

Word

A cross-linguistic typology

Edited by

R. M. W. Dixon

and

Alexandra Y. Aikhenvald

Research Centre for Linguistic Typology, La Trobe University



CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge CB2 1RP, United Kingdom

CAMBRIDGE UNIVERSITY PRESS

The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011–4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

<http://www.cambridge.org>

© In the work Cambridge University Press 2002

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2002

Printed in the United Kingdom at the University Press, Cambridge

Typeface Times 10/12 pt *System* L^AT_EX 2_ε [TB]

A catalogue record for this book is available from the British Library

ISBN 0 521 81899 0 hardback

Contents

| | |
|--|-----------|
| <i>List of contributors</i> | page viii |
| <i>Preface</i> | x |
| <i>List of abbreviations</i> | xi |
| 1 Word: a typological framework | 1 |
| R. M. W. DIXON AND ALEXANDRA Y. AIKHENVALD | |
| 1 The tradition | 2 |
| 2 Doing without 'word' | 3 |
| 3 What is a word? | 5 |
| 4 Confusions | 6 |
| 5 Some suggested criteria | 10 |
| 6 Phonological word | 13 |
| 7 Grammatical word | 18 |
| 8 Clitics | 25 |
| 9 Relationship between grammatical and phonological words | 27 |
| 10 Do all languages have words? | 32 |
| 11 The social status of words | 32 |
| 12 Summary | 34 |
| Appendix Sample outline account of phonological word and grammatical word in Fijian | 35 |
| References | 37 |
| 2 Typological parameters for the study of clitics, with special reference to Tariana | 42 |
| ALEXANDRA Y. AIKHENVALD | |
| 1 Parameters for clitics | 43 |
| 2 Words and clitics in Tariana | 57 |
| 3 Summary: the status of clitics | 70 |
| Appendix Additional issues concerning clitics, and parameters suggested for distinguishing between clitics and affixes | 71 |
| References | 75 |
| 3 The word in Cup'ik | 79 |
| ANTHONY C. WOODBURY | |
| 1 Introduction | 79 |
| 2 The language | 79 |

| | | |
|---|--|-----|
| 3 | A thumbnail typology | 80 |
| 4 | The word for 'word' in Cup'ik | 81 |
| 5 | Grammatical word | 81 |
| 6 | Enclitics in grammar | 89 |
| 7 | Phonological word | 91 |
| 8 | Summary of grammatical-word-phonological-word mismatches | 97 |
| 9 | Conclusion | 98 |
| | References | 98 |
| 4 | The word in Eastern/Central Arrernte | 100 |
| | JOHN HENDERSON | |
| 1 | Introduction | 100 |
| 2 | Brief summary of typology | 101 |
| 3 | Phonological word | 101 |
| 4 | Criteria for grammatical word | 107 |
| 5 | Clitics | 108 |
| 6 | Relationship between phonological and grammatical word | 112 |
| 7 | Complex predicates | 114 |
| 8 | Conclusion | 123 |
| | References | 123 |
| 5 | The eclectic morphology of Jarawara, and the status of word | 125 |
| | R. M. W. DIXON | |
| 1 | Introduction | 125 |
| 2 | Phonological word | 127 |
| 3 | Clause structure and verbal clauses | 129 |
| 4 | Predicate structure | 130 |
| 5 | Grammatical word and predicate – general discussion | 140 |
| 6 | The grammatical word in Jarawara | 143 |
| 7 | Summary of instances where the two kinds of word do not correspond | 147 |
| | Appendix List of miscellaneous, tense-modal and mood suffixes | 148 |
| | References | 152 |
| 6 | Towards a notion of 'word' in sign languages | 153 |
| | ULRIKE ZESHAN | |
| 1 | Words and signs: on psychological and cultural validity | 153 |
| 2 | Grammatical and phonological words in sign languages | 156 |
| 3 | Compounds and clitics in sign languages | 161 |
| 4 | Words and signs revisited | 167 |
| | References | 177 |
| 7 | Synchronic and diachronic perspective on 'word' in Siouan | 180 |
| | ROBERT RANKIN, JOHN BOYLE, RANDOLPH GRACZYK AND JOHN KOONTZ | |
| 1 | Typology | 180 |
| 2 | Grammatical words | 183 |
| 3 | Phonological words | 192 |
| 4 | Enclitics | 195 |
| 5 | Definitional problems | 199 |

| | |
|---|-----|
| Contents | vii |
| 6 Speakers' concepts of 'word' | 201 |
| 7 Conclusion | 202 |
| References | 203 |
| 8 What is a word in Dagbani? | 205 |
| KNUT J. OLAWSKY | |
| 1 Phonological word | 205 |
| 2 Grammatical word | 212 |
| 3 Clitics | 216 |
| 4 Relationship between grammatical and phonological word | 223 |
| Appendices | 224 |
| References | 226 |
| 9 The word in Georgian | 227 |
| ALICE C. HARRIS | |
| 1 Summary of the typology | 227 |
| 2 The grammatical word in Georgian | 228 |
| 3 The phonological word in Georgian | 232 |
| 4 Clitics | 236 |
| 5 Conclusion | 241 |
| References | 241 |
| 10 The word in Modern Greek | 243 |
| BRIAN JOSEPH | |
| 1 Introduction | 243 |
| 2 Some important preliminaries | 244 |
| 3 Various types of 'word' and relevant criteria and tests | 248 |
| 4 Summation regarding wordhood | 259 |
| 5 Conclusion | 261 |
| References | 263 |
| 11 What can we conclude? | 266 |
| P. H. MATTHEWS | |
| 1 Introduction | 266 |
| 2 The word as established in Latin | 267 |
| 3 How can general linguists help? | 271 |
| 4 'Clitics' | 276 |
| 5 Are units like words necessary? | 279 |
| References | 280 |
| <i>Index of authors</i> | 282 |
| <i>Index of languages and language families</i> | 285 |
| <i>Index of subjects</i> | 288 |

Contributors

Alexandra Y. Aikhenvald
Research Centre for Linguistic
Typology
La Trobe University
Victoria, 3086
Australia
e-mail:
a.aikhenvald@latrobe.edu.au

John Boyle
Dept of Linguistics
University of Chicago
Chicago, IL 60637
USA
e-mail:
jpboyle@midway.uchicago.edu

R. M. W. Dixon
Research Centre for Linguistic
Typology
La Trobe University
Victoria, 3086
Australia
no e-mail

Randolph Graczyk
PO Box 29
Pryor, MT, 59066
USA
e-mail: rgraczyk@acl.com

Alice C. Harris
Department of Linguistics

SUNY Stony Brook
Stony Brook, NY 11794
U.S.A.
e-mail: alice.harris@stonybrook.edu

John Henderson
Dept of Linguistics
University of Western Australia
WA, 6907
Australia
e-mail:
john.henderson@uwa.edu.au

Brian Joseph
Dept of Linguistics
Ohio State University
222 Oxley Hall
Columbus, OH 43210
USA
e-mail:
bjoseph@ling.ohio-state.edu

John Koontz
471, E Rainbow Ct
Louisville, CO 80027-2125
USA
e-mail: john.koontz@colorado.edu

P. H. Matthews
St John's College
Cambridge, CB2 1TP
UK
no e-mail

Knut J. Olawsky

Research Centre for Linguistic
Typology
La Trobe University
Victoria, 3086
Australia
e-mail:
olawsky@latrobe.edu.au

Robert Rankin

Dept of Linguistics
University of Kansas
Lawrence, KS 66045-2140
USA
e-mail: rankin@ukans.edu

Anthony C. Woodbury

Dept of Linguistics
Calhoun Hall 501
University of Texas
Austin, TX 78712
USA
e-mail: acw@mail.utexas.edu

Ulrike Zeshan

Research Centre for Linguistic
Typology
La Trobe University
Victoria, 3086
Australia
e-mail: u.zeshan@latrobe.edu.au

1 Word: a typological framework

R. M. W. Dixon and Alexandra Y. Aikhenvald

In this book we ask how ‘word’ should be defined. What are the criteria for ‘word’? Is ‘word’, as the term is generally understood, an appropriate unit to recognise for every type of language?

This introductory chapter first looks at what scholars have said about ‘word’, and then discusses the categories and distinctions which need to be examined. Chapter 2 suggests a number of typological parameters for the study of clitics. Following chapters then provide detailed examination of the notion of ‘word’ in a selection of spoken languages from Africa, North and South America, Australia, the Caucasus and Greece, together with a discussion of words in sign languages. The final chapter, by P. H. Matthews, asks what has been learnt from these general and particular studies.

This introduction begins by surveying the criteria that have been put forward for ‘word’, and suggests that one should sensibly keep apart phonological criteria, which define ‘phonological word’, and grammatical criteria, which define ‘grammatical word’. In some languages the two types of word coincide and one can then felicitously talk of a single unit ‘word’, which has a place both in the hierarchy of phonological units and in the hierarchy of grammatical units. In other languages phonological word and grammatical word generally coincide, but do not always do so. We may have a grammatical word consisting of a whole number of phonological words, or a phonological word consisting of a whole number of grammatical words. Or there can be a more complex correspondence between the two types of word with, say, a grammatical word consisting of all of one and part of another phonological word.

§1 summarises the tradition, §2 discusses linguists who would do without the word and §3 surveys opinions concerning ‘what is a word’. In §4 a number of confusions are discussed and then in §5 some suggested criteria are examined. The heart of the chapter is in §§6–8 – proposed definitions for phonological word and for grammatical word (and the status of clitics) – followed by (in §9) examination of the relationship between the two types of word. In §10 we ask whether all kinds of languages have words; in §11 there is brief discussion of the varying social status of ‘word’ in different languages, and then §12 provides a summary of the results of the introductory chapter. Finally, the appendix gives a

brief statement of the criteria for phonological word and for grammatical word (and their relationship) in a sample language, Fijian.

1 The tradition

Many writers have assumed that ‘word’ is a – or the – basic unit of language. Bolinger (1963: 113) comments: ‘Why is it that the element of language which the naive speaker feels that [they] know best is the one about which linguists say the least? To the untutored person, speaking is putting words together, writing is a matter of correct word-spelling and word-spacing, translating is getting words to match words, meaning is a matter of word definitions, and linguistic change is merely the addition or loss or corruption of words.’ Bolinger himself takes ‘word’ as a prime, commenting that it is ‘the source, not the result, of phonemic contrasts’.

And, as Lyons (1968: 194) comments: ‘The word is the unit *par excellence* of traditional grammatical theory. It is the basis of the distinction which is frequently drawn between morphology and syntax and it is the principal unit of lexicography (or “dictionary-making”).’ Indeed, for the Greeks and Romans the word was the basic unit for the statement of morphological patterns; they used a ‘word and paradigm’ approach, setting out the various grammatical forms of a given lexeme in corresponding rows and columns, with no attempt to segment into morphemes (Robins 1967: 25). (In fact Greek and Latin are fusional languages where it is not an easy matter to segment words into morphemes, without bringing in the impedimenta of underlying forms, morphophonological rules, and the like.)

Much that has been written about the word is decidedly eurocentric. It has sometimes been said that ‘primitive languages’ do not have words, an opinion which Lyons (1968: 199) explicitly rejects, partly on the basis of Sapir’s report that uneducated speakers of American Indian languages can dictate ‘word by word’.

However, it appears that only some languages actually have a lexeme with the meaning ‘word’.¹ Even in some familiar languages where this does occur it may be a recent development. For instance, in Old English the primary meaning of *word* was (a) for referring to speech, as contrasted with act or thought. There was a second sense, which may then just have been emerging: (b) what occurs between spaces in written language. In the development to Modern English (b) has become the major sense – the one used in this book – with sense (a) still surviving mainly in fixed phrases, e.g. *the spoken word*,

¹ Dixon (1977a: 88) states ‘every (or almost every) language has a word for “word”’; this is erroneous. Wierzbicka (1996, 1998) has ‘word’ as a universal semantic primitive, which is said to be realised in every language; this is equally erroneous.

the written word, the Word of God, a word of warning/advice/caution, Can I have a quick word with you? Similar remarks apply to corresponding terms in some other European languages, e.g. *mot* in French and *slovo* in Russian. German also has a noun, *Wort*, with these two senses, but there are here two plural forms – *Worte* for ‘speech’ and *Wörter* for ‘what is written between spaces’.²

Over on the other side of the world, *vosa* in Fijian is a verb meaning ‘speak, talk’ and also a noun, with several related senses: ‘language’, ‘talk, speech’ and ‘word’. It is likely that we have here a similar line of semantic development to *word* in English.

The vast majority of languages spoken by small tribal groups (with from a few hundred to a few thousand speakers) have a lexeme meaning ‘(proper) name’ but none have the meaning ‘word’. This applies to many languages from Australia (including Arrernte, chapter 4 in this volume), Amazonia (including Jarawara, chapter 5) and New Guinea with which we, or our colleagues, are familiar.

2 Doing without ‘word’

The idea of ‘word’ as a unit of language was developed for the familiar languages of Europe which by-and-large have a synthetic structure. Indeed – as will be shown below – some of the criteria for ‘word’ are only fully applicable for languages of this type.

What about languages from extreme ends of the typological continuum – those of an analytic or of a polysynthetic profile? Reviewing the first edition of Nida’s (1944) *Morphology*, Hockett (1944: 255) notes that Nida ‘devotes a chapter to the criteria by which words may be recognised. None of these criteria, nor any combination of them, gives any fruitful results with Chinese . . . the real implication is that THERE ARE NO WORDS IN CHINESE. The whole tradition of “words” as worked out with western languages is useless in Chinese.’ (However, a quite different opinion is expressed by the leading Chinese linguist, Chao, discussed in §10 and §11 below.)

Some of the polysynthetic languages of North America lack any unit that looks like the sort of word we are used to from European languages. Gray (1939: 146) presents his own definition of word as ‘a complex of sounds which in itself possesses a meaning fixed and accepted by convention’. (Note that this would, in fact, also be satisfied by a prefix such as *un-* or a phrase such as

² It is likely that all languages with an established (non-ideographic) orthographic tradition do have a word for ‘word’. Other languages tend to create such a term once they are exposed to writing. The interesting question is how many languages with no written tradition have a lexeme which corresponds to *word* in English, *mot* in French, etc.

The King of England.) Then, in a footnote, he reports a suggestion from H. J. Uldall that in polysynthetic languages (such as Maidu from north-east California) 'the word as such is not a relevant part of analysis'. This idea is repeated by Milewski (1951), who argues that in all languages of the world there are three kinds of morphological unit, morphemes, syntactic groups and clauses. However, Krámský (1969: 74–5) suggests that Milewski's 'syntactic group' in polysynthetic North American languages does satisfy criteria for 'word'. It is clear that the unit 'word' can be recognised in polysynthetic languages, it is just that it is much longer and more complex than the words linguists were used to at that time.

It is, however, important to distinguish between the structure of the predicate and the structure of the verb. This is discussed in §§5–6 of chapter 5, where it is shown that a language may have fairly complex predicate structure but relatively simple verb structure (as in English and Fijian), or simple predicate structure combined with somewhat complex verb structure (as in Dyirbal), or both complex predicate structure and complex verb structure (as in Jarawara).

There are those who consider 'word', as a general notion, not to be a basic category of language. The anthropologist Malinowski (1966: 11) insists that one should analyse utterances, not any smaller units of language taken out of their context of use. He can then say 'isolated words are in fact only linguistic figments, the products of an advanced linguistic analysis'.³

Other linguists accord a grudging role to 'word'. For Börgström (1954: 276) 'words are utterance-segments consisting of one or more morphemes. Assuming that there is a procedure for the demarcation of morphemes, it is possible, I believe, to formulate a set of distribution rules as a procedure for the demarcation of words.' Writing in the same year, Garvin (1954: 345) is less sure about this: 'in the present state of our techniques one may assume that we know how to isolate morphemes properly – that is, unequivocally and without unaccountable residue. It is not so certain that we know how to isolate words, and hence how to separate morphology from syntax.'

The extreme position is taken by Harris (1946: 161) who presents a procedure (illustrated for English and Hidatsa) for analysing utterances into morphemes: 'the method described in this paper will require no elements other than morphemes and sequences of morphemes, and no operation other than substitution, repeated again and again'. The unit 'word' does not feature in Harris' analysis.⁴

³ We also find (perhaps as a further reflection of Malinowski's position) Potter's (1967: 78) statement: 'unlike a phoneme or a syllable, a word is not a linguistic unit at all. It is no more than a conventional or arbitrary segment of utterances.'

⁴ We have noted one instance of *word* in this paper, but this is used in an informal rather than in an analytic sense. On page 166 Harris is discussing the English sentence *I know John was in* and talks of 'pronouncing its intonation twice, once over the first two words and again over the last three'.

3 What is a word?

Matthews commences the section 'What are words?' in the second edition of his seminal textbook *Morphology* (1991: 208) with: 'there have been many definitions of the word, and if any had been successful I would have given it long ago, instead of dodging the issue until now'.

Matthews mentions that the ancient grammarians simply had word as the smallest unit of syntax. But, he comments, to follow that line 'will only turn our larger problem back to front. If words are to be defined by reference to syntax, what in turn is syntax, and why are syntactic relations not contracted by parts of words as well as whole words?'

Some of the definitions suggested for word are horrifying in their complexity and clearly infringe the principle that a definition should not be more difficult to understand than the word it purports to define.⁵ There are useful surveys of definitions of 'word' in Rosetti (1947), Weinreich (1954), Ullmann (1957) and Krámský (1969).

Some definitions are simple and appealing. These include Sapir's (1921: 34) 'one of the smallest, completely satisfying bits of isolated "meaning" into which the sentence resolves itself' and Žirmunskij's (1966: 66): 'the word is the most concise unit of language, which is independent in meaning and form'. But each of these is essentially vague; they do not provide definite criteria for deciding 'what is a word' in a given language.

Sweet (1875/6: 474) suggests: 'we may, therefore, define a word as an ultimate or indecomposable sentence'. That is, anything which is a word can make up a complete sentence. Sweet offers as examples of this (from English) *Come!* and *Up?* (meaning 'Shall we go up?'). However, he is then concerned over what to do with forms like English *the* and *a*, which he terms 'half-words'.

Bloomfield (1933: 178) pursues a similar line in his definition: 'a word, then, is a free form which does not consist entirely of (two or more) lesser free forms; in brief, a word is a *minimum free form*' (his italics). This is probably the most oft-quoted definition of 'word' but it is, in fact, scarcely workable. There is further discussion in §5 below.

⁵ We can quote two rather extreme examples. Firstly, Longacre's (1964: 101) definition, which was conceived within the formal framework of tagmemics: 'a class of syntagmemes of a comparatively low hierarchical order, ranking below such syntagmemes as the phrase and the clause and above such syntagmemes as the stem (as well as above roots which have no external structure and are therefore not syntagmemes). It may be of greatly varied structure... Words tend to be rigidly ordered linear sequences containing tagmemes which (aside from those manifested by stems) are manifested by closed classes of morphemes unexpandable into morpheme sequences and giving only stereotyped bits of information.'

Krámský devotes a whole monograph to discussing 'word'. He surveys past definitions and then comes up with his own (1969: 67): 'the word is the smallest independent unit of language referring to a certain extra-linguistic reality or to a relation of such realities and characterised by certain formal features (acoustic, morphemic) either actually (as an independent component of the context) or potentially (as a unit of the lexical plan)'.

4 Confusions

The word ‘word’ is used in many ways in everyday speech, and in much linguistic discourse. It is important to make certain fundamental distinctions:

- (1) between a lexeme and its varying forms;
- (2) between an orthographic word (something written between two spaces) and other types of word;
- (3) between a unit primarily defined on grammatical criteria and one primarily defined on phonological criteria.

These are discussed, in turn, in §§4.1–3.

The (grammatical) word forms the interface between morphology and syntax. Morphology deals with the composition of words while syntax deals with the combination of words. One could imagine slightly different words being required as ideal units for these two purposes. That is, there could be a ‘morphological word’ and a ‘syntactic word’ which would perhaps generally coincide but might not always do so. We are not aware of this sort of distinction having been fully justified for any language;⁶ but it is certainly a possibility. (In chapter 7, Rankin et al. put forward the idea that the term ‘syntactic word’ could perhaps be used – in Siouan languages – for a type of word incorporating a relative clause, the whole constituting one phonological word.)

4.1 *Word and lexeme*

Consider the following examples, from English and Latin, of the root or underlying form of a lexeme and its inflected forms, as used in a sentence.

| | root or underlying form | inflected forms |
|-----|-------------------------|--|
| (a) | <i>look</i> | <i>look</i> present, non-3sg subject <i>looks</i> present, 3sg subject <i>looked</i> past <i>looking</i> participle |
| (b) | <i>lup-</i> ‘wolf’ | <i>lupus</i> nominative sg <i>lupō</i> dative/ablative sg <i>lupī</i> genitive sg, nominative pl etc. |

⁶ The possibility of this is mentioned by Di Sciullo and Williams (1987) without, however, the formulation of any explicit cross-linguistic or language-specific criteria. This question is also aired in Gak (1990). Dai (1998) establishes separate units ‘syntactic word’, ‘phonological word’, and ‘morphological word’ in Chinese. He suggests that a compound is one syntactic word and also one morphological word but that it may have different syntactic and morphological structures.

A number of other types of ‘word’ have been suggested. For example, Packard (2000: 7–14) lists: orthographic word, sociological word, lexical word, semantic word, phonological word, morphological word, syntactic word, and psycholinguistic word.

The term ‘word’ is sometimes used in reference to the root or underlying form, and sometimes in reference to the inflected forms. That is we hear, on the one hand things like ‘*look, looks, looked* and *looking* are forms of the same word’, and on the other hand things like ‘the lexeme *look* is realised as word-forms *look, looks, looked* and *looking*’.

Bally (1950: 287–9) is so concerned about this ambiguity of usage that he recommends abandoning the label ‘mot’ in French (and ‘word’ in English) and instead employing ‘sémantème’ for the root or underlying form and ‘molécule syntaxique’ for inflected forms. Lyons (1968: 197) prefers a different course. While recognising that in classical grammar ‘word’ was used to mean ‘sémantème’ he notes that modern usage tends to employ ‘word’ as a label for ‘molécule syntaxique’ and suggests standardising on this.

We have followed Lyons’ suggestion, of using ‘lexeme’ as the label for ‘root or underlying form’ and ‘(grammatical) word’ for ‘inflected form of a lexeme’. Note that Lyons uses italics for words and capitals for lexemes – thus, the word *looked* is the past tense form of the lexeme LOOK.

Lyons’ convention is useful from another viewpoint, for dealing with lexemes that involve two words. These include phrasal verbs in English such as MAKE UP, as in *I made the story up* and *I made it up*. Note that the words of this lexeme are mapped onto two non-contiguous syntactic slots – an inflected form of *make* goes into the verb slot while *up* follows the object NP.⁷ That is, the lexeme MAKE UP consists of two words, each of which has its own syntactic behaviour. If we had decided on ‘word’ as the label for lexeme, there would then be need for a separate notion of ‘syntactic word’. We would have had to say that the (lexical) word *make up* consists of two syntactic words, *make* and *up*. This is avoided by describing MAKE UP as a lexeme that consists of two (grammatical) words, an inflected form of *make* and the preposition *up*. (Similar remarks apply to phenomena such as separable preverbs in German and Hungarian.)

4.2 Orthographic word

In many language communities a word is thought of as having (semantic, grammatical and phonological) unity and, in writing, words are conventionally separated by spaces. (In §9 below we investigate the writing convention when phonological and grammatical criteria do not produce the same unit.)

Indeed, in his *Phonemics*, Pike (1947: 89) defines ‘word’ as ‘the smallest unit arrived at for some particular language as the most convenient type of

⁷ The *up* can move to the left over an object that is a full NP but not over a preposition – *I made up the story* but not **I made up it*. Note the distinction between a phrasal verb like *make up* and one like *pick on*, where the *on* must precede the object NP, e.g. *He picked on his brother* or *He picked on him* but not **He picked his brother on* or **He picked him on*. See Dixon (1982; 1991: 274–8).

grammatical entity to separate by spaces; in general, it constitutes one of those units of a particular language which actually or potentially may be pronounced by itself'. Pike here implies that the ideal orthographic convention is to write spaces between grammatical words. The first part of his definition is circular – spaces are written around a grammatical word and a grammatical word is what is felt to be appropriately written between spaces; that is, no explicit criterion for 'grammatical word' is provided. The second part of his definition is essentially Bloomfield's 'minimum free form', discussed in §5 and §7 below.

Writing conventions are unlikely to be absolutely consistent. In English, for instance, the convention is to write *cannot* as one word but the analogous *must not* as two. There appears to be no reason for this; it is just a convention of the language community.

The Bantu languages of southern Africa have a complex but agglutinative verb structure. Van Wyk (1967: 230) describes different conventions used in these languages for writing word divisions:

- (a) disjunctivism – 'according to which relatively simple, and, therefore, relatively short, linguistic units are written and regarded as words';
- (b) conjunctivism – 'according to which simple units are joined to form long words with complex morphological structures'.

He exemplifies with the Northern Sotho sentence 'we shall skin it with his knife'. The two ways of writing this are:

- (a) *re tlo e bua ka thipa ya gagwe*, according to the disjunctive system;
- (b) *retloebua kathipa yagagwe*, according to the conjunctive system.

Van Wyk does not provide an interlinear gloss. However, we have been able to ascertain that *re-* is the 1pl subject prefix, *-tlo-* is the future prefix, *-e-* is a 3sg object prefix, *-bua* is the verb root 'to skin', *ka-* is an instrumental prefix, *thipa* is the noun 'knife', *ya-* is a class 9 prefix (agreeing with the class 9 noun 'knife') and *gagwe* is 'his'.

In fact different orthographic strategies have been adopted for different Bantu languages. Northern Sotho, Southern Sotho and Tswana are written disjunctively while Zulu and Xhosa are written conjunctively. There is no inherent grammatical difference between these languages; it is just that different writing conventions are followed. In the conjunctive system spaces are written between grammatical words (which may be long); in the disjunctive system spaces are written between morphemes within grammatical words. This may have been influenced by the fact that some of the prefixes are bound pronouns and case-type markers, corresponding to free pronouns and prepositions in languages such as English and Dutch (the languages of the Europeans who helped devise these writing systems), which are there written as separate words.

The orthographic conventions used for a language tend to reflect what the language was like at the time when an orthography was first adopted. For

example, *knee* was pronounced with an initial *k* when English was first written. A language may undergo considerable changes, few of which get incorporated into the orthography. French, for instance, has shifted from a mildly synthetic structure to one bordering on the polysynthetic. A sentence such as *je ne l'ai pas vu* 'I have not seen it' can be considered a single word, on both grammatical and phonological criteria. But the language is – as a reflection of its history – written disjunctively, with the consequence that speakers will say that the sentence consists of five or six words (see Vendryes 1925: 87–8). This is one of the reasons why linguists have found it harder to decide 'what is a word' for French than for many other languages. (This point is further pursued by Matthews in chapter 11.)

4.3 *Grammatical and phonological aspects*

Before the idea (followed here) that one should deal separately with 'grammatical word' and 'phonological word' and then examine the relationship between the two units, there was confusion about exactly what a word is.

As Ullmann (1957: 46) points out 'since the word is the central element of the language system, it is natural for it to face both ways: not only is it the chief subject matter of lexicology, but it is dependent on phonology for the analysis of its sound-structure, and on syntax for the delimitation of its status in more complex configurations'. But is 'word' primarily a grammatical unit, with some phonological properties; or is it primarily a phonological unit, with some grammatical properties; or is it equally a unit in grammar and in phonology? Ideas have varied.

The majority opinion has been that 'word' is primarily a unit of grammar although, as Matthews (1991: 209) notes 'the word tends to be a unit of phonology as well as grammar. In Latin, for example, it was the unit within which accents were determined'. Jespersen (1924: 92) states 'words are linguistic units, but they are not phonetic units' and Bloomfield (1933: 181) agrees that 'the word is not primarily a phonetic unit', while Meillet (1964: 136) maintains: 'le mot n'admet pas, comme la syllabe, une définition phonétique; en effet la notion de mot n'est pas phonétique, mais morphologique et syntaxique'.

Lyons (1968: 200–1) puts it this way: 'we will continue to assume, with the majority of linguists, that in all languages the morpheme is the minimum unit of grammatical analysis. The question we have set ourselves therefore is this: how shall we define a unit intermediate in rank between the morpheme and the sentence and one which will correspond fairly closely with our intuitive ideas of what is a 'word', these intuitive ideas being supported, in general, by the conventions of the orthographic tradition?' He then adds (p 204): 'in many languages the word is phonologically marked in some way'.

Pike (1947: 90) makes a clear distinction between 'grammatical units', which include 'morphemes, words, clitics, phrases and utterances', and 'phonological

units', which include 'phonemes, syllables, stress groups, rhythm groups, intonation groups, utterances, and so on'. Halliday puts forward a similar view, having 'word' as one of the five 'units' in his grammatical theory, the full list being: morpheme, word, group, clause and sentence (see, for example, Halliday, McIntosh and Stevens 1964: 25).

Just a few linguists opt for the opposite position. Newman (1967: 182–3) begins his perceptive study of words and word classes in Yokuts with lists of phonological and grammatical criteria, stating 'morphological criteria serve to supplement the phonological features for delimiting the unit word'. And Wells (1947: 99) states 'because of their insufficiency, the phonemic criteria of a word must be supplemented, for every or nearly every language, by criteria of the second kind... the grammatical'.

Utilising phonological and grammatical criteria to define a single unit can, not unnaturally, lead to conflicts and ambiguities. Wells rightly states – working in terms of a single unit 'word' – 'in fact, the word is most solid as a unit in those languages where phonemic and grammatical criteria reinforce each other'.

An alternative position is to provide a set of criteria for deciding 'what is a word' that mix grammatical and phonological features, with no indication of what should be given priority when they do not provide the same result; see, for example, Bazell (1957: 25–6) and Chao (1968), discussed in §5.

We will – in §6 and §7 – suggest definitions for phonological word and for grammatical word, which should in each instance give a clear and unambiguous result. We will also, in §8, briefly discuss clitics, which may constitute a grammatical word but not an independent phonological word (clitics are discussed more fully in chapter 2). Before that it will be instructive to look – in §5 – at some of the types of criteria that have been put forward in the literature.

5 Some suggested criteria

In a short but classic discussion of 'the word' Bazell (1953: 67–8) states that 'criteria may be found which are either necessary, or sufficient, but not both'. If criterion X is necessary but not sufficient for defining 'word' this implies that all words show X but some other units show X as well. If criterion X is sufficient but not necessary this implies that any unit showing X is a word but there are also some words that do not show X.

Bazell then provides examples: 'the vowel-congruence [vowel harmony] of alternating morphs is a sufficient but not necessary criterion of word-unity in Turkish; the presence of at least one vowel is a necessary but not a sufficient criterion of word-status in English. The possibility of pause is a sufficient criterion, in most languages, of word-division'.

Lyons (1968: 200) paraphrases Meillet: 'a word may be defined as the unit of a particular meaning with a particular complex of sounds capable of a particular

grammatical employment'. He then points out that this may be a necessary but is by no means a sufficient criterion – a phrase such as *the new book* or affixes such as *un-* and *-able* (as in *unacceptable*) also have these properties.

In contrast, Bloomfield's well-known definition of 'word' as a 'minimum free form' is plainly sufficient but not necessary. As Matthews (1991: 210) points out 'Latin *et* "and" would normally be called a word, and so would English *my* or *the*. But are these words that could occur on their own?' They could do so in a kind of citation ('Did you mean *et* or *aut*?' '*Et.*') but so too could a part of a word. Matthews recalls having heard a dialogue: '(A) "Did you say *révisé* or *dévisé*?"'. (B) "Re".'

In his grammar of spoken Mandarin, Chao (1968: 146–7) suggests that 'the definition of a word as a minimum free form (free at both ends) has often been felt to be too drastic, and weaker conditions have been proposed instead. In languages with clear and regular phonological marking, it is fairly simple to find word boundaries without trying to find an isolated occurrence of the word as an independent utterance. For example, words in Latin can in most cases be marked off by the penultimate and antepenultimate stress rules. In the Wu dialects, compound words are recognisable from their tone sandhi, which are different within words from the tone sandhi between words . . . In Mandarin, stress and tonal patterns can sometimes be used to mark off words, but potential pauses are more generally available for this purpose'. Note that Chao is here adding one or more phonological criteria to Bloomfield's essentially grammatical criterion.

The possibility of pausing before and/or after has often been suggested as a criterion for 'word'. In the quotation just given Chao appears to consider it as necessary and sufficient for Mandarin, whereas Bazell quoted it as a 'sufficient criterion, in most languages' – that is, if one can pause on either side of a unit it must be a word, but there are some other words, in addition.

In a typical synthetic language a case could be made for 'potential pause' being a necessary but not a sufficient criterion. That is, pauses can be at the boundaries of units which are both (a whole number of) phonological and (a whole number of) grammatical words – and one always has the possibility of pausing at such a boundary – but there may also occasionally be pauses in the middle of a word (typically, at a morpheme boundary which is also a syllable boundary), for example *it's very un-* <pause, perhaps including *um*> *suitable*. (This is discussed further under (f) in §7.) It may be that in analytic languages, such as Chinese, pauses can only occur at word boundaries, never in the middle of a word. And it is certainly the case that the more polysynthetic a language is – that is, the longer its words tend to be – the more likelihood there is of a pause being made in the middle of a word, in addition to between words. This applies particularly to languages which are polysynthetic and agglutinative, less to those that are polysynthetic and fusional.

Where one may pause in natural speech is undoubtedly related to (but not necessarily identical to) where people do pause when dictating. Firth (1957: 5) suggests that one way of discovering the words of a language is 'by slow dictation, using any feeling for word-units the native may have'. Sapir (1921: 33–4) is more definite, stating: 'no more convincing test could be desired than this, that the naive Indian, quite unaccustomed to the concept of the written word, has nevertheless no serious difficulty in dictating a text to a linguistic student word by word'. However, Bloomfield (1933: 178) puts forward a contrary opinion: 'people who have not learned to read and write, have some difficulty when, by any chance, they are called upon to make word divisions'.

An explanation for these differences of opinion may well be that the various scholars were dealing with different types of language. When working on Dyirbal – a mildly synthetic and predominantly agglutinative language from northern Australia – Dixon found that speakers did dictate phrase-by-phrase, or more slowly word-by-word, or more slowly still syllable-by-syllable (never morpheme-by-morpheme). He then worked on Jarawara, a polysynthetic but basically agglutinative language from southern Amazonia. Here a verb form might involve six or more morphemes making up twelve or more syllables. When speakers dictate this language at a pace that the linguist can transcribe they tend to break up long words into feet (disyllabic units) and to pause between these. (A morpheme may span two feet, and a foot may span two morphemes.) As mentioned before, the longer the average length of word in a language, the more likelihood there is of pausing at some specifiable places within the word.

Sapir concludes that the unit word has 'psychological validity' for speakers of American Indian languages with whom he worked, presumably in a similar manner to speakers of English and other European languages – see the quotation from Bolinger at the beginning of §1. This is again a matter which may depend on the typological profile of the language involved; we return to it in §11.

The ideal situation, of course, would be for there to be one or more criteria for 'word' that would apply in all languages and be both necessary and sufficient in each. Vendryes (1925: 55–6) seeks such a universal criterion, just for 'phonetic word'. He assesses accent (or stress) as a possible candidate but finds that while this is fine for some languages it is not adequate for all: 'in certain languages the position of the accent is clearly decided by the word-ending; in others the accent falls upon the final or penultimate syllable, and in others again upon the beginning of the word. But these cases do not exhaust all the possibilities; there are tongues, indeed, in which the variable accent gives no indication of the word-ending. On the other hand, it may happen that there will be only one accent in a group of several words; or, conversely, a single word may have two. Greek and Sanskrit prove that Indo-European possessed what are called enclitics, short

words never used independently, but attached to the preceding word'. Note that Vendryes is here confusing two different kinds of unit. Although he sets out to discuss 'phonetic word', he then describes a clitic as a 'short word'. In fact a clitic is something which may have the status of a grammatical word but never that of a phonological word, being generally a stress-less element which attaches to some full phonological word (which does bear stress)⁸ – see §8 below, and chapter 2. Leaving aside such instances it is the case that stress is not a significant feature in some languages, and is not then available for deciding 'what is a word'. Vendryes appears to be correct in inferring that accent/stress is not a universal criterion for 'phonological word'.

We have seen that many discussions of 'word' combine grammatical and phonological criteria without any clear statement concerning the relative statuses of these two kinds of criteria. The most sensible course of action is to keep apart the two kinds of criteria and the units which they define.

6 Phonological word

It is clear that there is no single criterion which can serve to define a unit 'phonological word' in every language. Rather there is a range of types of criteria such that every language that has a unit 'phonological word' (which is probably every language in the world) utilises a selection of these.

We can offer the following definition:

A **phonological word** is a phonological unit larger than the syllable (in some languages it may minimally be just one syllable) which has at least one (and generally more than one) phonological defining property chosen from the following areas:

- (a) *Segmental features* – internal syllabic and segmental structure; phonetic realisations in terms of this; word boundary phenomena; pause phenomena.
- (b) *Prosodic features* – stress (or accent) and/or tone assignment; prosodic features such as nasalisation, retroflexion, vowel harmony.
- (c) *Phonological rules* – some rules apply only within a phonological word; others (external sandhi rules) apply specifically across a phonological word boundary.

Note that there is likely to be a close interaction between these types of features. For example, many phonological rules, under (c), operate in terms of stress

⁸ Quite a lot of the discussion of word suffers from not recognising the unit clitic and its status with respect to grammatical and phonological criteria for 'word'.

assignment within a word, under (b); the appearance of certain phonemes at certain positions within a phonological word, under (a), may be a consequence of the operation of certain phonological rules, under (c).

We can now briefly discuss these types of criteria, one at a time.

(a) *Segmental features*. In some Australian languages, for example, a root or suffix may have one or more syllables but every phonological word must involve at least two syllables. In Walmatjari (Hudson 1978: 37–43) a disyllabic verb root may take a zero tense–mood suffix, e.g. *luwa-∅* ‘hit!’ (the allomorph of imperative for the conjugation to which this verb root belongs is zero), whereas a monosyllabic root must take a suffix that is at least one syllable in extent, e.g. *ya-nta* ‘go!’ (here the imperative allomorph is *-nta*). In the Mbyá variety of Guaraní (Tupí–Guaraní branch of Tupí family) a monosyllabic root, when used without affixes, is obligatorily reduplicated in order to satisfy the requirement that each word have at least two syllables, e.g. root *hũ* ‘black’ becomes *hũʔhũ* as a complete word (Guedes 1991: 44, 49). In other languages each word must have at least two moras; thus, if a word is monosyllabic it must include a long vowel or a diphthong – this happens in Warekena (Aikhenvald 1998: 409) and in Fijian (Dixon 1988a: 25).

Looking now at segmental restrictions, there are languages in which a word-medial syllable may begin with a lateral but a word-initial syllable may not, e.g. the Western Australian language Yingkarta (Dench 1998: 15). One of the most common restrictions is that a word may not commence with *r* (this applies for Tariana, see §2.2 of chapter 2). And there are languages in which a word-medial syllable may end in consonant but every phonological word must be vowel-final, e.g. the Pitjantjatjara dialect of the Western Desert language of Australia (Dixon 1980: 209). Phonotactic possibilities sometimes vary for words of different types; for example, nouns and verbs may show different phonotactic possibilities. In §3.1 of chapter 10, Joseph finds that there are restrictions on final consonants in Modern Greek for words of native origin, but these do not apply to recent loans.

Trubetzkoy (1969: 275) notes that in some languages ‘certain distinctive oppositions’ occur only in initial or final position: ‘This is true, for example, for the aspirated occlusives of the Scottish-Gaelic dialect of Barra Island, the aspirated and recursive consonants of East Bengali, the recursive occlusives and emphatic palatalised consonants of Chechen’.

There are often different possibilities for sequences of phoneme types within a phonological word and across word boundaries; for example, a sequence of two vowels may only occur between words. In some Australian languages each word begins with a single consonant and ends with a vowel or a single consonant so that there can be a sequence of at most two consonants across a word boundary; however within a phonological word there can be a sequence of three consonants

(for example, *bulmbun* ‘mourning’ in Yidiɲ). In contrast, Zoque allows syllables beginning in CCC and CCCC only in word-initial position (Wonderly 1951: 116). In Estonian ‘only the first syllable of a word may begin with a vowel; every non-first syllable begins with a one mora consonant . . . If a word ending in a vowel is followed by a word beginning with a vowel, the occurrence of the sequence serves as a boundary marker, since only the first syllable of a word may begin with a vowel’ (Lehiste 1962: 179–80).

The realisation of vowel clusters between consecutive syllables may vary depending on whether the syllables belong to the same or different phonological words. In Fijian, for instance, certain vowel sequences are pronounced as diphthongs within a phonological word (e.g. the /oi/ in *boica* ‘smell’) but the same sequence across a phonological word boundary is pronounced as two distinct vowels e.g. *ilo.ilo* ‘glass’ (where ‘.’ indicates a phonological word boundary within a grammatical word, here the boundary of an inherent reduplication).

Quite apart from the possible positioning of phonemes within a word, their phonetic realisation often depends on position in a word. For example ‘in Japanese, where “g” initially is realised as the voiced obstruent g, and medially as a nasal ŋ, g is a positive and ŋ a negative non-phonemic boundary signal’ (Trubetzkoy 1969: 292). Similarly, the operation of certain phonological rules – see (c) below – can signal boundaries. For instance, in the Papuan language Yimas ‘the final nasal plus stop cluster simplification rule only applies at the end of words’ while ‘initial semivowel formation only applies at the beginning of words’ (Foley 1991: 80).

For the Arawak language Bare, Aikhenvald (1996) states that aspirated consonants are only found in word-initial position (most of them come from phonological rules which only apply at this place in the word, e.g. *me-haba* ‘3pl-fingernail’ → *m^heba* ‘their fingernails’). Thus, the presence of an aspirated consonant marks the beginning of a phonological word in Bare. And the presence of a nasalised vowel marks the end of a word, since this is the only structural slot in which nasalised vowels occur. In §3.2 of chapter 4, Henderson shows how the realisation of vowels at word boundaries in Arrernte constitutes a criterion for the recognition of these boundaries.

Meillet (1964: 137–40) has a useful discussion of processes applying at the ends of words in Indo-European languages (and see also Meillet 1970: 43–9). Trubetzkoy (1969: 273–97) provides an incisive discussion of boundary signals, mostly relating to the phonological word.

In some languages, words have special final features when followed by a pause. For example, in Warekena the occurrence of an *-hV* indicates that this must be the end of a phonological word that is followed by a pause (Aikhenvald 1996: 503; 1998: 411). In Semitic languages, such as Biblical Hebrew and Classical Arabic, words have distinctive forms when followed by a pause – see, for example, Gray (1934: 28–9) and Dresher (1994). The occurrence of

pausal forms is never likely to constitute a necessary and sufficient criterion for recognising a phonological word, but can be a useful concomitant feature.

(b) *Prosodic features.* In very many – but not quite all – languages, stress (or accent) provides one criterion for phonological word. Many languages have fixed stress – on the first or last or penultimate or antepenultimate syllable (or mora) of a phonological word. It should then be possible to ascertain the position of word boundaries from the location of stress. (For example, Olawsky shows, in §1.2 of chapter 8, that stress falls on the penultimate syllable in Dagbani; see also the examples given in Bloomfield 1933: 182 and Trubetzkoy 1969: 277–8.) The placement of stress may be linked to the segmental properties of phonemes; for example, in Latin stress falls on the penultimate syllable if it is long and on the antepenultimate if the penultimate is short.

In languages with contrastive stress there will generally be just one syllable with primary stress per word – see Weinreich (1954) on Yiddish, and Joseph and Philippaki-Warbuton (1987: 242–3) on Modern Greek. Although here phonological word boundaries cannot be deduced from the position of stress, one can tell from the number of stressed syllables in an utterance how many phonological words it contains (and one can deduce that a word boundary must lie somewhere between two stressed syllables).

However, in some languages stress placement may depend on a combination of morphological and phonological factors. In such cases stress may not be a useful criterion for phonological word.

A tonal system may relate to the syllable or to the phonological word – the latter applies in Lhasa Tibetan (see Sprigg 1955) and to the Papuan language Kewa (Franklin 1971, Franklin and Franklin 1978), for example.

A suprasegmental prosody such as nasalisation or retroflexion will have a syntagmatic extent, and this may be a phonological word. For example, Allen (1957) provides a prosodic account of aspiration in nominals for Hāṛautī (Rajasthani) in terms of the unit ‘word’. Among his conclusions is: ‘a breathy transition is never followed or preceded by another breathy transition within the same word’. Robins (1957) describes vowel nasality in Sundanese as having prosodic extent. A nasal consonant engenders nasalisation of a following vowel and of all subsequent vowels if separated from it only by a glottal stop or *h*; this continues until a word boundary is reached. (Robins points out that this applies to all nominal words except for loans and onomatopoeics.)

In Terena, an Arawak language, Bendor-Samuel (1966) describes how each word has one of three prosodies – nasalisation, yodisation (involving fronting and raising of all vowels, similar to vowel harmony) or neither nasalisation nor yodisation.

There is in Sanskrit a prosody of retroflexion which extends until the end of a word, under certain conditions. Allen (1951: 940) translates Pāṇini’s rule as:

‘*r*, *ṛ*, *r*, *ṣ*, in spite of intervening vowels, gutturals (including *h*), labials (including *v*), *y* and Anusvāra, change *n* to *ṇ* if followed by vowels, *n*, *m*, *y*, *v*’.

Vowel harmony is a prosody which operates over a certain syntagmatic extent, and this is often the phonological word. In Turkish, for instance, the vowels in certain types of word must either be all front or all back (Bloomfield 1933: 181; Waterson 1956). Trubetzkoy (1969: 285) mentions an associated phenomenon (found in Kazakh and a number of other Turkic languages) which he calls ‘synharmonism’ – a word can contain only front vowels and palatalised consonants or only back vowels and velarised consonants. This is also found in North-eastern Neo-Aramaic (Jastrow 1997: 352–3).

(c) *Phonological rules.* In many languages the optimum analysis involves recognising underlying forms for roots and affixes and then a number of phonological rules which apply to generate the surface forms. Each rule applies over a certain syntagmatic extent. Many rules apply just within the phonological word while some apply across a phonological word boundary.

We can first look at rules that only apply within a phonological word. In Hungarian, for instance, a rule of palatalisation assimilates dentals *d*, *t*, *l* or *n* to a following semi-vowel *j*, yielding the corresponding palatal sound, and the rule applies just within a phonological word (Kenesei, Vago and Fenyvesi 1998: 438, 440; Nespor and Vogel 1986: 123–4). In the Australian language Yidiñ (Dixon 1977a: 42–98; 1977b) some trisyllabic nominals are assigned an underlying form ending with a morphophoneme, e.g. *gajarrA* ‘brown possum’. There are the following rules that apply within a phonological word:

- (i) If a phonological word has an odd number of syllables then the penultimate vowel is lengthened.
- (ii) If a morphophoneme *A* is the last segment of a phonological word, it is omitted; otherwise it is realised as *a*.

We can compare what happens to *gajarrA* with zero suffix (for absolutive case) and with suffix *-gu* (for purposive case).

| | | | |
|-----|-----------------|-----------------|------------------|
| (1) | underlying form | <i>gajarrA</i> | <i>gajarrAgu</i> |
| | rule (i) | <i>gaja:rrA</i> | – |
| | rule (ii) | <i>gaja:rr</i> | <i>gajarragu</i> |

A root plus monosyllabic suffix (such as purposive *-gu*) forms one phonological word. But a disyllabic suffix always commences a separate phonological word. For example, *gajarrA* ‘brown possum’ plus privative suffix *-gimbal* ‘without’ gives *gajarrA.gimbal*, a single grammatical word that consists of two phonological words (again using ‘.’ for a phonological word boundary within a grammatical word). To this can be added purposive suffix *-gu*, which is part of the same phonological word as *-gimbal*. Rules (i) and (ii) then apply separately to the two phonological words within this grammatical word.

- (2) underlying form *gajarrA.gimbalgu*
 rule (i) *gaja:rrA.gimba:lgu*
 rule (ii) *gaja:rr.gimba:lgu*

If *gajarrAgimbalgu* had been one phonological word, it would consist of an even number of syllables. Rule (i) would not apply and the surface form would be **gajarragimbalgu*; the occurring form is, in fact, *gaja:rr.gimba:lgu*.

In §3.1 of chapter 4, Henderson refers to ‘prosodically conditioned allomorphy’. In effect, this involves phonological rules for the realisation of a vowel in a suffix (as *i* or *e*), depending on the number of preceding syllables within the phonological word. Similar rules in Jarawara are mentioned by Dixon in §2 of chapter 5.

In some languages the phonological rules that apply within a phonological word relate to stress or tone, and are thus an extension of (b).

Then there are some languages in which a special set of ‘(external) sandhi rules’ apply across word boundaries. In these languages word boundaries may be recognised partly by the operation of the sandhi rules. Allen (1972) is a detailed account of Sandhi in Sanskrit. Mutation in Celtic languages is a phenomenon of the same general type (see, for example, Gregor 1980: 149–57; Ball 1993: 9–10). (Rice 1990 has a useful discussion of types of phonological rule and the syntagmatic domains over which they apply.)

There can also be unusual, language-particular criteria for wordhood. Henderson (in §3.4 of chapter 4) describes how, in olden days, speakers of Arrernte used a play language style called ‘Rabbit Talk’. This involves relocating the initial syllable of a polysyllabic word to the end of the word, and thus indicates word boundaries.

Different types of criteria are relevant to defining the phonological word in different languages. And the relative importance and weighting of criteria differ from language to language. For example, in some languages a rule of vowel harmony may constitute a necessary and sufficient condition for recognising phonological words, whereas in others it may be sufficient but not necessary (see Bazell’s remarks on Turkish quoted in §5).

Sign languages employ a different medium of expression from their spoken cousins. Nonetheless, criteria similar to those discussed above have been enunciated. In §2 of chapter 6, Zeshan summarises Sandler’s discussion of phonological word in Israeli sign language, including phonological rules which operate within and across phonological words.

7 Grammatical word

For phonological word we could offer only a number of *types of* criteria, no one of which applies in every language. In the case of grammatical word it is

possible to put forward universal criteria, although tempered by a number of caveats.

A **grammatical word** consists of a number of grammatical elements which:

- (a) always occur together, rather than scattered through the clause (the criterion of cohesiveness);
- (b) occur in a fixed order;
- (c) have a conventionalised coherence and meaning.

A few comments are in order on the criteria.

Van Wyk (1968: 546) translates Reichling's (1935) 'two basic criteria for word identity' as 'internal immutability' and 'syntagmatic mobility' – these relate to our (a) and (b). Cohesiveness is a strong criterion. It is sometimes said that in Portuguese a pronominal clitic can intervene between verb root and future tense suffix (a putative exception to (a)). However, the facts are as follows. Future tense marking in present-day Portuguese has developed from a periphrastic form involving the verbal infinitive plus an inflected form of the verb 'have'. The form of the 'have' auxiliary first developed into an enclitic which can follow an object pronominal enclitic added to the infinitive form of a verb (such as *procurar* 'look for'), for example *procurá=lo=ei* 'I will look for it' (where = marks a clitic boundary). An alternative is to place the object pronoun before the verb ('it' is then just *o*), giving *eu (I) o (it) procurar-ei* 'I will look for it'. In this construction the future tense (plus 1sg subject) form *-ei* has evolved further, to be a suffix to the verb.⁹ The important point is that in *procurá=lo=ei* the *=ei* is a clitic, not a suffix.

We can illustrate the criterion of fixed order for Dyirbal, where there are two forms with similar meanings, *bulayi* 'two' and *jarran* 'two, each of two, a pair'. One could say either of:

- | | | | | |
|-----|--------------------|-------------|---------------|-----------------|
| (3) | <i>Ban</i> | <i>yibi</i> | <i>bulayi</i> | <i>bani-nyu</i> |
| | DETERMINER(fem) | woman | two | come-PAST |
| | The two women came | | | |
| | | | | |
| (4) | <i>Ban</i> | <i>yibi</i> | <i>jarran</i> | <i>bani-nyu</i> |
| | DETERMINER(fem) | woman | two | come-PAST |
| | The two women came | | | |

Dyirbal is a language with remarkably free word order. In (3) the four forms *ban*, *yibi*, *bulayi* and *baninyu* can be permuted and occur in any order (e.g. *yibi ban baninyu bulayi*). However in (4) *jarran* must follow *yibi*; here we can only

⁹ Verbs of the form *procurá=lo=ei* are still freely used in the Portuguese spoken in Portugal, but in Brazil they are confined to the written register and to a formal spoken style which deliberately reflects the conventions of writing (Prista 1966: 60–1).