

Chinese: A Language of Compound Words?

Giorgio Francesco Arcodia

Università degli Studi di Pavia

1. Introduction

In this paper we shall address the issue of the massive creation of complex words in Mandarin Chinese, a phenomenon which has led many linguists to regard Modern Chinese as a “language of compound words”. The issue of the massive creation of compound words in Chinese is parallel to another issue, that of the process of disyllabification of the preferred word-form in the historical development of the language. The development of the typical word-form from monosyllabic to disyllabic has in many cases led to incorrect judgments about the compound status of many newly created words.

In what follows, we shall first seek a working definition of compound and make some remarks about the concepts of ‘morpheme’ and ‘word’, and we shall see how these notions may be applied to the Chinese language (sections 2 and 3); we shall then address the issue of ‘compounding’ in Chinese from a synchronic point of view, clarifying what kind of complex word forms may be regarded as compounds (section 4). We shall then turn to diachronic issues, and we shall review some of the literature on the evolution of the Chinese word-form (among others, Feng, 2001 and Shi, 2002), from which we may draw a brief sketch of the developmental processes at issue here (section 5). We shall first review the common explanations which one can find in the literature for the development of complex word creation, and then we shall review the phonological/prosodic reason for disyllabification in the Chinese language, which provided the conditions for a number of changes, among which is the increase of compound words, drawing mainly upon analyses by Feng (2001) and Shi (2002), and we shall argue that the basic phonological unit in Modern Chinese is made up of two syllables (a foot, in Feng’s terminology; section 6).

We shall then introduce our hypothesis, namely that the creation of a large number of compounded words was caused by the interplay of a number of factors, which include the above mentioned phonological simplification and the fact that in the Chinese lexicon there are almost no elements which can act as word boundaries, that is, fusive and/or agglutinative inflectional markers; moreover, the abundance of lexical morphemes, endowed with a stable relationship between phonological and orthographic form, is also supposed to be a facilitating factor in complex word production. We shall compare the Chinese data with some examples of multi-word expressions from the Romance languages, a family where the phenomenon of compounding is not as widespread as in Chinese; our hypothesis is that the different development in the domain of word formation between these two languages / language groups is motivated by the tendency to analyticity in the expression of coordination and subordination relationships both in syntax and in word formation (section 7). Lastly, we shall put forward the hypothesis that the extent to which a language is analytic or synthetic in word formation might have consequences on the development of compounding / complex word creation (section 8).

2. Lexeme, word and compound

Fabb’s definition of “compound” (1998: 66) is representative of the simplest and possibly most common conception of this category that we can find in the literature on word formation: “a compound

* I would like to thank Emanuele Banfi and Nicola Grandi (Università di Milano-Bicocca) for reading a version of this paper and for their valuable suggestions. Needless to say, all the mistakes which may be found are the sole responsibility of the author.

is a word which consists of two or more words”. Unfortunately, as has been noticed by many morphologists, by such a definition any expression which is larger than a word but cannot be defined a phrase falls into the category of compound, “which runs the risk of becoming a ‘utility room category’” (Grandi, 2006: 31-32¹); quoting Dai’s words (1998: 125): “[c]omound’ (...) is used as a cover term for a collection of related, but not necessarily identical, phenomena in the literature, ranging from a word composed of two or more bound stems to a word consisting of two or more existing words”.²

One quite big problem with this definition is that it relies on another controversial notion, i.e., what a word is: “[G]iven the difficulty that there has been for many years in defining a word, it is not surprising that there should be difficulty with the borderline of compounding. Items which fit poorly into the category of word should also fit poorly in the category of possible compound element” (Bauer, 2005: 106-107); as a matter of fact, we do not find the word as a constituent for this construction in all definitions of compound: Bauer (1998: 404) talks about “stems” or, later on, about “lexemes” (Bauer, 2005: 105);³ Grandi (2006: 43; my translation) defines compound constituents as “two or more forms which, according to the judgment of native speakers, have lexical autonomy (...)”. In a recent article, Bauer (2006: 719) is even more cautious when he talks about what a compound is made of, and he uses the term *subword*, reflecting the difficulties that arise when one examines data from languages belonging to different morphological types, such as Chinese, where we have no markers for common categories, such as gender and number for nouns: “[t]he implication of this is that the forms in which the individual subwords appear may be differently defined in different languages; a citation form in one, a stem in another, a specific compounding form in yet a third, a word form in a fourth”; it is the latter definition which, in our opinion, may best reflect the complexity of the phenomenon of complex word creation.

The issue of providing a consistent definition of the category of compound in the world’s languages is clearly beyond the aims of this paper; in the remainder of this section we shall make clear the theoretical background of our work, that is, what are the notions of ‘lexeme’ and ‘word’ which we use.

A good working definition of lexeme may be found in Amiot (2005: 194): “a multi-stratal entity, underspecified for flexion, which is characterized by three properties: a phonological representation, a syntactic category,⁴ a semantic representation”.⁵ As to the word, Ramat (1990, revised in Ramat, 2005) opts for the characterization of the prototype of the word, which should possess the properties of cohesion (no morphological material can appear inside of it), autonomy (i.e. it may appear in isolation) and mobility (can be used in different positions in a sentence): the prototypical word, according to him, may be found in isolating languages, such as Chinese, where it is typically monomorphemic, without classifiers or gender markers, maximally symbolic and opaque. In their cross-linguistic study of the notion of word, Dixon & Aikhenvald (2004) separate the two notions of “phonological” and “grammatical word”: whereas for the grammatical word “it is possible to put forward universal criteria, although tempered by a number of caveats”, “[f]or phonological word we could offer only a number of *types of criteria*, no one of which applies in every language” (2004: 18-19; their italics). According to these two authors, a grammatical word consists of a number of “grammatical elements” that occur together (cohesiveness), occur in a fixed order and “have a conventionalized coherence and meaning”. For our pur-

¹ In the original, “categoria ripostiglio”.

² See also Bauer (2001: 704): “there is a tendency apparent in the literature to call any multi-word lexical item a ‘compound’”.

³ Haspelmath (2002: 85) calls compound constituents “base lexemes”.

⁴ Here by “syntactic category” Amiot means Noun, Verb, Adjective and possibly Adverb.

⁵ See also Montermini (forthcoming): “[d]ans les études de morphologie actuelles, il est convenu de reconnaître une unité que l’on appelle lexème et qui correspond à la notion de mot dans la grammaire traditionnelle et dans le sens commun. En général, les morphologues concordent à considérer le lexème comme une unité abstraite, qui n’apparaît jamais en tant que telle en syntaxe, et qui comporte plusieurs rubriques qui garantissent l’interface avec les différentes composantes de la grammaire. En simplifiant, on peut considérer que tout lexème comporte au moins quatre rubriques : une représentation concrète (phonologique ou graphique), une catégorisation syntaxique (en termes de catégorie lexicale – nom, verbe, adjectif – mais aussi de traits de sous-catégorisation, structure argumentale, etc.), des propriétés morphologiques (qui spécifient, par exemple, la classe de flexion, la structure du paradigme, etc.), et une représentation sémantique (...)”.

poses, the most important insight of Dixon and Aikhenvald is that in a given language, the phonological and the grammatical word may not always coincide: as we shall see, this seems to be the case in Modern Chinese.

3. ‘Morpheme’ and ‘word’ in Chinese

Much has been written on the peculiar relationship between syllable, morpheme and character in Chinese, and on the consequence that this configuration has on the perception of what is a word (to name but one work on this topic, Chao, 1968); here we shall only briefly review the essential points already outlined in the previous literature.

The most striking characteristic of the Chinese lexicon is the almost perfect correspondence between the syllable and the written character (汉字 *Hànzì*)⁶ which, in turn, 90% of the times represent a morpheme:⁷ we see thus a very strong tendency in Chinese to a matching between a certain unit of phonology (the syllable), a unit of meaning (the morpheme) and a unit of writing (the character). Moreover, as it is known, in Chinese there are almost no compulsory markers for categories such as gender and number for nouns and adjectives, or tense for verbs (except for a small set of aspect / mood particles, almost never obligatorily marked; see Li & Thompson, 1981), thus having no morphological marker of word boundary and no sharp distinction between root, lexeme and word.

In fact, the vast majority of Chinese morphemes have a lexical nature, and the great part of them are bound, which are termed by Packard (2000: 77-78) “bound roots”; these may be compared to the so-called “neoclassical constituents” of Standard Average European languages (henceforth, SAE), such as *philo-*, *-logy* or *-phobia*, having lexical (rather than grammatical) meaning and always bound to some other constituent. Let us compare two (nearly) synonymous morphemes, the first being a bound root, the second being a word (i.e. a syntactically free form):

- | | | | |
|-----|-----------------|---|---|
| (1) | 食 <i>Shí</i> | → | 食欲 <i>shíyù</i> eat+desire ‘appetite’ |
| | ‘to eat, food’ | | |
| (2) | 吃 <i>Chī</i> | → | 吃一个苹果 <i>chī yí-ge píngguǒ</i> eat one-CLASS apple ‘to eat an apple’ |
| | ‘to eat’ | | |

When we have such cases of synonymy, usually the bound version has a stronger ‘classical’ flavour, whereas the free form is more modern. According to Packard (2000: 79), most complex words in Chinese are created by combining bound roots, differently from what usually happens in Indo-European languages. The constituents of complex words in Chinese have a strong phonological and semantic stability, showing no significative erosion and maintaining structural and semantic transparency, whereas opacification is quite uncommon.

The characteristics described above led to a sort of overlapping of the notions of morpheme, word and character in Chinese and, as Chao (1968: 136) puts it, the character is the “sociological word” for the “Chinaman of the street”, that is, “that type of unit (...) which the general, nonlinguistic public is conscious of, talks about, has an everyday term for (...)”. Thus, polysyllabic words may well be analyzed by the average speaker of the language as a combination of characters, that is, generally speaking, morphemes.

Let us turn now to the issue of what kind of complex words have been regarded as compounds in the literature, and what is their status as far as the morphology-syntax boundary is concerned.

⁶ The main exception to this principle is the subsyllabic morpheme 儿 *r*, originally ‘child’ (with the reading *ér*), which is used as a “word-forming affix” (Packard, 2000): e.g. 沿儿 *yánr* ‘edge, brim’ or 把儿 *bàr* ‘handle’.

⁷ According to data in Yin (1984) (quoted in Wang, 1998: 3).

4. Criteria for identifying compounds in Chinese

Given the substantial typological difference between Indo-European languages and Mandarin Chinese (taken as a representative of Sinitic languages, even though the non-Mandarin dialects of the South of China show a slightly different behavior in the creation of complex words), it is no surprise that the application of a ‘Western’ notion such as that of ‘compound’ has caused much debate. According to Packard (2000: 81), who applies the “textbook definition” of compounding (see the quotation from Fabb, 1998 in sec. 1), compounding as such is a minor phenomenon in Chinese, and “bound root words”, that is, complex words formed by at least one bound root, should be regarded as the main output of word formation processes in Chinese morphology.⁸ This view contrasts with what has been put forward by many prominent Chinese linguists: according to Chao (1968), only words which contain at least one bound constituent are compounds; Dong (2004) considers compounding of bound roots (词根复合 *cígēn fùhé*) as the most typical pattern of word formation for Chinese, even though free word compounding is not uncommon, and the only kind of complex words which are of no interest for the morphologist are those which have lexicalized and thus become non-transparent, which may well be analyzed as monomorphemic (such as 知道 *zhīdao* ‘to know’, where only the first constituent, ‘to know’, contributes to the meaning of the whole word).

As we have shown in section 2, there is at present a widespread agreement on the fact that compounding is indeed the combination of lexical morphemes, which may be free words, roots, etc., and therefore we shall adopt Dong’s perspective, focusing rather on the borderline between morphology and syntax, i.e. between compounds and phrases. For this goal, we shall introduce Feng’s (2001) distinction between what he calls “syntactic compounds” (句法词 *jùfǎcí*, literally ‘syntactic words’) and “morphological compounds” (词法词 *cífǎcí*, lit. ‘morphological words’). Compare these two examples (quoted from Feng, 2001: 170; my glosses and translations).

| | | |
|---|---|--|
| (3) 大盘子 <i>Dàpánzi</i> big plate-WFA ‘big plate’ | *白大盘子 <i>bái dàpánzi</i> white big plate-WFA ‘white big plate’ | *很大盘子 <i>hěn dàpánzi</i> very big plate ‘very big plate’ |
| (4) 大褂儿 <i>Dàguàr</i> big gown-WFA ‘(big) gown’ | 白大褂儿 <i>bái dàguàr</i> white big+gown-WFA ‘white (big) gown’ | *很大褂儿 <i>hěn dàguàr</i> very big gown-WFA ‘very (big) gown’ |

In (3) we have an example of what Feng terms a “syntactic compound”: the construction 大盘子 *dàpánzi* ‘big plate’ cannot be modified by the adjective 白 *bái* ‘white’, as if it were a phrase (where the grammatical ordering for adjectives would be dimension + color), but cannot be modified by the adjective / verb intensifier 很 *hěn* ‘very’, as if 大 *dà* ‘big’ were part of the word; in other words, it is a complex word created according to syntactic rules, and therefore shows both properties of phrases and words. This case reminds me of the Italian *trasporto latte* (“milk transport”) type constructions, which have been studied by Bisetto & Scalise (1999) and by Lieber & Scalise (2006), although I am not quite sure whether the explanation for these constructions is the same. In (4) we have an instance of a typical compound, which apparently has the very same structure as the one in (3): 大褂儿 *dàguàr*, lit. ‘big gown’, a traditional Chinese piece of garment, can be modified by the adjective 白 *bái* ‘white’, as *dà* ‘big’ is part of the word, and therefore the rule for the ordering of adjectives (which is a syntactic rule) does not apply; also, being also a word it is not modifiable by the adverb *hěn* ‘very’, just as (3). In short, we have in Chinese a class of complex words, which resemble any other compounded form, but which

⁸ “Why has there been confusion on this issue of ‘compound’ in Chinese linguistics? (...) In Chinese, the early label for ‘two-syllable words’ was *fùhécí* <复合词> compound-word ‘compounded word’, because they were indeed words that were formed through the combination of morphemes as implied by the term *fùhé* ‘compound’. This term lent itself rather nicely to the English translation ‘compound word’ (...)” (Packard, 2000: 78).

are formed according to rules of syntax, and exhibit characteristics of phrases (they must comply with the adjective ordering rules) and of words (they cannot be modified by an intensifier for adjectives / verbs).

In the next section we shall turn to the diachronic issues, and, in light of the historical data, we shall add some more remarks on the notion of compound in Chinese in section 6.

5. Some aspects of the historical development of the Chinese lexicon

Despite the striking continuity of the relationship between syllable, character and morpheme,⁹ there has indeed been a typological lexical shift in the evolution from Old Chinese into Middle and, later, Modern Chinese. In (5) we give an example of Classical Chinese prose (from the *Analecta* of Confucius, adapted from Wang, 1998: 157):

- (5) 學而時習之 不亦樂乎
Xué ér shí xí zhī bù yì lè hū
 study and often practice this not also joy question
 ‘To study and often apply what one has studied, is it not joy?’

The lexicon of Old Chinese (ca. 1200 BC – 300 AD) has a strong tendency towards monosyllabicity; all the character / morphs in the quoted passage are actually words. Nevertheless, even in the remote past of the language we have already some multi-syllable expressions,¹⁰ some examples of which are given in (6) (from Pulleyblank, 2004: 1736) and (7) (from Feng, 1998: 243):

- | | |
|-----------------|---------------------------|
| (6) 百姓 | 寡人 |
| <i>Bǎixìng</i> | <i>guǎrén</i> |
| hundred-surname | lonely-person |
| ‘common people’ | ‘I’ (used by a sovereign) |
| (7) 恐懼 | 杀戮 |
| <i>Kǒngjù</i> | <i>shālù</i> |
| fear-terror | kill-slaughter |
| ‘to fear’ | ‘to massacre’ |

In the evolution towards the modern language, the lexicon has undergone a massive process of disyllabification: whereas before 200 BC disyllabic words represented roughly 20% of the lexicon (at least in the written style), in the modern language estimates are above 80% (Shi, 2002: 70-72),¹¹ and the disyllable is now regarded as the preferred word-form in the Modern Chinese lexicon. Given the fact that, as said in section 3, the syllable in Chinese largely coincides with the morpheme, and, therefore, almost all disyllables are made up of two lexical morphemes, it is not surprising that Chinese has been defined as a “language of compound words”, as quoted in the introduction. However, I insist that disyllabification and compounding are two distinct phenomena, albeit related and indeed interacting with each other; we shall now examine the mechanisms which led to this peculiar configuration of the Chinese lexicon.

⁹ There is, though, one quite big difference between Old and Modern Chinese: in Old Chinese we still find some syllable modification processes (and, therefore, sub-syllabic morphemes), some of which possibly are the remnants of rules which were productive at an earlier stage (see Baxter & Sagart, 1998, Pulleyblank 2000, among others).

¹⁰ Here I advisedly avoid the use of the term “compound”.

¹¹ Although the disyllabic forms were already present in texts from ca. 1000 BC, a stronger tendency to disyllabification developed during the 漢 *Hàn* period (206 BC - 220 AD).

6. How to account for the disyllabification of the language?

Two main explanations are commonly found in the literature for the phenomenon of disyllabification of the Chinese lexicon: the so-called ‘functional explanation’ and the ‘phonological explanation’.

According to the first (Cheng, 1992: 44, Packard, 2000: 265-266), in the transition from a primitive society to a feudal one (between 1000 and 300 B.C.), there was a pragmatic need to create new words for new referents: “the shift towards the use of disyllabic words occurred when free monosyllabic words combined into new disyllabic words both through compounding (...) and through abbreviation of longer phrases. The newly juxtaposed morphemes subsequently often lost their status as free words, undergoing semantic shift or reduction due to the general effects of lexicalization (...)” (Packard, 2000: 365). According to the ‘phonological explanation’, which is indeed very common in the literature and seem to enjoy a widespread consensus (see e.g. Wang, 1998, Lin, 2001, Shi, 2002), the simplification of the phonology of the language caused the loss of many distinctions, and many syllables which were once separated became homophonous: to avoid ambiguity, the size of the word was enlarged, by adding an extra syllable.

It is worth mentioning that the latter theory does not, in principle, exclude the functional account, and some linguists stress the fact there might have been an interplay of these two motivations: “the increasing complexity of the lexicon, together with the simplification of the phonological system, provides an endless force to advance the tendency to disyllabification” (Shi, 2002: 74). Packard (2000: 267) himself considers the phonological explanation to be more plausible, “because it involves two processes that remain operative in the modern language: the continued simplification of the Chinese phonological system (...) and the continuation of ‘compounding’ as a way of forming new words”.

Let us have a closer look to what happened in the phonology of Chinese in its evolution towards the modern language. In Ancient Chinese (around 1000 B.C.), the minimal syllable was CVC and the maximal one was CCCMVCCC,¹² with at least ten possible final consonants. In Middle Chinese (around 800 A.D.), the syllable structure was simplified to CV (minimal) and {C, S} V {C, S} (maximal), with no consonant cluster allowed in the coda and only two types of final consonants, three nasals and three stops; the stop final consonants did not survive in Old Mandarin (Feng, 1998: 224). In the modern language, the minimal syllable can be a vowel, which may be preceded by a semivowel and/or a single consonant; no consonant clusters are allowed, and the only possible codas are [n] and [ŋ] (Yip, 2000: 20). The evolution in syllable structure may be summarized as such:

| Stage | Minimal syllable | Maximal syllable | Final consonants |
|-----------------|------------------|------------------|------------------------------|
| Ancient Chinese | CVC | CCCMVCCC | at least ten different |
| Middle Chinese | CV | {C, S} V {C, S} | [m], [n], [ŋ], [p], [t], [k] |
| Modern Chinese | V | {C, S} V C | [n], [ŋ] |

Table 1: The evolution of syllable structure in Chinese

Modern Mandarin has only 405 syllables, which may be read in four tones¹³ (Lin, 2001: 27-29); among these, only 297 have a single meaning, and the rest of them have at least two separate meanings, often corresponding to different characters (Lin, 2001: 9 and 85):

- (8) 亿 易 译 异 义 艺
 ‘100,000,000’ ‘easy’ ‘translate’ ‘different’ ‘justice’ ‘art’

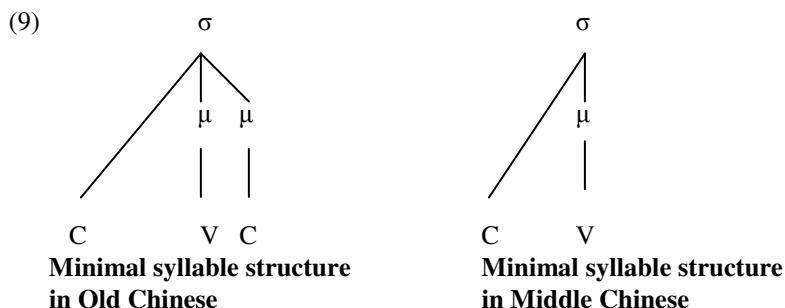
These characters are all read *yì* in the modern standard language, and are therefore ambiguous in their phonological form; they may even have polysemy in their written form, such as e.g. 易 *yì* which could mean both ‘easy’ (容易 *róngyì*, ‘easy’) and ‘change’ (易经 *yìjīng*, ‘Book of Changes’), but are normally unambiguous when used in a multimorphemic word.

Although the phonological explanation might seem quite straightforward, the best account for the establishment of the disyllabic unit as the preferred word-form involves the framework of Prosodic Morphology (McCarthy & Prince, 1993), and has been put forward by Feng (1998, 2001).

¹² Here C stands for consonant, V for vowel, M for medial and S for semivowel, following the usage of the source.

¹³ Although at the present stage not all syllables are actually attested in the four tones.

In the above mentioned theoretical framework, the Prosodic Word is the minimal independent unit of prosody, and it is realized by the foot (in Feng, 2001, 音步 *yīnbù*). It is given as a rule that a foot must be binary, either under syllabic or moraic analysis (McCarty & Prince, 1993: 43, quoted in Feng, 1998: 227 ff.): in the transition from Ancient Chinese to Middle Chinese the minimal syllable becomes CV (see Table 1), having only one mora, and the bimoraic foot is no longer possible. Words made of one minimal syllable only cannot constitute a foot any longer, and the new model for a foot (and, therefore, a prosodic word) is a two-syllable combination (adapted from Feng, 1998: 228):¹⁴



How does this interact with the creation of compounds? We have already pointed out that Classical Chinese lexicon was made mostly of monosyllabic words (sect. 4). The new disyllabic foot, therefore, coincides with what Feng calls a “two-word prosodic combination”; foot and phrase eventually merge together. If the two elements in a disyllabic prosodic word, which at this stage may still be regarded as a phrase, are used frequently, their relative order may become fixed, becoming thus an ‘idiomatized Prosodic Word’; the idiomatized Prosodic Word may in turn evolve into a compound through lexicalization, which involves a semantic shift of some sort.¹⁵ This process is summarized as follows:

(10) phonological change > disyllabic feet > disyllabic phrases > idiomatized Prosodic Words > compounds

What happens at this stage is that the prosodic requirement for two-syllable combinations becomes stronger, and the stronger it becomes, the more coordinating structures are created, often made of two synonymous or strongly related morphemes; this happens because coordinating structures are easier to build for prosodic purposes, as it is possible to add to a morph (quasi-)synonym without significantly changing its meaning, differently from what happens when a modifier is added to build a modifier-modified hierarchical compound (Feng, 1998: 223; glosses and translations adapted from the source):¹⁶

| | | | |
|----------------|----------------|----------------|----------------|
| (11) 衣裳 | 裳衣 | 圖書 | 書圖 |
| <i>yīshang</i> | <i>shāngyī</i> | <i>túshū</i> | <i>shūtú</i> |
| shirt-skirt | skirt-shirt | pictures-books | books-pictures |
| ‘clothes’ | | ‘publications’ | |

The two constructions in (11) could still be used with both orders at the time they were created, lacking one of the basic properties of the word (see sect. 1); later on, one of the two orders lexicalized (the former in both cases) and the words have survived as such until the present day.

¹⁴ Here σ stands for syllable, μ for mora.

¹⁵ According to Dong (2002: 37-40), in Chinese a phrase may lexicalize if it is disyllabic, the two elements occur together, it is frequent and undergoes semantic shift (改造 *gǎizào*, literally ‘reform, reconstruction’)

¹⁶ The picture here is somewhat simplified. Those compounds that nowadays look synonymic (i.e. made of synonymous constituents) actually evolve from other kinds of structures, such as collective constructions, where the two constituents are representative of an hyperonymous category: 皮革 *pígé* ‘leather’ (lit., furred animal leather-furless animal leather); see Waelchli (2005: 124).

Another way of forming disyllabic Prosodic Words is to add a so-called “dummy affix” (Lin, 2001: 82), that is, a suffix void of meaning, such as *-子 -zi*¹⁷ and *-头 -tou*, which may be added to a bound morph belonging to any word class to form a noun:

| | | | | |
|------|-------------|--------------|-----------------|---------------|
| (12) | 鼻子 | 梳子 | 想头 | 苦头 |
| | <i>Bízi</i> | <i>shūzi</i> | <i>xiǎngtóu</i> | <i>kǔtóu</i> |
| | nose-WFA | comb-WFA | think-WFA | suffering-WFA |
| | ‘nose’ | ‘comb’ | ‘idea’ | ‘suffering’ |

Packard (2000: 74) regards such elements as “word-forming affixes”, and regards words such as those in (12) as derived, as the affixes have the function of forming a Noun; however, given the fact that they often attach to morphemes which are basically already nouns (such as 鼻 *bí* in (12)) and considering the well-known categorial ambiguity of morphemes in Chinese (and in other typologically similar isolating languages), our opinion is that the noun forming function is parallel to the prosodic need to enlarge the size of the word to form a disyllabic prosodic unit. This is an example of a (disyllabic) word formation process which does not correspond to compounding.

Another sign of the tendency for the disyllable to be the preferred word form in the modern language is that expressions of three or more syllables tend to be abbreviated in two-syllable structures (lexical items, at least from the prosodical point of view):

| | | | | |
|------|----------------|----------------|---|-------------------|
| (13) | 劳动 | 保险 | → | 劳保 |
| | <i>láodòng</i> | <i>bǎoxiǎn</i> | | <i>láobǎo</i> |
| | labor-act | protect-risk | | labor-protect |
| | ‘labor’ | ‘insurance’ | | ‘labor insurance’ |

According to the figures in Sawyer (1995, quoted in Packard, 2000: 268), such abbreviated expressions account for 18,7% of new words in Chinese. They are frequently non-transparent, and they are used creatively also in Chinese communities abroad; (14) is a neologism coined in the Chinese community in Milan:

| | | | | |
|------|-----------------|---------------|---|------------------------------|
| (14) | 红线 | 地铁 | → | 红铁 |
| | <i>hóngxiàn</i> | <i>dìtiě</i> | | <i>hóngtiě</i> |
| | red-line | earth-iron | | red-iron |
| | ‘red line’ | ‘underground’ | | ‘red line underground train’ |

The latter expression in (14) is used colloquially to refer to the first line of the Milan underground, and this conventionalized meaning cannot be inferred from the meaning of the two constituents, meaning ‘red’ and ‘iron’.¹⁸

Having made clear which changes occurred in the phonology of Chinese and the interaction of such an evolution with prosody, as well as their consequences on word formation which go beyond the development of compounding (whatever definition we give to this label), let us now introduce our analysis of the interaction between syntax and morphology in complex word production.

7. Coordination and subordination in syntax and morphology

In Chinese syntax, as Chao (1968: 262-263) puts it, “the simplest and most frequent marker of coordination is zero”:

| | |
|------|-------------------------------|
| (15) | 他们卖桌子椅子 |
| | <i>tāmen mài zhuōzi yǐzi</i> |
| | he-PL sell table chair |
| | ‘They sell tables and chairs’ |

¹⁷ Historically, *-子 -zi* was a diminutive suffix, but now it has lost its evaluative meaning in almost all cases.

¹⁸ According to Packard (2000: 275-280), this is a productive way of coining “new morphemes” in Chinese, as an element may acquire a new meaning because of having been a constituent in such an abbreviated expressions.

- (16) 甜酸苦辣的东西他都吃 (...)
tián suān kǔ là de dōngxi tā dōu chī
 sweet sour bitter spicy STR PART thing he all eat
 ‘He eats sweet, sour, bitter and spicy things (...)’

We have seen in section 6 (11) that two-syllable (quasi-)synonymous phrases (see footnote 16), made of two juxtaposed morphemes / words, were often favoured for idiomatization (and eventually lexicalization) often in the development of the Chinese lexicon. Waelchli (2005: 215) claims that in the development of the morpho-syntax of a given language, a heavier form for the marking of coordination in syntax may appear, and coordinate phrases which have already become lexicalized do not participate in the development: indeed, this may be the case for Chinese.

If we compare Romance syntax,¹⁹ we see that here coordination is explicitly marked; in word formation, number and gender agreement may be regarded as a syntactic marker of coordination (Grandi, 2006; Arcodia, 2006; examples in Bisetto, Scalise & Guevara, 2004 and Olsen, 2000):

- (17) *studente lavoratore* → *studenti lavoratori*
 student worker student-PL worker-PL
 ‘working student’ ‘working students’ (It.)
- (18) *cantante-actor* → *cantante-actriz*
 singer actor singer actor-FEM
 ‘singer-actor’ ‘singer-actress’ (Sp.)

Subordination in Chinese syntax is marked by the structural particle 的 *de*; let us compare these two examples (examples quoted from Biasco et al., 2003: 101-102):

- (19) 说话人
shuōhuàrén
 speak-person
 ‘speaker’
- (20) 说话的人
shuōhuà de rén
 speak STR PART person
 ‘The person who is speaking’

Example (19) is a compound, whereas (20) is a phrase. In 19, 人 *rén* ‘person’ has a close tie with the determiner and has no referential function. The two different patterns, asyndetic and syndetic, are iconically motivated: the determination relationship is zero-marked when the connection between the constituents is conceptually tighter, and has an explicit morphological marking when the connection is looser. In Chinese, as in English for instance, constructions where two nouns have a zero-marked modification (attributive) relationship (both in phrases and in compounds) are used “to describe a common, well-known kind of thing that needs a special name” (Swan, 1980: 424 1)

In Romance syntax, there is a tendency to the analytical expression of subordinating relationships, differently from English:

- (21) *cibo per cani*
 food for dog-PL
 ‘dog food’

¹⁹ Here we took into consideration French, Italian and Spanish.

- (22) testa di ponte
 head of bridge
 ‘bridgehead’ (It.)

The conclusion we may draw from the picture outlined here is that in Chinese the development of a high number of complex words has been caused by the interaction of prosody and morpho-syntax: since the bulk of the Chinese lexicon is made up of monosyllabic lexemes, and given the absence of morphological markers of relational information (coordination, subordination, gender, number, etc.) which may hamper lexicalization, a great number of them eventually became (typical) compounds, whereas many structures which have an intermediate status between compounds and phrases (i.e. *syntactic compounds*) are perceived by speakers as words, mainly for prosodic reasons

Not all of these disyllabic forms are typical compounds, such as coordinating compounds made of (quasi-)synonymous members, which synchronically appear closer to simple, monomorphemic words, or words containing a ‘dummy affix’, which may be regarded as derivatives.

The Romance languages show a tendency to analyticity in the expression of coordination and subordination relationships both in syntax and in word formation, thus presenting an obstacle to the development of lexicalized phrases into words; also, in the languages of this group there has been no massive phonological simplification, and thus no motivation for the creation of anti-economical compounds such as the synonymous ones.

8. A hint for further research

From what we have seen about the interaction of morphosyntax and functional motivations in word formation, one is tempted to hypothesize that the correlation between syntheticity in word formation and high productivity of compounding in a language might be universal (or just very common), and that analyticity could hamper the production of typical compounds. Such a hypothesis, however, remains speculative, and should be tested on a proper language sample.

Abbreviations used in the glosses

CLASS: classifier
 WFA: word forming affix
 PL: plural
 FEM: feminine
 STR PART: structural particle

References

- Amiot, Dany 2005. Between compounding and derivation – elements of word-formation corresponding to prepositions. In W.U. Dressler, D. Kastovsky, O.E. Pfeiffer, F. Rainer, eds. *Morphology and its Demarcations*. Amsterdam – Philadelphia: John Benjamins. 183-195.
- Arcodia, Giorgio Francesco 2006. La natura morfologica dei composti coordinativi in cinese mandarino. *Studi italiani di linguistica teorica e applicata* 3:549-578.
- Bauer, Laurie 1998. Is there a class of neoclassical compounds, and if so is it productive?. *Linguistics* 36:403-422.
- Bauer, Laurie 2001. Compounding. In M. Haspelmath, E. König, W. Oesterreicher, W. Raible, eds. *Language Typology and Language Universals*. Berlin – New York: Walter de Gruyter. 695-707.
- Bauer, Laurie 2005. The borderline between derivation and compounding. In W.U. Dressler, D. Kastovsky, O.E. Pfeiffer, F. Rainer, eds. *Morphology and its Demarcations*. Amsterdam – Philadelphia: John Benjamins. 97-108.
- Bauer, Laurie 2006. Compound In K. Brown, ed.. *Encyclopedia of Language and Linguistics*. Oxford: Elsevier. 719-726 (II vol.).
- Baxter, William H. and Laurent Sagart 1998. Word formation in Old Chinese In J.L. Packard, ed. *New Approaches to Chinese Word Formation*. Berlin – New York: Mouton de Gruyter. 35-76.

- Biasco, Margherita, Wen Mao and Emanuele Banfi 2003. *Introduzione allo studio della lingua cinese*. Roma: Carocci.
- Bisetto, Antonietta and Sergio Scalise 1999. Compounding. Morphology and/or syntax? In L. Mereu, ed. *Boundaries of Morphology and Syntax*. Amsterdam – Philadelphia: John Benjamins. 31-48.
- Bisetto, Antonietta, Emiliano Guevara and Sergio Scalise 2004. The MorBo/Comp project. Talk presented at the conference *Word Structure and Lexical Systems: Models and Applications*. Pavia, December 16-17.
- Chao, Yuen Ren (赵元任) 1968. *A Grammar of Spoken Chinese*. Berkeley – Los Angeles: University of California Press.
- Cheng, Xiangqing 1992. 先秦双音词研究 (Researches on pre-Qin disyllabic words). In X. Cheng, ed. *先秦汉语研究 (Researches on pre-Qin Chinese)*. Jinan: Shandong Jiaoyu Chubanshe. 44-112.
- Dai, John X. 1998. Syntactical, phonological and morphological words in Chinese. In J.L. Packard, ed. *New Approaches to Chinese Word Formation*. Berlin – New York: Mouton de Gruyter. 103-133;
- Dixon, Robert M.W., Alexandra Y. Aikhenvald 2002. Word: a typological framework. In R.M.W. Dixon, A.Y. Aikhenvald, eds. *Word. A Cross-Linguistic Typology*. Cambridge: Cambridge University Press. 1-41.
- Dong, Xiufang 2002. 詞彙化: 漢語雙音詞的衍生和發展 (*Lexicalization: Genesis and Development of Disyllabic Words in Chinese*). Chengdu: Sichuan Renmin Chubanshe.
- Dong, Xiufang (董秀芳) 2004. 汉语的词典与词法 (*Chinese Lexicon and Morphology*). Beijing: Beijing Daxue Chubanshe.
- Fabb, Nigel 1998. Compounding. In A. Spencer, A. Zwicky, eds. *Handbook of Morphology*. Oxford: Blackwell. 66-83.
- Feng, Shengli (冯胜利) 1998. Prosodic structure and compound words in Classical Chinese. In J.L. Packard, ed. *New Approaches to Chinese Word Formation*. Berlin – New York: Mouton de Gruyter. 197-260.
- Feng, Shengli (冯胜利) 2000. 论汉语“词”的多维性 (On the multidimensionality of “word” in Chinese. *Dangdai Yuyanxue* 3:161-174.
- Grandi, Nicola 2006. Considerazioni sulla definizione e la classificazione dei composti. *Annali dell'Università di Ferrara. Sezione di Lettere*:31-52 <http://eprints.unife.it/annali/lettere/grandi.pdf>.
- Haspelmath, Martin 2002. *Understanding Morphology*. London: Arnold.
- Hopper, Paul J. and Elizabeth C. Traugott 1993. *Grammaticalization*. Cambridge: Cambridge University Press.
- Jiang, Shaoyu (蒋绍愚) 2005 [1989]. 古汉语词汇纲要 (An outline of Classical Chinese lexicon). Beijing: Shangwu Yinshuguan.
- Li, Charles N. and Sandra A. Thompson 1981. *Mandarin Chinese. A Functional Reference Grammar*. Berkeley: University of California Press.
- Lieber, Rochelle and Sergio Scalise 2006. The Lexical Integrity Hypothesis in a new theoretical universe. *Lingue e Linguaggio* V.1:1-26.
- Lin, Hua 2001. *A Grammar of Mandarin Chinese*. München: Lincom Europa.
- McCarthy, John J. and Alan S. Prince 1993. *Prosodic Morphology I – Constraint Interaction and Satisfaction*. University of Massachusetts and Rutgers University.
- Montermini, Fabio forthcoming. La composition en italien dans un cadre de morphologie lexématique. *Lexique*.
- Olsen, Susan 2000. Composition In G. Booij, C. Lehmann, J. Mugdan, eds. *Morphologie / Morphology I*. Berlin – New York: Walter de Gruyter. 897-916.
- Packard, Jerome L. 2000. *The Morphology of Chinese. A Linguistic and Cognitive Approach*. Cambridge: Cambridge University Press.
- Pulleyblank, Edwin G. 2004. From Archaic Chinese to Mandarin. In G. Booij, C. Lehmann, J. Mugdan, eds. *Morphologie / Morphology II*, Berlin – New York: Walter de Gruyter. 1730-1740.
- Ramat, Paolo 1990. Definizione di ‘parola’ e sua tipologia. In M. Berretta, P. Molinelli, A. Valentini, eds. *Parallela 4. Morfologia / Morphologie*. Tübingen: Gunter Narr. 3-15.
- Ramat, Paolo 2005. Per una definizione di ‘parola’. In P. Ramat, ed. *Pagine Linguistiche*, Roma – Bari: Laterza.
- Sagart, Laurent 1999. *The Roots of Old Chinese*. Amsterdam – Philadelphia: John Benjamins.
- Sawer, Michael 1995. Handling neologisms in teaching and learning modern Standard Chinese. *Australian Review of Applied Linguistics* 12:203-228.
- Scalise, Sergio 1992. Compounding in Italian. *Rivista di Linguistica* 4:175-99.
- Shi, Yuzhi (石毓智) 2002. *The Establishment of Modern Chinese Grammar. The Formation of the Resultative Construction and its Effects*. Amsterdam – Philadelphia: John Benjamins.
- Swan, Michael 1980. *Practical English Usage*. Oxford: Oxford University Press.
- Ten Hacken, Pius 2000. Derivation and compounding. in G. Booij, C. Lehmann, J. Mugdan, eds. *Morphologie / Morphology I*, Berlin – New York: Walter de Gruyter. 349-359.
- Waelchli, Bernard 2005. *Co-Compounds and Natural Coordination*. Oxford: Oxford University Press.
- Wang, Fusheng 1998. *La formazione delle parole nella lingua cinese contemporanea*. Trieste: Università degli

Studi di Trieste – Scuola Superiore di Lingue.

Yin, Fuyong 1984. 汉语语素的定量研究 (Quantitative research on Chinese morphemes). *Zhongguo Yuwen* 5:338-347.

Yip, Po-Ching (叶步青) 2000, *The Chinese Lexicon. A Comprehensive Survey*, London – New York: Routledge.

Selected Proceedings of the 5th Décembrettes: Morphology in Toulouse

edited by Fabio Montermini,
Gilles Boyé, and Nabil Hathout

Cascadilla Proceedings Project Somerville, MA 2007

Copyright information

Selected Proceedings of the 5th Décembrettes: Morphology in Toulouse
© 2007 Cascadilla Proceedings Project, Somerville, MA. All rights reserved

ISBN 978-1-57473-421-8 library binding

A copyright notice for each paper is located at the bottom of the first page of the paper.
Reprints for course packs can be authorized by Cascadilla Proceedings Project.

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, e-mail: sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Arcodia, Giorgio Francesco. 2007. Chinese: A Language of Compound Words? In *Selected Proceedings of the 5th Décembrettes: Morphology in Toulouse*, ed. Fabio Montermini, Gilles Boyé, and Nabil Hathout, 79-90. Somerville, MA: Cascadilla Proceedings Project.

or:

Arcodia, Giorgio Francesco. 2007. Chinese: A Language of Compound Words? In *Selected Proceedings of the 5th Décembrettes: Morphology in Toulouse*, ed. Fabio Montermini, Gilles Boyé, and Nabil Hathout, 79-90. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #1617.