

level intentionally, because to do otherwise would have resulted in great delays as I tackled problems of detail, and also because the resulting exposition has allowed me to make the points and address the issues I wish to focus on. There are also likely to be logical lacunae and analytical abysses in the interplay of ideas that I have forged in putting this work together. I invite the reader to point these out, and to offer suggestions and criticism.

**Packard, Jerome (2000). *The Morphology of Chinese: A Linguistic and Cognitive Approach*. Cambridge: Cambridge University Press.**

## 2 | Defining the word in Chinese

### 2.1 What is a 'word': different views

For speakers of some languages, the 'word' is a robustly intuitive notion. But it seems that no matter what the language, we have a hard time providing an exact definition that encompasses all and only those entities that our intuition tells us are words (see, e.g., Anderson 1985b: 153–4). This means that the concept 'word' is nothing if not elusive, and suggests that perhaps there is no concept of word that is universally applicable. Indeed, if there is no cross-linguistic, or universal psycholinguistic evidence for the existence of the word, then we may well doubt the validity of the word as a primitive natural language construct. It could a priori be the case that there is really no such thing in absolute terms as the 'word', and that it is just an artifactual linguistic construct that happens to coincide with salient units intermediate between morphemes and phrases that happen to appear in many of the world's languages.

There is another reason why the possibility that the 'word' is a derived rather than primitive construct may occur to us: words are definable using several disparate linguistic criteria. For some of these criteria considered in isolation, the label 'word' seems strangely inappropriate, since words so defined seem overly abstract, with nothing very 'word-like' about them. Let us take a look at these criteria to see if any of them are closer than others in providing an accurate portrayal of 'word'.

#### 2.1.1 Orthographic word

Probably the most popular conception of the word (especially in languages such as English) is that of the 'orthographic word', that is, the word as defined by writing conventions. It is easy for an English speaker (or a pigeon, for that matter) to segment a written English text into words strictly by the visual appearance of the text, i.e., by picking out the written material that occurs between the spaces. Speakers of English therefore have a strong 'intuition' as to what is and is not a word in spoken language, partly as an effect learned through experience

with orthography: in producing written English the speaker/writer must put the spaces in their proper place. This, of course, raises the question of what criteria are used to decide where the spaces go in the first place. It turns out that the criterion that is closest to the orthographic word in English is remarkably close to that of the 'syntactic word' (see 2.1.7 below).

In deciding for the purposes of this study what are words in Chinese, we could safely eliminate the orthographic word for reasons having little to do with Chinese *per se* – namely, that orthographic words are usually defined using non-orthographic criteria. That is, items are usually selected for membership in the 'orthographic word' category based upon linguistic properties other than the nature of the orthography. In any case, the orthographic word has no relevance specifically for Chinese, since Chinese orthography segments written texts into characters, which generally represent morphemes rather than 'words'.<sup>1</sup>

## 2.1.2 Sociological word

The term 'sociological word' may be attributed to Chao (1968: 136), and describes a concept that native speakers use to refer to linguistic units of a certain size. Chao defines it as 'that type of unit, intermediate in size between a phoneme and a sentence, which the general, non-linguistic public is conscious of, talks about, has an everyday term for, and is practically concerned with in various ways' (Chao 1968: 136–8). The sociological word is the familiar 'word' in English, and in Chinese, it is the 字, meaning either the Chinese written character or the Chinese spoken morpheme. The concept of the sociological word will be further discussed in 2.2.

## 2.1.3 Lexical word

Another common conception of 'word' we might call the *lexical word* (termed the *listeme* by Di Sciullo and Williams 1987: 1), which incorporates the 'listedness' characteristic of lexical items. That is, the lexicon is traditionally seen as that component of the grammar that contains

<sup>1</sup> Of course the orthographic definition of 'word' does work, albeit tautologically, for romanized Chinese, since in romanized Chinese the goal is generally to put spaces between words rather than between morphemes.

all that is not predictable, and must therefore be stored in a memorized list. To that extent, 'words' are those idiosyncratic, arbitrary pairings of sound and meaning that cannot be generated by rule 'on line' that we file away in memory for use in the performance of a speech act.

The 'listedness' criterion is neither sufficient nor necessary to define 'word', because it is common to have both 'listed' items that are not words (e.g. idiomatic phrases or 'listed syntactic objects', Di Sciullo and Williams 1987: 5) and words that are not 'listed' (e.g. large numbers of complex words in languages such as Turkish or Italian that are productively constructed using members of affixation paradigms, and are not likely to be stored away as 'listemes'). The concept of the lexical word is popular because it most closely comports with the idea of 'listing as a dictionary entry' that is popularly taken to be a defining criterion for 'word', and because it overlaps almost completely with the orthographic word discussed above.

The lexical definition of 'word' is not useful as a defining concept in our investigation of Chinese for just this reason: the 'listedness' criterion fails to include many Chinese words created by rule (see 7.4.1) and improperly includes many things approximating Di Sciullo and Williams' 'listed syntactic objects'. So while it will be interesting to keep this notion in mind – especially when it comes to the time to consider the structure of the Chinese lexicon – for the time being we will set aside the concept of lexical word.

## 2.1.4 Semantic word

A definition using semantic criteria is one of the most traditional ways of characterizing the notion of 'word'. The *semantic word* is sometimes equated with the idea of a 'unitary concept'. Sapir (1921/1949: 25) portrayed the word as 'the outward sign of a specific idea, whether of a single concept or image or of a number of such concepts or images definitely connected into a whole'. Baxter and Sagart (1997: citing Dowty, Wall and Peters 1981) characterize the semantic word as the 'basic expression' of formal semantics, a form with a semantic value such that such expressions may combine to form complex expressions, but may not be further decomposed into subexpressions (Baxter and Sagart 1997).

The semantic definition of word is one that strongly appeals to intuition – many people probably feel they have an idea of what a

'basic concept' might be, even if it is not uniquely definable either within or among speakers. However, the notion of semantic word is only minimally useful, because reducing concepts to their semantic primitives is a notoriously difficult exercise. Even if it were possible to come up with a list of such semantic primitives, examining them independently of their phonological form actually gets us no closer to defining 'word', since the concept of 'word' crucially requires reference to phonetic form. And once we relate those semantic primitives to phonological forms, what we get is a minimal pairing of form and meaning – an entity that is closer to the traditional morpheme than to the word.<sup>2</sup>

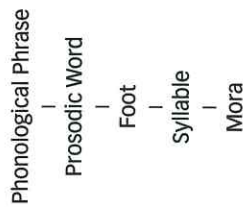
### 2.1.5 Phonological word

The *phonological word* is a 'word-sized' entity that is defined using phonological criteria. Chao (1968: 153–4) considers the existence of potential pauses – the places in a sentence where it is possible to pause naturally – to be a phonological criterion for the definition of word boundaries in Chinese (for a more general application of the concept, see Anderson 1985b: 150–2). But 'word' as defined by the phonological criterion of potential pause turns out to be of little use, since, like the orthographic and lexical definitions of 'word', this criterion turns out largely to be based upon other (i.e., syntactic, morphological or prosodic phonological) criteria. That is, the reason 'pauses' cannot go where a speaker feels it is inappropriate to place them is because their placement would violate the constituency of a syntactic, a morphological or a (otherwise defined) phonological word.

More recently the phonological definition of word has been based upon the domain of phonological rule application, or the output of a phonological rule. Dai (1997) gives examples of phonological word boundaries, defined by the application of a phonological rule. Baxter and Sagart (1997) give examples of accent (Czech) and sandhi (Sanskrit) phenomena, as well as stress units in Swahili, Polish and ancient Greek conditioned by independently defined word boundaries. The phonological word has also been characterized in prosodic terms, with Duanmu (1997) using phonological tone and stress evidence to distinguish words and phrases in modern Chinese.

<sup>2</sup> Thus we do refer to semantics when defining the morpheme, and make use of semantic criteria when we discuss the concept of 'semantic head' in 5.5.2.

Figure 1 Prosodic hierarchy



Another characterization of words in prosodic terms – the *prosodic word* – has been defined as an abstract constituent of a given level of prosodic phonological structure, located in the prosodic hierarchy between the phonological foot and the phonological phrase (see figure 1), with the prosodic word overlapping in many cases with the word as independently defined using other criteria (Selkirk 1980; Nespor and Vogel 1982, 1986; Inkelas 1989, 1993; McCarthy and Prince 1993; Feng 1997). In this theory, any instance of prosodic word must contain at least one foot, and every foot must in turn be bimoraic or bisyllabic. Thus, a prosodic word must contain at least two syllables or one bimoraic syllable. Feng (1997) applies this concept of prosodic word to ancient Chinese word formation, arguing that the prosodic word was important in the historical development of bisyllabic Chinese words.

We will not focus on words defined phonologically, because, while phonological structure may indeed be sensitive to and correspond to word-sized entities as independently defined elsewhere, and phonology does provide another important piece of evidence that converges on the construct *word*, nonetheless the other types of evidence correlate better with speakers' intuitions of what words are.

### 2.1.6 Morphological word

The *morphological word* may be understood as the result or 'output' of a word-formation rule. Di Sciullo and Williams (1987: 1) see morphological words (their term is *morphological objects*) as the set of items comprising morphemes and the output of the rules of morpheme combination. Anderson (1985b: 153) defines morphological word as 'a base together with the expression of the [grammatical] categories appropriate for its part-of-speech class'. Dai (1997: 112) has applied the

concept of morphological word to Chinese, meaning 'the maximal domain to which morphological rules may refer' also, 'the domains or outputs of compounding and affixation processes'.

My own definition of morphological word applied to Chinese is the proper output of word-formation rules in the language. Words defined in this way overlap to large extent albeit not completely with the set of wordlike entities defined using other criteria. For example, in the word *māotóuyīng* 猫头鹰 'cat-head-hawk' (with the structure  $[[N^0 N^0] N^0]_{NP}$ ); it is a 'cat-headed hawk' and not a 'hawk-headed cat', the constituent \**māotóu*- 'cat-head' is a morphological word because it is formed from the bona fide Chinese word-creating rule  $N^0 \rightarrow N^0 N^0$ , including the proper bracketing that results from the output of that rule (see 5.4 and table 23), as seen from the existence of words like *huòshān* 火山 'fire-mountain volcano', *bīnghé* 冰河 'ice-river glacier' and *mǎxióng* 马熊 'bear-bear brown bear'. However, unlike *huòshān*, *bīnghé* and *mǎxióng*, \**māotóu*- may not occur independently in an utterance (i.e., it may be a morphological word but it is not also a 'syntactic word', see 2.1.7 and the discussion of examples in chapter 5).

The morphological word – the 'morphological object' of Di Sciullo and Williams – turns out to be an important construct for Chinese, because there exists a clear, discrete set of word-formation rules in the language, the output of which does not overlap completely with the set of wordlike entities derived using other criteria. That is, all proper outputs of bona fide morphological rules are morphologically legal (and are therefore 'morphological words'), but some entities formed by these rules must be augmented with additional information before they can appear freely in utterances. However, while we will use the concept of morphological word in the analysis of Chinese that follows, it will not constitute our primary point of departure.

### 2.1.7 Syntactic word

A *syntactic word* is a form that can stand as an independent occupant of a syntactic form class slot, in other words, a syntactically free form, commonly designated in the literature as  $X^0$ . This is probably the most common current linguistic characterization of the notion 'word', and seems to serve as the basis for identifying the 'orthographic' and 'lexical' words discussed above (in languages with orthographies and dictionaries so designed, at any rate). In their X-bar analyses of word

structure, Sadock (1991: 27) and Selkirk (1982: 6) use the  $X^0$  designation to represent words as the minimal units of syntax, as do Di Sciullo and Williams, who term them 'syntactic atoms' (1987: 1).

Defining a syntactic word presumes that we can identify basic form class categories, and then use native speaker judgments to determine what entities are able to minimally occupy the category slots within utterances. This notion of syntactic word, as we shall see, will be one we crucially rely on in our description of Chinese words.

### 2.1.8 Psycholinguistic word

I use the term *psycholinguistic word* to refer to a portion of language at roughly the 'word' level of linguistic analysis that is (albeit perhaps not consciously) salient and highly relevant to the operation of the language processor, but does not necessarily match up consistently with any of the notions of *word* defined in traditional linguistic terms: it is a conception of 'word' as described *vis-à-vis* the operation of the language processor. The construct so defined could be a cognitive compilation of, e.g., phonological/prosodic, semantic, morphological and syntactic knowledge, with the relative proportions of such knowledge at any given point in processing time being dependent upon linguistic task demands or the state of the language processor.<sup>3</sup>

The existence of a psycholinguistic word is plausible from the perspective of psychology, as explained in a discussion of the contents of the lexicon by Henderson, who says that 'the lexicon of linguistics enjoys a different ontological status than psychology' and that 'it is perilous to assume that the boundaries of the linguist's hypothetical "lexicon" are congruent with that of the psychologist' (Henderson 1989: 35). With that as background, it is clear that the psycholinguistic word involves the 'psychological reality' of the linguistic construct 'word',<sup>4</sup> in the sense that it seeks to hold constant that part of 'word' that is the most 'psychologically real' at any given point in language processing.

<sup>3</sup> Di Sciullo and Williams (1987: 15–21), in a section entitled 'the psychological lexicon', discuss the storage of lexical items, but seem to equate the concept to that of 'listed object' or 'listeme'. Also, Hu (1985: 68–70) discusses an entity he calls the 'psychological word' (*xīnlǐcí* 心理词), but uses it to refer to the consensus of Chinese adults on the identity of words as extracted from written character texts.

<sup>4</sup> Hoosain (1992) discusses the psychological reality of words in Chinese.

One could imagine, for example, that of the semantic, phonological and grammatical information that composes the 'word', the type of information that is 'most active' at a given, fixed, point in the time course of language production (say, 500 milliseconds prior to onset of articulation) might be semantic, phonological or grammatical, depending upon whether linguistic task demands more heavily implicate the 'message' (roughly corresponding to the meaning; see Garrett 1988 and Levelt 1989), the sound or the syntax. Given this conception, the *psycholinguistic word* would have, e.g., a phonological identity in some contexts and a semantic identity in others, depending on the linguistic task.

The notion of the psycholinguistic word is quite intriguing and may in the end turn out to be quite relevant, but since little research has been done to determine what its properties might be in *any* language – much less Chinese – we are better off delaying its discussion until we have a firmer grasp on the concept *word* as understood based on traditional criteria.

## 2.2 The Chinese concept of 'word'

The 'word' is a clear and intuitive notion in English, because in the culture of English speakers the concept of the 'word' is particularly salient and robust. For example, there are *crossword* puzzles, English speakers find themselves searching for the right *word*, all writing is divided up into *words*, and what are searched for in dictionaries or databases are usually *words*. This is what Chao called the sociological word (Chao 1968: 136–8; see 2.1.2): the unit that the society and culture takes to be the salient, critical subcomponent of an utterance. For the English language and culture the concept is highly intuitive, and speakers of English might assume that the concept 'word' is universal because it is so salient in the culture and tradition of the English language.

In Chinese, however, the *word* is by no means a clear and intuitive notion. In Chinese language and culture, the clear and intuitive notion – the *sociological word* – is the *zì* 字. The term *zì* actually has two distinct meanings in popular usage: it can mean either a morpheme in the spoken language, or it can mean a written Chinese character (Hoosain

1992: 112).<sup>5</sup> But most speakers of Chinese do not distinguish between these two meanings of *zì* when they use the term: to these speakers, the *zì* as morpheme and the *zì* as written character are *one and the same thing*. This is due to the tacit assumption that the spoken *zì* (morpheme) can always be visually rendered with a written *zì* (character).

The status of the *zì* as the 'sociological word' in Chinese is just as salient as the status of the 'word' as the 'sociological word' in English. For example, in Chinese there are *zì* puzzles, Chinese speakers find themselves searching for the right *zì*, all writing is divided up into *zì*, and what are searched for in dictionaries or databases are usually *zì*. In the linguistic study of the Chinese language, the *zì* was considered to be the basic, primary unit of linguistic analysis as late as the 1920s (see Duanmu 1997; Packard 1997b). Further evidence of the salience of the *zì* as the sociological word is seen from the fact that Chinese speakers will often use the term *zì* to refer to a single two-character (two-syllable or two-morpheme) form, running counter to the usual equation of *zì* with 'character'.

There is a term in Chinese for 'word' as distinct from character, namely, *cí* 词. This term – which is used mostly as a technical term by specialists in language and linguistics – may be considered the 'syn-tactic word' in Chinese. For example, *hóngniǎo* 红鸟 red-bird 'red bird' consists of two characters and two morphemes, and it is also two *cí*, a noun and a modifying adjective. This is evident because of the generalizability and productivity of the two constituents: we can substitute virtually any adjective and any noun and it still retains its compositional meaning. In contrast, the term *hónghuā* 红花 red-flower 'safflower' (a Chinese medicinal term) consists of two characters and two morphemes, but it is one *cí*<sup>6</sup> because of its lack of compositionality: we cannot substitute another adjective for *hóng* or another noun for *huā* without losing the idiomatic meaning (example from Zhang 1985: 64).

Chinese scholars have written a great deal on how to distinguish *cí* from the smaller *zì* and the larger 'word groups' (*cíyǔ* 词组) or phrases (*duǎnyǔ* 短语 or *leyǔ* 短语), most often basing their decisions on

<sup>5</sup> There are, in addition, technical terms for morpheme: *cí* 词素 'word-element' and *yǔsù* 语素 'language-element'.

<sup>6</sup> The string *hóng huā* may also be considered two *cí* (meaning 'red flower'), and so it has both idiomatized and non-idiomatized readings, as with the English *black bird* and *blackbird*.

semantic or syntactic criteria. Wang (1953; citing Shuxiang Lü), for example, in addition to using semantic criteria (generally corresponding to 'basic unit of meaning'), defines a *cí* in Chinese as the 'smallest independently useable part of language' or 'that part of the sentence that can be used independently'.<sup>7</sup>

This issue of identifying the word arises as a practical problem in China because of the need to segment *pinyin* alphabetic phonetic writing into discrete orthographic units. *Pinyin* orthography is not written as an unsegmented string of letters nor is it segmented according to the syllable (as with the character orthography). Rather, scholarly authorities in China in principle define the unit of orthographic parsing to be the *cí* (Committee on Chinese Phonetic Orthography 1985). But in fact, the *cí* as defined for purposes of segmenting *pinyin* orthography and the *cí* as defined using grammatical principles are not the same thing. The former is termed the 'formal word' (or 'orthographic word') and the latter the 'theoretical word' (Zhang 1985: 62; citing Shuxiang Lü). The difference between the two is that the 'theoretical word' is, in essence, the *cí*, while the 'formal word' is based on the *cí* but undergoes further subjective redefinition based on length, conventional usage etc., focussing on ease of popular use (Zhang 1985: 62). This means, for example, that there is a de facto length constraint on the 'formal word' that does not apply to the 'theoretical word'.

The way I define the Chinese 'word' in the present work closely comports with the notion 'theoretical word' discussed above. The implication is that the notion of word presented here will have relatively little bearing on proposals such as those presented for orthography as seen in *Basic Rules of Chinese Phonetic Writing Orthography* (Committee 1985), because the latter is concerned with issues such as ease of use. The proposals in the present work could however be relevant to the parsing of texts into words as performed by computer algorithm, since, as we shall see, the present work is rule based, and depends on the identities and properties of constituent morphemes.

### 2.2.1 The reality of the 'word' in Chinese

The possibility that the 'word' is merely an artificial construct or epiphenomenon certainly occurs to speakers of Chinese, since –

<sup>7</sup> 词是‘语言的 最小的独立运用的单位’；词是‘句子所由组成的各个可以独立运用的部分’ Wang (1953: 3).

metalinguistically speaking – the 'word' in Chinese does not appear to be a particularly intuitive notion. Knowledge of the Chinese language, along with the 'culture of language' that accompanies that knowledge, suggests to Chinese speakers that the notion 'word' is a concept that comes from the West and so is based on the structure of western-type languages. Therefore, the intuition of many a Chinese speaker is that words simply do not exist in Chinese, and that the hearer simply 'gets the meaning' of an utterance as it unfolds, without it necessarily being parsed into word-sized units. This is suggested by some investigators in Chinese psycholinguistics as well, with Hoosain, for example, saying that 'a greater proportion of multimorphemic words in Chinese (compared with English) is not necessarily listed in the lexicon but instead have meanings *arrived at in the course of language use*' (Hoosain 1992: 126; my italics). Hoosain (1992: 126–8) also expresses doubt about whether Chinese speakers in fact possess an inherent conception of what a word is. Hoosain claims that a 'fluidity' of the word–morpheme boundary exists in the minds of Chinese speakers (1992: 118–20), and gives several reasons why this is true in Chinese more than in other languages.

Firstly, Hoosain says that the knowledge of classical Chinese language (in which many morphemes that are now bound were originally free) is variable among the native speaker population, and may affect speakers' judgments about what is and is not a word. Secondly, there is great variation among Chinese speakers in their knowledge of other Chinese dialects, which is important because, according to Hoosain, the bound–free status of morphemes differs across dialects. Thirdly, there is great variation in the bound–free status of morphemes according to context. Finally, Hoosain says that morphemes are more versatile in Chinese than in other languages, implying that morphemes may be inherently more indeterminate with respect to their bound–free status in Chinese.

All of these reasons that Hoosain uses to explain the fluidity of the word–morpheme boundary in the mind of the Chinese native speaker involve within- and among-speaker variation on the bound–free status of certain morphemes caused by various factors (see also Tang 1995: 196–8). However, the only thing this really suggests is that such speakers may be uncertain in their *metalinguistic judgment* about the status of certain morphemes. Metalinguistic judgments about one's language and how the language actually 'works' are two different

things. In particular, while speakers may be uncertain in making metalinguistic judgments about certain words and morphemes, it is unlikely that such speakers would be uncertain about the proper use of such forms. This becomes clear once we recognize that for a single morpheme to have both bound and free usages means that such a morpheme actually requires two separate entries in the mental lexicon: one as a free morpheme and one as a bound morpheme. Especially under these circumstances, native speakers may be uncertain when confronted with a metalinguistic choice, but will be totally clear on proper usage within a given context (see further discussion in 3.4.1).

Hoosain also cites several experiments purporting to show that the word is not a perceptual gestalt, and therefore not a psychologically real entity for native Chinese speakers. However, those results conflict with those of several more recent studies (Zhang and Peng 1992; Zhou and Marslen-Wilson 1994, 1995; Liu and Peng 1997; Taft and Zhu 1997), all of which suggest that two-morpheme words are indeed stored, retrieved and perceived as gestalt units. Also, Chinese characters were the experimental stimuli used in the studies cited by Hoosain to infer the properties of words. As we shall see in 7.5, it is likely that such results are confounded by the perceptual characteristics of Chinese characters as distinct from the words and morphemes those characters represent in the natural speech lexicon.

### 2.3 How we will define 'word' in Chinese

In this work, the syntactic definition of word will be used as the basis for analysing Chinese words. We begin with the syntactic definition as a first step in isolating wordlike units for analysis, for several reasons.

First, the syntactic definition is the one that most closely comports with the intuitive notion of 'word' among native speakers of Chinese, as evidenced by the fact that the Chinese technical term for 'word' (*cí* 词) is very close to the notion as defined using the syntactic definition. Also, aside from expressions which derive from Classical Chinese and different registers of use (such as literary vs. colloquial, standard vs. local dialect, individual variation, etc.), there is a surprising degree of unanimity among Chinese native speakers as to which entities are able to occupy a syntactic form class slot independently (see, e.g., Hu 1985: 69, who cites a study that found over 85 per cent agreement

on word boundaries). Where there is less than complete unanimity, it is likely that there are in fact two independent identities that coexist separately on a continuum in transition between, e.g., 'bound and free' or 'word and phrase' (see 4.3.4.3.2.1).

Second, the syntactic definition of 'word' motivates the concept in most other languages. It is the concept that Anderson refers to when speaking of the intuition of 'something real about the organization of the sentence' (1985b: 150), and says that sentences seem to be composed of such independent isolable entities. Third, some of the criteria for defining word discussed above – for example, the orthographic and lexical definitions, and the potential pause – are based upon the syntactic word. Finally, it will make sense for us to give a basic characterization of words using the syntactic definition because, as we shall see below, in Chinese the internal components of words are best understood and analyzed within a framework that complements the notion of 'syntactic word' as a basic defining concept.

The assumption of the existence of the syntactic word follows a universalist argument, which assumes that the word is biologically hard-wired and psychologically real, and has a tendency in natural language to 'weaken' the status of individual component morphemes, undermining their ability to function as free forms (see 6.1.1 and 6.2.2). Since it is generally recognized that sentence syntax contains the rules by which we produce and comprehend meaningful language, we must presume that utterances are segmented into minimal units that the syntax can manipulate. The constituents that are moved about by rules of syntax are nouns, verbs, etc., and the smallest occupant of one of those constituent slots is what we are theoretically defining as a 'word'. In the case of Chinese, these constituents cannot be morphemes, because morphemes are in no sense the units that are manipulated by syntactic rules to produce a comprehensible sentence or utterance (see discussion in 7.3.1 and 7.3.2). The *zì* or morpheme serves as a subpart of those entities that are the smallest things that can occupy a syntactic slot. Sometimes the *zì* can occupy it alone, but sometimes – indeed most of the time – the *zì* cannot occupy that slot by itself. But there are things that can minimally occupy those slots, and we have given them a name: they are called *words*.

To summarize, this work critically assumes that the linguistic construct of the syntactic 'word', rather than being an artifact of western linguistic analysis, is real and fundamental to the nature of language,

and therefore exists as a real linguistic construct universally used in producing and understanding utterances. To believe otherwise for Chinese, we would have to assume that the Chinese language is not so much 'word-based' as based on something else, with the most viable candidate being the morpheme. As we shall see in 7.3, such an assumption finds little empirical support.

## 3 | Chinese word components

### 3.1 Describing the components

#### 3.1.1 Possible descriptions

Having decided to use the 'syntactic' definition as our means of isolating Chinese words as units of analysis in this study, let us now consider how to understand the inner constituents of those words. We must determine which properties of word constituents will give us the greatest insight into Chinese words, and especially into the relationship between words' inner and outer properties. The characteristics of gestalt Chinese words are, as we shall see, related to the characteristics of the components that make them up. The question is: Of the many ways to characterize word components, which will give us the best insight into the properties of the gestalt word? Some of these characterizations apply poorly or inconsistently, and so provide an inadequate understanding of word components and their relation to the words they compose.<sup>1</sup> Let us consider some of the ways that have been used to describe Chinese word component morphemes.

#### 3.1.1.1 Relational description

Chinese words can be characterized by the general type of relationship that obtains between the two morphemes that make up the word. An example of this type of approach is Xia (1946) *Methods of Composing Two-character Words* (Pan, Yip and Han 1993: 38). Xia uses 'meaning limiting' (xiányì 限义) to describe two-morpheme nouns whose reference is 'limited' by having the first morpheme be a modifier of the noun morpheme on its right. He uses the term 'oppositional' (fānduì 反对) to describe words composed of opposing concepts. 'Modification' (fuzhuàng 副状) describes a general relationship of modification between the two word constituents. Finally, Xia uses the term 'cause-effect'

<sup>1</sup> Note that the question of how word components may be characterized is orthogonal to the discussion in 2.1 of how words themselves are to be defined. For example, it is conceivable that words are definable using syntactic criteria (see 2.1.7) but that the components of words are best described using a semantic analysis.