the Beans on Understanding and Memory for Idioms in Conversation", *Memory and Cognition*, 8: 149-56.

----- (1986), "Skating on Thin Ice: Literal Meaning and Understanding Idioms in Conversation", *Discourse Processes*, 9: 17-30.

Ginzburg, R. S., Khidekel, S. S., Knyazeva, G. Y., and Sankin, A. A. (1979), *A Course in Modern English Lexicology* (2nd edn., Moscow: Vuisshaya Shkola).

Gläser, R. (1986a), *Phraseologie der englischen Sprache* (2nd edn., 1990) (Leipzig: VEB Verlag Enzyklopädie).

----- (1986b), "A Plea for Phraseo-Stylistics", in D. Kastovsky and A. Szwedek (eds.), *Linguistics across Historical and Geographical Boundaries*. In Honour of Jacek Fisiak, i. Linguistic Theory and Historical Linguistics (Berlin: Mouton de Gruyter), 41-52.

----- (1986c), *"Wissenschaftssprache in der Erwachsenenhildung -dargestellt an Lehrbriefen der Open University in Grossbritannien",* in T. Bungarten (ed.), *Wissenschaftssprache und Gesellschaft: Aspekte der wissenschaftlichen Kommunikation und des Wissentransfers in der heutigen Zeit* (Hamburg: Edition Akademion), 333-49.

----- (1988a), "The Grading of Idiomaticity as a Presupposition for a Taxonomy of Idioms", in W. Hüllen and R. Schulze (eds.), *Understanding the Lexicon: Meaning, Sense and World Knowledge in Lexical Semantics* (Tübingen: Max Niemeyer), 264-79.

----- (1988b), "Variationen von Phraseologismen im Text", *Wissenschaftliche Zeitschrift der Wilhelm-Pieck-Universität Rostock: Theoretische und Angewandte Sprachwissenschaft,* 37: 48-52.

----- (1990), *Fachtextsorten im Englischen* (*Forum für Fachsprachenforschung* 13; Tübingen: Gunter Narr Verlag).

Glassman, L., Grinberg, D., Hibbard, C., Meehan, J., Reid, L. G., and van M.-C. Leunen (1992), *Hector. Connecting Words with Definitions* (SRC Report 92a; Palo Alto, Calif.: Digital Equipment Corporation Systems Research Center).

Goffman, E. (1981), *Forms of Talk* (Oxford: Basil Blackwell).

Granger, S. (1993), *The International Corpus of Learner English*, in Aarts *et al.* (1993), 57-71.

----- (1994), "The Learner Corpus: A Revolution in Applied Linguistics", *English Today*, 10/ 3: 25-9.

----- (1996), "From CA to CIA and back: An Integrated Contrastive Approach to Bilingual and Learner Computerized Corpora", in K. Aijmer, B. Altenberg, and M. Johansson (eds.), *Languages in Contrast: Text-Based Cross-Linguistic Studies*

Cross-Linguistic Studies (Lund Studies in English 88; Lund: Lund University Press), 37-51.

----- and Tyson, S. (1996), "Connector Usage in the English Essay Writing of Native and Non-Native EFL Speakers of English", *World Englishes*, 15/ 1: 17-27.

Gréciano, G. (1983), *Signification et dénotation en allemand: La Sémantique des expressions idiomatiques* (Recherches linguistiques 9; Paris: Klincksieck).

Greenbaum, S., and Svartvik, J. (1990), "The London-Lund Corpus of Spoken English", in J. Svartvik (ed.), *The London-Lund Corpus of Spoken English: Description and Research* (Lund: Lund University Press), 11-45.

Grefenstette, G. (1994), *Corpus-Derived First, Second and Third-Order Word Affinities*, in Martin*et al.* (1994), 279-90.

Grimes, ' J. (1990), "Inverse Lexical Functions", in J. Steele (ed.), *MeaningText Theory:, Linguistics, Lexicography and Implications* (Ottawa: University of Ottawa Press), 350-64.

Gross, M. (1981), *"Les Bases empiriques de la notion de prédicat sémantique", Langage*, 63: 7-52.

Guthrie, J., Guthrie, L., Wilks, Y., and Aidinejad, H. (1991), "SubjectDependent Cooccurrence and Word Sense Disambiguation", in *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics* (Berkeley and Los Angeles: University of California Press), 146-52.

Halliday, M. A. K. (1966), "Lexis as a Linguistic Level", in C. Bazell, J. Catford, M. A. K. Halliday, and R. Robins (eds.), *In Memory of J. R. Firth* (London: Longman), 148-62.

----- (1985), *An Introduction to Functional Grammar* (London: Edward Arnold).

----- and James, Z. L. (1993), "A Quantitative Study of Polarity and Primary Tense in the English Clause", in J. M. Sinclair, M. Hoey, and G. Fox (eds.), *Techniques of Description: Spoken and Written Discourse: A Festschrift for Malcolm Coulthard* (London: Routledge), 32-66.

Hanania, E., and Gradman, H. (1977), "Acquisition of English Structures: A Case Study of an Adult Native Speaker of Arabic in an EnglishSpeaking Environment", *Language Learning*, 27/ 1: 75-91.

Hausmann, F. J. (1979), *"Un dictionnaire des collocations est-il possible?", Travaux de linguistique et de littérature,* 17: 187-95.

----- (1985), *"Kollokationen im deutschen Wörterbuch: Ein Beitrag zur Theorie des lexikographischen Beispiels",* in H. Bergenholtz and J. Mugdan (eds.), *Lexikographie und Grammatik* (Tübingen: Max Niemeyer), 118-29.

----- (1989), *Le Dictionnaire de collocations,* in Hausmann*et al.* (1989-91), 1010-19.

----- Reichmann, O., Wiegand, H. E., and Zgusta, L. (1989-91), *Wörterbücher/Dictionaries/Dictionnaires,* 3 vols. (Berlin: Walter de Gruyter).

Healey, A. (1968), "English Idioms", *KIVUNG* 1/ 2: 71-108.

Heid, U. (1992), *"Décrire les collocations: Deux approches lexicographiques et leur application dans un outil informatisé"*

et leur application dans un outil informatisé', *Terminologie et traduction,* 2/ 3: 523-48.

Heid, U., and Freibott, G. (1991), *"Collocations clans une base de données terminologique et lexicale", META* 36/ 1: 77-91.

Heylen, D. (1993), *Collocations: Final Report of the ET-10/75 Study on the Lexicalisation of Semantic Operations* (Utrecht: Utrecht University, Stichting Taaltechnologie).

Hockett, C. (1958), *A Course in Modern Linguistics* (New York: Macmillan).

Holmes, J. (1984), "Modifying Illocutionary Force", *Journal of Pragmatics*, 8: 345-65.

Hornby, A. S., Gatenby, E. V., and Wakefield, H. (1942), *Idiomatic and Syntactic English Dictionary* (published by Oxford University Press, 1948, as A Learner's Dictionary of Current English; later, in 1952, retitled The Advanced Learner's Dictionary of Current English) (Tokyo: Kaitakusha).

Howarth, P. (1993), "A Phraseological Approach to Academic Writing", in G. Blue (ed.), *Language Learning and Success: Studying through English* (London: Macmillan), 58-69.

----- (1996), *Phraseology in English Academic Writing: Some Implications for Language Learning and Dictionary Making* (Lexicographica, Series maior 75; Tübingen: Max Niemeyer).

Hutchinson, T., and Waters, A. (1987), *English for Specific Purposes: A Learning-Centred Approach* (Cambridge: Cambridge University Press).

Ilgenfritz, P., Stephan-Gabinel, N., and Schneider, G. (1989), *Langenscheidts Kontextwörterbuch französich-deutsch* (Berlin: Langenscheidt).

Ilson, R., and Mel'čuk, I. (1989), "English BAKE Revisited (BAKE-ing an ECD)", *International Journal of Lexicography*, 2/ 4: 325-45.

Iordanskaya, L., Kim, M., and Polguère, A. (1994), "Some Procedural Problems in the Implementation of Lexical Functions for Text Generation"', in L. Wanner (ed.), *Lexical Functions* (Amsterdam: Benjamins, 1994).

Jackendoff, R. (1995), "The Boundaries of the Lexicon", in M. Everaert, E. J. van der Linden, A. Schenk, and R. Schreuder (eds.), *Idioms: Structural and Psychological Perspectives* (Hillsdale, NJ: Lawrence Erlbaum Associates), 133-69.

Johansson, S. (1991), *'Times Change, and so do Corpora'*, in Aijmer and Altenberg (1991), 305-14.

Kant, I. (1994), *Kritika chistogo razuma* (Moscow: Mysl').

Katz, J., and Postal, P. (1963), "Semantic Interpretation of Idioms and Sentences Containing Them", *MIT Research Laboratory of Electronics: Quarterly Progress Report*, 70: 275-82.

Keith, H. (1983), *Processes and Patterns in Physical Geography* (London: University Tutorial Press).

Keller, E. (1979), "Gambits: Conversational Strategy Signals", *Journal of Pragmatics*, 3/ 3: 219-38.

Kiefer, F., Kiss, G., and Pajzs, J. (1992) (eds.), *Papers in Computational Lexicography: Complex '92* (Budapest: Linguistics Institute, Hungarian Academy of Sciences).

Kjellmer, G. (1991), "A Mint of Phrases", in Aijmer and Altenberg (1991), 111-27.

----- (1994), *A Dictionary of English Collocations: Based on the Brown Corpus* (Oxford: Clarendon Press).

Klappenbach, R. (1968), "Probleme der Phraseologie", *Wissenschaftliche Zeitschrift der Karl-Marx-Universität*, 17/5: 221-7.

Kozlowska, C., and Dzieranowska, H. (1982), *Selected English Collocations* (Warsaw: Pafistwowe Wydawnictwo

Naukowe).

----- ----- (1988), *Selected English Collocations* (2nd edn., Warsaw: Państwowe Wydawnictwo Naukowe).

Kozyreva, M., Davies, R., and Johnson, C. (1995), *Russian-English Collocational Dictionary* (Moscow: Russkii yazyk).

Krashen, S., and Scarcella, R. (1978). "On Routines and Patterns in Language Acquisition and Performance", *Language Learning*, 28/2: 283-300.

Kunin, A. V. (1955), *English-Russian Phraseological Dictionary* (2nd edn., 1967; 3rd edn., 1984) (Moscow: Russkii yazyk).

Lakoff, G., and Johnson, M. (1980), *Metaphors We Live By* (Chicago: University of Chicago Press).

Lewis, M. (1993), *The Lexical Approach: The State of ELT and a Way Forward* (London: Language Teaching Publications).

Lindner, C. (1994), "Unnaturalness in Advanced Learners" English: A Corpus-Based Feasibility Study', unpublished MA dissertation (AlbertLudwigs-Universität Freiburg).

Lipka, L. (1974), "Probleme der Analyse englischer Idioms aus struktureller und generativer Sicht", *Linguistik und Didaktik*, 20: 274-85.

----- (1991), *An Outline of English Lexicology* (Forschung & Studium Anglistik 3) (2nd edn., 1992) (Tübingen: Max Niemeyer).

Long, T. H., and Summers, D. (1979) (eds.), *Longman Dictionary of English Idioms* (London: Longman).

Lotman, Y. M. (1992), *Izbrannuie stat'i*, i (Tallinn: Aleksandra).

McCarthy, M. J. (1990), *Vocabulary* (Oxford: Oxford University Press).

Mackin., R. (1978), "On Collocations: Words shall be known by the company they keep", in P. Strevens (ed.), *In Honour of A. S. Homby* (Oxford: Oxford University Press), 149-65.

McMordie, W. (1954), *English Idioms and How to Use them* (London: Oxford University Press).

----- (1978), *English Idioms and How to Use them*, ed. J. Seidl (4th rev. edn., London: Oxford University Press).

Makkai, A. (1972), *Idiom Structure in English* (Janua linguarum, Series maior 48; The Hague: Mouton).

Martin, W., Meys, W., Moerland, M., ten E. Pas, van P. Sterkenburg, and Vossen, P. (1994) (eds.), *Euralex 1994 Proceedings* (Amsterdam: Euralex).

Mathews, M. M. (1951), *A Dictionary of Americanisms on Historical Principles* (Chicago: University of Chicago Press).

Mel'čuk, I. (1974), *Opyt teorii lingvisticheskikh modelei 'Smysl ⇔ Tekst' [An Outline of the Meaning-Text Theory of Linguistic Models]* (Moscow: Nauka).

Mel'čuk, I. (1981), "Meaning-Text Models: A Recent Trend in Soviet Linguistics", *Annual Review of Anthropology*, 10: 27-62.

----- (1982), *Towards a Formal Language of Linguistics* (Munich: W. Fink Verlag).

----- (1988a), *Dependency Syntax: Theory and Practice* (Albany, NY: SUNY Press).

----- (1988b), "Semantic Description of Lexical Units in an Explanatory Combinatorial Dictionary: Basic Principles and Heuristic Criteria", *International Journal of Lexicography*, 1/3: 165-88.

----- (1989), "Semantic Primitives from the Viewpoint of the Meaning-Text Linguistic Theory", *Quaderni di semantica*, 10: 65-102.

----- (1992a), "CHANGER et CHANGEMENT en français contemporain (étude sémantico-lexicographique)", *Bulletin de la Société de linguistique de Paris*, 87:161-223.

----- (1992b), "Paraphrase et lexique: La Thfiorie Sens-Texte et le Dictionnaire explicatif et combinatoire", in Mel'čuk *et al.* (1992), 9-58.

----- (1993), *Cours de morphologie générale*, i (Montréal: Les Presses de l'Université de Montréal).

----- (1996), "Lexical Functions: A Tool for the Description of Lexical Relations in a Lexicon", in Wanner (1996), 37-102.

----- and Polguère, A. (1987), "A Formal Lexicon in the Meaning-Text Theory (or How to Do Lexica with Words)", *Computational Linguistics*, 13: 261-75.

----- and Wanner, L. (1994), "Towards an Efficient Representation of Restricted Lexical Cooccurrence", in Martin *et al.* (1994), 325-38.

----- (1996), "Lexical Functions and Lexical Inheritance for Emotion Lexemes in German", in Wanner (1996), 209-78.

----- and Zholkovsky, A. (1984), *Explanatory Combinatorial Dictionary of Modern Russian* (Vienna: Wiener slawistischer Almanach).

----- ----- (1988), "The Explanatory Combinatorial Dictionary", in M. Evens (ed.), *Relational Models of the Lexicon* (Cambridge: Cambridge University Press), 41-74.

----- Arbatchewsky-Jumarie, N., Elnitsky, L., Iordanskaja, L., and Lessard, A. (1984), *Dictionnaire explicatif et combinatoire du franvais contemporain: Recherches lexico-sémantiques*, vol. i (Montréal: Les Presses de l'Université de Montréal).

----- ----- Dagenais, L., Elnitsky, L., Iordanskaja, L., Lefebvre, M.-N., and Mantha, S. (1988), *Dictionnaire explicatif et combinatoire du français contemporain: Recherches lexico-sémantiques*, vol. ii (Montréal: Les Presses de l'Université de Montréal).

----- ----- Iordanskaja, L., and Mantha, S. (1992), *Dictionnaire explicatif et combinatoire du français contemporain: Recherches lexico-sémantiques*, vol. iiii (Montréal: Les Presses de l'Université de Montréal).

Michiels, A. (1975), "Idiomaticity in English", *Revue des langues vivantes*, 43/2: 184-99.

----- (1982), "Exploiting a Large Dictionary Database", unpublished Ph.D. thesis (Université de Liège).

Mitchell, T. F. (1971), "Linguistic Goings-On: Collocations and other Matters Axising on the Syntagrnatic Record", *Archivum linguisticuln*, 2: 3-69.

Mohle, D., and Raupach, M. (1989), "Language Transfer of Procedural Knowledge", in H. Dechert and M. Raupach (eds.), *Transfer in Language Production* (Norwood, NJ: Ablex Publishing Corporation), 195-216.

Moon, R. E. (1988), "Time and Idioms", in Snell-Homby (1988), 107-15.

----- (1992a), "Textual Aspects of Fixed Expressions in Learners" Dictionaries', in Arnaud and Béjoint (1992), 13-27.

----- (1992b), ' "There is Reason in the Roasting of Eggs": A Consideration of Fixed Expressions in Native-Speaker Dictionaries', in Tommola *et al.* (1992), 493-502.

----- (1998), *Fixed Expressions and Text: A Study of the Distribution and Textual Behaviour of Fixed Expressions in English* (Oxford Studies in Lexicology and Lexicography; Oxford: Clarendon Press).

Morgan, J. (1978), "Two Types of Convention in Indirect Speech Acts", in P. Cole (ed.), *Syntax and Semantics, ix: Pragmatics* (New York: Academic Press), 261-80.

Morozenko, V. V. (1986), "Dictionaries for Teaching Reading Comprehension, and their Classification", in Murphy (1986), 70-80.

Müller, R. (1993), *Phraseologismen in englischen Fachtexten der Humanmedizin* (Leipziger Fachsprachen-Studien 6; Frankfurt am Main: Peter Lang).

Murphy, M. J. (1986) (ed.), *Proceedings of the Fourth Joint Anglo-Soviet Seminar on English Studies* (London: The British Council).

Nattinger, J. R. (1988), "Some Current Trends in Vocabulary Teaching", in Carter and McCarthy (1988), 62-82.

----- and DeCarrico, J. S. (1992), *Lexical Phrases and Language Teaching* (Oxford: Oxford University Press).

Newmeyer, F. J. (1974), "The Regularity of Idiom Behaviour", *Lingua*, 34: 327-42.

Norrick, N. R. (1985), *How Proverbs Mean: Semantic Studies in English Proverbs* (Berlin: Mouton).

Nunberg, G., Sag, I., and Wasow, T. (1994), "Idioms", *Language*, 70/3: 491-538.

Palmer, H. E. (1933), *Second Interim Report on English Collocations* (Tokyo: Kaitakusha).

----- (1938), *A Grammar of English Words* (London: Longmans Green).

----- (1942), "Category Two of English Linguistic Symbols: Nature and Extent of an Enquiry now Proposed" (unpublished memorandum).

Pawley, A. (1985), "On Speech Formulas and Linguistic Competence", *Lenguas modernas* (Chile) 12: 84-104.

----- and Syder, F. H. (1983), "Two Puzzles for Linguistic Theory: Nativelike Selection and Nativelike Fluency", in J. C. Richards and R. W. Schmidt (eds.), *Language and Communication* (London: Longman), 191-225.

Peters, A. (1977), "Language Learning Strategies: Does the Whole Equal the Sum of the Parts"?, *Language*, 53/3: 560-73.

Peters, A. (1983), *The Units of Language Acquisition* (Cambridge: Cambridge University Press).

Polzer, K. (1989), "Phraseologismen in englischen Werbetexten", student dissertation (University of Leipzig).

Potebnya, A. A. (1905), *Iz zapisok po uorii slovesnosti* (Khar'kov: M.V. Potebnya).

Procter, P. (1978) (ed.), *Longman Dictionary of Contemporary English* (3rd edn., 1995; London: Longman).

Pustejovsky, J., Bergler, S., and Anick, P. (1993), "Lexical Semantic Technique for Corpus Analysis", *Computational Linguistics*, 19/2: 331-58.

Quirk, R., and Mulholland, J. (1964), "Complex Prepositions and Related Sequences", *English Studies*, 45: 64-7.

------ Greenbaum, S., Leech, G., and Svartvik, J. (1985), *A Comprehensive Grammar of the English Language* (London: Longman).

Raupach, M. (1983), "'Analysis and Evaluation of Communication Strategies", in C. Faerch and G. Kasper (eds.), *Strategies in Interlanguage Communication* (London: Longman), 199-209.

Renouf, A., and Sinclair, J. (1991), "Collocational Frameworks in English", in Aijmer and Altenberg (1991), 128-43.

Richards, J. C. (1983), "Communicative Needs in Foreign Language Learning", *ELTJ* 37/2: 111-20.

Ridout, R., and Witting, C. (1969), *English Proverbs Explained* (London: Pan Books).

Rose, J. H. (1978), "Types of Idioms", *Linguistics*, 203: 55-62.

Rudzka, B., Channell, J., Ostyn, P., and Putseys, Y. (1981), *The Words You Need* (London: Macmillan).

----- ----- ----- ----- (1985), *More Words You Need* (London: Macmillan).

Ruhl, C. E. (1975), "Primary Verbs", in *The 1st LACUS Forum* (Columbia, SC: Hornbeam), 436-45.

----- (1978), "Alleged Idioms with hit", in *The 5th LACUS Forum* (Columbia, SC: Hornbeam), 93-107.

Sapir, E. (1964), "Conceptual Categories in Primitive Languages", in D. Hymes (ed.), *Language in Culture and Society* (New York: Harper & Row), 128-50.

Schiffrin, D. (1987), *Discourse Markers* (Cambridge: Cambridge University Press).

Scholz, C. (1986), "Phraseologische Einheiten, ihre Struktur und Funktion im englischen Gegenwartsroman", student dissertation (University of Leipzig).

Schulz, A. (1988), "Phraseologismen in englischen Lehrbuchtexten", student dissertation (University of Leipzig).

Simpson, J. A., and Speake, J. (1992) (eds.), *The Concise Oxford Dictionary of Proverbs* (2nd edn., Oxford: Oxford University Press).

Sinclair, J. (1987), "Collocation: A Progress Report", in Steele and Threadgold (1987), ii. 319-31.

----- (1991), *Corpus, Concordance, Collocation* (Oxford: Oxford University Press).

Sinclair, J. (1994), ' *Trust the Text*', in M. Coulthard (ed.), Advances in Written Text Analysis (London: Routledge), 12-25.

----- Hanks, P. W., Fox, G., Moon, R. E., and Stock, P. F. (1987) (eds.), *Collins Cobuild English Language Dictionary* (2nd edn., 1995) (Glasgow: Harper Collins).

----- and Moon, R. E. (1989) (eds.), *Collins Cobuild Dictionary of Phrasal Verbs* (London: Collins).

----- ----- (1995) (eds.), *Collins Cobuild Dictionary of Idioms* (London: Harper Collins).

Smadja, F. (1991), "Macrocoding the Lexicon with Co-occurrence Knowledge", in *Zemik* (1991), 165-89.

----- (1993), "Retrieving Collocations from Text: Xtract", *Computational Linguistics,* 19/1: 143-77.

Smith, W. G. (1935), *The Oxford Dictionary of English Proverbs* (3rd edn. rev. by F. P. Wilson 1970) (Oxford: Clarendon Press).

Snellgrove, L. E. (1979), *The Early Modern Age* (Longman Secondary Histories; London: Longman).

Snell-Homby, M. (1988) (ed.), *ZüriLEX '86 Proceedings* (Tübingen: Francke Verlag).

Sorhus, H. B. (1977), To Hear Ourselves -- Implications for Teaching English as a Second Language, *ELTJ* 31/3: 211-21.

Steele, R., and Threadgold, T. (1987) (eds.), *Language Topics: Essays in Honour of Michael Halliday*, ii (Amsterdam: John Benjamins).

Stein, G. (1991), The Phrasal Verb Type to have a look in Modern English, *IRAL* 29: 1-29.

----- and Quirk, R. (1991), "On Having a Look in a Corpus", in Aijmer and Altenberg (1991), 197-203.

Stenström, A.-B. (1984), *Questions and Responses in English Conversation* (Lund: Lund University Press).

Strdssler, J. (1982), *Idioms in English: A Pragmatic Analysis* (Tübingen: Gunter Narr).

Svartvik, J., and Quirk, R. (1980), *A Corpus of English Conversation* (Lund Studies in English; Lund: CWK Gleerup).

Swales, J. (1990), *Genre Analysis: English in Academic and Research Settings* (Cambridge: Cambridge University Press).

Tarone, E. (1983), "On the Variability of Interlanguage Systems", *Applied Linguistics*, 4/2: 142-64.

Teliya, V. N. (1981), *Tipui yazuikovuikh znacheniy: Svyazannoe znachenie slova v yazuike* (Moscow: Nauka).

----- (1993), ' *Kul' tumo-natsional' nuie konnotatsii frazeologizmov (ot mirovideniya k miroponimaniyu)',* in N. I. Tolstoy (ed.), *Slavyanskoe yazuikoznanie. XI Mezhdunarodnuiy s'ezd slavistov* (Moscow: Nauka), 302-14.

Ter-Minasova, S. (1992), "The Freedom of Word-Combinations and the Compilation of Learners' Combinatory Dictionaries", in Tommola *et al.* (1992), 533-9.

Thompson, D. (1995) (ed.), *The Concise Oxford Dictionary of Current English* (9th edn., Oxford: Clarendon Press).

Tomaszczyk, J. (1981), "Issues and Developments in Bilingual Pedagogical Lexicography", *Applied Linguistics*, 2/3: 287-96.

Tommola, H., Varantola, K., Salmi-Tolonen, T., and Schopp, J. (1992) (eds.), *Euralex '92 Proceedings* (Series translatologica A, I; Tampere: University of Tampere).

Tristram, H. (1978), *Linguistik und die Interpretation englischer literarischer Texte* (Tübingen: Max Niemeyer).

Tutin, A. (1992), "ètudes des anaphores grammaticales et lexicales dans la perspective de la génération automatique dans des textes de procedures", unpublished Ph.D. thesis *(Université de Montréal).*

Van J. Roey (1990), *French-English Contrastive Lexicology: An Introduction* (Louvain-la-Neuve: Peeters).

Vereshchagin, E. M., and Kostomarov, V. G. (1983), *Yazuik i kul'tura: Lingvostranovedenie v prepodavanii russkogo yazuika kak inostrannogo* (Moscow: *Russkiy yazuik*).

Vinogradov, V. V. (1947), ' *Ob osnovnuikh tipakh frazeologicheskikh edinits v russkom yazuike*', in A. A. Shakhmatov, *1864- 1920. Sbornik statey i materialov* (Moscow: Nauka), 339-64.

Wanner, L. (1996) (ed.), *Lexical Functions in Lexicography and Natural Language Processing* (Amsterdam: John Benjamins).

----- and Bateman, J. (1990), "Lexical Co-occurrence Relations in Text Generation", in *Proceedings of the 5th International Workshop on Natural Language Generation* (Dawson, Pa.).

Weinert, R. (1995), "The Role of Formulaic Language in Second Language Acquisition: A Review", *Applied Linguistics*, 16/2: 118-205.

Weinreich, U. (1969), "Problems in the Analysis of Idioms", in J. Puhvel (ed.), *Substance and Structure of Language* (Berkeley and Los Angeles: University of California Press), 23-81.

Weisgerber, J. L. (1929), *Muttersprache und Geistesbildung* (Göttingen: Vandenhoeck & Ruprecht).

Werlich, E. (1976), *Text Grammar of English* (Heidelberg: Quelle & Meyer).

Werner, S., and Werner, T. (1990), "Art, Struktur und Funktion von Phraseologismen im Background Article der Zeitschrift Marxism Today", student dissertation (University of Leipzig).

Whorf, B. L. (1956), *Language, Thought and Reality: Selected Writings of Benjamin Lee Whorf*, ed. J. B. Carroll (New York: Wiley).

Wierzbicka, A. (1982), "Why can you have a drink when you can't ★have an eat?", *Language*, 58: 753-99.

----- (1992), *Semantics, Culture and Cognition* (New York: Oxford University Press).

Wilks, Y., Fass, D., Guo, C.-M., McDonald, J., Plate, T., and Slator, B. (1989), A Tractable Machine Dictionary as a Resource for Computational Semantics, in *Boguraev and Briscoe* (1989), 193-228.

Willis, D. (1990), *The Lexical Syllabus: A New Approach to Language Teaching* (London & Glasgow: Collins ELT).

Wolf, C. (1979)., *Kein Ort. Nirgends* (7th impression, 1988) (Berlin &

Weimar: Aufbau Verlag). English translation by Jan van Heurck, No Place on Earth (NewYork: Farrar, Straus & Giroux, 1982).

Wong L. Fillmore (1976), "The Second Time Around: Cognitive and Social Strategies in Second Language Acquisition", unpublished Ph.D thesis (Stanford University).

Yakovleva, E. (1974), "The Prosody of the Qualificative Word-Combination", in Akhmanova (1974), 91-109.

Yorio, C. (1980), "Conventionalized Language Forms", *TESOL Quarterly*, 14: 433-42.

----- (1989), "Idiomaticity as an Indicator of Second Language Proficiency", in K. Hyltenstam and L. K. Obler (eds.), *Bilingualism across the Lifespan* (Cambridge: Cambridge University Press), 55-72.

Zernik, U. (1991) (ed.), *Lexical Acquisition: Using On-Line Resources to Build a Lexicon* (Hillsdale, NJ: Lawrence Erlbaum Associates).

Zgusta, L. (1971), *Manual of Lexicography* (Janua linguarum, Series maior 39; The Hague: Mouton).

Z+ˇolkovskij, A., and Mel'čuk, I. (1967), *"O semanticheskorn sinteze [On Semantic Synthesis]", Problemy kibernetiki*, 19: 177-238.

[This page intentionally left blank.]

Index

Note: Page references in *italics* indicate tables and figures.
