Verbum XXXIV, 2012, nº 2, 233-256

NEOCLASSICAL COMPOUNDING BEYOND EUROPE: THE CASE OF EAST ASIA*

Giorgio Francesco ARCODIA
Università di Milano-Bicocca

Bianca BASCIANO
Università Ca' Foscari Venezia

RÉSUMÉ

Cet article fait l'hypothèse que la notion de 'composition néoclassique' peut être étendue au chinois moderne et à d'autres langues de l'Asie orientale. L'article traite de deux phénomènes liées : le premier est l'existence de couples de morphèmes synonymes dans la langue chinoise, dont l'un est non autonome et correspond à un formant néoclassique, tandis que l'autre est libre et appartient à la langue moderne (par exemple 食 shí 'manger' et 飲 yǐn 'boire' vs. 吃 chī and 喝 hē). Le second est la diffusion de morphèmes chinois comme éléments de formation de mots dans des langues des pays l'Asie orientale qui étaient sous l'influence de la culture chinoise, comme le vietnamien, le japonais, le coréen, ainsi que quelques dialectes chinois.

ABSTRACT

In this paper, we shall propose an extension of the notion of 'neoclassical compound' to the morphology of Modern Mandarin Chinese and other languages of East Asia. For our discussion, two related phenomena are of particular interest: the existence in Chinese of synonymic morphemes, one of which is bound and has a classical flavour, while the other is free and has a more 'modern' connotation (e.g. 食 shí 'eat' and 飲 yǐn 'drink' vs. 吃 chī and 喝 hē), and the diffusion of Chinese morphemes as bound word constituents in languages of East Asia which have come under the influence of Chinese culture, i.e. Vietnamese, Japanese, Korean, as well as in some Chinese dialects.

---

* Traditional characters have been used as a default for Chinese. The romanisation systems used are *Hanyu Pinyin* for Standard Mandarin Chinese, Hepburn for Japanese and the Revised Romanisation for Korean; for other Chinese varieties, the transcriptions are given as provided by the sources. The glosses follow the general guidelines of the Leipzig Glossing Rules. For academic purposes, Giorgio F. Arcodia is responsible for sections 3 and 4, Bianca Basciano is responsible for sections 1, 2 and 5.

1. INTRODUCTION

The phenomenon of neoclassical compounding has attracted much attention in the literature on word formation (see e.g. Scalise 1984, Bauer 1998, Olsen 2000, among others). This is not only because in Standard Average European (henceforth: SAE) languages neoclassical constituents are nowadays productively used also to create 'ordinary' words (besides technical and scientific terminology; see e.g. Iacobini (2010) on Italian) but, also, because they possess 'hybrid' properties, lying somewhere between lexemes and affixes, posing a problem for morphological theory. Thus, for instance, Naumann & Vogel (2000) point out that items such as Eng. *anthropo* are bound and do not correspond to any free word in the lexicon of the language, just as affixes, but nonetheless bear a recognisable lexical meaning (here, 'man'); moreover, if we treat *anthropo* as an affix, then we would have to admit that *anthropology* is a word made of two affixes and no lexical root (and cf. the case of *anthropic*).

The problems in the treatment of neoclassical compounds which we briefly outlined here are met again in the study of the morphology of Modern Standard Mandarin Chinese (henceforth: SMC). The 'typical' word of SMC is made of two lexical morphemes, each corresponding to a syllable and to a character in writing, as 電腦 *diàn-nǎo* 'computer', lit. 'electric(ity)-brain'; larger words, as well as monomorphemic words, are also attested (e.g. 投影器 *tóu-yǐng-qì* 'throw-image-instrument, projector', 書 *shū* 'book'), but the bimorphemic / disyllabic word is by far the preferred model (see the statistics in Yip 2000: 17-18). Hence, it is not surprising that Chinese has been defined as a « language of compounded words » (Lin 2001: 62). However, the vast majority of Chinese lexical morphemes are actually bound, i.e. they cannot occur in a syntactic slot if they are not 'completed' as words by other morphological material. Packard proposes that Chinese bound roots are akin to English 'Latinate' stems, i.e. neoclassical constituents, whereas 'true' compounds are only those composed of free morphs, as e.g. 冰山 *bīng-shān* 'iceberg', lit. 'ice-mountain' (2000: 77-78). For our purposes, two related phenomena are of particular interest:

a) the existence of synonymic morphemes, one of which is bound and has a classical flavour, while the other is free and has a more 'modern' connotation, as 食 *shí* 'eat' and 飲 *yǐn* 'drink' *vs.* 吃 *chī* and 喝 *hē*; 食 *shí* and 飲 *yǐn* are never used as such in Modern Chinese, and are rather employed as word constituents (食品 *shípǐn* 'foodstuff', 飲料 *yǐnliào* 'beverage(s)')
b) the diffusion outside the Mandarin-speaking world of Chinese morphemes as bound word constituents in languages of East Asia which, at some point in their history, have come under the influence of

Chinese culture, i.e. Vietnamese, Japanese, Korean, as well as in Chinese dialects.

Both phenomena are akin to neoclassical compounding in SAE languages. The situations described are parallel to that of pairs such as It. *−fero* 'bearing, producing' (< Lat. *-fer-*) in e.g. *fruttifero* 'fruit-bearing', 'interest-bearing' vs. the ordinary lexeme *portatore*, or *idr -o-* 'water' (< Gr. *hýdōr*) in e.g. *idrorepellente* 'water-repellent' vs. *acqua*. That is, both in Europe and in the Asian area of Chinese influence, including China itself, lexical morphemes from the prestige 'classical' languages are used as bound constituents in word formation, both for the production of technical and high-flown vocabulary and, also, in the creation of ordinary words, also in combination with other words. In the East Asian case, however, the language involved is only one, i.e. Chinese, mostly in the form of Classical Chinese, the written language modelled after the Confucian classics; this 'frozen' variety was « the principal vehicle of culture and civilization for the whole of East Asia for many centuries » (Pulleyblank 1995: 3), thus playing a role similar to that of Ancient Greek and Latin in Europe. Moreover, following Ceccagno & Basciano (2009) and Basciano & Ceccagno (2009), we shall show that some Chinese bound roots have a specialized meaning and form complex words mainly belonging to a specialistic lexical domain and/or to a formal register, as 家 *jiā* in 語言學家 *yǔyán-xué-jiā* 'language-study-specialist, linguist', and can thus be considered akin to neoclassical constituents, as we shall see.

In this paper we shall discuss the phenomena introduced above, comparing SAE neoclassical compounds with two classes of Chinese compounds and with compounds formed by Chinese morphemes in the languages of East Asia, as well as in some Chinese dialects. The paper is organised as follows. Firstly, we shall provide an overview of the aspects of Chinese morphology which are crucial to our analysis, discussing Chinese complex words akin to neoclassical compounds. In §3, we shall discuss the role of morphemes and complex words of Chinese origins in Japanese, Korean and Vietnamese. In §4, we shall briefly illustrate the issue of strata in the lexicon of Chinese dialects, extending the notion of neoclassical compounding to Taiwanese Southern Min, the most important dialect of Taiwan. Lastly, we shall summarise the main conclusions of our analysis.

2. THE MORPHOLOGY OF MODERN MANDARIN CHINESE: MORPHEMES, ROOTS, WORDS, COMPOUNDS

The vast majority of Chinese syllables corresponds to a morpheme[1], represented by a single character in writing: e.g. 馬 *mǎ* 'horse', 水 *shuǐ*

[1] We understand 'morpheme' as the smallest meaningful unit, be it lexical (e.g. Eng *dog*, *give*) or grammatical (e.g. Eng. *the*, *-ed*), and 'word' as 'syntactic word', i.e. an item able to occupy a syntactic slot. The 'root' is the part of the word bearing lexical meaning

'water', -著 *-zhe* 'durative aspect marker'. Thus, the great majority of Chinese morphemes are monosyllabic; DeFrancis (1984) proposes the term *morphosyllabic* to describe the tendential correspondence between syllable and morpheme in Chinese. However, Chinese has a small number of polysyllabic morphemes too, e.g. 玻璃 *bōli* 'glass', 葡萄 *pútao* 'grape'; most of them originated as loanwords, the pronunciation of which was adapted to Chinese phonology. These represent the only exception to the 1:1 correspondence between syllable and morpheme in the language[2].

2.1 Free and bound roots

In SMC we find both grammatical and lexical free and bound morphemes. However, the great majority of SMC morphemes are lexical and correspond to roots, either free or bound, and hence the most important distinction is that between free and bound roots (see e.g. Packard 2000, Yang 2003, Basciano & Ceccagno 2009, Ceccagno & Basciano 2009). Free roots are those which can be used to fill a syntactic slot, i.e. they are (syntactic) words, as e.g. 狗 *gǒu* 'dog', 來 *lái* 'come'. In contrast, bound roots have a lexical meaning but cannot fill a syntactic slot; they must be combined with another root, word or affix to form a word. For example, the root 桌 *zhuō* 'table' is a morpheme but not a word, since it cannot be used independently in a sentence, and it is not an affix either, since it is positionally free and bears lexical meaning; it must combine with other morphemes in order to form a word, as e.g. in 書桌 *shū-zhuō* 'book-table, (writing) desk', 飯桌 *fàn-zhuō* 'food-table, dining table', 桌面 *zhuō-miàn* 'table-surface, tabletop'. The free form for 'table' in SMC is 桌子 *zhuōzi*, where 子 *zi* is an empty suffix, no longer productive in the modern language, consistently building nouns.

According to Packard (2000), about 70% of SMC roots are bound. Most bound roots in the modern language were monosyllabic words (i.e. free roots) in the classical language, which lost their independence in the historical development of the language. In the evolution towards the modern language, the Chinese lexicon underwent a massive process of disyllabi-

---

obtaind after removing all affixes (inflectional and/or derivational), which cannot be analyzed into smaller parts. A root is always a morpheme; it can also be a word, as e.g. Eng. *car*, but not always, as e.g. It. *palla* 'ball', where *pall-* is the (bound) root, while *-a* is an inflectional affix. For further details see the discussion in the following sections.

2 The correspondence between syllable and morpheme is so strong in SMC that sometimes syllables of polysyllabic morphemes are reanalysed as morphemes, forming complex words (see Basciano & Ceccagno 2009: 109-112). For example, the word 葡糖 *pútáng* 'glucose; dextrose' is formed by 葡 *pú*, which, as seen above, is a syllable(/character) but not a morpheme, forming the disyllabic morpheme/word 葡萄 *pútao* 'grape', and 糖 *táng* 'sugar'. However, in the complex word 葡糖 *pú-táng* 'glucose', 葡 *pú* "absorbs" the meaning of the disyllabic morpheme 葡萄 *pútao* 'grape', so that it should be interpreted as 'grape-sugar, glucose'.

fication: while before 200 BCE disyllabic words accounted for about 20% of the lexicon (at least in the written style), in the modern language, as mentioned earlier, they are more than 80% (see Shi 2002: 70-72), and the disyllabic word became the preferred word form (we will get back to this later). The prevalence of bound roots in modern Mandarin is arguably related to this process; according to Dai (1990), the frequent usage of roots in compounding over time has led to the loss of syntactic independence for many of them.

However, the distinction between free and bound roots is not always clear. Many bound roots can be used as free roots under specific circumstances (see Yang 2003). First of all, we may remark that some bound roots can appear as free words in some specific syntactic contexts, as, for instance: in the sequence 'number + classifier + noun', e.g. 一隻鴨 *yī zhī yā* 'one CLF duck, one duck', where 鴨 *yā* normally is a bound root, the corresponding free form of which is 鴨子 *yāzi* (see Basciano & Ceccagno 2009); before a place word (方位詞 *fāngwèicí*), as e.g. 杯裡 *bēi-li* 'glass-inside, in the glass' (where 杯 *bēi* is a bound root, cf. the word 杯子 *bēizi* 'glass'; cf. also 杯子裡 *bēizi-li* 'glass-inside'; see Yang 2003). Moreover, bound forms, which cannot be used independently as words in speech, are often used as free forms in the written language, which retains some characteristics of Classical Chinese[3]. For instance, the bound root 毆 *ōu* 'beat' cannot occupy by itself a syntactic slot, but must combine with other roots; the corresponding free form for 'beat' is the compound word 毆打 *ōudǎ* 'beat-hit/beat, beat'[4]. However, in writing you may find this bound morpheme occupying a verbal syntactic slot by itself, as e.g. in the following sentence from a newspaper article (Yang 2003: 205):

(1) 中國一男子在東京機場被毆
Zhōngguó yī nánzǐ zài Dōngjīng jīchǎng bèi ōu

China one man at Tokyo airport PASS beat
'A Chinese man was beaten at Tokyo airport'

---

3 See DeFrancis (1984:193), « [m]uch of Chinese writing incorporates many elements alien to speech at times to such an extent as to make it incomprehensible when read orally. For more reasons than one this might be called unspeakable writing ».

4 The use of synonymic compounds is part of the disyllabification process briefly mentioned above. According to Feng (1998), the prosodic requirement for two-syllable combinations became stronger and stronger, and the stronger it became, the more coordinating structures were created, often made of two synonymous or strongly related morphemes; coordinating structures, indeed, are easier to build for prosodic purposes, since it is possible to add to a morpheme a (quasi-)synonymous item without significantly changing its meaning.

According to Yang, this may be explained by the prevalence of monosyllabic words in Classical Chinese, which still shows up in Modern Chinese writing[5].

2.2 Compound words

In SMC we find the following kinds of words:

a. monosyllabic monomorphemic words, i.e. words consisting of a single morpheme / syllable / character, as e.g. 書 *shū* 'book', 貓 *māo* 'cat', 去 *qù* 'go'
b. polysyllabic monomorphemic words (see §2), as 玻璃 *bōli* 'glass', 咖啡 *kāfēi* 'coffee', 馬拉松 *mǎlāsōng* 'marathon'
c. polysyllabic multimorphemic words (compound words), i.e. words consisting of more that one morpheme / syllable, as 電腦 *diàn-nǎo* 'electricity-brain, computer', 足球 *zú-qiú* 'foot-ball, football', 洗衣機 *xǐ-yī-jī* 'wash-clothes-machine, washing machine'.

As mentioned in the introduction, compound words are by far the most common kind of words in Modern Chinese: about 80% of SMC words are compounds (Xing 2006: 117), the bimorphemic / disyllabic word being the preferred model. However, according to some, only those words formed by free roots/words are true compounds. For instance, Dai (1992) assumes that productive compounding in Mandarin involves the combination of words; in contrast, root-based compounds are simply listed in the lexicon. So, for instance, a complex word like 書包 *shū-bāo* 'book-bag, schoolbag' is a compound, since it is formed by two free roots/words, while a complex word like 蟻王 *yǐ-wáng* 'ant-king, queen ant' is not a compound, since 蟻 *yǐ* is a bound root (cf. 螞蟻 *mǎyǐ* 'ant'). Similarly, Packard (2000: 80-81) distinguishes between compound words, i.e. words formed by two root words, as e.g. 冰山 *bīng-shān* 'ice-mountain, iceberg', and bound root words, i.e. those formed by a root word plus a bound root or by two bound roots, as e.g. 電腦 *diàn-nǎo* 'electricity-brain, computer'.

2.2.1 Bound roots, bound root words and neoclassical compounds

Packard (2000) believes that Chinese bound roots are similar in a way to Latinate stems in SAE. As a matter of fact, they have lexical rather than grammatical meaning, but are also bound; thus, both Chinese bound roots and Latinate stems share this hybrid nature (see also Packard 1998: 27, fn. 30 and 32). However, Packard also highlights that Chinese bound roots are positionally less restricted than Latinate stems, since they can occur freely either to the left or to the right of a complex word, as 友 *yǒu* 'friend' e.g. in

5 Note that many SMC bound roots can be used as free roots in other dialects (Yang 2003: 207).

工友 *gōng-yǒu* ‘work-friend, fellow worker’, 友愛 *yǒu-ài* ‘friend-love, friendly affection’, 友邦 *yǒu-bāng* ‘friend-nation, friendly nation’. Actually, Latinate stems in SAE too exhibit a certain degree of flexibility (Pirani 2008, Basciano & Ceccagno 2009); see e.g. Italian *morfologia* ‘morphology’ *vs. antropomorfo* ‘anthropomorphous’, *musicologo* ‘musicologist’ *vs. logopedista* ‘speech therapist’, *idrofilo* ‘hydrophilic’ *vs. filantropo* ‘philanthropist’.

Pirani (2008) compares SMC bound roots and Latinate stems in SAE, which she both terms ‘semi-words’, highlighting that they derive from a ‘classical’ language, i.e. Classical Chinese and Greek / Latin. She divides them into two different groups. The first group is formed by ‘semi-nouns’, involved in root compounding, i.e. semi-words selecting bound morphemes, which can be found either to the left or to the right of a complex word and can change their meaning according to the context. They are employed to form scientific and technical vocabulary. According to Pirani (2008: 268; 272), examples of semi-nouns are: SMC 子 *zǐ* ‘child, son’[6] (e.g. 電子 *diàn-zǐ* ‘electricity-particle, quantum’, 子音 *zǐ-yīn* ‘small.element-sound, consonant’); It. *fono* ‘phono’ (e.g. *fonologia* ‘phonology’, *microfono* ‘microphone’). The second group is formed by ‘semi-affixes’, involved in derivational processes, i.e. a restricted group of semi-words that select free morphemes. Semi-affixes in SMC can only appear to the right of a complex word, while semi affixes in European languages are exclusively used either as left-hand constituents (prefixes) or as right-hand constituents (suffixes). They always bear the same meaning and are not exclusively used for the creation of a technical and scientific vocabulary. According to Pirani (2008: 269, 272), examples of semi-affixes are SMC 機 *jī* ‘machine’ (e.g. 飛機 *fēi-jī* ‘fly-machine, airplane’, 電話機 *diànhuà-jī* ‘telephone-machine, telephone (set)’) and It. *−teca* ‘deposit’ (e.g. *paninoteca* ‘sandwich bar’, *discoteca* ‘discotheque’)[7].

Ceccagno & Basciano (2009) and Basciano & Ceccagno (2009) recognise the similarities between Mandarin bound roots and Latinate stems in European languages: they are bound forms (like affixes) with a full lexical meaning (like words), which take part in morphological processes but cannot independently occupy a syntactic slot. However, they also highlight that there are some important differences between these two kinds of roots. First of all, SMC bound roots are actively used to form complex words belonging to any area of the lexicon; they are not confined to any specific area, to a

6 Note that, besides ‘child’, ‘son’, 子 *zǐ* has a number of (related) meanings, including ‘young, small’, ‘sthg. small and hard’, ‘constituent’.

7 Pirani’s classification has some problems. First of all, it is not always the case that what Pirani calls ‘semi-affixes’ select free morphemes, see e.g. Italian *ludoteca* ‘gamesroom’, *emeroteca* ‘newspaper library’, where both *ludo* and *emero* are bound roots. Moreover, we can observe that the SMC bound root 機 *jī* ‘machine’ can be found to the left too, as e.g. 機鏟 *jī-chǎn* ‘machine-shovel, mechanical shovel’, 機槍 *jī-qiāng* ‘machine-gun, machine gun’.

technical or specialised language or to a formal register. Consider, for example, the bound root 椅 *yǐ* 'chair' (cf. the word 椅子 *yǐzi* 'chair') used as a constituent in complex words: 輪椅 *lún-yǐ* 'wheel-chair, wheelchair', 搖椅 *yáo-yǐ* 'shake-chair, rocking chair', vs. 椅背 *yǐ-bèi* 'chair-back of a body / object, the back of a chair', 椅墊 *yǐ-diàn* 'chair-cushion, chair cushion' (see also §2.1).

However, Ceccagno & Basciano (2009: 17-18) and Basciano & Ceccagno (2009: 116) remark that bound roots do not always come from Classical Chinese. As a matter of fact, new bound roots can be created, mainly through a process of reanalysis (see fn. 2). Consider, for example, the loanword 酒吧 *jiǔbā* 'bar (for alcoholic drinks)', a hybrid form, where 吧 *bā* is the phonetic adaptation of the English *bar*, while 酒 *jiǔ* 'alcohol' provides semantic information. The second syllable, 吧 *bā*, is not a morpheme by itself; however, it appears in some compound neologisms with the meaning of 'bar': e.g. 網吧 *wǎng-bā* 'net-bar, Internet café', 吧台 *bā-tái* 'bar-counter, bar counter' (see also Packard 2000: 280-283).

Furthermore, bound roots are not restricted in number as Latinate stems in the languages of Europe; actually, as said before, most SMC roots are bound. Ceccagno & Basciano (2009: 19-21) and Basciano & Ceccagno (2009: 116-117) remark that SMC bound roots are akin to lexical roots in a language such as Italian, which has fairly rich inflection (see also Packard 2000: 77). Differently from English, where lexical roots are generally free, i.e. they are (also) words (*dog*, *cat*, *house*), in Italian the great majority of lexical roots are bound and must be combined with an inflectional affix in order to form words: e.g. the word *libro* 'book' is formed by the root *libr-* 'book' plus *-o*, an inflectional affix ([+masculine], [+singular]). In Chinese, bound roots must combine with other roots (either free or bound) or with (derivational) affixes in order to form words; since most SMC roots are bound, words tend to have a complex structure, very often containing more than one lexical morpheme.

Bound roots are active in word formation processes in the same way as free roots, and they are productively used to form compound words; moreover, bound and free roots are formally identical, the only difference being that bound roots cannot occupy a syntactic slot, as said earlier. Sproat & Shih (1996), on the basis of a corpus-based study, show that (bound) root compounding is generally productive (*contra* Dai 1992; see above, §2.2). Therefore, there seems to be little reason to consider compounds of free forms and compounds of bound roots or made of a combination of bound and free roots as formed by different processes. This seems to be especially true if we consider that the definition of compounding is problematic in general, and the exact nature of the constituents of compounding often depends on the characteristics of a given language (for an overview see Lieber & Štekauer 2009: 4-8, Scalise & Vogel 2010: 5-6). Compounds can be formed by roots (either free or bound), words or even phrases in some

cases (see the so-called 'phrasal compounds'); Bauer (2006: 719) states that the elements forming a compound are best understood as lexemes, and that

> « [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 ».

Thus, if compound constituents can have different shapes in different languages, especially in relation to the morphological profile of the language at issue, it is not surprising that we have a high number of complex words formed by bound lexical roots in Mandarin.

Shall we conclude that Chinese has nothing similar to Latinate stems and neoclassical compounds in SAE? According to us, there are at least two types of compounds that can be compared to neoclassical compounds. The first type was proposed in Ceccagno & Basciano (2009: 22-27) and Basciano & Ceccagno (2009: 117-120). These compounds contain bound roots which have a specialised meaning and which, when combining with other roots or words, form complex words belonging to a specialistic lexical domain and/or to a formal register. One such example is the root 家 *jiā:* as a free root it means 'house, family', while as a bound root it means 'specialist, person who has a certain knowledge or is engaged in certain special business'[8], as in 藝術家 *yìshù-jiā* 'art-specialist, artist', 畫家 *huà-jiā* 'paint-specialist, painter', 政治家 *zhèngzhì-jiā* 'politics-specialist, statesman / politician'. Another example is the root 學 *xué*, which is both a free root meaning 'study' and a bound root meaning 'branch of learning, subject of study', forming technical vocabulary, as e.g.: 社會學 *shèhuì-xué* 'society-study, sociology', 數學 *shù-xué* 'number-study, mathematics', 詞彙學 *cíhuì-xué* 'lexicon-study, lexicology'. Basciano & Ceccagno (2009: 118) note that this kind of bound roots have the following properties: they tend to form nouns; they form complex words belonging to specific fields; they tend to combine with roots belonging to a specific lexical category (mainly nouns, even though sometimes they can combine with roots from different lexical categories); they occupy a fixed position in complex words. Given these selectional and distributional properties, we can say that these bound roots are more similar to affixes[9], if compared to other bound roots[10].

---

8 Originally, 家 *jiā* indicated a school of thought (e.g. 法家 *Fǎjiā* 'Legalist school', 儒家 *rújiā* 'Confucian school'), and then it was extended by metonymy to indicate an individual belonging to a certain trade/profession (Wang 1980 [1957]: 230).

9 An anonymous reviewer pointed out that, since 家 *jiā* has a fixed position, is productive and builds lexemes which have a consistent meaning, a possible assumption is that it is engaged in a process of grammaticalization; it is no longer a true lexeme, but not yet a true suffix. A similar point of view concerning the neoclassical constituent *-logue* in French (e.g. *anthropologue*) may be found in Dal & Amiot (2008).

Moreover, note that this kind of bound roots can be combined with one another, as long as another root or word is also present, as e.g. 語言學家 *yǔyán-xué-jiā* 'language-study-specialist, linguist' (cf. 語言學 *yǔyán-xué* 'language-study, linguistics).

Besides this kind of bound roots, we propose that in SMC there is another group of word formatives, quite different from those presented above, which can be regarded as akin to Latinate stems in SAE languages. In SMC, there are pairs of synonymic morphemes, one of which is bound and has a classical flavour, while the other is free and has a more 'modern' connotation, as the above mentioned 食 *shí* 'eat' (cf. 吃 *chī*) and 飲 *yǐn* 'drink' (cf. 喝 *hē*). These roots were used as free roots in Classical Chinese and can be still used as such in writing, e.g. in newspapers, the style of which retains some characteristics of the Classical language (see §2.1 and fn. 3). Consider the example in (2):

(2) 少吃羊肉，多吃瘦肉 [...] 多食鴨肉 [...][11]
shǎo chī yáng-ròu, duō chī shòu-ròu [...] *duō shí yā-ròu* [...]
less eat sheep-meat more eat lean-meat more eat duck-meat
'Eat less mutton, eat more lean meat [...] eat more duck [...]'

In the same sentence, both 吃 *chī* and 食 *shí* are used. In the spoken language, the normal word for 'eat' is 吃 *chī*, not 食 *shí*. Thus, 食 *shí* can be used as a free root only in writing; in the spoken language it is a bound form, meaning 'meal, related to food' and can appear only in complex words, such as: 食品 *shí-pǐn* 'meal-product, food', 食堂 *shí-táng* 'meal-hall, dining room', 豬食 *zhū-shí* 'pig-meal, pig feed/swill'. The same goes for the root 飲 *yǐn*, which is a verb ('drink') that can be found as a free form only in writing. In the spoken language, the free form 喝 *hē* is generally used, while 飲 *yǐn* is employed as a word constituent with the meaning of 'drink (noun)'[12], such as: 飲料 *yǐn-liào* 'drink-material (raw), drink, beverage', 飲品 *yǐn-pǐn* 'drink-product, soft drinks', 餐飲 *cān-yǐn* 'eat-drink, food and beverages / restaurants, bars, coffee houses, and tea rooms'. Another example is the root 函 *hán* 'letter' used mainly in writing; the corresponding spoken form is 信 *xìn*. However, 函 *hán* can be easily found as a constituent

---

[10] Note, for example, that the root 學 *xué* can also be a bound root with the meaning of 'school' or 'learning, knowledge'. With these meanings it behaves as a normal bound root, combining freely with other roots, attaching either to the left or to the right (see Basciano & Ceccagno 2009: 119): see e.g. 上學 *shàng-xué* 'go.to-school, attend school', 中學 *zhōng-xué* 'middle-school, middle school', 學區 *xué-qū* 'school-district, school district'.

[11] From 春天易上火巧用饮食灭 - 多食蔬菜水果补充维生素 [Extinguish the spring heat with an appropriate diet - Eat more vegetables and fruits to supplement vitamins], 人民网 *People's Daily Online*, 6 April 2012: http://health.people.com.cn/GB/17587916.html (accessed: 04/07/2013).

[12] In verb-object compounds, it can also be found in its verbal meaning, e.g. 飲冰 *yìn-bīng* 'drink-ice, cool oneself by drinking ice water'.

in complex words, such as: 公函 *gōng-hán* 'public-letter, official / collective letter', 函告 *hán-gào* 'letter-inform, inform by letter'.

As can be seen from the examples above, those formatives behave as all other bound roots as far as their position in the complex word is concerned; i.e. they can be either left-hand or right-hand constituents. Note also that these items can combine with one another, as e.g. 飲食 *yǐn-shí* 'drinks-meal, food and drink/diet'. To sum up, these bound roots have the following characteristics:

a. they are drawn from a written classical language, and often can be still used as free forms in writing
b. in the spoken language, they are usually bound and have a free equivalent
c. they are positionally free, i.e. they can be either left-hand or right-hand constituents in complex words
d. they are not necessarily confined to technical or specialised vocabulary, but are also commonly found in ordinary words.

3. CHINESE MORPHEMES OUTSIDE CHINA: JAPAN, KOREA, VIETNAM

Modern Japanese, Korean and Vietnamese are characterised by a very high share of words made of morphemes borrowed from Chinese, which are termed, respectively, 'Sino-Japanese', 'Sino-Korean' and 'Sino-Vietnamese' vocabulary (henceforth, 'S-J', 'S-K' and 'S-V'); the proportion of 'Sino-' words in the lexica of these languages is comparable to that of Latinate vocabulary in Modern English (see Shibatani 1990, Sohn 1999, Nguyen 1997). Japan, Korea and Vietnam were all under the strong cultural influence of China; moreover, at some point in their history, Classical Chinese was the language for formal writing *par excellence*, the 'Latin of East Asia' (Ames 1998). Early Korean, Japanese and Vietnamese *literati* mostly wrote in this language, rather than in their own native tongues (see Lee & Ramsey 2011, Shibatani 1990, Nguyen 1997); thus, finding a high number of Chinese words in the lexica of these languages is not surprising.

What is more interesting, in our perspective, is that many of those words are not Chinese *stricto sensu* but, rather, they are composed of morphemes of Chinese origin, just as neoclassical compounds in SAE languages are composed of morphemes of Greek and/or Latin origin, but often are not actual Greek or Latin words. In order to 'modernise' the lexicon, a huge number of words entered the Chinese lexicon between the XIX and the XX centuries; however, many of them were actually first used in Japan and then borrowed into Chinese, as e.g. 電話 *denwa* 'telephone' (lit. 'electric-speech'; SMC *diànhuà*) or 主義 *shugi* 'doctrine' ('main-significance'; SMC *zhǔyì*, Masini 1993; see also Lippert 2001). A word as 'philosophy' was first coined in Japanese as 哲学 *tetsugaku* (lit. 'clear / sagacious-study') and then exported to China (哲學 *zhéxué*), Korea (철학 *cheolhag*) and Vietnam (*triết*

học; Ames 1998, Lippert 2001); several S-K and S-V words come from S-J, rather than from Chinese itself (Sohn 1999, Bisang 2001, Alves 2007). Moreover, some words coined outside China were actually never exported, as S-K 편지 *pyeonji* 'letter' (lit. 'comfortable-paper') or 총각 *chonggak* 'bachelor' (lit. 'all-horn'; Sohn 1999:104)[13].

Note that in the creation of S-J, S-K and S-V vocabulary items the word formation conventions of Chinese, rather than those of native morphology, are followed. Thus, for instance, even though Vietnamese noun-noun compounds have the modified-modifier order, S-V compounds have the opposite order: compare *nước mắt* 'tear(s)', lit. 'water-eye' and *giáo sư* 'teacher' ('teach-master'; see Nguyen 1997); in Japanese, a SOV language, native compounds containing a verbal constituent and its object are right-headed, as 人殺し *hito-goroshi*, lit. 'person-kill(ing)' *vs.* 殺人 *satsu-jin* 'kill-person', both meaning 'manslaughter' (Shibatani 1990: 240; see also Kageyama 2009). The same happens for neoclassical compounds in a Romance language such as Italian: compare *cruciforme* 'cross-shaped' (Lat. *crucis*, genitive of *crux* 'cross' and *−forme*, from *forma* 'shape') and *pesce spada* 'swordfish' (lit. 'fish-sword').

Let us now turn to a brief overview of Chinese vocabulary in the three languages at issue here.

3.1 Sino-Japanese

Despite being genetically unrelated to Chinese, the Japanese language owes to this language not only the writing system, but also a huge portion of its vocabulary, as said before. The relationship of Japanese with the Chinese language and the Chinese writing system is very complex and has changed much throughout history (see Frellesvig 2010: 258-293 for an overview); here, we shall focus on the issues pertaining to S-J vocabulary, for reasons of space.

The (Classical) Chinese language was introduced in Japan around 400 CE, although Chinese words may have entered the country as early as the first century; throughout the years, many 漢語 *kango* (lit. 'Chinese words') were gradually accepted into spoken dialects (Shibatani 1990: 145). The fact that S-J words and morphemes came to Japan at different times is reflected in their pronunciation, and we may distinguish between three different strata (Shibatani 1990: 120-121; Frellesvig 2010: 274-276). 吳音 *go'on*, from the name of the 吳 *Wú* state in ancient China (229-280), represent the oldest stratum, associated with the borrowings of the fifth and sixth century. 漢音 *kan'on*, i.e. the 'Chinese' readings, are associated with the borrowings of the

[13] The corresponding SMC words 便紙 *biànzhǐ* and 總角 *zǒngjiǎo* mean, respectively, 'toilet paper' and 'child's hair twisted into a knot', and are believed to have no connection with the S-K terms.

seventh and eighth century, when Japanese envoys studied in Chang'an, then the capital of the Chinese Tang dynasty (618-907). 唐宋音 *tō-sō'on* (or simply 唐音 *tō'on*), from the name of the Chinese dynasties Tang and Song (960-1279), are the readings associated with later borrowings, mostly terms related to Zen buddhism[14].

However, much of the S-J vocabulary in current use is not directly derived from the premodern contacts with Chinese but, rather, was coined in the struggle for the modernisation of the Japanese vocabulary in the Meiji period (1868-1912) and beyond. Following a long period of rejection of foreign influence[15], during this time Japan strove to modernise the country by importing not only technology, but also ideas from the West; many of these new notions did not have a corresponding term in Japanese (or Chinese), and hence many new words had to be coined, especially by translators of Western works. We mentioned before (§3) 電話 *denwa* 'telephone' and 主義 *shugi* 'doctrine': other examples of S-J words coined in the Meiji period, and exported towards China, Korea, and Vietnam, include e.g.:

(1) a. S-J 社会 *shakai* 'society' (lit. 'organised.body-meeting')
b. SMC 社會*shèhuì*
c. S-K사회 *sahoe*
d. S-V *xã hội*

(2) a. S-J 国家 *kokka* 'state' ('country-family')
b. SMC 國家 *guójiā*
c. S-K *gukga*
d. S-V *quốc gia*

(3) a. S-J 文学 *bungaku* 'literature' ('text-study')
b. SMC 文學 *wénxué*
c. S-K *munhak*
d. S-V *văn học*

As pointed out by Frellesvig (2010: 409-410), the ready availability of S-J morphemes to coin neologisms which « are short and in some cases provide educated readers with semantic clues », and have « the important function of giving the impression that these words were part of and belonged

[14] As many characters / morphemes were included in borrowings from more than one layer, in Modern Japanese we have multiple readings for a vast number of them, as e.g. 行 'to go', 'to perform', which has the *go'on* reading *gyō* (as in 行政 *gyōsei* 'administration'), the *kan'on* reading *kō* (移行 *ikō* 'migration') and the *tō-sō'on* reading *an* (行火 *anka* 'bed warmer'). In compounds, readings from different strata may be combined. For instance, the *go'on* compound 食堂 *jikidō*, originally a dining room in a Buddhist temple, was 'redrafted' as *shokudō* for (any) 'dining hall', in which *shoku* is the *kan'on* reading for 食, whereas 堂 *dō* retains its *go'on* reading (Frellesvig 2010: 409).

[15] Especially the period known in Japanese history as 鎖国 *sakoku* 'locked country' (1633-1853).

in an intellectual tradition » contributed to the « successful » modernisation of the Japanese lexicon.

Note, however, that not all S-J words which were coined in this period were actually created *ex nihilo*. The word 社会 *shakai*, for instance, was already attested in Chinese with another meaning, namely 'meetings on the occasion of two season festivals, in honour of local deities'; however, it was in Japan that the word got its modern meaning, and was reimported as such in China (Masini 1993: 195). As to 文学 *bungaku*, this word is attested in Confucian *Analects* with the meaning 'literary acquirements', and was used by the Italian Jesuit Giulio Aleni in the XVII century with the modern meaning 'literature'; however, some considered it to be a word of Japanese coinage because it was arguably through the influence of Meiji Japanese that the term 文學 *wénxué* spread into Chinese (Masini 1993: 204-205; on the Chinese origin of 文学 *bungaku* see also Lippert 2001).

Lastly, it is worth pointing out that S-J words were not only coined to fill lexical gaps; there actually are several synonymic pairs in the language, as e.g. native 言葉 *kotoba* vs. S-J 言語 *gengo*, both for 'language', 受け取り *uketori* vs. 領収書 *ryōshūsho* 'receipt', 形 *katachi* vs. 形態 *keitai* 'form' (Shibatani 1990: 146). As expected, the S-J 'versions' are commonly found in the written language and are part of academic vocabulary, whereas the native equivalents are generally perceived as colloquial.

3.2 Sino-Korean

Chinese words and characters were in use in the Korean peninsula at least as early as the first century CE; Classical Chinese was the language of early writings in Korea, and remained the most prestigious written code until the XIX century (Lee & Ramsey 2011: 287). During the Middle Korean period (X-XVI centuries), especially since the establishment of a Chinese-style civil service examination in 958, the language was flooded with words of Chinese origin; at this stage, Chinese loanwords included many terms from everyday vocabulary, as e.g. 우산 *usan* (from 雨傘 *yǔ-sǎn*, lit. 'rain-umbrella'), which ousted the native term 슈릎 *syulup* (Lee & Ramsey 2011: 236). By the time of Late Middle Korean (XV century), some early Chinese borrowings were relatively soon assimilated and treated as native vocabulary (written in the native *Hangul* sillabary rather than in Chinese characters), as e.g. 차판 *chapan* 'food, side dishes' (from 茶飯 *chá-fàn*, lit. 'tea-food'; Lee & Ramsey 2011: 235). Note that the reading tradition of S-K is not as composite as that of S-J (or S-V; see §3.3), and is mostly based on Northern

Chinese (Chang'an) readings of the seventh and eighth century (as Japanese 漢音 *kan'on;* Sohn 1999: 103; Lee & Ramsey 2011: 69)[16].

In the Early Modern Korean period (XVII-XIX centuries), many more S-K words, also from the basic vocabulary, replaced native terms, as e.g. 산 *san* 'mountain' (SMC 山 *shān*) and 강 *gang* (SMC 江 *jiāng*). Moreover, during this time many 'Western' neologisms were introduced into Korean with the 'mediation' of Chinese, as e.g. 자명종 *jamyeongjong* 'alarm clock' (SMC 自鳴鐘 *zì-míng-zhōng*, lit. 'self-ringing-clock').

Just as for Japanese, revolutionary changes in the lexicon of Korean occurred with the Contemporary period, i.e. starting from the end of the XIX century, when the language absorbed thousands of new words to designate new 'Western' notions; the material for these neologisms came from Classical Chinese and/or Japanese, as said before. Examples include the above mentioned 철학 *cheolhag* 'philosophy', 사회 *sahoe* 'society', 국가 *gukga* 'country', 문학 *munhak* 'literature', as well as many others (Lee & Ramsey 2001: 302); below are two more examples, with their SMC; S-J and S-V equivalents:

(4) a. S-K 정치 *jeongchi* 'politics'
b. SMC 政治 *zhèngzhì*
c. S-J 政治 *seiji*
d. S-V *chính trị*

(5) a. S-K 경제 *gyeongje* 'economy'
b. SMC 經濟 *jīngjì*
c. S-J 経済 *keizai*
d. S-V *kinh tế*

Some S-J words were accepted in the Korean lexicon, but not in China, as e.g. 영화 *yeong-hwa* 'film', lit. 'project-image', from S-J 映画 *eiga* (compare SMC 電影 *diàn-yǐng*, lit. 'electric-shadow'; Sohn 1999: 105). Also, as mentioned earlier (§3), a number of S-K words were coined in Korea and were never 'exported' to China, Japan or Vietnam

Despite the fact that, as said before, the introduction of Chinese borrowings sometimes caused the loss of native vocabulary, we often have synonymic pairs, just as seen above for Japanese. This is fairly common in the basic vocabulary, as e.g. native 사람 *salam* 'person' or 달걀 *dalgyal* 'egg' vs. S-K 인간 *ingan* (SMC 人間 *rénjiān* 'human world', but S-J 人間 *ningen* 'human being, person') and 계란 *gyeran* (from 雞 *jī* 'chicken' and 卵 *luǎn* 'egg, ovum'). As expected, when a 'doublet' is present, SK words tend to sound more formal and abstract, prestigious and even pedantic, and are

16 However, some direct Chinese loans of the XV and XVI century have a different reading, arguably (somehow) reflecting the pronunciations of the time in Northern China (Lee & Ramsey 2011: 237-238).

found in writings and speech of a formal nature, whereas native words belong to the colloquial register; for instance, whereas the native term 일자리 *iljari* 'workplace' is used in reference to low-paid jobs, S-K 직장 *jigjang* (from 職 *zhí* 'profession' and 場 *chǎng* 'place'; S-J 職場 *shokuba*) « implies a white-collar job » (Sohn 1999: 105-106). As to bound vs. free status, usually S-K morphemes are bound, whereas their native equivalents are free, as e.g. 천 *cheon* (SMC 天 *tiān;* compare 천문학 *cheonmunhak* 'astronomy', SMC 天文學 *tiān-wén-xué*, lit. 'sky-text-study') vs. 하늘 *haneul*. However, if there is no equivalent native words, the S-K item is usually free, as e.g. 차 *cha* 'tea' (SMC 茶 *chá;* Sohn 1999:106). Whereas the former are analogous to neoclassical constituents, the latter look more like 'prestige' loanwords, just as French words in English.

3.3 Sino-Vietnamese

Vietnam was under Chinese rule for more than one thousand years (111 BCE - 939 CE), and Classical Chinese was used as a written language for high culture functions and a medium for education throughout this time and beyond, even up to the beginning of the XX century (Nguyen 1997). However, the number of words which were borrowed from Chinese was relatively small until the seventh century and, moreover, in this early period the items borrowed were mostly related to everyday life, as e.g. *đèn* 'lamp' (SMC 燈 *dēng*), *giường* 'bed' (SMC 床 *chuáng*), or *dao* 'knife' (SMC 刀 *dāo*; Bisang 2001: 195); these early loanwords are hardly recognised as such by language users today, and are often considered part of native vocabulary (Alves 2007). Borrowing increased significantly between the seventh and the tenth century; the pronunciation of the borrowings of this period was based on Tang dynasty Chinese (again, as Japanese 漢音 *kan'on* and S-K), and provided the basis for modern S-V pronunciation (Bisang 2001). When a morpheme was 'borrowed twice', i.e. with the older and with the Tang pronunciation, the tendency is for the older form to be used as a free morpheme, and for the newer form to be used as a bound morpheme, as e.g. *gan* (free) and *can* (bound) 'liver' (SMC 肝 *gān*; Nguyen 1997: 37-38).

After the end of Chinese rule, lexical borrowing did not end, but rather increased, especially since Chinese-style administration and Confucianism remained in place; the difference is that, starting from this time, words were generally introduced through written sources, and their pronunciation did not reflect the Chinese language of the time, but rather the 'frozen' (Tang dynasty) S-V pronunciation (Nguyen 1997, Bisang 2001). Then, just as in Chinese, Japanese and Korean, a huge number of S-V words were introduced in recent times as 'vocabulary of modernisation', many of which came from Japanese, rather than Chinese, as the above mentioned *xã hội* 'society', or *khoa học* 'science' (S-J 科学 *kagaku*; compare SMC 科學 *kēxué*, S-

K과학 *gwahak*); S-V words which entered the language during this latter period number in the thousands (Alves 2007: 354).

Vietnamese, however, differs from Japanese and Korean in one respect. Although S-V compounds generally follow the Chinese modifier-modified word order rule, in recent times there has been a tendency towards the 'nativisation' of the order of consituents in complex words. Thus, for instance, S-V *Âu Châu* 'Europe', from 歐洲 *Ōuzhōu* ('Europe-continent') is being replaced by *Châu Âu* (Alves 2007: 355); *viện trưởng* 'leader of an institute' ('institute-leader', SMC 院長 *yuànzhǎng*) is also attested as *trưởng viện*, with the native morpheme order (Bisang 2001: 199). Note that this variation in constituent order is not always neutral. For instance, words with the S-V morpheme *tính* 'nature' (SMC 性 *xìng*) tend to be understood as nouns if they follow the native order, and as modifiers if they have the Chinese order: compare *tính ác* 'badness, evil' and *ác tính* 'of a bad/evil character' (Bisang 2001: 200).

3.4 Summary

In short, morphemes of Chinese origin in Japanese, Korean and Vietnamese are a very close equivalent of SAE neoclassical constituents for the following reasons:

a. they are drawn from a written classical language, i.e. Classical Chinese, the role of which is comparable to that of Latin and Greek in Europe, i.e. a written *lingua franca* used for high-culture functions[17]
b. they are usually bound, and often have a free native equivalent
c. the words formed with them sound more formal, abstract than their native equivalents, and they are the normal choice for the creation of technical and high-flown vocabulary, but they may be used also in ordinary words
d. they are used independently in the creation of words in all of the languages considered, and the neologisms containing them may circulate freely
e. they follow word formation rules of the source language, regardless of any difference with 'native' word formation rules (but cf. the case of S–V vocabulary)
f. their phonological form did not undergo the sound changes expected in Chinese, i.e. their development is independent from the source language.

Also, as is apparent from the examples given before, the diffusion of S-J (and Chinese) neologisms in SMC, Korean and Vietnamese led to a convergence in the area of learned vocabulary analogous to that of neoclassical

[17] Here we are simplifying things a bit, as Latin could also be used as a spoken language by European *élites*, at least until the seventeenth century.

compounds in SAE languages; compare e.g. English *psychology*, Italian *psicologia*, Polish *psychologia*, Latvian *psiholoģija* and Hungarian *pszichológia*.

4. STRATA IN THE MORPHOLOGY OF CHINESE DIALECTS

The term 'Chinese' is generally used in English, as well as in many other European languages (French *chinois*, German *Chinesisch*, etc.), to refer to SMC, i.e. the national language of China, also referred to as 'Mandarin' or 'Mandarin Chinese'. However, 'Chinese' is also sometimes used as a synonym of 'Sinitic', and languages other than the national standard are referred to as '(Chinese) dialects'. The use of the term 'dialect' here may be misleading: in the Sinitic context, when we talk about dialects we are not dealing with dialect *of* SMC, but rather of varieties *related to* SMC, just as e.g. Dutch is related to English, or French is related to Portuguese; each Chinese 'dialect', thus, may be viewed as a distinct object for comparison (Norman 2003: 72)[18]. The differences among dialects are particularly evident in the domains of phonology and lexicon, but they also involve morphology and syntax, as may be expected for any language family (see Yue 1993, 2003 for an overview).

We nevertheless believe that the term 'dialect' is appropriate here because of the fact that no Sinitic variety other than SMC (and, in a sense, Cantonese) has the status of a standard language and, as such, the relation between Standard and dialects is asymmetrical, so to say. This is not true only in today's China. Since long before the emergence of SMC, the 'official' written language of China was the above mentioned Classical Chinese, which could be understood by the cultivated *élites*, whereas dialect writing has always been marginal in the history of China (again, with the exception of some form of Cantonese; see Snow 2004, Wu 2005). Another written code, 'Vernacular Chinese' (白話 *báihuà*) was used for popular literary genres at least since the Tang dynasty (618-907); it was mostly based on specific prestige dialects (especially Jiang-Huai and Northern Mandarin; see Coblin 2000). This meant that all the Chinese who spoke a different dialect had to write in a language which they did not speak (Chen 1999: 82-83); even today, for instance, many schoolgirls and schoolboys in Hong Kong learn written standard Chinese in school, but the spoken language they are exposed to is often (Hong Kong) Cantonese.

This kind of diglossic asymmetry between a standard language (be it Classical Chinese, Vernacular Chinese or SMC) and Chinese dialects led to stratification at all levels of language. At the phonological level, there often

---

[18] Needless to say, SMC is also subject to significant diatopic variation, just as any other language; the local variants of SMC are usually termed 地方普通話 *dìfāng Pǔtōnghuà*, lit. 'regional *Putonghua* (common language)' (Chen 1999: 42).

is a distinction between the 'colloquial reading(s)' (白讀 *báidú*) and the 'literary reading(s)' (文讀 *wéndú*) of characters. The former represent the normal evolution of the morpheme in the dialect at issue, whereas the latter reflect the pronunciation of (some form of) Mandarin; we may interpret this as a distinction between a 'native' and a 'borrowed' stratum (Lien 2004). In the lexicon of a dialect such as Taiwanese Southern Min (henceforth: TSM), there are as many as three different chronological strata: the Han dynasty (206 BCE - 202 CE) and the Nanbeichao (420-581) strata constitute the colloquial stratum, whereas the Tang (Chang'an, VII-VIII century) stratum is the basis for literary readings (Lien 2004: 310). Thus, for instance, the character 遠 'far' (SMC *yuán*) has the colloquial reading *hng*^{7} and the literary reading *oan*^{2}; the character 命 'life' (SMC *mìng*) has *mia*^{7} and *bing*^{7}; for 老 'old' (SMC *lǎo*), we have the colloquial reading *lau*^{7}, the 'semi-literary' reading *lau*^{2}, and the literary reading *lo*^{2} [19].

The stratification of the TSM lexicon is reflected also in word formation. Literary *lo*^{2} is used as a bound morph, as in 老翁 *lo*^{2}-*ong*^{1} 'old man', whereas both *lau*^{7} and *lau*^{2} are free. Moreover, *lau*^{7} and *lau*^{2} are further distinguished by their semantics: *lau*^{7}, the colloquial variant, has retained its original 'old' meaning, but semi-literary *lau*^{2} is used either as a verb, meaning 'be good at' and 'rot, spoil', or as an evaluative prefix, as in 老細 *lau*^{2}-*soe*^{7} 'pal', used when addressing a younger friend. Thus, each of the reflexes of 老 'old' has a different meaning and function (Lien 2004: 314-316). What is interesting, in our perspective, is that the literary variant of the morpheme is bound, whereas the colloquial one is free, and the semi-literary variant developed a bound use: thus, we see a connection between 'classic' and bound, on the one hand, and between 'modern / native' and 'free', on the other hand, just as for neoclassical compounding in SAE languages.

This connection is even clearer when the literary and the colloquial versions of a morpheme are not cognates: such pairs or groups of morphs are *paradigmatically* related, rather than etymologically related. A nice example is 人 *jin*^{5} vs. 儂 *lang*^{5}, i.e., respectively, the literary (i.e. imported from the Tang *koinè*) and the colloquial morphs for 'man' (e.g. 證人 *ching*^{3}-*jin*^{5} 'prove-person, witness', 播音儂 *po*^{3}-*im*^{1}-*lang*^{5} 'broadcast-person, broadcaster'; Lien 2004: 325-327), which have different distribution and functions. 人 *jin*^{5} may be attached only to literary items, and clearly prefers monosyllabic stems, whereas 儂 *lang*^{5} combines both with colloquial and with literary stems, and shows some preference for disyllabic stems; given the tendency towards disyllabification both of SMC and of TSM (see above, §2.1-2.2), monosyllabic stems are associated with the 'older' lexicon, whereas disyllabic stems are associated with the modern lexicon.

Other pairs of literary and colloquial morphs in the (broadly defined) agentive domain are 師 *su*^{1} vs. 師傅 *sai*^{1}-*hu*^{7} / 師 *sai*^{1}, both for 'master', and

[19] In the transcription of TSM, superscript numbers indicate tone categories.

家 *ka*^{1} vs. *ke*^{1}, both for 'person' (cf. SMC 家 *jiā* discussed above, §2.2.1). From the semantic point of view, nouns with the literary suffix 師 *su*^{1} usually refer to 'people engaging in mental work who enjoy higher social status' (e.g. 律師 *lut*^{8}*-su*^{1} 'law-master, lawyer'), whereas nouns with the colloquial 師(傅) *sai*^{1}(*-hu*^{7}) refer to 'someone who engages in lowly and menial work' (塗水師傅 *tho*^{5}*-chui*^{2}*-sai*^{1}*hu*^{7} 'mud-water-master, bricklayer'; Lien 2004: 331-332). Also, they select different bases, as only 師(傅) *sai*^{1}(*-hu*^{7}) attaches to verb-object compounds (拍刀師傅 *phah*^{4}*-to*^{1}*-sai*^{1}*hu*^{7} 'strike-knife-master, cutler'), and requires a disyllabic base; again, the literary variant prefers monosyllabic bases, although this is not as strict as for 人 *jin*^{5} discussed above (cf. 藥劑師 *ioh*^{8}*-che*^{3}*-su*^{1} 'medicine-master, pharmacist'). Lastly, the 'abbreviated' colloquial form 師 *sai*^{1} may be used « as an affective suffix to suggest intimate rapport and camaraderie with the addressee » (for a skilled worker: 阿善師 *A*^{1}*-Sien*^{7}*-sai*^{1} 'A-Sien'; Lien 2004: 333), a function which is never associated with 師 *su*^{1}. As to 家 *ka*^{1} and *ke*^{1}, the former is used in « cultivated words » (企業家 *khi*^{3}*giap*^{8}*-ka*^{1} 'enterprise-person, enterpreneur'), and is a pure agentive suffix[20], whereas the latter is used in « popular and demotic words », and has undergone extension of meaning; compare e.g. 管家 *koan*^{2}*-ke*^{1} 'manage-person, head-servant, steward', 頭家 *thau*^{5}*-ke*^{1} 'head-person, boss, master, husband' and 少年家 *siau*^{5}*lien*^{5}*-ke*^{1} 'youth-person, youth' (Lien 2004: 333-334).

Lien (2004: 334-335) summarises the characteristics of colloquial and literary formatives in TSM as such:

a. colloquial morphs tend to be free, whereas literary words « are bound and often found in compounds or frozen expressions »
b. colloquial morphs tend to be found in basic or popular words, whereas literary morphs tend to be found in technical or cultivated vocabulary
c. colloquial words often denote concrete entities, whereas literary words commonly denote abstract entities.

This characterisation is very close to that of neoclassical compounds in SAE languages (except, perhaps, for c., since in scientific domains neoclassical compounds are often used to name concrete entities)[21]. Unfortunately, we do not have such detailed analyses for other Chinese dialects and, hence, we do not know whether this functional specialisation for literary and colloquial forms is found in other dialects. Some hints in this direction come from other Southern Min varieties. For instance, in the Leizhou dialect (Guangdong province), neologisms from SMC such as 土改 *t'u*^{42}*kai*^{42} 'land reform' and 生產 *seŋ*^{33}*saŋ*^{42} 'produce, production' tend to use literary readings, whereas in ordinary words colloquial readings are usually chosen (e.g. 新婦 *sem*^{33}*pu*^{33} 'bride'; Cai 1993: 50[22]). The same is said to be true for

20 Compare our treatment of SMC 家 *jiā* (§ 2.2.1).

21 We would like to thank an anonymous reviewer for pointing this out to us.

22 Here superscript numbers indicate tone height (on a 1 to 5 scale).

Zhangzhou (Fujian province). Moreover, in this dialect there are also some pairs akin to those seen above for TSM: for instance, 指 'point (to/out)' is *tsi*^{3} in compound words, in which it has the derived meaning 'to lead', whereas the colloquial reading *ki*^{3} is used when this verb conveys its 'proper' meaning; 轉 'to turn' is *tsuan*^{3}, again, in compound words, whereas the colloquial variant *tuī*^{3} is used only in combination with monosyllabic verbs, and the colloquial variant *tun*^{3} is used only in the meaning 'come back', showing semantic / functional specialisation (Zhang 1999). More data is needed, however, to properly assess the extent of this phenomenon across Sinitic.

5. CONCLUSIONS

In this article we proposed an extension of the notion of neoclassical compounding to the East Asian linguistic world. We highlighted that there seem to be two kinds of compounds in Modern Chinese which share most defining features with SAE neoclassical compounds.

Also, we showed how Chinese bound lexical morphemes in Japanese, Korean and Vietnamese have the same role of neoclassic constituents in SAE languages. The key features of these word formatives are being bound (often *vs.* a native free equivalent), reflecting a frozen pronunciation from some earlier historical stage of Chinese, and being somehow perceived as more formal and/or 'abstract'. Also, they tend to circulate freely among Modern Chinese, Japanese, Korean and Vietnamese, with different pronunciations, analogously to neoclassical compounds in SAE languages.

Lastly, we discussed the notion of strata in the lexica of Chinese dialects, and we showed that the partition between 'literary' and 'colloquial' word formatives in Taiwanese Southern Min (and, possibly, in other Min dialects) partly overlaps with that between bound / formal and free / colloquial, similarly to what was shown for Modern Chinese, Japanese, Korean and Vietnamese. Further research is needed to find out whether the tendencies outlined here for Taiwanese Southern Min are found also in other Chinese dialects.

REFERENCES

ALVES M. J. (2007). Sino-Vietnamese grammatical borrowing: an overview. In: Y. Matras, J. Sakel (eds), *Grammatical borrowing in cross-linguistic perspective*. Berlin-New York: Mouton de Gruyter, 343-361.

AMES R. T. (1998). East Asian philosophy. In: E. Craig (ed), *Routledge encyclopedia of philosophy*. London: Routledge. Retrieved June 25, 2013, from http://www.rep.routledge.com/article/G218.

BASCIANO B., CECCAGNO A. (2009). The Chinese language and some notions from Western linguistics. *Lingue e Linguaggio* 8.1, 105-136.

BAUER L. (1998). Is there a class of neoclassical compounds, and if so is it productive? *Linguistics* 36, 403-22.

BAUER L. (2006). Compound. In: K. Brown (ed), *Encyclopedia of Language and Linguistics*, volume 2. Oxford: Elsevier, 719-726.

BISANG W. (2001). Syntax / morphology asymmetry in Vietnamese – a consequence of contact by writing between Vietnamese, Chinese and Standard Average European Languages. In: B. Igla, T. Stolz (eds), *Was ich noch sagen wollte... A multilingual Festschrift for Norbert Boretzky*. Berlin: Akademieverlag, 189-201.

CAI Y. (1993). 海康方言志 [The Haikang dialect]. Guangzhou: Zhongshan Daxue Chubanshe.

CECCAGNO A., BASCIANO B. (2009). *Shuobuchulai: la formazione delle parole in cinese*. Bologna: Serendipità.

CHEN P. (1999). *Modern Chinese. History and sociolinguistics*. Cambridge: Cambridge University Press.

COBLIN W. S. (2000). A brief history of Mandarin. *Journal of the American Oriental Society* 120, 537-552.

DAI J.X. (1990). Historical Morphologization of Syntactic Words: Evidence from Chinese Derived Verbs. *Diachronica* 7.1, 9-46.

DAI J.X. (1992). *Chinese Morphology and Its Interface With the Syntax*. Ph.D. dissertation, Ohio State University.

DAL G., AMIOT D. (2007). La composition néoclassique en français et ordre des constituants. In: D. Amiot (éd.). *La composition dans les langues*. Arras: Artois Presses Université, 89-113.

DEFRANCIS J. (1984). *The Chinese Language: Facts and Fantasy*. Honolulu: University of Hawaii Press.

FENG S. (1998). Prosodic structure and compound words in Classical Chinese. In: J. Packard (ed), *New approaches to Chinese Word Formation*. Berlin and New York: Mouton de Gruyter, 197-260.

FRELLESVIG B. (2010). *A history of the Japanese language*. Cambridge: Cambridge University Press.

IACOBINI C. (2010). Elementi formativi. In: R. Simone (ed), *Enciclopedia dell'italiano*. Roma: Treccani.

KAGEYAMA T. (2009). Isolate: Japanese. In: R. Lieber, P. Štekauer (eds), *The Oxford handbook of compounding*. Oxford: Oxford University Press, 512-526.

LEE K.-M., RAMSEY S.R. (2011). *A history of the Korean language*. Cambridge: Cambridge University Press.

LIEBER R., ŠTEKAUER P. (2009). Introduction: status and definition of compounding. In: R. Lieber, P. Štekauer (eds), *The Oxford handbook of compounding*. Oxford: Oxford University Press, 3-18.

LIEN C. (2004). Competing Morphological Changes in Taiwanese Southern Min. In: H. Chappell (ed.), *Chinese grammar. Synchronic and diachronic perspectives*. Oxford: Oxford University Press, 309-339.

LIN H. (2001). *A grammar of Mandarin Chinese.* Munich: Lincom Europa.

LIPPERT W. (2001). Language in the modernization process: the integration of Western concepts and terms into Chinese and Japanese in the nineteenth century. In: M. Lackner, I. Amelung, J. Kurtz (eds), *New terms for new ideas. Western knowledge and lexical change in Late Imperial China*. Leiden-Boston-Köln: Brill, 57-66.

MASINI F. (1993). *The formation of Modern Chinese lexicon and its evolution towards a national language: the period from 1840 to 1898.* Berkeley: Project on Linguistic Analysis, University of California.

NAUMANN B., VOGEL P.M. (2000). Derivation. In: G. Booij, C. Lehmann, J. Mugdan, (eds), *Morphologie-Morphology*. Berlin-New York: Mouton de Gruyter, 929-943.

NGUYEN D.-H. (1997). *Vietnamese*. Amsterdam-Philadelphia: John Benjamins.

NORMAN J. (2003). The Chinese dialects: phonology. In: G. Thurgood, R. J. LaPolla (eds), *The Sino-Tibetan languages*. London: Routledge, 72-83.

OLSEN S. (2000). Composition. In: G. Booij, C. Lehmann, J. Mugdan (eds), *Morphologie-Morphology*. Berlin-New York: Mouton de Gruyter, 897-916.

PACKARD J. (1998). Introduction. In: J. Packard (ed), *New Approaches to Chinese Word Formation*. Berlin and New York: Mouton de Gruyter, 1-34.

PACKARD J. (2000). *The morphology of Chinese: A linguistic and cognitive approach.* Cambridge: Cambridge University Press.

PIRANI L. (2008). Bound Roots in Mandarin Chinese and Comparison With European "Semi-Words". In: M. K. M. Chan, H. Kang (eds), *Proceedings of the 20th North American Conference on Chinese Linguistics* (NACCL-20), volume 1. Columbus, Ohio: The Ohio State University, 261-277.

PULLEYBLANK E.G. (1995). *Outline of Classical Chinese grammar.* Vancouver: University of British Columbia Press.

SCALISE S. (1984). *Generative morphology*. Dordrecht: Foris.

SCALISE S., VOGEL I. (2009). Why compounding? In: S. Scalise, I. Volgel (eds), *Cross-Disciplinary Issues in Compounding.* Amsterdam and Philadelphia: John Benjamins Publishing, 1-18.

SHI Y. (2002). *The Estabilishment of Modern Chinese Grammar. The Formation of the Resultative Construction and Its Effects*. Amsterdam and Philadelphia: John Benjamins.

SHIBATANI M. (1990). *The languages of Japan.* Cambridge: Cambridge University Press.

SNOW D.B. (2004). *Cantonese as written language: the growth of a written Chinese vernacular.* Hong Kong: Hong Kong University Press.

SOHN H.-M. (1999). *The Korean language.* Cambridge: Cambridge University Press.

SPROAT R., SHIH C. (1996). A Corpus-Based Analysis of Mandarin Nominal Root Compounds. *Journal of East Asian Linguistics* 5, 49-71.

WANG L. (1980[1957]). 汉语史稿 [Draft history of the Chinese Language]. Beijing: Zhonghua Shuju.

WU Y. (2005). *A synchronic and diachronic study of the Chinese Xiang dialects*. Berlin-New York: Mouton de Gruyter.

XING J.Z. (2006). *Teaching and Learning Chinese As a Foreign Language: A Pedagogical Grammar*. Hong Kong: Hong Kong University Press.

YANG X. (2003). 汉语语素论 [A theory of Chinese morphemes]. Nanjing: Nanjing Daxue Chubanshe.

YIP P. (2000). *The Chinese lexicon. A comprehensive survey*. London: Routledge.

YUE A. (1993). *Comparative Chinese dialectal grammar. Handbook for investigators.* Paris: EHESS-CRLAO.

YUE A. (2003). Chinese dialects: grammar. In: G. Thurgood, R. J. LaPolla (eds), *The Sino-Tibetan languages*. London: Routledge, 84-125.

ZHANG Q. (1999). 漳州市志 - 方言 [Zhangzhou City Gazetteer - Dialect]. Beijing: Zhongguo Shehui Kexue Chubanshe.