Introduction: The many faces of phraseology

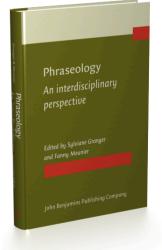
- Sylviane Granger | Université catholique de Louvain, Belgium
 Fanny Meunier | Université catholique de Louvain, Belgium
- doi https://doi.org/10.1075/z.139.04gra

Pages xix–xxviii of

Phraseology: An interdisciplinary perspective

Edited by Sylviane Granger and Fanny Meunier

2008. xxviii, 422 pp.



© John Benjamins Publishing Company

This electronic file may not be altered in any way. For any reuse of this material written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).

For further information, please contact rights@benjamins.nl or consult our website at benjamins.com/rights

Introduction

The many faces of phraseology

Sylviane Granger and Fanny Meunier

Phraseology is pervasive in all language fields and yet despite this fact – or perhaps precisely because of it – it has only relatively recently become established as a discipline in its own right. It is usually presented as a subfield of lexicology dealing with the study of word combinations rather than single words. These multi-word units (MWUs) are classified into a range of subtypes according to their degree of semantic non-compositionality, syntactic fixedness, lexical restrictions and institutionalization. Long regarded as a peripheral issue, phraseology is now taking centre stage in a wide range of fields, from natural language processing to foreign language teaching and now, 25 years after the publication of Pawley & Syder's (1983) seminal article, it is gradually acquiring the place it deserves in linguistic theory.

The impetus for this volume came from an interdisciplinary conference on phraseology entitled *Phraseology 2005*. The Many Faces of Phraseology organized in Louvain-la-Neuve in October 2005. The 170 participants, gathered from a wide variety of countries and specialist research areas, were there to scrutinize the field of phraseology from a wide range of perspectives. Three volumes emanated from the conference: a volume in French entitled *La phraséologie dans tous ses états* edited by Catherine Bolly, Jean René Klein and Béatrice Lamiroy (Cahiers de l'Institut de Linguistique de Louvain, Peeters, 2005), and two volumes in English, one entitled *Phraseology in Foreign Language Learning and Teaching* (Meunier, F. & S. Granger eds. 2008) entirely devoted to the role played by phraseology in L2 learning and teaching, and the current volume that purports to stress the multi-faceted nature of phraseology.

Our aim in organizing the conference was to take stock of the fast growing role of phraseology in a wide range of linguistic disciplines, a development that undoubtedly has a great deal to do with **corpus linguistics** research, which has both demonstrated the key role of phraseological expressions in language and also provided researchers with the automated methods of extraction and analysis they need to study them successfully, a development which has led to a major expansion of the field. Whereas previously phraseology had encompassed the study of only the most fixed and opaque

multi-word units, it now covers a much wider range of lexical units, many of which display a high degree of syntactic variability and semantic compositionality.

The current phraseological boom is evidenced by the numerous specialist publications and conferences on the subject. There are many niche areas of research buzzing with activity. It would seem however, that there is very little contact between these different areas of activity. Natural language processing researchers are often unfamiliar with work related to the typology of phraseological expressions. Researchers trying to draw up rigorous phraseological typologies are often equally unfamiliar with work being carried out in the automatic extraction of phraseological units. Similarly, there is very little contact between psycholinguistic researchers attempting to define the role of phraseology in language acquisition, comprehension and production and educational researchers aiming to give phraseology a higher profile in language teaching. In general terms, corpus linguistics studies describing phraseological expressions in large computer corpora are undeservedly little known. This lack of contact between different areas of phraseological research is problematic for two reasons: first, it means there is a very real chance of researchers 'reinventing the wheel'; second, and more importantly, it increases the likelihood of researchers coming up with erroneous data analyses. The aim of the conference was thus to enable researchers working in the field of phraseology to meet other researchers studying the same types of expressions from perhaps quite different perspectives. The current volume is meant to reflect this interdisciplinary dimension. Most of the chapters in the volume are based on presentations made at the conference. There are, however, a series of specially commissioned chapters that aim to give an overview of the different perspectives on phraseology.

The volume targets both would-be and experienced phraseologists and provides readers with a variety of ways into the field. For those who are intrigued by the general idea of phraseology without fully knowing what it encompasses, the volume provides a rich overview and introduces a wide range of methodological approaches. For those already experienced phraseologists, the volume is an invitation to look at the field with different eyes. Despite its wide scope, the volume does not claim exhaustivity, however. For a fully comprehensive view of phraseology, the reader is referred to other volumes (notably the other two volumes emanating from the conference and Burger et al.'s recent two-volume handbook of phraseology).

One particularly appealing aspect of the volume is the range of languages it covers: not only English, but also Dutch, French, German, Japanese, Russian and Spanish. This plurilingual perspective provides insights into the workings of phraseology in different languages and, at the same time, into a wide range of phraseological traditions: Anglo-Saxon, East European, French, German, but also Asian.

The volume opens with a preface to the volume by the late **John Sinclair**, who was a keynote speaker at the conference. John had too many other commitments to write a chapter but he kindly agreed to write the preface to the volume. Although he was already very unwell and would have had every reason not to write the preface, he sent us the first draft of his text on 7 March, exactly six days before he died. It must

have been one of the very last pieces he wrote and we have decided to publish it 'as is', despite the changes which inevitably took place during the editing phase.

The volume contains four major sections. Section 1, 'Phraseology: theory, typology and terminology' sets the scene: it delimits the field, introduces major categories of word combinations and the terms used to refer to them as well as highlighting the role played by phraseology in several linguistic theories. Section 2, 'Corpus-based analyses of phraseological units' turns theory into practice by means of a series of corpus-based case studies of different categories of word combinations. Section 3, 'Phraseology across languages and cultures', focuses on a crucial perspective in phraseology, viz. the contrastive perspective, cast in two different lights: purely linguistic and cultural. Section 4, as its title 'Phraseology in lexicography and natural language processing' suggests, brings together lexicographical and natural language processing perspectives, two perspectives which used to be separate but are now moving progressively closer. At the end of the volume, we have reproduced the extended abstract of John Sinclair's plenary presentation at the *Phraseology 2005* conference, whose title "The phrase, the whole phrase and nothing but the phrase" seems to sum up his legacy to linguistics quite perfectly.

The volume contains six overview chapters: Gries on phraseology and linguistic theory, Granger & Paquot on categorization and terminology, Colson on crosslinguistic approaches, Piirainen on cross-cultural issues, Moon on phraseology and lexicography and Heid on phraseology and natural language processing (NLP). These chapters provide an excellent starting point for researchers who are not particularly familiar with phraseological studies. The other chapters tackle more specific aspects of phraseology, particular theoretical approaches, methodologies, research frameworks or particular categories of word combinations. While each chapter has been classified into one of the four sections, it is worth noting that many are at the intersection of two or more sections. The volume is interdisciplinary, not only because it brings together studies from different disciplines but also, encouragingly, because several studies are intrinsically interdisciplinary.

The next section of this introduction briefly outlines each chapter and highlights some major trends emerging from the volume.

Section 1 starts with an overview chapter by Stefan Gries, which tackles phraseology from three major angles: definition, theory and identification. As regards the definition of phraseological units, Gries deplores the general lack of rigour and argues convincingly that the only way of ensuring the comparability of phraseological studies is to make the criteria used to define phraseological units maximally explicit. To this end, he proposes six parameters: nature and number of elements, frequency of occurrence, distance between elements, lexical and syntactic flexibility, semantic unity and non-compositionality. As regards theory, he demonstrates the growing role played by the notion of phraseologism, from TGG, where it has been on the fringe, to the more recent linguistic frameworks of cognitive linguistics and construction grammar and the methodological paradigm of corpus linguistics, where it is much more central. As to the identification of phraseologisms, the author is critical of the lack

of sophisticated statistical techniques and advocates a greater use of NLP techniques and more generally, increased dialogue between linguistic trends and disciplines. In the second chapter of the section, Sylviane Granger & Magali Paquot aim to disentangle the phraseological web. They first present the main traditions of phraseological studies and address the differences in the approaches adopted. The authors argue that the variations in scope are a result of the field's fuzzy borders with four neighbouring disciplines (semantics, morphology, syntax and discourse). They then introduce some influential typologies of phraseology used in English lexicology and give an overview of the categories of multi-word units used in data-driven phraseological studies. The final sections of the article offer suggestions for reconciling the different approaches and propose a most welcome clarification of the terminology. The following two chapters delve more deeply into two theoretical frameworks within which phraseological studies can be conducted. Willy Martin highlights the contribution of frame semantics to phraseology. He shows how a representation of the meaning of lexical items in terms of conceptual semantic frames can help distinguish between the different types of word combinations. Focusing more particularly on lexical collocations, he distinguishes between type-bound collocations like koffie malen (E. 'grind coffee'), token-bound collocations like slappe koffie (E. 'weak coffee') and in-between cases like koffie drinken (E. 'drink coffee'), which are essentially type-bound but intrude into the token-bound category because coffee is a prototypical drink. This qualitative frame-based approach is a useful way of interpreting the data extracted automatically from corpora and can therefore be viewed as a valuable complement to quantitative corpus-based approaches. Marija Omazić takes the standpoint of cognitive linguistics and establishes the role of two theories – metaphor theory and conceptual integration theory - in the processing of figurative phraseological units. The two theories are presented as complementary. The theory of metaphor accounts for the processing of a wide range of conventional phraseological units like to burn with love while the conceptual integration theory caters for modified units like be born with a wooden (instead of silver) spoon in one's mouth. A range of attested modified idioms are used to describe the factors involved in the interpretation of these idioms and the different stages of the 'unpacking process' that underlies the processing of idiom modifications. The final two chapters in the section focus on two major features of phraseological units: non-compositionality and fixedness. In Chapter 5 Maria Helena Svensson attempts to get to grips with the highly complex notion of non-compositionality, a recurrent criterion in phraseological studies, but nevertheless often ill-defined or simply taken for granted. To help clarify the notion, Svensson suggests breaking it down into four scaled dimensions: motivation/non motivation, transparency/opacity, analyzability/unanalyzability and literal/figurative meaning. These notions are closely related but not exact equivalents and therefore need to be clearly distinguished in phraseological studies. The interaction of a series of other notions such as prototypicality, salience and frequency is also discussed as well as related notions like encyclopedic non-compositionality. In conclusion, Svensson argues against using non-compositionality as an all-embracing term and suggests specifying which aspect of non-compositionality is intended. In

Chapter 6 Gill Philip tackles the notion of fixedness, a crucial notion which has long been considered as the defining feature of phraseological units together with non-compositionality. Recent corpus-based studies have shown, however, that beside their canonical forms, so-called 'fixed phrases' display a wide range of variants and that variation within phraseological units is the rule rather than the exception. The chapter focuses on multi-word units referred to as 'phraseological skeletons' which include collocational frameworks, lexico-grammatical frames and semi-prepackaged phrases. The challenge posed by these types of phrases is that they are very difficult to extract automatically as variants are largely unpredictable. In her chapter Philip states the case for phrase deconstruction and suggests ingenious search strategies for extracting variants of idiomatic phrases, thereby paving the way for further exploration of the role of creativity in phraseology.

Section 2 contains a number of case studies which differ in the types of word combinations they analyze but share the characteristic of being corpus-based. One of the interests of the section is that it illustrates a range of different corpus methodologies that can be used to identify and analyze phraseological units in corpora. The opening chapter by Pierre J. L. Arnaud, Emmanuel Ferragne, Diana M. Lewis & François Maniez analyzes the little explored field of incipient lexicalization of phraseological units on the basis of Adj + N sequences extracted from the British National Corpus (BNC). All Adj + N sequences containing a highly frequent central adjective are extracted and further categorized syntactically. Analysis of the structures points to varying degrees of linear fusion of certain sequences which manifest themselves as syntactic recategorization as Adi, N or Adv, loss of compositionality and loss of semantic transparency. Interestingly, no evidence of phonological coalescence was found, which suggests that syntactic and semantic shifts may be interdependent but phonological change may be independently motivated. In Chapter 8 Kay Wikberg makes use of the same corpus to throw light on the category of phrasal simile which figures in all typologies of multi-word units but had not previously featured in a large-scale corpus-based investigation. The concept of simile is defined and set apart from the neighbouring concepts of literal comparison and metaphor. Automatic extraction of four simile patterns from the BNC is followed by a time-consuming and complex procedure aimed at singling out figurative occurrences. The analysis of the selected similes highlights interesting differences between the patterns in terms of frequency, lexicalization and register. Figurativeness also lies at the heart of the following chapter by Hans Lindquist & Magnus Levin which offers a detailed analysis of the two frequent body part nouns foot and mouth. The analytic framework used is a combination of concepts from cognitive linguistics and methods from corpus linguistics. Here too, the corpus used is the BNC supplemented with British, American and Australian newspaper data. Ngrams of different lengths are extracted and further scanned to single out the phrases that display linguistic integrity. The minute linguistic analysis to which the phrases are submitted highlights the rich phraseology displayed by the two nouns and draws attention in particular to the major role played by metonymy and metaphor in the creation and extension of new phrasal patterns. In Chapter 10 Geoffrey C. Williams introduces

the notion of collocational resonance, a notion that draws on his own work on collocational networks and Hoey's lexical priming. The result could be considered as a phraseological manifestation of the literary notion of intertextuality. Using a corpusdriven methodology based on the z-score co-occurrence statistical measure, Williams analyzes the effect of resonance of New Testament formulae with God as the central node and compares the collocational networks from those extracted from the works of Shakespeare and the BNC. The study highlights interesting changes from contextual to restricted collocations, a gradual shift to formulaic use and loss of religious resonance. The study opens new vistas in understanding the link between language and collective memory. In the last chapter of the section, Margaret Maclagan, Boyd Davis & Ron Lunsford offer an innovative perspective on a little researched field, that of the role of phraseology in pathological speech. They analyze a range of multi-word units in a corpus of speech of people with Alzheimer's disease on the basis of a pre-established list of multi-word units supplemented with phrases noted by the analysts. The study shows that multi-word units are pervasive in impaired speech and serve an important social-interactional function. Extenders like things like that or all that sort of thing help impaired speakers maintain the appearance of competence. As the corpus is longitudinal, the authors are able to trace the evolution in the use of these units, from perfect mastery to reduced pragmatic appropriateness and semantic range.

Section 3, devoted to cross-linguistic and cross-cultural approaches to phraseology, opens with two overview chapters. The first, by Jean-Pierre Colson, focuses on studies in cross-linguistic and contrastive phraseology. Colson criticizes the frequent lack of theoretical framework in many of the studies in the domain and argues that proper theoretical foundations could be put in place drawing from the fields of cognitive semantics and corpus linguistics. After a critical presentation of the strengths and weaknesses of those two approaches, the author calls for greater interaction between contrastive and translation studies, as the concept of phraseology is still notably absent from studies on translation theory and practice. Colson also stresses the need for more phraseologically-oriented cross-linguistic and contrastive research on non-European language families in order to assess the universality of phraseology. In the second overview chapter, Elisabeth Piirainen demonstrates the relevance of studying conventional figurative language to reveal its cultural content and analyses this connection between figurative language and culture from various angles. She presents a typology of the cultural elements underlying phrasemes (such as for instance textual dependence, pre-scientific conceptions of the world or aspects of material culture). She then highlights the fact that various types of phrasemes are unequally affected by aspects of culture and also illustrates some of the manifestations of culture in language starting either from source concepts (like taurine phraseology in Spanish) or from semantic fields. Finally, she tackles the disputed link between phraseology and the worldview of a language community, and also addresses issues in historical and etymological research. Five other chapters follow the overview chapters, two with a cultural focus and the other three with a linguistic contrastive one. Annette Sabban, in her chapter on the culture-bound nature of phraseology, starts with a discussion of terminological issues

related to the terms 'culture', 'culture-specific' and 'culturally-bound'. She then focuses on idioms as particularly relevant in examining the phenomenon and deals with some of the problems that may arise in the interpretation of phrasemes in terms of culture, including differences in cultural knowledge and in speakers' motivation of idiomatic meaning. Sabban also insists on the importance of distinguishing between concepts in language (i.e. results of modes of thinking which may no longer be relevant to the users of a language) and current concepts of thought (which may no longer coincide with concepts in language). In Chapter 15, Elisabeth Piirainen's second contribution to the volume, the author examines phraseology in a European framework in the light of a cross-linguistic and cross-cultural research project on widespread idioms, i.e. idioms displaying similar lexical structure and figurative meaning in various languages, be they genetically and geographically close or distant. The chapter addresses one of the issues mentioned in her overview chapter, viz. the fact that classical phraseology too often associates a (national) language and its linguistic or cultural specificities with space. The author first criticizes the use of terms such as 'Europeanism' or even 'universalism' to refer to cross-linguistic similarities of idioms in only a few (often rather closely related) languages. She then presents the Widespread Idioms project which aims to systematically investigate the similarities among idioms in as many languages as possible, with a view to either corroborating or refuting the often claimed uniformity of European phraseologies based on a common European cultural heritage. Preliminary results are presented and possible explanations for widespread cross-linguistic similarities are put forward. As for the three contrastively-oriented chapters of Section 3, they all use corpus methodology, at least in part. In Chapter 16 Christelle Cosme & Gaëtanelle Gilquin present the results of a contrastive corpus analysis of the free and bound uses of the English preposition 'with' and its intuitive French equivalent 'avec'. Cosme and Gilquin stress the very low mutual translatability of the prepositions and partly attribute this lack of equivalence to polysemy and phraseology factors. The authors also propose concrete pedagogical applications of their findings, notably in bilingual lexicography and foreign language teaching methodology. In the next contrastive chapter, Priscilla Ishida analyses Japanese and English anger idioms. After an introduction on past approaches to the cross-linguistic analysis of idioms, she presents a 4-step method which focuses on both the L1 and L2 semantic networks and which is complemented by textual and discourse analysis. Ishida demonstrates that whilst full semantic correspondence is rare, many Japanese and English idioms overlap partially in meaning. The final part of the chapter also addresses semantic correspondence and the translation problems they pose. The last contrastive chapter, by Olga Mudraya, Scott S. L. Piao, Paul Rayson, Serge Sharoff, Bogdan Babych & Laura Löfberg deals with the automatic extraction of reliable lists of multilingual phraseological units. The authors focus on the translation equivalents of phrasal and light verbs in English and Russian and address a series of issues such as the different morpho-syntactic structures and the varying semantic properties. The use of the English Semantic Tagger developed at Lancaster University is illustrated and the results of several case studies are presented. Mudrava and her colleagues also stress the value of cross-linguistic corpus-based studies for theoretical and applied studies in contrastive linguistics, multilingual lexicon extraction and language teaching.

Section 4 groups together lexicography and natural language processing (NLP). This combination seems rather natural firstly because the lexicon constitutes a key component of any NLP tool and, secondly, because the fast growing development of electronic dictionaries allows for the direct integration of NLP technology. As was the case in Section 3, the section opens with two overview chapters, the first by Rosamund Moon on dictionaries and collocation, and the second by Ulrich Heid on an overview of phraseology and NLP. Moon's chapter starts with the analysis of the collocational behaviour of the three English words, river, rivet and riven, in large corpora. She then compares her results to the collocational representations of these words in monolingual dictionaries for natives and for learners, in bilingual dictionaries and in specialized dictionaries of collocations. Through her analysis, Moon provides the readers with a diachronic and synchronic perspective on the place of phraseology in dictionaries. She offers a balanced and critical approach to the challenges met by lexicographers and to the benefits and limits of corpus-based lexicography. Heid's overview chapter on phraseology and NLP begins with a definition of 'computational phraseology' and a discussion of the value of the term multi-word expressions (MWEs), a term often preferred to phraseology in NLP circles. After a brief discussion of the role of phraseology in NLP applications such as machine translation or natural language generation, Heid addresses an impressive number of issues linked to multi-word expressions: issues of formal representation and annotation (lexical, morphosyntactic, syntactic and semantic idiosyncrasies and variation), aspects of lexicographical representation for human users, frequency and productivity issues, and finally, issues related to the (semi)automatic identification and classification of MWEs. The problems of identification, classification and annotation of MWEs are further explored in Cornelia Tschichold's contribution. She criticizes the lack of good computational tools for the study of the variability of MWEs in large corpora. After a section on data collection issues, she focuses on lexicographic work per se and on the creation of a lexicographic database. She describes the WordManager-PhraseManager, a language independent system that can help lexicographers and linguists to carry out a tailor-made annotation of the internal structure of MWEs. The system can capture the variability of MWEs and hence favours maximum reusability of the phraseological database. Brigitte Orliac also deals with the automatic extraction of specialized collocations and presents a method for extracting verb plus noun collocations in a specialized corpus of computer science texts. After a discussion of various extraction methodologies (where statistical methods usually precede linguistic analysis or filtering), she proposes a method based on the use of the lexical functions of the Meaning-Text theory put forward by Mel'čuk. The collocation extractor, Colex, combines linguistic annotation and statistical methods, with linguistic annotation taking chronological precedence. The statistical measures used by Colex in later stages of the analysis make it possible to distinguish between bound and free combinations. In the last NLP-oriented chapter of the volume José-Manuel Pazos Bretaña & Antonio Pamies Bertrán also analyse combined statistical

and grammatical criteria for the retrieval of phraseological units in corpora. They first assess the success of purely statistical methods and measures for the automatic extraction of phraseological units and show that whilst these methods can be effective and provide the user with lists of n-grams in decreasing order of frequency, the number of unexpected non-phraseological combinations is too high. To remedy this lack of precision in the results, the authors propose a preliminary three-step linguistic annotation of the data (lemmatization, grammatical filtering and part-of-speech tagging) and demonstrate that a combined approach (preliminary linguistic treatment followed by statistical analysis) yields much better results than a purely statistical one.

Although this volume contains a highly diverse range of articles, some key ideas emerge. First, as regards the scope of the field, the volume brings support to the wide view of phraseology. It recognizes the importance of a whole range of largely compositional multi-word units over and above the more traditional types. This wider view is in keeping with the corpus approach to phraseology which is very much in evidence in the volume. The corpora used to investigate phraseological units in the different chapters of the volume are extremely varied: monolingual vs. bilingual, written vs. spoken, large representative corpora like the British National Corpus vs. small locally-collected corpora, native vs. learner corpora, impaired vs non-impaired speech, etc. However, as many types of phraseological units are highly infrequent (cf. Moon 1998), corpus data are often complemented by other types of data, such as newspaper databases or the web. Second, the volume shows that multi-word units can be extracted using a wide range of methods and that the number and types of units extracted varies greatly according to the method used. The starting-point can be word forms, lemmas, syntactic structures or variable patterns. Alternatively, using a more corpus-driven approach, multi-word units can be extracted via the application of statistical tests, such as mutual information or the t-score or a combination of statistical tests and linguistic criteria. A third major thread running through the volume is that there is a limit to the information that can be derived from corpora and that more controlled data types like elicitation techniques may prove very useful. Surprisingly perhaps in a volume where corpora are so much in evidence, the word intuition or introspection is often used, even in the most NLP-oriented chapters. As pointed out by Lindquist (cf. Chapter 9), "corpus-drivenness has its limits" and recourse to intuition and subjective judgement is needed. This shows that that the two approaches to phraseology, the traditional approach and the quantitative approach, should not be viewed as conflicting but as complementary. Rather than replace older approaches by newer ones, one should examine how they can be made to cross-fertilize, with the more traditional phraseologists learning about interesting new techniques and new types of units, and proponents of the distributional approach learning to dissect the results of their queries in linguistically interesting ways.

John Sinclair has been a tremendous source of inspiration for us like for so many researchers around the world. It therefore seems quite natural to give him the last word by quoting the concluding words of an article he wrote 40 years ago but which still has a thoroughly modern resonance today: "[t]he theory of lexis opens up exciting areas

for describing language more accurately and more usefully. The practical problems are immense, and no secret has been made of them here, but the results that they promise are, possibly because of their novelty, no less fascinating than those of any other branch of linguistics" (Sinclair 1966: 429).

References

- Burger, H., D. Dobrovol'skij, P. Kühn & N. R. Norrick (eds.) (2007). *Phraseologie/Phraseology. Ein internationales Handbuch zeitgenössischer Forschung / An International Handbook of Contemporary Research*. Vols 1 & 2. Berlin: Mouton de Gruyter.
- Moon, R. (1998). Fixed Expressions and Idioms in English. A Corpus-based Approach. Oxford: Clarendon Press.
- Pawley, A. & F. Syder (1983). Two puzzles for linguistic theory: Native-like selection and native-like fluency. In J. Richards & R. Schmidt (Eds.), *Language and Communication*, 191–226. London: Longman.
- Sinclair, J. McH. (1966). Beginning the Study of Lexis. In C. E. Bazell, J. C. Catford, M. A. K. Halliday & P. H. Robins (eds.) *In Memory of J. R. Firth*, 410–430. London: Longman.