Preface

The proposition that there is a correlation between language and culture or culture-specific ways of thinking can be traced back to the views of Herder and von Humboldt in the late 18th and early 19th centuries. It was most explicitly formulated, however, by the German-American linguist and anthropologist Edward Sapir in various publications from 1929 onward (re-published posthumously in 1949 under the title Selected writings of Edward Sapir in language, culture and personality), and in the writings of his pupil Benjamin Lee Whorf (re-published posthumously in 1956 as Language, thought, and reality: Selected writings of Benjamin Lee Whorf). The Sapir-Whorf hypothesis, as it came to be called, expresses the notion that different languages lead their speakers to different conceptualizations of the same extralinguistic reality, which seems to be most evident in the way that reality is segmented by the lexicon.

Even though few linguists would fully agree with a strict reading of the Sapir-Whorf hypothesis today, it is generally accepted that a language, especially its lexicon, influences its speakers’ cultural patterns of thought and perception in various ways, for example through a culture-specific segmentation of the extralinguistic reality, the frequency of occurrence of particular lexical items, or the existence of keywords or key word combinations revealing core cultural values. Nevertheless, the exact workings of the link between language and culture are still poorly understood. The few specific theoretical frameworks that do exist are often felt to be inadequate, and the research methodology is only insufficiently developed (it is telling, in this context, that the methods employed by Whorf in particular seem to have had serious shortcomings).

The aim of this volume, then, is to explore the cultural dimension of a wide range of preconstructed or semi-preconstructed word combinations in English. These include highly opaque multiword units of the *kick-the-bucket* type, collocations, irreversible binominals, phrasal verbs, compounds, metaphorical expressions, similes, proverbs, familiar quotations, catchphrases, clichés, slogans, expletives, and discourse markers such as politeness formulae – all of which have been subsumed under *phraseology*, or under *idiom* in the Anglo-American linguistic tradition. The volume is divided into four sections, focusing on particular lexemes (e.g. *enjoy* and its
collocates), types of word combinations (e.g. proverbs and similes), use-related varieties (such as the language of tourism or answering-machine messages), and user-related varieties (such as Aboriginal English or African English). The assignment of the papers to these sections is, of course, not always clear-cut: Many of the papers address issues pertaining to more than one section, and the dividing lines between the sections are therefore permeable, rather than rigid. The sections are preceded by a prologue, tracing the developments in the study of formulaic language, and followed by an epilogue, which draws together the threads laid out in the various papers, and ends with a résumé of the research questions raised in the prologue. The epilogue also draws attention to good academic practice in a way that, it is hoped, will encourage other researchers to conceptualize their projects carefully, both in terms of procedures and assumptions, and also in terms of the potential theoretical import of their work.

There are a number of important works, published in the past fifteen years, that explore the relation between language and culture in general, but the study of the relation between English phraseology and culture in particular has been largely neglected. This volume is the first book-length publication devoted entirely to this topic. It should be of interest to all those interested in phraseology (or idiomatology) and variational linguistics, and to those interested in the interface between language and culture, which is a particular concern in cognitive linguistics and anthropological linguistics.

I would like to thank Penny Lee for reading an earlier draft of the typescript and making valuable comments toward its improvement. I am also grateful to Karin Stettler for competently preparing the typescript and for her endless patience over the years, and to Susanne Jantos for doing the hard work of compiling the index.

*Paul Skandera*
*Innsbruck, Austria*
*October 2006*

## Contents

Preface
Paul Skandera

### Prologue

Developments in the study of formulaic language since 1970:
A personal view
Andrew Pawley

### Focus on particular lexemes

Reasonably well: Natural Semantic Metalanguage as a tool for the study of phraseology and its cultural underpinnings
Anna Wierzbicka

Australian perceptions of the weekend: Evidence from collocations and elsewhere
Bert Peeters

Enjoy!: The (phraseological) culture of having fun
Monika Bednarek and Wolfram Bublitz

Hot, heiß, and горячий: A case study of collocations in English, German, and Russian
Doris Schönefeld

### Focus on types of idioms

Collections of proverbs and proverb dictionaries: Some historical observations on what's in them and what's not (with a note on current "gendered" proverbs)
Charles Clay Doyle
Yankee wisdom: American proverbs and the worldview of New England
Wolfgang Mieder

Similes and other evaluative idioms in Australian English
Pam Peters

**Definitely maybe:** Modality clusters and politeness in spoken discourse
Svenja Adolphs

**Focus on use-related varieties: Registers**

Lexical developments in greenspeaking
Melina Magdalena and Peter Mühlhäusler

The phraseology of tourism: A central lexical field and its cultural construction
Andrea Gerbig and Angela Shek

Idiomaticity in a cultural and activity type perspective: The conventionalization of routine phrases in answering-machine messages
Karin Aijmer

**Focus on user-related varieties: Dialects and ethnolects**

Greetings as an act of identity in Tristan da Cunha English:
From individual to social significance?
Daniel Schreier

Multiword units in Aboriginal English:
Australian cultural expression in an adopted language
Ian G. Malcolm and Farzad Sharifian

Fixed expressions as manifestations of cultural conceptualizations:
Examples from African varieties of English
Hans-Georg Wolf and Frank Polzenhagen

Varieties of English around the world: Collocational and cultural profiles
Christian Mair

**Epilogue**

Formulaic language in cultural perspective
Penny Lee

Index
Developments in the study of formulaic language since 1970: A personal view

Andrew Pawley

There continues to be a need for a model of natural discourse that pulls together the diverse cognitive and social factors responsible for the shape of language. (Chafe 1996: 49)

1. Introduction

A few years ago Anthony Cowie observed that, whereas in 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" except in dictionary-making (Cowie 1998a: 18), it "has now become [a] major field of pure and applied research for Western linguists" as it had earlier for scholars in Eastern Europe (Cowie 1998a: 1).¹

In this essay I will review developments in the study of phraseology, focusing on the period since 1970. However, instead of "phraseology" I prefer to speak of the study of "conventional expressions" or "formulaic language". The class of (speech) formulae in its broadest sense is taken here to subsume all conventional multiword expressions and also to include single word expressions that serve speech act functions, such as Hello! and Thanks!

The date of 1970 is cited as a boundary chiefly because it was about that time that a number of structural linguists began to pay close attention to conventional expressions and formulaic language, a field already cultivated by practitioners of several other disciplines. By 1970, it is fair to say, scholars in at least nine different disciplines – literary studies, folklore studies, social anthropology, neurology, experimental psychology, educational psychology, microsociology, the teaching of English as a foreign language and lexicography – had done significant research on aspects of formulaic language. During the 1970s linguists established their own research agen-
those of written language. She began to think about this question while teaching English language and literature at high school. Her main interest in conversation was in its social dynamics and in how the conventions governing face-to-face meetings, of being in the company of others, shape speech behaviour. However, she noted that one ingredient of conversation that distinguished it from formal written discourse seems to be its more frequent use of lively, colloquial expressions. In the early 1960s she compiled a sizeable dictionary of Antipodean English colloquial phrases with notes on their contexts of use. In 1971 Syder and I began to record and transcribe a sizeable corpus of English conversational speech with the idea of looking at its phonological, grammatical, lexical and sociolinguistic characteristics.

Around that time I was asked to write some materials for the new English language syllabus for senior forms in New Zealand secondary schools. This syllabus, devised by John Pride, called for a strongly sociolinguistic approach. In the course of “reading up” for this task and for the conversation project we came across more and more references to formulaic uses of language. There was a network of connections that criss-crossed a number of disciplines in the humanities and the social and biological sciences. It was exciting to discover that scholars in diverse fields concerned with human behaviour or physiology had independently concluded that conventional speech played an important role in their particular domain.

2. Some research traditions which investigated formulaic language prior to 1970

A rough taxonomy follows of the various categories of researchers who during the 1960s or earlier did work on one or another aspect of formulaic language. At least eight distinct research traditions can be distinguished.¹

(1) Literary scholars working on epic sung poetry. Most famously, Milman Parry in the 1920s and early 1930s did groundbreaking work on the role of formulae in the poems attributed to Homer. He and his student, Albert Lord, caught in the nick of time the South Slavic tradition of composing singing epic poems in public performances, which still flourished before World War II. Studying the extraordinary skills of the illiterate Yugoslav singer-composers provided them with a living laboratory in which to test hypotheses about the composition and transmission of Homeric poetry.
Parry himself was heir to an established international tradition of Homeric studies and to a Slavic tradition of work on modern epic poetry. Even so, Parry and Lord as a team can be considered the Darwin of research into Homer, inasmuch as they brought about a paradigm shift. Homer’s epics were once considered written literature of the highest form. Parry put into a rigorous and testable form his theory that the poems came from an oral tradition (Parry 1928, 1930, 1932). Following Parry’s early death Lord continued their research and wrote the definitive work (Lord 1960).

Parry (1930: 80) defined a formula as “a group of words which is regularly employed under the same metrical conditions to express a given essential idea”. Parry and Lord recognised the paradox that formulae are at the same time memorised and flexible, and have the dual functions of ensuring that performance be fluent yet contain creative variations. Formulae may show special word order, enabling a word sequence to be adapted to the metrical requirements of a half-line of verse; they also show changes in rhythm and intonation different from those of ordinary speech. A “substitution system” is a group of formulae which show lexical substitutions expressing the same basic structure and idea, or which express the same basic idea with varying numbers of syllables, enabling the poet to meet a range of different metric conditions.

(2) *Anthropologists and folklorists concerned with ritual speech and song.* They represent a partly distinct tradition in the study of oral formulaic genres (Bauman 1975, 1986; Bauman and Sherzer 1974). There is a large literature on this field and I will refer to just a couple of representative works. In 1935 Malinowski published *Coral Gardens and their Magic*, the most linguistic of his several ethnographic accounts of the Trobriand Islanders of New Guinea. He observed that the magical incantations of Trobriand Islanders are composed of fixed formulae, spoken with singsong intonation and distinctive rhythm and pitch, with meanings that can only be interpreted non-literally, and with the purpose of trying to control the supernatural. Notice, once again, a combination of features attributed to formulae: fixed words, distinctive intonation, distinctive rhythm, and pragmatic function separate from literal meaning. Following a long tradition in Indo-European studies of collecting and comparing variants of particular folk tales and poems, Iona and Peter Opie published *The Lore and Language of Schoolchildren* in 1959 (their second major work, *The Singing Game*, appeared in 1985). The Opies’ study demonstrated the role of 6 to 10-year-old children as carriers and creative manipulators of a remarkably persistent tradition of nursery rhymes, game chants and sayings, rich in formulaic units of a kind quite similar to Trobriand magical incantations.

Dell Hymes’ seminal essay on the ethnography of speaking and rituals of encounter (Hymes 1962) sparked a blaze of research in the 1960s by American linguistic anthropologists which focussed on “performance routines and on detailed ethnographic observation of how people actually use language” (Finnegan 1992: 42–43). Hymes (1968: 126–127) acknowledged that “a vast proportion of verbal behaviour consists of recurrent patterns, …[including] the full range of utterances that acquire conventional significance for an individual, group or whole culture”.

(3) *Philosophers and sociologists concerned with ordinary language use as strategic interaction.* The role of speech routines as a key ingredient in social competence began to receive systematic study in diverse quarters in the 1960s. Philosophers of ordinary language and sociologists studying everyday encounters were in different ways concerned with the strategic use of utterances to perform speech actions. Among the sociologists Erving Goffman was an important pioneer in the study of face-to-face talk, as were the ethnomethodologists, whose work gave rise to the dominant approach to conversation analysis pioneered by Sacks and Schegloff. In particular, Goffman drew attention to the social norms that govern people’s behaviour in public, norms that underlie established discourse structures and motivate choices of conversational moves. Some of his writings (e.g. Goffman 1971) also attend to the conventional forms of words and gestures used to carry out such moves.

The work of linguistic philosophers such as Austin (1962), Searle (1969) and Grice (1975) was to play a key role in the development of linguistic pragmatics in the 1970s. Austin and Searle used “speech act” broadly to refer to any utterance that performs a discourse function over and above those of referring and predicating, e.g. the functions of greeting, farewelling, introducing, welcoming, complimenting, insulting, apologising, complaining, criticising, refusing, blaming, cursing, forbidding, promising, lamenting, warning, naming, performing marriage, declaring a meeting open, and so on. Such utterances typically take the form of conventional expressions. Following Lyons (1968: 178) I will refer to conventional expressions for performing speech acts as a subclass of “situation-bound expressions” (on this topic see also Coulmas 1979, 1981; Keeskes 2000; Kiefer 1996).
(4) Neurologists and neuropsychologists, concerned with localisation of language functions in the brain. Work on brain-damaged patients by Paul Broca in the 1860s demonstrated left hemisphere dominance in speech processing. Soon after John Hughlings Jackson concluded that certain kinds of severe aphasia to Broca’s area in the left hemisphere knocked out creative or “propositional” speech but left speakers with “automatic” speech, the ability to recall familiar expressions and texts. Since about 1960 there has been a dramatic expansion of research on language functions in the brain, with important contributions by scholars trained in linguistics. The journal Brain and Language, established in 1969, carries a considerable literature touching on this subject.

(5) Psychologists concerned with learning and speech processing. Lashley’s (1951) paper on the problem of serial order in behaviour proved to be one of the most influential in mid-20th century psychology, arguing against a rigidly behaviourist account of how of elements of behaviour are connected and generated. Lashley argued that an essential characteristic of most serially ordered behaviour is that it conforms to a kind of “schema of action”, a central determining event which selects elements and determines their order before generation. In the 1950s and 60s experimental studies of speech for different cognitive tasks showed that familiar or repeated word strings pattern differently from novel strings in terms of frequency and placement of hesitations and other variables, with greater fluency corresponding to automatisation or “chunking” of familiar strings (Goldman-Eisler 1968; Rochester 1973).

(6) Research in educational psychology. At the end of the 1950s Basil Bernstein (1958, 1960, 1961) put forward a controversial two-part hypothesis connecting patterns of language use with patterns of thinking. First, he distinguished two varieties of spoken English, initially called restricted and elaborated codes (later private and public languages), based partly on the frequency of what he called “precoded” or memorised utterances (formulæ presenting stereotyped ideas or with highly contextualised functions) vs. “now-coded” utterances (freshly-minted, seeking to formulate original thoughts). In this matter he was much influenced by Goldman-Eisler’s work (see (5) above). Second, Bernstein related these two putative varieties of English to class differences in habits of thinking and attitudes. In this second point, he was influenced by the work of Vygotsky. Bernstein’s provocative proposals suffered from conceptual confusions and methodological weaknesses which stopped them from being fully explored. Bernstein’s ideas can be placed within a wider tradition of work on language, culture, cognition and personality by educational psychologists, anthropologists and linguists, one strand of which is concerned with the question of whether people’s customary ways of talking correlate with particular perceptions and worldviews (e.g. Cole and Scribner 1974; Enfield 2002; Grace 1987; Lucy 1992; Wierzbicka 1986).

(7) Grammarians. Grammarians of earlier generations had acknowledged that conventional expressions play a part in language. Jespersen (1922), for example, distinguished between “free” and “fixed” expressions. However, it was not until the 1960s that grammarians began seriously to contemplate the difficulties that semi-productive constructions create for models of language that posit a sharp boundary between syntax and lexicon. Chomsky noted that derivational morphology is a grey area (Chomsky 1965: 184–192). Early attempts to derive nominal compounds by transformational rule were seen to be unsatisfactory (Zimmer 1964, 1971). Several influential works on the grammar and semantics of idioms appeared, including Chafe (1968), Fraser (1970), Makkai (1972) and Weinreich (1969). A few grammarians briefly considered the significance of a wider range of conventional expressions than canonical idioms, e.g. Lyons (1968) briefly discusses “situation-bound expressions” (see 4.4) and Weinreich (1969) discusses “familiar expressions”.

(8) Phrasal dictionaries of English. Such dictionaries have been around for several generations, but until recently did not handle this domain very systematically. To the English-speaking public, the best known works of the mid-20th century were perhaps Eric Partridge’s dictionaries of slang, cliches and catch phrases. However, technically much better grammatical treatments of multiword expressions appeared in small English phrasal dictionaries aimed at foreign learners, beginning with the pioneering work on collocations of H.E. Palmer (1933, 1938) and A.S. Hornby (Hornby et al. 1942). For a detailed history of English dictionaries for foreign learners see Cowie (1999). Cowie (1998b: 210) refers to Palmer and Hornby as “the founding fathers of EFL lexicography”. Their work strongly influenced the handling of phrasal units in general-purpose dictionaries designed for EFL learners in the decades that followed (Cowie 1998c). However, until the 1970s most large phrasal dictionaries were fairly primitive affairs, containing a fixed phrase, a definition and perhaps an illustrative sentence.
Weinreich (1969) and Cowie (1981, 1998b) have drawn English speakers' attention to the fact that in Eastern Europe, especially the Soviet Union, phraseology has been a prominent field in linguistics and lexicography since the late 1940s. In the typology of “phraseological units” devised by Vinogradov (1947) and refined by Amosova (1963) and by Mel'čuk three major categories are distinguished: pure idioms, figurative idioms and restricted collocations. (As the terms for these categories vary from author to author, it is convenient to use those terms adopted by Cowie 1998a.) Pure idioms, e.g. beat around the bush, and chew the fat, are expressions whose literal meanings give no clue to their idiomatic meaning, and figurative idioms, e.g. hold water, steal someone’s heart, and run the gauntlet, are expressions whose idiomatic meanings can be regarded as figurative extensions of their literal meanings. Restricted collocations, e.g. pay one’s respects / a compliment / court, and meet the demand / needs / requirements are word combinations where the sense assigned to one partner (pay and meet in the above examples) is “bound to” or governed by its association with another word or phrase. The East European scholars also recognised an important functional and structural distinction between “word-like” expressions and “sentence-like” expressions, such as I beg your pardon and I must apologise.

For the most part, these studies of formulaic language in a range of disciplines represented separate islands of research, with comparatively little interchange of ideas between them. A coherent discipline (or subdiscipline) is recognisable as such by having a core subject matter, research questions and research methods that are generally accepted, and by having some of the associated institutional trappings, such as regular conferences, courses that give it a place in the curriculum, textbooks and readers, even its own journals. Phonology, syntax, and historical linguistics, for instance, were clearly disciplines in this sense. The study of formulaic language was not. Nevertheless, the sum of the diverse findings in these islands of research was impressive. The following is a brief summary of the more important conclusions that had emerged by about 1970.

1. In the typology of conventional expressions it is useful to distinguish (a) between word-like and sentence-like expressions, (b) between fixed and flexible (semi-productive) expressions, (c) between idioms and restricted collocations, (d) between true idioms and figurative idioms, and (e) between certain types of restricted collocations.

2. Certain specialised speech genres, where performers sustain exceptional fluency and coherence over extended stretches of discourse, depend almost entirely on formulae, including both fixed and flexible types.

3. In both everyday conversation and specialised genres, sentence-like formulae are used to achieve strategic (social, magical, etc.) ends.

4. Familiar combinations of words and ideas are spoken more fluently than novel combinations.

5. There is some evidence that once word combinations are well learned or automatised they are processed in the right hemisphere, whereas the production and comprehension of novel combinations are processed in the left hemisphere.

6. Those conventional expressions which reflect semi-productive morphological and syntactic processes present a problem for grammarians.

3. The 1970s: Linguists establish research agendas in formulaic language

On to the 1970s. What are the grounds for regarding that decade as a watershed for work on formulaic language in Western linguistics and lexicography?

Two factors stand out. First, a significant number of linguists began to study formulaic language. (I include those lexicographers who brought a training in linguistics to the task of compiling dictionaries of idiomatic expressions.) Second, the main lines of most of the research programs in this field currently pursued by linguists began to take shape.

The chief business of descriptive linguistics has long been grammatical and phonological analysis. A central concern of theoretical linguistics in the 1960s and 70s was discovering universal principles of language structure to serve (among other purposes) as grounds for choosing between competing models of language and language acquisition. In the 1960s formulaic expressions were generally regarded as marginal to the proper subject matter of descriptive or theoretical linguistics. Several mindsets sustained that view. First, formulaic speech was considered to constitute only a small part of the normal output of native speakers. All the emphasis then was on the power of syntax to generate an infinite number of sentences. We learnt to chant the mantra that “most of the sentences that speakers utter are novel”. Second (and here I follow Bolinger 1976), theoretical linguists were after a monolithic model, first of “language”, later of “grammar”, and bits that did not fit comfortably tended to be shunted off to one side. Idioms
are marginal to grammar because they are either frozen or grammatically or semantically anomalous. And well-formed formulae (such as proverbs, cliches and many restricted collocations) are of no interest to grammarians because they are merely conventional or commonplace expressions. Grammar is about rules and structure, and gives no place to the notions of "established usage" or "frequency of use" (although generative grammarians have sneaked conventionality on board in the guise of "selectional restrictions" in syntax and "blocking" rules in morphology).

In the 1,100 page A Grammar of Contemporary English speech formulae and formulaic constructions are described as "something of a museum of oddments" (Quirk et al. 1971: 411). To their credit, the authors include a brief but thoughtful discussion of several types of minor constructions such as those exemplified by *How do you do?*, *Why get so upset?*, *To think I was once a millionaire*, *May the best man win*, *Least said, soonest mended* and *I beg your pardon*. Each is considered to be formulaic, because it is grammatically anomalous (and in some cases also because it is used in "stereotyped communicative situations"). However, Quirk et al. devote less than three pages to this extremely numerous and diverse class of utterances. And that is probably more than most grammars did (and do).

A certain degree of reductionism is essential to scientific progress. The problem is always to know when it is time to extend the subject matter, to tackle the bits that don't fit the theory, and when it is time to modify or abandon the theory. After linguists had looked closely at formulaic constructions for a while these no longer looked quite so odd and marginal. Thus, Fillmore, Kay and O'Connor (1988: 501) were to conclude, in a paper in *Language*, that "the realm of idiomaticity in a language includes a great deal that is productive, highly structured and worthy of grammatical investigation".

This shift of perspective was well under way (in at least some quarters of linguistics) during the 1970s. During that period small but valiant bands of lexicographers painstakingly assembled evidence regarding the lexical and grammatical variability of thousands of formulae. Stimulated by the work of philosophers of ordinary language, grammarians wrote many papers on speech act semantics and pragmatics, some of which showed the conversational uses of certain conventional forms of words and construction types. From first and second language acquisition circles there came a steady flow of works treating the role of speech formulae or prefabs (pre-fabricated expressions) in language learning. Several projects investigating the encoding of thoughts in spontaneous speech were begun, each concerned with the role of familiar expressions in enhancing fluency, allowing the encoder to overcome the severe limitations of human short term memory. Neuropsychology attracted some able scholars with a training in linguistics, who investigated the neurological status of automatic speech.

Within a few years formulaic language had gained a modest degree of status as a subject matter in linguistics. A foot was wedged in the establishment door when a course on speech formulae, taught by Charles Fillmore and Lily Wong Fillmore, was offered for the first time at the 1977 Linguistic Institute sponsored by the Linguistic Society of America. Another indicator was the appearance of analytic papers (e.g. Ferguson 1976) and doctoral theses tackling particular issues. The quantity of recent publications on the role of prefabs in L2 acquisition was sufficient to draw a review article by Krashen and Scarcella (1978) and a collection of papers on conversational routines appeared (Coulmas 1981).

By the end of the 1970s linguists working on formulaic language were pursuing a range of research questions fairly similar to those that engage us today. Some of these research questions were inherited from earlier work in other disciplines, as we saw earlier. However, linguists and lexicographers have brought new perspectives and methods to bear on these and their enquiries in turn generated some new questions.

There were questions concerning taxonomy and description:

1. Identification. How to identify well-formed conventional expressions in text?
2. Classification. How to classify conventional expressions? What structural and functional criteria are relevant?
3. Transcription. How to represent speech formulae in transcriptions of speech? What details are relevant to record?
4. Description of variability. What substantive concepts and notational devices are needed to describe the variability found within certain (more or less) productive formulae?
5. The composition of pragmatic speech formulae. What range of features sets apart discourse-strategic speech formulae (situation-bound expressions), such as *How are you?*, from word-like conventional phrases.
6. Oral formulaic genres. Which features of oral formulaic genres set them apart from "ordinary" language?
7. Prevalence in "ordinary" language. How formulaic are "ordinary" genres of speech and writing?

There were also general questions about linguistic competence, language use and language change that implicate formulaic language:
8. **Speech production and comprehension.** How do competent speakers of a language (or genre) manage to talk more fluently than (it appears) their limited processing capacity should permit? How do simultaneous interpreters manage to translate utterances with minimal time lag, and sometimes even finish ahead of the speaker?

9. **Idiomaticity, or native-like selection.** How do speakers know which grammatical strings are native-like?

10. ** Appropriateness.** How do people know “the right thing to say” in particular social contexts?

11. **Language acquisition.** How do children and adults acquire language? What role do conventional expressions play in first language learning? Why are certain types of formulae particularly difficult for mature L2 learners to master?

12. **Grammaticalisation.** What is the origin of grammar? How do syntactic and morphological constructions come about?

There were questions to do with models of language:

13. **The grammar-lexicon boundary.** Where to put the “phrase book” in a grammar-lexicon description?

14. **View of language.** By admitting the machinery needed to describe speech formulae, are we not changing the goals of linguistics and indeed changing our view of what a language is?

And finally, there were other issues, including:

15. **Localisation of function.** To what extent are conventional and free expressions processed in different parts of the brain? What do aphasics use speech formulae for?

4. **Progress on particular research questions since 1970**

The rest of this paper will comment on the progress that has been made in dealing with these questions. It is necessary to be selective. To review work on all issues and projects in the field would be an immense task. I will say nothing about the large body of research on locating language functions in the brain (reviewed in van Lancker 1987, 1997; Wray 2002) or about recent speculations on the role of formulaic utterances in the origins of grammar.

4.1. **Oral formulaic genres**

Any course on formulaic language would do well to begin with a study of one or more of highly developed oral formulaic genres, whether it be the singing of epic poetry in the Balkans or New Guinea (Rumsey 2001) or the selling routines of professional auctioneers (Kuiper 1996). In these we find long stretches of discourse where almost every utterance is a formula.

Oral formulaic speech traditions show five features that, taken together, distinguish them from other discourse genres:

(a) very strict discourse structure rules, specifying the topics proper to the discourse and their order of occurrence. For example, in stock auctions (Kuiper and Hago 1984) there are four compulsory immediate constituents: (1) **Description of the lot,** (2) **Search for the first bid,** (3) **Calling the bids,** and (4) **Sale.** Most of these constituents in turn may consist of several constituents, e.g. **Description** can consist of **Provenance + Number, History, Preparation,** and **Potential.**

(b) a very high concentration of speech formulae (usually 90 percent or more of clauses), each associated with a particular discourse context or range of discourse contexts.

(c) special grammatical rules applying to formulae.

(d) special prosodic or musical patterns.

(e) exceptional fluency, i.e. fewer than average unplanned pauses within clauses.

These findings have come out of an impressive body of work on formulaic genres in English built up over the past 25 years by Koenraad Kuiper and his associates at the University of Canterbury in New Zealand. Kuiper began by studying auctioneers’ sales talk at livestock auctions in North Canterbury and then, with students and colleagues, went on to look at other kinds of auctions and various kinds of sports commentary and later, at a wider range of spoken and written genres (e.g. Flindell 1991; Hickey 1991; Hickey and Kuiper 2000; Ji et al. 1990; Kuiper 1992, 1996; Kuiper and Austin 1990; Kuiper and Hago 1984, 1985; Kuiper and Flindall 2000; Kuiper and Tan 1989; Kuiper and Tillis 1986). Kuiper and his collaborators outline a framework for constructing generative descriptions of oral formulaic discourse. The descriptions are intended to be generative in two senses. First, they seek to be explicit, defining in a precise manner the object of inquiry and its structure. Second, they seek to be predictive, formulating rules for the production of acceptable utterances or texts which go beyond the corpus of recorded examples. The discourse structure is hierarchical and
can be formally represented by context-free rewrite rules (with a few extra notational conventions). Individual formulae are indexed to discourse contexts.

Kuiper and Hago (1984) sought an explanation as to why the sale talk of livestock auctioneers has these characteristics. They concluded that the oral formulaic technique has evolved to allow the performer to maintain abnormal fluency while also achieving acceptable standards of content and delivery. They note the close parallels with the Yugoslav oral poets in the need to retain the attention of a mobile audience, in the heavy load placed on the short term memory, in the dense employment of formulae, in the methods by which neophyte practitioners learn their craft and become virtuosos. There are differences: the auctioneer interacts with his audience during the performance. There is less creativity of imagery in the auctioneer’s talk. However, it is characteristic of both types that performers do not rely on verbatim recall. Perfect recall of long stretches of text requires exceptional concentration and can detract from other facets of performance (note that recall of text by stage actors is a very different task). A more efficient technique is to draw on memorised chunks but to be able to vary the text somewhat.

Hickey (1991) found that the New Zealand Meteorological Office scripted weather forecasts are highly formulaic, both in discourse structure and in choice of expressions. The general synopsis is always followed by district forecasts. The district forecasts and the kinds of weather information given for each always occur in the same order and the specific information is always couched in formulaic terms. Hickey and Kuiper (2000) conclude that in this case the speech routines are created for the listeners’ benefit, so that the only variables listeners have to pay attention to relate to the weather itself. Wray (2002: 79–83) found that people’s complaints about the incomprehensibility of British weather forecasts on Radio 4 reflected the lack of a strict order in presenting information.

In Auckland, following Kuiper’s lead, my students and I looked at the structure of various kinds of radio sports commentaries, e.g. on cricket (Pawley 1991), rugby (Brown 1987), and children’s playground rhymes (Smith 1982) and found that these conform to the basic oral formulaic pattern. For other kinds of oral formulaic genres, see Finnegan (1977) and Rumsey (2001).

4.2. Identifying conventional expressions

Given that most conventional expressions are not idioms but are well-formed grammatically, how does one identify them in text? This is not a trivial problem. For example, there was debate over whether Wong-Fillmore (1976) had correctly identified “formulae” in her thesis, or had cast the net too wide and included some free constructions.

What defines a multi-word string as being conventional, formulaic, lexicalised? It has long been agreed that there are clear cases in principle: true idioms and stereotyped situation-bound utterances, and (in English) to a lesser extent verb + particle and nominal compounds. The disagreement has mainly been over where to draw the boundary between free (or novel) expressions and restricted collocations. In corpus studies, problems arise in picking out well-formed formulaic expressions. Automatic search procedures rely on a lookup list. The quality of results depends on the quality of the list plus the routine’s ability to recognise discontinuous expressions.

Cowie (1998b) and Hausmann (1979) have criticised some corpus linguists for their “insistence...that frequency of co-occurrence is the only significant measure of ‘conventionality’ in language” (Cowie 1998b: 226).

It is now widely acknowledged that conventionalisation (and therefore lexicalisation) of multi-morphemic and multi-word expressions is a matter of degree. A list of 27 diagnostics for identifying lexicalised expressions was proposed in Pawley (1986). That paper stresses that the importance of applying not only the usual range of syntactic and semantic tests, but also a variety of tests concerning social status (indicating, e.g. whether speakers recognise a particular word combination as the authorised “name” or “term” for a recognised entity), pause placement, prosodic patterns, writing conventions, ellipsis of final constituents, collocational restrictions, and so on. Lists of diagnostics are also proposed in Cowie and Mackin (1975), Makkai (1972), Moon (1998a), Zgusta (1971). Wray (2002) and Wray and Namba (2003) give a checklist of criteria for identifying formulaic chunks in a corpus.

4.3. “Transcribing” speech formulae

The task of describing speech formulae, or any other form of speech begins with the task of “transcribing”. It has been a failing of much research on formulaic language by linguists and lexicographers that it treats only
the words and meanings of formulae. Anyone who wants to understand how spoken language works has to pay attention also to a range of suprasegmental and paralinguistic features, which in particular contexts may include tempo, volume, timbre, pauses and hesitations, fillers, false starts, laughter, feedback, accompanying gestures (Crystal and Davy 1969). They must try to construe the social context and motives of the participants. Scholars working on performance in oral formulaic genres have similarly been aware of the need to attend to many channels of behaviour. Bauman (1975) proposes that three levels of analysis are needed for the study of performance: textual, contextual and sociocultural. Musical scores offer a better analogy to transcribers than ordinary written language, as far as the non-segmental dimensions of speech are concerned. Conversation analysts soon realised the key role of transcription and developed notational systems for indicating at least some of these details (e.g. Jefferson 1985; Stenström 1994).

4.4. Classifying formulae. The division between formulae with word-like functions and formulae with discourse functions

Various taxonomies of conventional expressions are possible, according to the choice of criteria. We referred earlier to the widely-accepted classification developed by the Soviet phraseologists: pure idioms, figurative idioms, restricted collocations. Since then, more fine-grained taxonomies of idioms and restricted collocations, based on syntactic, semantic and functional criteria, have been proposed by various scholars, including Fernando (1996), Makkai (1972), Mel'čuk (1988, 1998), Moon (1998a), Natttinger (1980) and Wray (2002). The details are too intricate for us to take up here (but see also 4.7, 4.8).

There is a fundamental difference between lexeme-like and sentence-like formulae, a distinction recognised by the East European phraseologists (Gläser 1986) and by Palmer (1942). Minimal lexical units are bundles of perhaps just three inherent components: form, meaning and grammatical context (part of speech category). Certain conventional expressions resemble lexical units in most respects. Phrasal expressions such as nasty piece of work, break up, feel antipathy towards, shirk responsibility, pearly white, out of bounds, for good (for ever) differ from minimal nouns, verbs, adjective and adverbs, chiefly in having internal grammatical and semantic structure.

However, many conventional expressions cannot satisfactorily be treated as large lexical units. I refer to those formulae that have strategic functions in discourse and social interaction, such as Hello!; Pleased to meet you; Dear X, Here's to XI; I declare this (meeting, etc.) closed; If it's good enough for X, it's good enough for Y, or Give me a break, will you? Expressions of this type are quintessential speech formulae because they are tried and true ways of doing things, standard recipes for achieving social purposes. A partial description by Palmer of this major category appears in Bongers (1947).

More than any unit of language this class of expressions shows the need for a model of discourse that integrates the diverse cognitive, social and historical factors responsible for shaping language (to paraphrase Chafe's words quoted at the head of this paper). Such expressions are sometimes referred to as "situation-bound expressions". We might also call them "pragmatic formulae" or "discourse-strategic formulae". They are bundles of seven or eight components: (a) segmental phonology, (b) music — when spoken they require a certain tempo, rhythm, melody, voice quality, etc., (c) grammatical category, (d) grammatical structure, often a mini-grammar, (e) idiomaticity constraints, (f) literal meaning, (g) pragmatic function, associated with a particular place in discourse structure or with a particular social context or purpose, and often (h) body language. (The musical component applies also to written formulae insofar as people know how these should sound when read aloud.) Thus, the expression (I'm) pleased to meet you should be marked for its position as B's response in an introduction sequence, after someone has introduced A to B or after A has introduced himself, and for its function as ritually acknowledging A's status as a new associate. It should be spoken with "bright" tone, with main contour stress on meet, with eye contact with the addressee, with accompanying handshake (under certain conditions). It belongs to a construction type which allows limited lexical variation while keeping the function intact, e.g. for pleased a few near synonyms can be substituted (e.g. glad, delighted).

There have been surprisingly few studies of the full array of attributes exhibited by pragmatic formulae. Fragmentary accounts can be found in Pawley (1985, 1986, 1991, 2001), Pawley and Syder (1983a). Certain of the features have been noted in work on conversational routines (e.g. Goffman 1971; Lane 1978; Coulmas 1981; Aijmer 1996), and in work on construction grammar, referred to in 4.7 below. The ways in which formulae are indexed to discourse contexts in certain kinds of highly structured discourse has been made explicit in the work on Kuiper and his collaborators.
Wray (1999) and Wray and Perkins (2000) distinguish three main interactional functions of formulae, along with subtypes: (i) manipulation of others (e.g. commands, requests, politeness markers, bargains), (ii) asserting separate identity (e.g. story-telling skills, turn claimers, personal turns of phrase), and (iii) asserting group identity (e.g. group chants, ritual texts, proverbs, forms of address, hedges).

4.5. How prevalent are formulae in “ordinary” speech and writing?

How prevalent are formulae in “ordinary” speech and writing, i.e. discourse that is not obviously highly formulaic? Which types of formulaic expression are most frequent? Can genres or text types usefully be distinguished on the basis of the types of formulae they exhibit? To what extent does “ordinary discourse” resemble oral formulaic styles in having strict discourse structure rules with sentence-level formulae indexed to these?

Becker (1975) and Bolinger (1976) concluded, presumably on the basis of experience rather than careful statistical analysis, that prefabricated expressions are pervasive in language. The creation of electronically-stored corpora and sophisticated computer software has led to an explosion of large corpus-based studies over the past 20 years, and these have confirmed the pervasiveness of phrasal units in various genres of discourse, written and spoken, and have facilitated taxonomic studies of such units.

But just how pervasive? Altenberg (1998: 101-102) found that the London-Lund corpus of spoken English of just over half a million words contains over 201,000 recurrent word-combinations. He estimates that over 80 percent of the words in the corpus form part of a recurrent word combination in one way or another.

However, it is not necessary to study large corpora to show that everyday speech and writing depends heavily on conventional expressions. Analysis of quite small samples of spoken or written text are sufficient to show this (Cowie 1991, 1992, 2004; Cowie and Howarth 1996; and also (using a larger corpus) Howarth 1996). For instance, Cowie (1992) examined two articles in The Times and found that about 35–45 percent of all sequences of a given structural type consisted of restricted collocations. He noted a good deal of creative adaptation in the use of collocations. Cowie (1991) scanned two articles in The Observer for collocations of transitive verb plus object noun and found that about 40 percent of the instances were commonplace restricted collocations, but in this sample there was little creative manipulation of collocates. Using a large corpus of written language, Moon (1998a) found that around 40 percent of phrasal lexemes consisted of verb plus complement.

A separate question is how many conventional expressions are known to the average native speaker of a language. Altenberg (1998) notes that the London-Lund corpus of spoken English contains over 68,000 distinct recurrent expressions. There remains the problem of when a recurrent expression can be considered to be lexicalised rather than free, or rather to what degree it is lexicalised (see discussion in 4.2).

4.6. The role of conventional expressions in fluent speech production and comprehension

At one level, the problems and solutions faced by all spontaneous speakers resemble those of auctioneers or commentators doing play-by-play reports on rapid sports. Formulae contribute to fluency and coherence. However, there is a significant difference. In these particular highly formulaic genres, there is little demand for syntactic integration across clauses. Most of the events and situations can be described by separate clauses and the order and content of things to be said is limited. In conversational speech, complex sentences are common and the order and content of topics is less predictable.

In the 1970s, Frances Syder and I and our assistants transcribed a corpus of around 300,000 words of spontaneous speech. The pattern of pause placements that Syder and I observed led us to propose “the one clause at a time hypothesis”, which holds that in a single planning act it is not possible for a speaker to encode novel lexical combinations across independent clause boundaries (Pawley and Syder 1983a, 2000). When a speaker commits himself to a multi-clause utterance he gambles on being able to formulate a fluent and coherent continuation. Yet to attain native-like fluency, say in a narrative or debate, one must routinely be able to manage such continuations. The risks are reduced by the availability of a large store of familiar construction types and speech formulae, some of which span two or more clauses. However, when native speakers attempt complex sentence structures without being able to fall back on such formulae their speech is typically hesitant, though not necessarily incoherent. This conclusion is consistent with that of Goldman-Eisler (1968), who found higher levels of hesitation when speakers have to explain the point of a cartoon rather than describe the sequence of events. In the 1970s, Wallace Chafe and his stu-
dents carried out an experimental study of the relation between idea units and the flow of speech (Chafe 1979, 1980). This led him to propose (Chafe 1987, 1994) an encoding restriction on units of fluent speech that is even more severe than the “one clause at a time constraint”, namely the “one new concept at a time constraint”.

Wray (1992: 160) offers an explanation of why listeners should have a low tolerance to non-fluency in terms of the “focusing hypothesis”. This hypothesis is based on the view that the brain cannot easily focus on two things at once. When processing speech, speakers and listeners have a choice of processing strategies: analytic or holistic. They normally prefer a holistic strategy because it is more economical, requiring minimal attention to be paid to low-level operations while paying attention to the larger unit. If listeners are processing clauses holistically, they are likely to be disturbed by mid-clause dysfluencies which require a shift of focus to a more atomistic level of analysis.

Wray and Perkins (2000) note that formulaic sequences do not have to be generated and so free the encoder to attend to concurrent tasks. They go on to identify different tasks performed by different kinds of phrases. First, a wide range of ready-made phrases increase speed and fluency of production. Second, there are formulae that buy time, e.g. fillers, turn-holders, discourse organisers, and repetitions. Third, there are mnemonics and memorised texts, which gain and retain access to information otherwise unlikely to be remembered.

4.7. The role of formulae in idiomatic command of a language. Rethinking the grammar-lexicon boundary

In a course he taught at the University of Hawaii in 1975 and in two subsequent books, George Grace (1981, 1987) argued that the key problem in understanding how language works is understanding how it is used to say particular things, for describing what happened, or who did what to whom. It seemed that the dominant linguistic theories did not formally recognise that anything is being said at all. This line of enquiry leads us to ask: What things can be said in a language? What apparatus do we need to describe established ways of talking about things? For example, why do languages have clauses? Clauses are good for saying things. They are well adapted to specifying conceptual events and situations. Speaking a language idiomatically is a matter of conforming to established ways of saying things.

Grace pointed out the limitations of what he called the “grammar-lexicon” view of language for handling idiomaticity and other facets of language. That phraseological expressions straddle the grammar-lexicon boundary awkwardly is part of the problem but only one part. There is something fundamentally wrong with dividing language up in this way. He proposed instead a basic division between what is said (content) and how it is said (form).

I attended that 1975 course and read Grace’s ensuing series of “ethno-linguistic notes” and as I did so, a number of things that had puzzled Frances Syder and myself in our conversational studies fell into place. Next year we wrote the first draft of “Two puzzles for linguistic theory: native-like selection and native-like fluency” (Pawley and Syder 1983a). The first puzzle stems from the fact that there are many grammatically possible ways of expressing the same essential idea but most are not native-like. How does the native speaker know which are the “normal” ways and which are odd? It seems obvious that knowledge of specific conventional expressions is an important ingredient. However, it is also clear that knowledge of some quite abstract construction types must be part of the idiomatic command of a language. For example, English speakers conventionally tell the time by naming an hour and specifying a number of minutes before or after that hour. Ways of referring to the minutes are highly conventionalised. When using the *It's past* *H* formula, the minutes can be grouped in units of fives, tens or a quarter or half an hour, but not in thirds of an hour. One speaks of “half past one”, but not of “half to two”, or of “a quarter past one” but not of “a third past one”. One says “20 to two” but not “40 past one” or “half past one plus 10”. We suggested (Pawley and Syder 1983a: 216–217) that every dictionary entry for a complex lexical form of literal meaning has its own mini-grammar.

Later, drawing on ideas in Grace (1981, 1987), Kuiper and Hago (1984) and Syder (1983), I argued that to know a language we need to command many different “subject matter codes”. A subject matter code consists of “conventions shared by members of a speech community that specify, in more or less detail, what things can be said about a particular topic, how these things are said, idiomatically, and when and why they are said, appropriately. That is to say, it is a code for binding linguistic content with form, context and purpose” (Pawley 1991: 339). While discourse genres with restricted subject matters, such as the language of weather forecasts, auctions or courtroom trials, provide the most highly structured examples of subject matter codes, “normal” talk or writing in any domain will conform to quite strict conventions of this type.
The challenge of making sense of selectional restrictions and semantic interpretations in grammatical constructions has led some grammarians to question models that treat grammatical competence as an autonomous system. Fillmore, Kay, Goldberg and their associates argue for a definition of "construction" that includes not only syntactic, but also lexical, semantic and pragmatic information, i.e. which can subsume much in the domain of idiomaticity (Croft 2001; Fillmore 1984; Fillmore, Kay and O'Connor 1988; Fillmore et al. 1999; Goldberg 1995; Goldberg (ed.) 1996; Lambrecht 1984; Tomasello 2003). After investigating the properties of the let alone construction (as in I wouldn't give five dollars for that, let alone ten), Fillmore et al (1988: 534) conclude that

in the construction of a grammar more is needed than a system of general grammatical rules and a lexicon of fixed words and phrases... [A] large part of a language user's competence is to be described as a repertory of clusters of information including, simultaneously, morphosyntactic patterns, semantic interpretation principles to which these are dedicated... and in many cases, specific pragmatic functions in whose service they exist. ... A language can associate semantic information with structures larger than elementary lexical items and can associate semantic interpretations with structures larger than and more complex than those definable by means of single phrase structure rules.

Restricted collocations and idioms have figured in a variety of recent studies done under the rubric of "event structure". Levin, Rappaport, Havav and others have investigated the semantics and grammar of verbs (simple and complex) and their arguments (e.g. Levin 1993; Levin and Rappaport 1995). They argue that the meaning of a verb largely determines its argument structure and, therefore, the core syntax of clauses. With particular complex predicates, such as wipe off, wipe clean, shoot up, shoot down, shoot dead, go down, the question arises whether the predicate should be regarded as a lexical unit or as a syntactic string. Using a somewhat different framework, Talmy (e.g. Talmy 1985, 2000) has compared constraints on the combination of verbs and their satellites in various languages. In their Pattern Grammar Hunston and Francis (2000) attempt to do something similar for English using the Cobuild corpus.

4.8. Conventional expressions in children's language learning


Since then there have been many further studies, taking up, particularly, the issues raised by Wong Fillmore and Peters. These are reviewed in some detail by Wray (2000, 2002). Wray and Perkins (2000) identify four stages in the use of formulae in first language acquisition: the child (1) begins with a fully holistic strategy, (2) moves to a strongly analytic phase as grammar and lexicon are acquired, (3) makes more and more use of processing shortcuts, through fusion of sequences that were once analytically processed, and (4) by late teenage years, reaches a balance that favours holistic processing, with a preference for metaphorical over literal interpretations.

4.9. Conventional expressions in language learning by adults

Most adult second language learners seem to have particular difficulty with certain kinds of formulaic language, not only during early stages of learning, even when they are otherwise completely fluent. The literature on these matters is fairly limited (see reviews by Yorio 1989; Weinert 1995; Wray 2000, 2002). Yorio (1989: 68) concludes that "Unlike children [adult L2 learners] do not appear to make extensive use of prefabricated, formulaic language, and when they do they do not appear to be able to use it to further their grammatical development."

However, some kinds of idiomatic expression present more problems than others. Howarth (1996) compared the written English proficiency of native and non-native university students, concentrating on verb + object collocations. He found that idioms present fewer problems for non-natives than restricted collocations do, presumably because idioms tend to be relatively fixed and salient. Non-natives make more sparing use of restricted collocations than natives (23 vs. 38 percent of total collocations) and make
many errors. However, collocational errors do not appear to correlate closely with other measures of a learner’s proficiency. He suggests that EFL teachers, knowing little of the phraseological mechanisms of the language, lead learners to believe that English has two categories: free combinations and idioms, with little awareness of the middle ground occupied by restricted collocations.

Granger (1998) compared native and fluent non-native speakers’ use of conventional phrases in a corpus of writings. Her non-natives had French as their mother tongue. She finds sharp differences. Non-natives overused individual amplifiers, like completely and totally, in places were idiomaticity conventions demand a specific adverb. When using certain “sentence-builder” formulae which can be formulated either as passive (“It is claimed that”) or active (“I / we / you claim that”) constructions, they massively overused the active. And non-natives overused certain lexical phrases, such as the fact that, and as far as is concerned, a pattern which Granger viewed as showing over-reliance on a limited repertoire of “fixed anchorage” points. Finally, non-natives are much less able to detect deviations from standard collocations.

How should restricted collocations and other prefabricated expressions be taught to foreign learners? Cowie would include texts in which restricted collocations are first identified then imitated and finally judiciously varied. Methods must take account of “striking evidence of stability and repetition in [multiword] vocabulary use” (Cowie 1991: 114). Granger (1998) warns against the current vogue of basing EFL programs on first language models of learning, which would favour procedures that make heavy use of prefabs, and against relying on generic teaching materials, put together without regard to differences between the mother tongues of the learners. She calls for teaching materials to be based on language-specific contrastive research.

4.10. Phrasal expressions in (mainly English) lexicography

The Euralex bibliography of phraseology www.imu.uni-stuttgart.de/euralex/bibweb/ shows that of the productive scholars at work in the 1970s, the majority were Eastern Europeans. Around 1970 several major publishing houses of English dictionaries began to get serious about phraseology. The penny had dropped that a very large part of the native speaker’s linguistic knowledge consists of phrasal expressions of one sort or another and that in the ever growing market of foreign language learners of English there was room for dictionaries that deal solely with idiomatic phrases and that give these systematic treatment. Some of the best lexicographers in the business put their minds to it and within a few years high quality phrasal dictionaries of English began to appear.

The new generation of phrasal dictionaries improved on their predecessors in various ways. A finer-grained taxonomy of types of idiomatic expressions was worked out. The main advances have been in the treatment of the largest class of multiword units, which are often termed “restricted collocations”, in contrast to pure idioms and figurative idioms. In restricted collocations at least one element has a specialised sense that occurs only in combination with the other element(s), e.g. break in break one’s fall, meet in meet the demand, chequered in a chequered career / history.

The Oxford Dictionary of Current Idiomatic English (ODCIE) vol 1 (Cowie and Mackin 1975) was the first large dictionary of English phrases produced by native speakers. Basically it treated phrasal verbs – multiword units consisting of a verb and a particle and / or a preposition, e.g. back away, fall through, size up, abide by, run into, take to, put up with, set up as, take out on. But some of its headphrases contained additional elements, as in come in handy, come down to earth, come between somebody and something, when one’s ship comes in, come into sight / view, come out of the blue, come to a dead end, come to terms with. A second volume in the ODCIE series followed (Cowie, Mackin and McCaig 1983), dealing with idioms (pure and figurative) and with those restricted collocations that are invariable (break one’s journey, curry favour) or which display limited variation (a chequered career / history).

Technically, the two ODCIE works were superior to earlier English phrasal dictionaries in various ways. Headphrases were richly illustrated with well contextualised examples. Variability in semi-productive expressions is expressed in a clear and precise way. Variation is usually shown by a stroke between alternate fillers of a slot, the collocates. Thus we find the headphrases: Let into a / the secret; have an ear / eye / nose for; have the best / worst of; and keep a tab / tabs / a tag on. Optional elements are shown in parentheses, e.g. keep in touch (with), protect (against / from), stop (dead) in one’s tracks. ODCIE2 was the first English dictionary to distinguish between pure idioms (like spill the beans), figurative idioms (do a U-turn, keep a clean sheet), and restricted collocations (break one’s journey).

Other publishers soon followed suit: the Longman Dictionary of English Idioms appeared in 1982, Selected English Collocations (Kozłowska and
Dzienkowalska) in 1982, *The BBI Combinatory Dictionary of English* (Benson, Benson and Ison) in 1986 and Kjellmer's three volume *Dictionary of English Collocations* in 1994. The BBI Combinatory Dictionary supplies information about word combinations on several levels, including syntax (complementation patterns of verbs). Lexical collocations are arranged by grammatical patterns, e.g. transitive verb + noun (commit treason), adjective + noun (strong tea).

In spite of impressive progress in this field phrasal expressions continue to be a hard nut for lexicographers, partly for the reasons outlined by Moon (1998a). There are problems of placement and ordering. Should a phrase such as (be-TENSE not) the be-all and end-all be alphabetised under be, not, the, or be-all? Should the phrase be given a separate entry or be registered as a subentry of one or more of its constituents? Most dictionaries place them as secondary entries under one or more single-word headwords with cross-references. Often, adequate data are lacking on the variability and use of such expressions. Which is the canonical form of a variable expression? How to represent variability? When are related formulae to be considered exponents of one basic formula or more than one (Stubbins 2001)?

None of the phrasal dictionaries of English give due attention to the special features of situation-bound expressions. "Speech-act expressions" are discussed in the introduction to the *ODCIE2*, but in the dictionary entries these are not given separate labels. The reader can, however, pick them out because the definitions of such phrases contain information about discourse function, as in (4), and sometimes also about discourse context, as in (5):

(4) how many times / how often do I have to do sth? complaint that one's opinions, statements, requests have not been heeded or remembered; complaint that one has heard sth more often than is necessary or desirable.

(5) how about it / that? (informal) exclamatory question asked about sth just done, discussed or discovered; challenging remark, often made to sb with whom one disagrees, or whose behaviour one wishes to change.

The definition given for how about it / that? points to two distinct discourse uses for this expression. A closer examination shows that, in fact, at least three functionally and formally distinct expressions have been combined in a single entry. The three distinct functions are associated with different prosodic patterns. How about it? should be spoken with a melody that signals 'this is a challenge (to deliver on a promise or proposition'). How about that, on the other hand, may be associated with at least two quite different prosodic patterns and functions. In one of its uses the phrase should be spoken with a tune that (by stressing and raising the pitch on the second syllable of about) expresses marveling at or being pleasantly surprised by a discovery, similar to Isn't that amazing? In another use the speaker says, in effect 'I draw your attention to this matter and I want your opinion on it'. Here the main contour stresses are on how and that and there is no main stress on about. As this example shows, a shortcoming of the *ODCIE* is the lack of information about the various features here lumped together under the labels "music" and "body language".

The accounts given of function and context for situation-bound expressions tend to be skimpy. A look through *ODCIE2* shows that the number of expressions whose discourse functions and / or discourse contexts are indicated is probably less than five percent of the 7,000 entries. A good many other expressions that are certainly used as situation-bound expressions are included but their functions are not made explicit.

I referred above to useful devices introduced by recent phrasal dictionaries in order to describe variation in the lexical core of conventional expressions, i.e. those elements that are lexically fixed or that allow a very limited range of lexical choices. However, describing the lexical core of such expressions is not the hardest part of the job of describing their formal structure. More difficult is to establish which other lexical elements can enter into construction with the lexical core. Often the choice of subject or object of a verb, or of elements modifying the verb or modifying one of the nouns, is highly restricted. Sometimes there are severe restrictions on the choice of grammatical elements, such as determiners or tense or aspect markers. In such cases the *ODCIE* gives a list of the choices in each such variable element, placing the list(s) of possible fillers immediately after the definition and before the illustrative examples, using S for subject, O for object, V for verb, A for adverb, det for determiner, and so on. For example:

(6) hold water be sound, valid... S: theory, argument, explanation, excuse, suggestion,...

This tells us that it is idiomatic to say: *the theory holds water; the argument holds water, or his excuse holds water, but not, say, the talk holds water or the novel holds water*, because talk and novel are not listed as possible sub-
jects. What about that idea holds water? Idea is not included in the list of possible subjects but it seems acceptable. There is other information that we are not given. Are there other restrictions on what you can do with phrases, e.g. can you negate hold water? The answer is yes. In fact, this expression usually appears in the negative. Can you vary the tense freely? Can you add modifiers? It turns out you can say the theory doesn’t / didn’t hold water, but it is not quite idiomatic to say the theory isn’t holding much water any more, or the theory will hold some water tomorrow.

The next example is a noun phrase which typically occurs with one of just a few verbs:

(7) a howling success [Comp (NP)] (informal) a very great success, sth that receives much praise and (loud) acclamation V: be; turn into; make sth. [A triangle enclosing an exclamation mark is placed after V to indicates that the set of collocates is restricted not open.]

Igor Mel’čuk and his Russian associates have done pilot studies of a special kind of phrasal lexicon which they term an “explanatory combinatorial dictionary” (ECD) (Mel’čuk 1988, 1998; Mel’čuk and Zolokovsky 1984). Its headwords are idioms and restricted collocations and it aims to rigorously describe collocational restrictions and syntactic patterns associated with such phrasal units.

Some of the technical shortcomings of current phrasal dictionaries can be put down to limitations of budget and market requirements. Commercial lexicographers work within tight constraints of time. They are simply not able to do really fine-grained research on the behaviour of every one of the thousands of headwords or headphrases included. The books they produce must be affordable and their contents accessible to readers with a limited knowledge of the technical vocabulary of linguistics and lexicography. The more important point is that works such as the ODICE have gone some way towards developing the apparatus needed to do adequate descriptions of situation-bound expressions (SBEs).

However, not all the shortcomings of contemporary dictionaries of SBEs can be put down to time constraints. In particular, I doubt that we yet have the analytic and descriptive tools to handle all aspects of SBEs. As Austin and others pointed out some time ago, speech acts, like language use in general, are bits of social behaviour. To deal with speech functions and discourse contexts we need a well-articulated model of social behaviour, in which the rules governing face to face talk are seen to be subordinate to more general conventions of social conduct. And we need an apparatus for describing the music and body language that is part of everyday speech.

4.11. Grammaticalisation

A good case can be made that formuiality is responsible for the very existence of grammar, for the emergence of grammatical constructions. Since the 1970s there has been a great deal of work on “grammaticalisation” but much of this has focused on how lexical words change into grammatical functors. There has been rather less done on how relatively loose discourse structure conventions are prone to develop into tighter multiclause sentence structures which in turn may become compacted into single clause structures which in due course may reduce to word-like structures. A pioneer in this regard has been Talmy Givón, who gave us the slogan “today’s morphology is yesterday’s syntax” (Givón 1971: 413) and proposed the grammaticalisation path: discourse > syntax > morphology > morphophonemics > zero. In a number of works (e.g. Givón 1979a, b) he has insisted that the processes of clause combining should be considered part of grammaticalisation, a view that has gained fairly general acceptance (Hopper and Traugott 1993).

A striking example of discourse structure becoming clause structure, with the aid of speech formulae, is seen in certain verb serialising languages. There are some such languages that allow short narratives, reporting several successive events, to be compacted into single clauses with complex predicates. Such narrative serial verb constructions, containing a string of up to five or six bare verb stems under a single intonation contour, are found, for example, in certain languages of New Guinea (Durie 1997; Pawley and Lane 1998). The single clause narratives represent recurrent types of event sequences and are highly schematic in form; that is to say, they are formulae.

5. Conclusions

The quantity of research on the structure and use of formuialic expressions has multiplied over the past 30 years. Has this led to the emergence of a new subdiscipline with its own core research agenda and methods, terminology, conferences, journal, textbooks, etc.?
In lexicography, to a degree, this has happened. As Cowie points out, phraseology now occupies a prominent place in lexicographical theory and practice. The problems posed by restricted collocations and idioms have stimulated dictionary-makers to think hard about how to produce linguistically more sophisticated analyses and to hold conferences where they can compare ideas. There are some differences in preferred methods and terminologies but there is much common ground. Regular conferences now take place under the auspices of the European Society for Phraseology (Europhras). There have also been four International Symposia on Phraseology (Leeds, German, Stuttgart, Rome). Readers have been published in English, German, Russian and Spanish.

In linguistics generally there is no comparable cohesion, though we may note the recent launching of an electronic journal devoted to constructions, <http://www.phil-fak.uni-duesseldorf.de/constructions>. Perhaps it is a good thing that “formulaic language” has not become a field in itself, a specialised branch of linguistics. The beauty of conventionalisation is that it is a process which touches on many different facets of language structure and use: language learning, encoding and decoding, idiomaticity, grammaticalisation and the grammar-lexicon boundary, the workings of conversation and discourse, how and why particular speech genres arise, and so on. Thus, the study of the structure and use of formulaic expressions is most usefully pursued not as an end in itself but as one facet of a range of more general problems to do with language.

Two opposing views have dominated Western thinking about what languages are. There is the formal, mathematical view, which sees any natural language as an algorithm for specifying an infinite number of grammatically and phonologically well-formed strings. And there is the humanistic perspective, in which a language is regarded chiefly as a means for encoding a particular culture or world view, represented, for example, by the things that people say (or write) to achieve particular social and communicative purposes. To me, the most satisfactory thing about the study of speech formulae, especially productive speech formulae, is that this provides a broad bridge that spans these two views. In doing so it pushes us towards adopting a model of language that can have the best of both worlds.

Notes

1. An earlier version of this paper was presented at a colloquium on formulaic language at the American Association of Applied Linguists conference, Vancouver, March 11-14, 2000. I am indebted to Wally Chahe, Tony Cowie, Kon Kuiper, and Alison Wray for much useful discussion and for thoughtful comments and corrections on the revised version.

2. There were still other domains of scholarship which treated formulaic language which I have not mentioned, e.g. work on text concordances, translation and decipherment.

References

Aijmer, Karin

Altenberg, Bengt

Amosova, Natalija N.

Austin, John L.

Bauman, Richard

Bauman, Richard

Bauman, Richard, and Joel Sherzer

Becker, Joseph D.

Benson, Morton, Evelyn Benson, and Robert Ison
Bernstein, Basil

Bernstein, Basil

Bolinger, Dwight

Bongers, Herman

Brown, Dennis

Chafe, Wallace

Chafe, Wallace

Chafe, Wallace

Chafe, Wallace (ed.)

Chafe, Wallace

Chafe, Wallace

Chafe, Wallace
1996  Beyond beads on a string and branches in a tree. In Conceptual

Coulmas, Florian

Coulmas, Florian (ed.)

Cowie, Anthony P.

Cowie, Anthony P.

Cowie, Anthony P.

Cowie, Anthony P.

Cowie, Anthony P.

Cowie, Anthony P.

Cowie, Anthony P.

Chomsky, Noam

Clark, Ruth

Code, Chris

Cote, Michael, and Sylvia Scribner

Cowie, Anthony P.  
2004 Phraseology in formal academic prose. In Jan Aarts, Inge de Moorink, and Herman Wekker (eds.), Studies in English Language Teaching in Honour of Flor Aarts, Amsterdam: Rodopi, 43–56.

Cowie, Anthony P. (ed.)  

Cowie, Anthony P., and Peter Howarth  

Cowie, Anthony P., and Ronald Mackin  

Cowie, Anthony P., Ronald Mackin, and Isabel R. McCaig  

Crystal, David, and Derek Davy  

Downing, Pamela A.  

Du Bois, John  

Durie, Mark  

Enfield, Nicholas J. (ed.)  

Ferguson, Charles  

Fernando, Chiara  

Fillmore, Charles J.  

Givón, Talmy  

Givón, Talmy  

Givón, Talmy  

Gläser, Rosemarie  

Goffman, Erving  

Goldberg, Adele E.  

Goldberg, Adele E. (ed.)  

Goldman-Eisner, Frieda  

Grace, George W.  


Lyons, John

Makkai, Adam

Mel'ëük, Igor

Mel'ëük, Igor

Mel'ëük, Igor, and Alexander K. Zholkovsky

Moon, Rosamund E.

Moon, Rosamund E.

Nattinger, James R.

Opie, Iona, and Peter Opie

Opie, Iona, and Peter Opie

Palmer, Harold E.

Palmer, Harold E.

Parry, Milman

Parry, Milman
Parry, Milman

Pawley, Andrew

Pawley, Andrew

Pawley, Andrew

Pawley, Andrew

Pawley, Andrew

Pawley, Andrew, and Jonathan Lane

Pawley, Andrew, and Frances Hodgekiss Syder

Pawley, Andrew, and Frances Hodgekiss Syder

Pawley, Andrew, and Frances Hodgekiss Syder

Peters, Ann

Peters, Ann
1983  *The Units of Language Acquisition*. Cambridge: Cambridge University Press.

Quirk, Randolph, Geoffrey Leech, and Jan Svartvik

Riggenbach, Heidi (ed.)

Rumsey, Alan

Searle, John

Stenström, Anna-Brita

Stubbs, Michael

Syder, Frances Hodgekiss

Talmy, Leonard

Talmy, Leonard

Tomasetto, Michael

van Lancker, Diana
van Lancker, Diana

Vinogradov, Viktor V.

Weinert, Regina

Weinreich, Uriel

Wong-Fillmore, Lily

Wray, Alison

Wray, Alison

Wray, Alison

Wray, Alison

Wray, Alison, and Kazuhiro Namba

Wray, Alison, and Mick Perkins

Yorio, Carlos A.