VOCABULARY

APPLIED LINGUISTIC PERSPECTIVES

'... recommended for anyone who has occasion to wonder about words and their meanings.'
EFL Gazette

RONALD CARTER

Vocabulary

From reviews of the first edition:

'A timely state-of-the-art overview of the whole subject . . . the author's stance is broad . . . the work is eminently readable and coheres as a whole extremely well . . . an extremely wide range of research and literature is covered. This is certainly recommended for anyone who has occasion to wonder about words and their meanings.'   *EFL Gazette*

'*Vocabulary* is and deserves to be a widely-read and widely-used textbook in undergraduate courses on Semantics, Lexicology, and Corpus-Linguistics, and on diploma and postgraduate TESOL courses such as Teaching Vocabulary.'   Andrew Goatly, *Lingnan University, Hong Kong*

How do we teach and learn vocabulary? How do words work in literary texts? How are words treated in dictionaries? In this book, Ronald Carter provides the necessary basis for the further study of modern English vocabulary with particular reference to linguistic descriptive frameworks and educational contexts. *Vocabulary: Applied Linguistic Perspectives* includes an introductory account of linguistic approaches to the analysis of the modern lexicon in English and discusses key topics such as vocabulary and language teaching, dictionaries and lexicography and the literary, stylistic study of vocabulary.

This *Routledge Linguistics Classic* includes a substantial new introductory chapter situating the book in the current digital age, covering changes and developments in related fields from lexicography and corpus linguistics to vocabulary testing and assessment as well as additional new references.

Vocabulary: Applied Linguistic Perspectives has been widely praised since first publication for the breadth, depth and clarity of its approach and is a key text for postgraduate students and researchers studying vocabulary studies within the fields of English Language, Applied Linguistics and Education.

Ronald Carter is Professor of Modern English Language at the University of Nottingham, UK, and has written and edited more than 30 books, including most recently *Language and Creativity* (2004). He is the author and editor of 12 books with Routledge, and has edited and co-edited five Routledge series.

Vocabulary
Applied Linguistic Perspectives

Ronald Carter

Routledge
Taylor & Francis Group

Contents

List of figures and tables

Figures

Tables

Acknowledgements

Preface to the first edition

This book is a guide to some ways in which linguistic insights into the structure of the modern English lexicon might be applied. Applied linguistics is generally associated with second- and foreign-language teaching and this book contains an extensive review of developments in vocabulary teaching. The book will, however, demonstrate that the possibilities for application can be broader in scope. In addition to brief discussions of mother-tongue early language development and a number of social issues in vocabulary use, such as the relationship between lexis and ideology, the book examines some applications of lexicological description in three main areas: language learning and teaching, lexicography, and literary-text study.

Although there are (in the third part of the book) two detailed case studies in which particular arguments are presented relevant to work in what might be termed 'lexical stylistics', the orientation of this book is mainly descriptive. It seeks to *report* on research, *review* developments in lexical description and *comment* on specific points of intersection between linguistic insight and areas of application. No description is entirely neutral, however, not least because, in a relatively wide-ranging book such as this, selection is inevitable and selection is itself a form of evaluation. For example, the importance of studying lexis in discourse and of using informants to assist description of lexical associations is stated regularly, both explicitly and implicitly, throughout the book. But, wherever possible, a descriptive overview constitutes the main design.

This is not an apology. There are very few books or studies of lexis in relation to applied linguistics and it is hoped that to an extent this book helps to clear some ground from a number of different perspectives at the same time as indicating areas in which work needs to be done. However, I should have liked to have devoted more space to the social, cultural and political issues which the study of vocabulary necessitates, in particular in the following areas: the vocabularies of world Englishes, and the implications of this for lexicographic practice and for stylistic analysis of international literatures in English; sociolinguistic and historical accounts of lexical change, particularly with reference to dialect; the lexicology of sexism, ageism and racism – areas which are beginning to attract the attention of language researchers. I also

recognize that more space could have been given to psycholinguistic accounts of the lexicon, especially in relation to language learning. Such omissions have to be seen in relation to the biases which any 'perspective' entails but also to the need to describe the chosen areas of application in as much detail as possible.

Applied linguistics is a growth area in linguistics. There is an ever increasing number of books, journals, courses, conferences and professional associations which appropriate the term. However, although discussions in Kaplan (1980), Brumfit (1981), Crystal (1981) and Widdowson (1981; 1985) are notable exceptions, debate about the aims, goals and procedures for applying linguistics is not especially prominent. This book does not seek explicitly to address such issues; the domain of applied linguistics is not made any less undertheorized as a result of its appearance. And readers expecting to find a course book in lexical semantics may also be disappointed. The field is a wide and complex one, and work is reviewed here in which aspects of lexis are selected for what is judged to be their relevance to the above mentioned specific contexts of applications. The book is a practical guide to what has been done, what is being done and, where feasible, to *how* it is done. In this respect the book is not programmatic but this is not to say that programmatic as well as procedural issues are not of considerable importance to applied linguistics as a discipline.

The book is divided into three main parts. The first part, which consists of four chapters, outlines some basic issues and undertakes analysis of lexis in increasingly larger units of language – moving from the forms of words to the roles of words in discourse contexts. The second part consists of three chapters and is devoted to discussion of applied linguistic issues raised by vocabulary being 'used' in language learning and teaching contexts, in lexicography (especially with regard to learner dictionaries) and as a component in the stylistic analysis of literature. In all three domains vocabulary is of particular relevance and has not received as much attention as it merits. The third part of the book consists of two studies in lexical stylistics in which particular analytical procedures – based mainly on informant analysis – are shown to be applicable to fuller and more precise description of the role of lexis in the style of literary texts and in the way style levels can be marked in lexicographic entries.

It is hoped that these applied linguistic perspectives will provide some basis for further study and 'application' to be made. Vocabulary and vocabulary use are now beginning to attract the attention of applied linguists after a period of relative neglect. It is hoped that this book may contribute to this renewal of attention. And finally it is hoped that a frequently used word in the English lexicon – *he* – will be read to mean he/she wherever it occurs in the book.

<div align="right">
Ronald Carter
Nottingham
1987
</div>

Preface to the second edition

This second edition of *Vocabulary* updates recent progress in the last ten years with particular reference to the momentous changes in the computational analysis of vocabulary. There are updated and re-worked sections on vocabulary and language teaching; dictionaries and lexicography; and the literary stylistic study of vocabulary. And there are newly written parts of chapters on the relationship between vocabulary, grammar and discourse. The emphasis remains squarely on the description of vocabulary and on the basis that the more we know about how vocabulary works the better our applications are likely to be. The second edition does not, any more than the first edition, pretend to extensive coverage of all aspects of vocabulary but rather offers *perspectives*. But they are perspectives, of course, which inevitably entail the limitations which accompany the viewing point. In a field as dynamic as vocabulary studies has been since the publication of the first edition of this book in 1987 it is always difficult to know where to begin, and where and when to end. In taking these decisions, it is inevitable that where you stop might be where others would start.

Ronald Carter
Nottingham
January 1998

Vocabulary and applied linguistics: recent past and nearer future

This chapter has two main purposes: to reflect on major changes in the field of vocabulary studies since the publication in 1998 of the second edition of *Vocabulary: Applied Linguistic Perspectives*; and to offer, as a complement to this necessarily largely retrospective view, a brief view of some likely directions that the field of vocabulary might take in the future.

In the past decade vocabulary studies have become an even more highly active field of research and application. One chapter cannot, of course, capture all of this diversity and richness; this chapter is organised therefore around key landmarks. These are:

1   Formulaic language: everyday words and chunks and things.
2   Developments in vocabulary learning and teaching.
3   Creativity and literary vocabularies.
4   Vocabulary and the spoken language: a prospective view.

Vocabulary: Applied Linguistic Perspectives is oriented more extensively to sociocultural approaches to vocabulary study. Although material from a more cognitive and psycholinguistic perspective is not ignored, it is not fully treated in the book or in this chapter.

Where is corpus linguistics? The answer is pretty much everywhere in this chapter. Such have been the developments in the use of corpora that the relationship of vocabulary to corpus linguistics does not justify any separate treatment; as with pretty much all current work in vocabulary from an applied linguistic perspective, corpus linguistic approaches are ubiquitous throughout this chapter.

1 Formulaic language: everyday words and chunks and things

From word to chunk: everyday vocabulary

The single word has served vocabulary studies well and will continue to do so. It remains very much the popular conception of how vocabulary is constituted and learned. And just as vocabulary is represented for many people as

single words, so too is it often the unusual or exceptional words that commonly attract attention. Most popular and to a lesser extent many academic books on words and phrases concentrate on the more exotic or uncommon words and phrases and include discussions of their etymology or derivation. Examples would be Liberman's *Word Origins: And How We Know Them* (2005), Hitchings, *The Secret Life of Words* (2009) or Cook's *It's All in a Word* (2009).

This section focuses both on common words and on the patterns that go beyond the confines of a single word. Units consisting of more than one word, such as phrasal verbs, compounds and idioms, have become the subject of increasing attention and the last decade has seen even more attention to formulaic vocabulary patterns or 'chunks'. Some forty years ago, both Halliday (1966) and Sinclair (1966) foresaw the development of computational analysis of texts as a way of getting at the common collocations of a language, and both, in different ways, have fulfilled that vision, especially Sinclair (1991; 2004). The advent of corpus linguistics has enabled linguists to verify these earlier, mainly intuition-based notions in actual, attested language use on a large scale.

When it comes to common (rather than uncommon) words it is the patterns created by banal, everyday words that are most difficult to identify by intuition alone and within recent vocabulary research it is just these words that computers have been very good at disclosing. Common verbs such as *get*, *go*, *turn* and so on display distinct preferences for what they combine with. Things *turn* or *go grey*, *brown*, *white*; people *go* (but not **turn*) *mad*, *insane*, *bald*, *blind*. The shifts in emphasis from the single word to collocations to pairs or groups of words as integrated chunks of meaning and usage have now become a much more fully accepted aspect of vocabulary description and pedagogy (Lewis, 2000; Altenberg and Granger, 2001; Willis, 2003; McCarthy and O'Dell, 2005). Clearly, for the learner of any second or foreign language too, learning the ways in which words form chunks is not a luxury; such patterning permeates even the most basic, frequent words and is needed for successful communication.

Formulaic language: fixed words in fixed places?

Although a considerable amount of work has been done on formulaic language in the last decade, research has tended to be widely distributed across a number of fields (child L1 acquisition, psychology, corpus linguistics). This diversity is illustrated by the wide variety of terminology (see Carter, 1998, ch 4; Wray, 2002, p. 9; Schmitt and Carter, 2004; Schmitt, 2010) that is found for the various sorts of formulaic language:

chunks	*formulaic speech*	*multiword units*
collocations	*formulas*	*prefabricated routines*
conventionalized forms	*holophrases*	*ready-made utterances*
clusters	*lexical bundles*	

Schmitt and Carter (2004, p. 3) illustrate the extent of terminological differences:

> formulaic sequences can be long (*You can lead a horse to water, but you can't make him drink*) or short (*Oh no!*), or anything in between. They are commonly used for different purposes. They can be used to express a message or idea (*The early bird gets the worm* = do not procrastinate), functions ([*I'm*] *just looking* [*thanks*] = declining an offer of assistance from a shopkeeper), social solidarity (*I know what you mean* = agreeing with an interlocutor), and to transact specific information in a precise and understandable way (*Wind 28 at 7* = in aviation language this formula is used to state that the wind is 7 knots per hour from 280 degrees). They realize many other purposes as well, as formulaic sequences can be used for most things society requires of communication through language. These sequences can be totally fixed (*Ladies and Gentlemen*) or have a number of 'slots' which can be filled with appropriate words or strings of words, for example, ([someone/thing, usually with authority] *made it plain that* [something as yet unrealized was intended or desired]).

What descriptions of these various forms reveal is that fixedness is a key feature of formulaic language, allowing such language to be memorized and used as wholes, rather than being newly created for each use. On first inspection, all formulaic language looks as if it is completely fixed, but this is not the case. Of course, idioms are usually cited as examples. Thus, corpus evidence shows that *to have forty winks* occurs almost exclusively as that exact phrase, and not as variations such as **to have thirty nine winks*, **to have a short wink.* In other words, if we want to use an idiom to express the notion 'to have a short sleep' we can only use the intact idiom *to have forty winks* and not some variation. However, much formulaic language is not fixed in this way, and in fact allows for a surprising amount of flexibility (see also Biber *et al.*, 1999; de Cock, 2000). Chapter 3 of this book (in part) and especially Moon (1997; 2010) illustrate this:

not touch someone/something with a bargepole	(British vs. American English)
not touch someone/something with a ten foot pole	
burn your boats	(varying a lexical component)
burn your bridges	
cost an arm and a leg	(verb variation)
pay an arm and a leg	
spend an arm and a leg	
charge an arm and a leg	

every cloud has a silver lining	(truncation)
silver lining	

break the ice	(transformation)
ice-breaker	
ice-breaking	

In fact, it seems that once a piece of formulaic language becomes well known in a speech community, it can be creatively adapted and still be comprehensible (see Carter, 2004, Chs 3 and 4 and section 3 below for further discussion).

Although in several studies the term 'cluster' captures the way words combine and coalesce, the term *chunk* is preferred here as a way of indicating the necessary pre-formulated and flexibly fixed nature of this phenomenon.

The main examples here will be drawn from spoken English, not least because issues of fluent production and reception are foregrounded in talk. As with most high-frequency phenomena, their core contribution to language use is subliminal and not immediately accessible to the intuition of the native speaker or fluent user. This is once again where a corpus comes in.

Looking at corpus data

Using an almost 5 million word sample of North American English conversation from the Cambridge International Corpus (CIC), research underlines that certain chunks occur with greater frequency than some common single words.

The table suggests that many high-frequency chunks (*I don't know*; *something like that*) are more frequent and more central to communication than even very frequent lexical words (*friend*) or very frequent function words such

Table 0.1 High frequency chunks and single words

you know	45,873
really	20,838
I think	17,046
people	11,984
kind of	9,962
and then	8,971
I don't know	8,074
where	7,851
their	6,487
something like that	1,027
friend	1,014
I don't know if	999
a lot of people	759
under	743

as *their*, *where* or *under*. Several chunks isolated in this way are, of course, syntactic fragments and their occurrence is probably due to the regularity of the content-world itself; for example, a fragment such as *and then* generates a common temporal sequence. (For more extended studies of such patterning, see Carter, 2006 and O'Keeffe, McCarthy and Carter, 2007, Chs 2 and 3.)

However, it is in pragmatic categories rather than in syntactic or semantic ones that we are likely to find the reasons why many chunks are so frequent. Pragmatic categories refer to the creation of speaker meanings in context, including such functions as discourse marking, the preservation of face and the expression of politeness, hedging, and purposive vagueness, which create a world of speakers and listeners interacting in real time rather than a purely propositional world, concerned with the content of what is said. For example, some of the most frequent chunks are discourse-markers, e.g. *You know*, *I mean I guess*, *(Do) You know what I mean. You know*, the most frequent chunk, is an important token of projected shared knowledge between speaker and listener. *I mean* is also commonly used when speakers need to paraphrase or elaborate.

Speakers also use indirect forms to perform speech acts such as directives (e.g. commands, requests, suggestions, etc.) to protect the face of their addressees, and the chunks reveal common frames for such acts. Indirectness is also important in the polite and non-face-threatening negotiation of attitude and stance. Chunks in this category include *Do you think*, *I don't know if*, *What do you think*, *I was going to say*. For example:

> [Student talking with a teacher: CANCODE corpus]
>
> **I don't know if** you've seen my magazine article yet but **I was going to say** can I have an appointment with you tomorrow?

Some of the most frequent chunks have a related hedging function, i.e., they modify propositions to make them less assertive and less open to refutation. These include: *I think*, *kind of*, *I don't know*, *I don't think*, *a little bit*. Carter and McCarthy (2006, Appendix 1) provide a range of further examples and frequency lists for 2, 4 and 6 word chunks. Also relevant here is work on listener language and the uses of feedback chunks or response tokens by listeners (items such as *yeah, I've got you*, *really*, *exactly*, *oh, I see*, *that's interesting*, etc.) to indicate their involvement and engagement with the ongoing discourse. (See McCarthy and Carter, 1994; Carter and McCarthy, 2006, pp. 188ff; O'Keeffe, McCarthy and Carter, 2007, pp. 140–58 for a range of examples.)

Chunks expressing purposive vagueness and approximation are also common. Vagueness is central to informal conversation and its absence can make utterances too direct and pedantic or sometimes even impolite, especially in references to number and quantity. Vagueness also enables speakers to refer

to semantic categories in an open-ended way which calls on shared knowledge to fill in what is otherwise referred to obliquely. Such tokens include: *a couple of*, *and things like that*, *and stuff like that*. (See a wide range of research reported in Cutting, 2007.)

What all such chunks show is the ubiquity of interactive meaning-making in everyday conversation and the degree to which speakers engage continuously with one another in an interpersonal dimension. The addition of these chunks to the vocabulary list of any language should not be seen as an optional extra, since the meanings they create are extremely frequent and necessary in discourse, and are fundamental to successful interaction. It is clear and, thanks to extensive corpus research, much clearer now (in 2012) than in discussions in earlier chapters of this book that chunks are best viewed as being evidence of single linguistic choices rather than assembled piece by piece (or word by word) at the moment of speaking.

2 Developments in vocabulary learning and teaching

Some of the questions raised concerning formulaic vocabulary patterns or chunks in recent research include the following:

- If the use of ready-made chunks is central to speaking a language fluently, how can they be most effectively presented and practised in language classrooms and teaching materials?
- How do learners typically process chunks when they encounter them?
- How can learners become aware of chunks and recognise potential chunks when they listen or read?
- Chunks are clearly linked to the world and culture of the native speaker. Why should learners who do not necessarily wish to sound like native speakers need to acquire them? (See note 2 and Prodromou, 2008 for further discussion.)

(See Schmitt, 2010, p. 146 for a more extensive list of accompanying research questions and suggested projects.)

Chunks and fluency

Many of the examples cited above make fluency in speech more of a reality. One of the features of chunks not already discussed, where the evidence has, of necessity, been the purely printed evidence of corpus output on a computer screen, is, however, that chunks have phonological unity; put simply, they need to be said at a fast speech rate. Typically, chunks occupy a single intonation unit (or *tone unit*, separated here by //, characterized by one strong stressed syllable, marked here in bold capitals) and the rest of the chunk is much reduced:

// he's **SHY** // *you know what i* ***MEAN*** //
// they sell **JEW**ellery // *and* ***THAT*** *sort of thing* //
// the **ROOM** was // *a* ***BIT*** *of a* // **MESS** actually //
<div align="right">(O'Keeffe, McCarthy and Carter, 2007, pp. 76–7)</div>

Provided the *chunk* is said with a fast tempo, the utterance will sound natural; the opposite, a fast message with a slow chunk, will sound completely unnatural. The appropriate use of a smooth, quickly uttered chunk can transform even a lower-level speaker's fluency.

Although chunks can be drilled for speed in isolation, naturally it is also necessary to incorporate them into sentences and longer utterances for more sustained practice. Presentation of chunks in spoken language can most naturally be done by raising awareness of them through listening and noticing activities. Practice can also take the form of re-inserting chunks into dialogues from which they have been removed (see McCarthy, McCarten and Sandiford, 2005–6 for numerous examples in the *Touchstone* course book series). High-frequency chunks should not necessarily be obligatory components of the learner's productive repertoire. Much depends on the learner's language learning needs. It may be that in some cases receptive mastery is more important than productive competence; but under no circumstances can chunks be ignored.

Processing chunks: a more psycholinguistic perspective

Wray (2002, p. 9) defines a *formulaic sequence* as 'a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored or retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar'. Wray (2000; 2008) stresses the non-analytical nature of formulaic language in native speaker competence. Attempts by teachers and textbooks to encourage the analysis of chunks by learners are, in Wray's words, 'pursuing native-like linguistic usage by promoting entirely *un*native-like processing behaviour' (2000, p. 463, her emphasis). This is certainly the case. However, Spöttl and McCarthy (2004) suggest counterarguments: (i) there is psycholinguistic evidence that, even among native speakers, at least some degree of literalness or at least metaphoric awareness is retained in the processing of figurative expressions (Gibbs and O'Brien, 1990; Gibbs, 1994). Learners may be even more inclined to analyse and may see it as an important part of the learning process. Receptive mastery may indeed gain from an occasional analytical approach (see Burns, 2001 and Sealey and Thompson, 2004 – the latter reference for application in mainly first language, primary school contexts). It might also be added that the more the learner has successfully acquired a repertoire of chunks, the easier it becomes to reflect and analyse them at a later stage, so that certain aspects of grammatical acquisition may flow from the knowledge and use of chunks, rather than vice-versa.

Chunks may not necessarily be acquired in an 'all-or-nothing' manner (Schmitt and Carter, 2004, p. 4); in other words, the absorption and learning of the meaning and appropriate use of a chunk may be gradual and only apparent over time and after a number of exposures, just as with grammatical structures or single words (see also Conklin and Schmitt, 2008). Within the limits of working memory, fluency is achieved through automaticity of language production, although exactly what automaticity means varies in the literature (Dörnyei, 2009). There is evidence that the use of chunks 'frees up' the cognitive processing load so that working memory can be allocated to other aspects of speaking, such as for planning content, for organizing 'strategic vocabulary' and for vocabulary for discourse management (Girard and Sionis, 2004; McCarten, 2010). Much of automatic processing also depends on the access of memory. Studying the neurobiological correlates of fluency, Dewaele (2002) states that fluency in both L1 and L2 depends on the existence, accessibility and retrieval of procedural knowledge from long-term memory. Fluent speech results when the appropriate or required language is present and accessible in long-term memory and there is sufficient storage of chunks in working memory.

In a study of learners' fluency development after a period abroad, Towell *et al*. (1996) conclude that the observed temporal aspects of fluent speech, i.e. higher rate of speech and longer uninterrupted runs, are a result of more efficient proceduralized knowledge. This could be due to automatization of the processes of speech production connected to the use of ready-made chunks of language. Cameron and Larsen-Freeman (2008) note that, although native speakers can theoretically produce infinitely many utterances that conform to the syntactic rules of the language, they rarely do, and rather tend to mix and match frequently used chunks (Kirk and Carter, 2010; McCarthy, 2010).

Chunks and writing

Although the main focus in this chapter so far has been on the spoken language and on the relationship between chunks and fluency, it is important to understand that chunks occur in all kinds of written contexts too. Fluent writing is of a different order from fluent speaking (and would require a separate chapter) but writing too depends on the writer being able to draw on a range of ready-made, off-the-peg items that act as structuring devices such as prefaces or as concluders. In this respect impersonal passive voice structures with *it* are common and are likely to be recalled and written as whole chunks. For example:

> *It should be remembered that*
> *It should be emphasized that*
> *It will be argued that*
> *It should be stressed that*