The 生 *shēng/sheng* Complex Words in Chinese between Morphology and Semantics

Emanuele BANFI   Giorgio Francesco ARCODIA
Università degli studi di Milano-Bicocca
emanuele.banfi@unimib.it   giorgio.arcodia@unimib.it

Abstract

In the long history of the Chinese language we may find a number of complex words with the morph 生 *shēng*, meaning "to be / to be by nature", as one of their constituents. For some instances of 生 *shēng* (also in the neutral tone) in such words, this morph may be regarded as a suffix-like formative, derivating nouns (such as e.g. in 先生 *xiānsheng* "sir, mr.", lit. "born earlier", 學生 *xuésheng* "student", "study + scholar", 醫生 *yīshēng* "physician", "medicine + professional"), verbs (出生 *chūshēng* "to be born", "go out + be born") and adverbs (好生 *hǎoshēng* "quite, exceedingly", "quite + very"). There are also instances, albeit rare and non-productive, of 生 *shēng* being used as a prefix-like element, specifically as an intensifier preceding verbs of perception (as in e.g. 生疼 *shēng téng* "very painful").

Here we shall first analyze and describe, basing on historical data, the nature and the importance of complex words with 生 *shēng*; in a typological-comparative perspective, we shall then turn to the examination of "parallel" complex words of some Indo-European languages, where one of the consitutents is a grammatical / lexical form which may be traced back to one of the two I.E. roots $*h_1es-$ / $*s-$ and $*b^hweh_2-$, whose original meaning is "to be / to be by nature", just as the Chinese morph under examination here.

Excluding, for obvious reasons, genealogical kinship between Sino-Tibetan and Indo-European languages, the fact that they have parallel morphological phenomena hints to the possibility of the existence of "semantic universals" operating at a cognitive level.

0. Introduction

Modern Chinese, although it has often been regarded as a typical isolating language, indeed has a number of (proto-?)morphological phenomena, mainly (but not limited to) complex words made of lexical morphemes. In this paper we shall analyse, both from the diachronic and from the synchronic perspective, complex words where one constituent is the morph(eme) 生 *shēng* / *sheng* "to be (born), to be by nature", which has had several historically attested semantic values as a constituent of polymorphemic words (and, to a lesser extent, as an affix / affixoid).

In the second part of our paper we shall turn to the analysis of "parallel" word-forms attested in some Indo-European languages; those words have as one of their constituents a morph with a (nearly) analogous meaning as Ch. 生 *shēng*, which, although having different phonological (and, needless to say, ortographic) shapes for different languages, may be in all cases traced back to one of the I.E. roots $*h_1es-$ / $*s-$ // $*b^hweh_2-$. From the typological point of view, the interest in such comparison lies in the fact, that Chinese and I.E. languages in the field of word formation, seem to have a fundamental difference: whereas in the former language the morphematic structure of complex words is, normally, completely transparent (even for many older word-forms), mainly because of the logographic nature of the Chinese writing system, in the I.E. languages, as it is known (see e.g. Hopper 1990), it is quite common for morpheme boundaries to become opaque in the historical development of a word, thus making "originally" compound or derivate words look like simple words in the syncrhonic stage; we shall provide several examples for this in the following sections.

As we can definitely n exclude a genealogical (or contact) explanation for the existence of parallel structures in complex words from Chinese and I.E. languages, such similarity suggests the possibility of "meaning universals" or, rather, common cognitive strategies operating in the

speakers' mind when it comes to "recruiting" morphemes for complex word building (also, grammaticalization: see Heine & Kuteva 2002).

1. Theoretical background. What is a complex word?

As mentioned in the introduction, our research is centred on the analysis of *complex words*, and we shall now explain what we include under this label.

In the literature on word formation, a *complex word* is a word-form[1] containing either one lexical morpheme and a derivational affix, or more than one lexical morpheme / *subword* (Bauer 2006[2]), and, optionally, a derivational affix, including *derivatives* and *compound words*; this notion, as it is known, is a descriptive label which is synchronic in nature, i.e. applicable to a word-form caught at a certain stage of its development. Here we shall adopt a rather different approach: for the purposes of our analysis, we shall apply the notion of complex word to the diachrony of the language: what may be regarded as a simple word now might be the evolution of a complex word, made up of more than one lexical / derivational morpheme, as the following Latin adjectives:

(1) *prō-b-us* "honest" < *prō-bh-os "(one) who is, by nature, before, in a prominent position"
super-b-us "arrogant" < *super-bh-os "(one) who is by nature above".

The Latin words have the appearance of simple words, containing only one lexical morpheme. However, by enclosing the "b" between hyphens we wanted to highlight the presence of a derivative form of the I.E. root *b^{h}weh$_2$- "to be, to be by nature" (see **0.**), already "blended" in the root in the Latin adjectives in (1), the reconstructed "pre-Latin" forms of which are those on the right side; by comparing the Latin forms and the reconstructed ones, we may notice the structural opacity of the former ones. In the Chinese lexicon, instead, owing to the peculiar nature of the writing system and of the morphology of the language there is an almost absolute transparency in word formation, differently from what typically happens in the Indo-European languages (see Belardi 1990). Let us examine a Chinese example:

2) 渡 *dù* "to cross (a sheet of water)" + 船 *chuán* "boat" = 渡船 *dùchuán* "ferry"

Needless to say, the meaning of a Chinese complex word is not always predictable: when a word has a "deep" level of lexicalization[3] (such as e.g. 壓根 "completely", lit. "press+root"; example in Packard 2000:222), it may acquire a "non-transparent" meaning; nevertheless, the morpheme structure of the word will still be visible.

Having made clear what we mean by "complex words", we shall turn now a brief discussion of some issues concerning the application of the notion of "morphology" to the Chinese language.

2. Some remarks on multi-morphemic words in the history of Chinese

---

[1] On the notion of "word", see Ramat (1990 and 2005) and the recent work by Dixon & Aikhenvald (2004).

[2] The classical definition of the compound, as it is known, regards only words as possible constituents for it (see e.g. Fabb 1998); Haspelmath (2002) and prefers the label "lexeme" for the forms which take part in the creation of compound words. The definition of what consitutes a "lexeme" or a "lexical morpheme" as a complex word constituent may well depend on the language one considers (especially, depending on the morphological type, such as agglutinating *vs.* inflectional, etc.), as it was pointed out by Bauer (2006:719): "[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".

[3] See Packard (2000:216-225), where a categorization of five "lexicalization types" is proposed for Chinese.

As we already mentioned in the introductory paragraph, Modern Chinese *has* morphology, and the number of works inquiring into word formation phenomena in this language is, nowadays, vast[4].

What is, possibly, not yet widespread known is that Chinese has always had multi-morphemic words, even in the old stages of its development[5]; in (3) we provide an example of a sub-sillabic suffix present in Old Chinese, the*-s* suffix, usually deriving nouns from adjectives or verbs (Baxter & Sagart 1998:54):

(3) 傳 *chuán* "transmit" < *driwen* < *drjon →
傳 *zhuàn* "a record" < *drjwenH* < *drjon-s (*"something transmitted")

This kind of sub-sillabic affixes, however, do not survive up to the modern language (and, in fact, they had already started deteriorating in Old Chinese). The model for morphological phenomena which "won the challenge of history", rather, is the agglutination of morphemes (mostly, monosyllabic), typically bearing lexical meaning, as we can see in this Old Chinese word (from Pulleyblank 2000:1736):

(4) 百 *băi* "100" + 姓 *xìng* "surname" = 百姓 *băixìng* "commmon people".

The spread of this typology of word-formation processes lead a linguist such as Lin (2001:62) to define Chinese as "a language of compounded words"[6].

In the huge number of complex words in Modern Chinese we may notice a few instances where one of the consituent morphemes (most frequently, the rightmost one) seems to have become grammaticalized in a derivational morpheme, showing no formal difference with the corresponding lexical morpheme (if present), but having apparently undergone some sort of semantic shift towards "bleaching" of meaning (Hopper & Traugott 2003): this seems to be the case, e.g., for 吧 *bā* "bar" (a loanword from English), the meaning of which has "faded", deriving words such as 棋吧 *qíbá* "chess club", where it means something like "meeting place". However, at present there is no agreement in the literature on the topic on whether such cases should be regarded as instances of grammaticalization[7]: here we shall not discuss the issue any further and we shall just stress the fact that the notions of "affix", "derivation" and "compounding" are controversial, and here we shall not always be using them advisedly.

3. A case study: complex word forms with 生*shēng/sheng*

Having made explicit the theoretical and empirical basis of our research, let us now turn to the analysis of our historical data on complex words containing the morpheme 生 *shēng / sheng*[8].

---

[4] To name but a few, Packard (1998 and 2000), Pan, Ye & Han (2004) and Dong (2002, 2004). On the possibility that a "typological drift" is under way in Chinese word formation, see Banfi (2005) and Banfi & Arcodia (forthcoming).

[5] "To speak of morphological processes in Old Chinese may surprise some readers, for there is a widespread belief that early Chinese had only impoverished morphology if it had any at all." (Baxter & Sagart 1998:35).

[6] According to the figures in Shi (2002:70-71), around 80% of the words in the Chinese lexicon are disyllabic, and the vast majority of those are, also, bimorphemic (on the genesis of disyllabic word-forms and on their supposed compound nature, see Feng 1998 and Arcodia 2007).

[7] For a summary of the issue of affixation in the Chinese linguistic bibliography, see Pan, Ye & Han 2004).

[8] Data on complex words containing the morpheme 生 *shēng/sheng* have been drawn mainly, for Chinese, from the dictionaries 漢語大詞典 *Hànyŭ Dàcídiǎn* and 現代漢語詞典 *Xiàndai Hanyŭ Cídiǎn*, from the reverse lexicon 現代漢語逆序詞典 *Xiàndai Hanyŭ Nìxù Cídiǎn* and from the corpora of the *Academia Sinica* (古漢語語料庫 *Gŭhànyŭ yŭliàokù* for Classical Chinese, 近代漢語標記語料庫 *Jìndài*

In the modern language, the character 生 *shēng/sheng* is a morph, lexical and bound (although it is worth remembering that, for Chinese, the bound-free status may depend on the micro-syntactic context, rather than being a stable feature of the morpheme itself; Zhou & Marslen-Wilson 1994); in Old Chinese, it was a free morph, a lexeme which could be used both as an independent word form and as a complex word constituent (*subword*; see **1.**). The 漢語大詞典 *Hànyǔ Dàcídiǎn* dictionary lists almost 50 semantic values for 生 *shēng/sheng*; below are the main semantic areas:

   i) to give birth to, to grow, life, all one's life, living, to cause, to burn fuel;
   ii) unripe, raw, rough, unknown / unfamiliar, stiff / unnatural, intensifying prefix for perception verbs / adjectives;
   iii) student / pupil, intellectual, male role in Chinese opera, suffix for professionals;
   iv) suffix for some adverbs.

In the following sentences (5 and 6), we may see two instances from the Classical language of a "shifted" meaning being assigned to 生 *shēng / sheng* (semantic area "i"), in its usage as a free word:

(5) "to have by nature, inborn" (天生、生來 *tiānshēng* (sky / nature + *sheng*), *shēnglái* (*sheng* + come); attested at least since the 3rd century BCE)

   石生而堅, 芷生而芳

   shí   shēng ér   jiān   zhǐ   shēng ér   fāng
   stone innate CONJ hard Dahurian angelica innate CONJ fragrant

   "Stone is hard by nature and Dahurian angelica is fragrant by nature" (文字 上德 *Wénzi Shàngdé*)

(6) "occupation, living" (生業 *shēngyè* (*sheng* + occupation), attested at least since the first century BCE)

   勃以織薄曲為業

   Bó   yǐ   zhī   bóqǔ wèi shēng
   Bo PREP weave reeds as living

   "Bo's living comes from bamboo trays weaving" (史記 絳侯周勃世家 *Shǐjì*, "Records of
   the Grand Historian", *Jiàng hóu Zhōu Bó shìjiā*)

As we mentioned in the introduction, in Modern Chinese □ *shēng / sheng* chiefly acts as a complex word constituent, a bound morph; let us see some examples of modern words containing it. The morpheme 生*shēng/sheng* may be used as a suffix(oid), bearing the meaning "student, intellectual" (7) , also forming nouns for occupations (8; meaning area "iii"):

(7)	大學生	初中生	儒生
	dàxuéshēng	*chūzhōngshēng*	*rúshēng*
	"university student"	"jun. sec. school student"	"Confucian scholar"

---

(8)   醫生   陰陽生
       yīshēng   *yīnyángshēng*
       "doctor"   "astrologer"

We also have instances of 生 *shēng/sheng* as a right-hand constituent in complex verbs (meaning area i):

(9) 再生   派生   增生
     zàishēng   *pàishēng*   *zēngshēng*
     "be reborn, regenerate"   "derive"   "proliferate,
multiply"

We have a few words where 生 *shēng/sheng* is used as a suffix-like formative for some adverbs (meaning area iv):

(10)   好生   怎生
            hǎoshēng   *zěnshēng*
            "quite, exceedingly"   "how, in which way"

The morph 生 *shēng/sheng* may also be the left-hand constituent in complex nouns, carrying the meaning "unknown" or "raw, rough" (area "ii"):

(11) 生詞   生魚   生藥
      shēngcí   *shēngyú*   *shēngyào*
      "new word"   "raw fish"   "unprocessed
medicine"

Moreover, 生 *shēng/sheng* may be used as a (proto-derivational?) intensifying prefix(oid), only with a restricted set of perception verbs ans adjectives (area ii):

(12)   生怕   生疼
            shēngpà   *shēngténg*
            "(very) afraid"   "(very) painful"

Not all of these word-formation patterns, however, are equally worth investigating on, from the morphological point of view: those in 8, 10 and 12 seem to be no longer productive and, moreover, 生 *shēng/sheng* is used as an intensifier and as a constituent in adverbs (10 and 12) only in a handful of words, fully lexicalized. In what follows, we shall rather focus on complex words where 生 *shēng/sheng* seem to act as a suffix(oid) meaning "student" (see the examples in 7), as such pattern seem to be still productive nowadays; in the next section, we shall provide some historical data on this word-formation model.

3.1 Some diachronical considerations on complex words containg 生 shēng/sheng

In table 1 we summarized data on those complex words where 生 *shēng/sheng* conveys the meaning "student, intellectual" or "professional"[9], which, as we shall see in the following paragraph, fit well in the comparison with the Indo-European complex words under examination here.

---

[9] The main sources for the data were the dictionaries 漢語大詞典 *Hànyǔ Dàcídiǎn* and the 古漢語大詞典 *Gǔhànyǔ Dàcídiǎn*, as well as web corpora (see footnote 8).

The 生 *shēng/sheng* Complex Words in Chinese between Morphology and Semantics

Table I. Some complex words with 生 *shēng/sheng* and their attested meanings.

Period → Word ↓	4th – 3rd cent. BCE	漢 *Hàn* dinasty (206 BCE– 220 CE)	南北朝 *Nánběi Cháo* - Northern and Southern Dynasties (420-589)	唐 *Táng* dynasty (618-907)	宋 *Sòng* dynasty (960-1279)	明－清 *Míng* and *Qīng* dynasties (1368-1644 and 1644-1911)	From 1912 to the present day
先生 *xiānsheng*	The elders / father and elder brothers (論語 *Lùnyǔ*); Senior erudite / scholar (孟子 *Mèngzǐ*)。	Teacher, master (禮記 *Lǐjì*); Literate / scholar (史記 *Shǐjì*); Title for physiognomists, diviners, singers, medical practitioners and *Feng Shui* experts (史記 *Shǐjì*)。					Antiquate term for people doing secretariat and supervision jobs (三元裡人民抗英斗爭史料 *Sānyuánlǐ rénmín kàng Yīng dòuzhēng shǐliào*); Used for men of high position in some field or area (解放日報 *Jiěfàng Rìbào*).
學生 *xuéshēng / xuésheng*			Student in a school (后漢書 *Hòu Hàn shū*).				
生 *lǐnshēng*						Student supported by the state (醒世姻緣傳 *Xǐngshì yīnyuán zhuàn*)	
監生 *jiànshēng*						Student in the Imperial Academy (金瓶梅 *Jīnpíngméi*)	
門生 *ménshēng*			Second-generation disciple (后漢書 *Hòu Hàn shū*); Servant, follower (后漢書 *Hòu Hàn shū*).	Humble term for candidates for the Imperial exams (唐白居易 *Táng bái jū yì*). An official's aide who is engaged in medical studies (唐典太醫署 *Táng liùdiǎn tàiyī shǔ*).	Humble term for reccomended candidates for the Imperial exams (朝野類要升轉 *Cháoyě lèiyào shēngzhuǎn*). A person with a knowledge of pharmacy whose activity is healing (書事 *Shūshì*).	Assistant to a high officer / advisor of a powerful person (禮部尚書張公墓志銘 *Lǐbù shàngshū Zhānggōng mùzhì míng*)	
醫生 *yīshēng*							
新生 *xīnshēng*	Newborn (莊子 *Zhuāngzǐ*)	Beginner Confucian scholar (風俗通 皇霸三王 *Fēngsútōng Huángbà sānwáng*).					Newly enrolled student (光明日報 *Guāngmíng rìbào*)
書生 *shūshēng*			Confucian scholar, student, scholar (后漢書 *Hòu Hàn shū*).				

The path which led to the acquisition of the "student, intellectual" meaning for the morpheme 生 *shēng/sheng* is quite peculiar, and appears to be very different from the one of the I.E. roots which shall be discussed here. According to Jiang (2005 [1989]:90), the meaning of 生 *shēng/sheng* developed from "to be born, to grow" (meaning area i, **3.**) to "student" through compound shortening / truncation (簡縮 *jiǎnsuō*): "By 'shortening' we mean the expression of a meaning, which originally was proper of a compound or phrase, by a single morpheme (a Chinese character)" (my translation). By the process outlined above, the morph 生 *shēng/sheng* takes on the "new" meaning out of having been a constituent in a complex word which, originally, posessed that semantic value. Of course, this happens also in I.E. languages, albeit maybe not so often:

(13) **Homo**sexual → **homo**phobia

In the example above, a morpheme *homo-*[10] "one, same", becomes a "substitute" for "homosexual": we could analyse those instances in several different ways (treating them as analogical formations, as clipping, etc.), but such a discussion is far beyond the aims of the present paper.

Let us illustrate the steps by which the value of 生 *shēng/sheng* evolved from "to be born" into "student" in Jiang's account (2005 [1989]:90):

(a) In the 秦 *Qín* (221 - 226 BCE) and 漢 *Hàn* (206 BCE – 220 CE) dynasties, 生*shēng* is used as the short form for 先生 *xiānshēng*, lit. "born earlier", a respectful address →

(b) in the *Mencius* (孟子 *Mèngzǐ*, 3rd cent. BCE), 先生 *xiānshēng* is used in the sense of "senior erudite / scholar" (cfr. table I) →

(c) the short form 生 *shēng* later (at least since the 史記 *Shǐjì*, 1st cent. BCE) becomes equivalent to "(Confucian) scholar, student", no matter what the age was→

(d) this meaning is preserved until today: 畢業生 (degree+*shēng*) "graduate student", 招生 *zhāoshēng* (recruit+*shēng*) "recruit students".
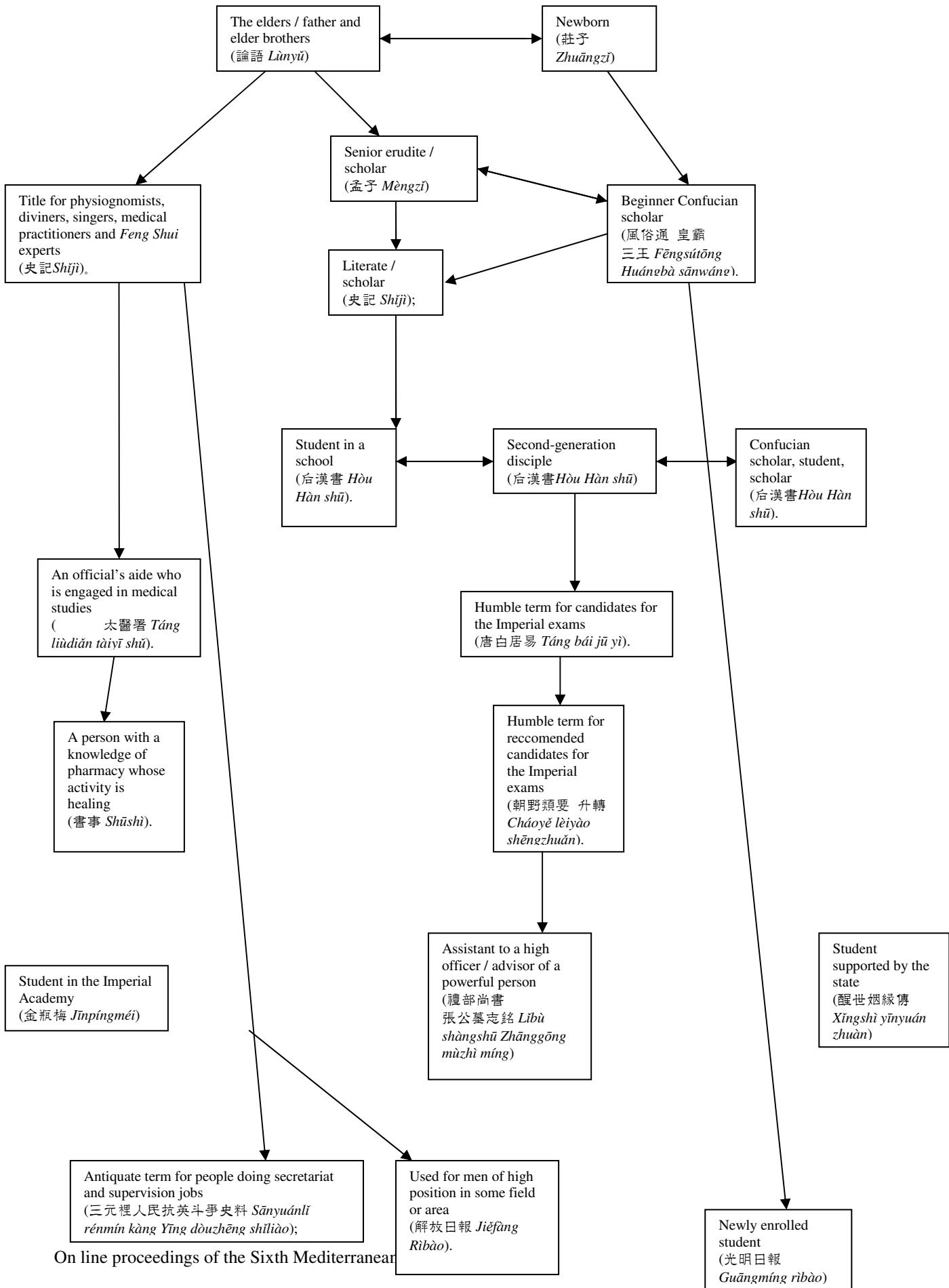
To sum up:

     i) **To be born > to be born earlier / be older**
    > ii) **person deserving respect**
    > iii) **scholar, medicine man, diviner**
    > iv) **student, doctor**

As the evolution of 生 *shēng/sheng* was driven by the complex words in which it appeared as a constituent, as we made clear above, it is interesting to analyse the history of those words. In table II, we provide a two-dimension representation of the semantic shifts which occurred in the 生 *shēng/sheng* complex words: along the vertical axis, historical relationships among different occurrences of a single word (or of "cognate" words) are shown, whereas the arrows (horizontal axis) make the semantic connections between different words at the synchronic level visible.

---

[10] From the Ancient Greek prefix ὁμο-, connected to the adjectival lexemes ὁμος / ὁμοῖος "same, similar" (see DELI 1999).

The 生 *shēng/sheng* Complex Words in Chinese between Morphology and Semantics

Table II. Some complex words with 生 *shēng/sheng* and their attested meanings.

The elders / father and elder brothers (論語 *Lùnyǔ*)

Newborn (莊子 *Zhuāngzǐ*)

Senior erudite / scholar (孟子 *Mèngzǐ*)

Beginner Confucian scholar (風俗通 皇霸 三王 *Fēngsútōng Huángbà sānwáng*).

Title for physiognomists, diviners, singers, medical practitioners and *Feng Shui* experts (史記 *Shǐjì*)。

Literate / scholar (史記 *Shǐjì*);

Student in a school (后漢書 *Hòu Hàn shū*).

Second-generation disciple (后漢書 *Hòu Hàn shū*)

Confucian scholar, student, scholar (后漢書 *Hòu Hàn shū*).

An official's aide who is engaged in medical studies (太醫署 *Táng liùdiǎn tàiyī shǔ*).

Humble term for candidates for the Imperial exams (唐白居易 *Táng bái jū yì*).

A person with a knowledge of pharmacy whose activity is healing (書事 *Shūshì*).

Humble term for reccomended candidates for the Imperial exams (朝野類要 升轉 *Cháoyě lèiyào shēngzhuǎn*).

Student in the Imperial Academy (金瓶梅 *Jīnpíngméi*)

Assistant to a high officer / advisor of a powerful person (禮部尚書 張公墓志銘 *Lǐbù shàngshū Zhānggōng mùzhì míng*)

Student supported by the state (醒世姻緣傳 *Xǐngshì yīnyuán zhuàn*)

Antiquate term for people doing secretariat and supervision jobs (三元裡人民抗英斗爭史料 *Sānyuánlǐ rénmín kàng Yīng dòuzhēng shǐliào*);

Used for men of high position in some field or area (解放日報 *Jiěfàng Rìbào*).

On line proceedings of the Sixth Mediterranean

Newly enrolled student (光明日報 *Guāngmíng rìbào*)

3.2 Some remarks on the Sino-Japanese complex words containg 生 shēng/sheng (sei)

To conclude our discussion on the evolution of 生 *shēng/sheng*, we shall provide some data on the "destiny" of the corresponding Sino-Japanese morpheme (pronounced *sei*).

According to the 岩波広辞苑 *Iwanami Kōjien* Japanese dictionary, 醫生 *yīshēng* "doctor" (see table I) in Japanese stands for "medicine student" (pron. *isei*). In the Japanese lexicon, there is a number of complex words where 生*sei* has the meaning of "student"; below are some examples:

(14)　　　　　　**院生**　　　　　　　　　　**越年生**

　　　　　　　　insei　　　　　　　　　　　*etsunensei*

　　　"postgraduate school student"　　　　"second year student"

(15)　　　　　　**寄宿生**　　　　　　　　　**寮生**
　　　　　　　　kishukusei　　　　　　　　*ryōsei*
　　　"boarding student"　　　　　　　"dormitory student, boarder"

(16)　　　　　　**給費生**　　　　　　　　　**自費生 / 私費生**

　　　　　　　　kyūhisei　　　　　　　　　*jihisei* / *shihisei*
　　　　"fellowship student"　　　　　　"self-supporting student"[11]

At a first glance, one might think that 生 *shēng* / *sei*, in Japanese, has undergone a grammaticalization process: this morph, in fact, appears in suffixal position in a series of complex words with a consistent meaning. Let us postpone the discussion on this to the concluding remarks; in the following paragraph, we shall introduce our Indo-European complex words data.

4. Parallel complex words in the Indo-European languages

In the preceding paragraph we illustrated the diachronical evolution of the Chinese morpheme 生 *shēng/sheng* "to be / to be by nature", including instances where, apparently, this morph has (partially?) grammaticalized into an affix(oid) (see examples 7, 8, 10, 12).

What has been observed in Chinese complex word formation, exemplified by the morpheme 生 *shēng/sheng*, lead us to make a comparison with similar developments in the I.E. languages, where there is a relevant number of word formation patterns, including a constituent semantically connected to the idea of "to be / to be by nature" (see par. **1.** for our definition of "complex word").

In I.E. languages, the most interesting case is that of some complex words, the structure of which may be traced back to a matrix "lexical root + V", which in some linguistic traditions go back to participial forms. In their semantic quality of "verbal nouns", participles share formal features of the nominal and verbal categories, thus representing a particularly interesting form of "transition" between these two points of the semantic *continuum*. An example for this could be the Latin word *studēns, studēntis* < *stud-ent-s, stud-ent-is: it is formed by a lexical root *stud- "to study" with a suffix *-ent, derivated from an allomorphic form of the I.E. root *es- / *h₁es- "to be, to exist". In other words, from a semantic perspective the structural matrix of Lat. *studēns, studēntis* is "the one who is in the dimension of studying". This structural matrix is

---

[11] Compare the words in 16 to Chin. 公費生 *gōngfèishēng* "fellowship student" (lit. "student at public expenses) and 自費生 *zìfèishēng* "self-supporting student".

actually analogous, in the synchronic perspective, to the Chinese "學生 *xuéshēng*-type" constructions.

Based on these remarks, we shall now examine a series of complex words drawn from I.E. languages that involve a morpheme in the semantic area of "to be / to be by nature".

As widely known, from the very beginning of research in Indo-European linguistics, comparison within members of the family made it possible to recognize two very productive roots, both conveying the meaning of "to be / to exist by nature": these roots are *$h_1$es- / *s- and *b^hweh$_2$-, diffused throughout I.E. languages and may be exemplified by the double forms (jointly expressed in a paradigm) such as Skr. (*ásti / bháviti*), Gr. (ἐστί / ἔφυ), Lat. (*est / fuīt*), in Germanic languages (Ger. *ist / bin*; Engl. *is / to be*), in the Slavic group (Russ. *jest' / byt'*), and, as a single form, in the Baltic languages (Latv. *búti*). In this section we shall analyse the occurrence of such roots (including their allomorphs) in word formation; in many of these cases, it is appropriate to talk about grammaticalization, albeit sometimes only partial.

First of all, it should be noticed that the two roots under examination here, as far as the formation of allomorphic paradigms of "to be" (such as Lat. *esse / sum / fuī*), the existential verb *par excellence*, are widely attested in I.E. languages through a number of word-forms which were originally participles and later evolved to the *status* of (somewhat) complex autonomous words, such as:

- Gr. τὰ ὄντα "the things which exist > truth", Lat. *entia* "existing things", derived from a participle form with apophonic variation (*ent- / *ont-);

- Hitt.. *ašant-* "existing > true, real", Skr. *sánt-* "existing, real", Av. *hant-* "existing > true, real", Lat. *sōns, sont-is* "the one who is (at the centre of a judgement) > guilty", O. Icel. *sannr* "existing > true", O.Engl. *sōð* "existing > true, actual". The Latin verb *sentire* (lit. "to identify oneself with sthg." > "to feel") is quite interesting since it involves a special case of morphological redundancy: it is derived from a participle form (with apophonic variation) *(h_1e)s-ent / *(h_1e)s-ont-, in which the base form of the verb "to be" is followed by a morpheme connected with the the present participle form of the very same verb (compare Lat. *essentia*; *4.1*).

In what follows, we shall analyse instances of (partial?) grammaticalization of the I.E. roots discussed above, providing examples of their usage (or, rather, of the usage of a form derived from them) as complex word constituents.

4.1 The I.E. roots *$h_1$es- / *s- and *b^hweh$_2$- in word formation

Participle forms (with apophonic variation) *(h_1e)s-ent / *(h_1e)s-ont- and *ent- / *ont- combined with prepositions yield nouns such as Lat. *ab-sent-ia* "assenza", *prae-sent-ia* "presenza" (related to the corresponding verb forms *abesse, praeesse*), Gr. παρουσία "presence" (< παρά + *ὀντ-jα, realated to the verb παρειμί). In combination with verbal roots of the *(e)d-ont- / *(e)d-ent- type, one finds words such as Hitt. *adant-*, Gr. ὁδοὺς, ὁδόντος, Church Sl. *ĕd-ęt*, Skr. *da(n)t*, Avest. *dantan-*, Lat. *dēns, dēntis*, O. Ir. *dét*, Goth. *tunþus*, O. H. Germ. *zand*, Latv. *dantis* "tooth". The structural matrix behind these words consists of a lexical root "to eat" + *ent- / -ont- (a morpheme of the present participle of the verb "to exist (by nature)") > "that is eating". Another interesting case is this peculiar instances of morphological redundancy: Latin *essentia* (< *esse + *-ent-ja, derived from an unattested participle *essēns, following the lexical pattern of *patientia < patiens, sapientia < sapiens*) and *sententia* (instead of the expected *sentientia,* derived from *sentiō* or, rather, from a participial base *sent- + *-ent-ja).

From the root *b^hweh$_2$-, Latin has developed an intervocalic morpheme –b- (regular derivation *-bh- > -b-), which may be found in a series of adjectives formed by:

a) preposition + *-b-, such as *prō-b-us* (< *prō-bh-os, lit. "(one) who is, by nature, before, in a prominent position" > "honest"; compare. Ved. *pra-bhúḥ* "eminent, prominent"), *super-b-us* (< *super-bh-os, lit. *super-bh-os "(one) who is by nature above" > "arrogant");

b) nominale/adjectival base + *-b-, such as *acer-b-us* (<*akro-bho-s < *ak- "to be sharp", lit. "which is sharpened (by nature) > sour".

This pattern is also found in a high number of Greek adjectives (where *-bh- > *-ph-). In such structures the first element may be on of the following:

a) a preposition, as in ἐμ-φυ-ής "innate, natural" and προσ-φυ-ής "naturally attached";
b) an adjective, as μεγαλο-φυ-ής "of noble nature";
c) an adjective acting as an adverb, such as in εὐρυ-φυ-ής "vast + to be (by nature) > developing";
d) an adverb, as in the adjective δι-φυ-ής "dual, double in nature";
e) a prefixoid, such as in the adjective αὐτο-φυ-ής "born by oneself, natural".

It is also worth noticing that the morph *-b- may be found also in the verbal paradigm of Latin and Italic languages, both as an aspect marker indicating past imperfective event (the type of Lat. *am-ā-b-a-m, mon-ē-b-a-m*, with its Oscan equivalent *fu-fa-ns* "they were") and as a marker of imminent / progressive event in Latin future verb-forms (the type of *am-ā-b-ō, mon-ē-b-ō*, the Faliscan equivalent of which is *pi-pa-fo* "I will drink").

The primary meaning of the morpheme *-ent- / *-ont- (and of its allomorph *-(V)nt-, zero degree of the root) must have been something like "who / what is (by nature)": later on, with a rather predictable semantic shift, it came to mean "related to, connected to" > "belonging to". This semantic variation may explain some features of a set of I.E. complex words: among these, we find collective / plural proper names in Anatolian (the Hittite type *utneyant-/udneyant-* "population of a region" < *utne/udne* "region"; *tuzziyant* "troops, army" < *tuzzi* "troop") with possible continuations, according to Solta (1958: 8-9) – following Kretschmer's and Oikonomides's thesis – in the plural / collective forms found in the Modern Greek dialects of Ponto (such as βρεχή, ἡ "rain" / βρεχάντας, τά "rains", χαρά, ἡ "joy" / χαράντας, τά "joy, wedding".

The most visible among the semantic values of one of the forms of the morpheme *-ent- / *-ont- / -*(V)nt- is that of participle marker (active, middle and passive voice), which can be considered to be located at the boundaries between the categories of adjectives proper and verbal adjectives. According to Adrados, Bernabé & Mendoza (1996: 321-323), there are serious problem in estabiliing clear boundaries betweeen the above mentioned categories: some ancient Latin denominative adjectives formed with the morpheme *-(a)nt-* (from the formal point of view, present participles of Classical Latin) used in poetry are particularly interesting in this respect, such as e.g. *animāns* "endowed with the spirit of life", *stellāns* "provided with stars", *comāns* "provided with hair", being in fact transitional between verbs and nouns / adjcetives (see also Hofmann, Szantyr & Leumann 1972: 589).

Yet, the most important semantic features of the morpheme *-ent- / *-ont- / -*(V)nt-, as a participle marker, is the clear insensitivity to diathesis. This typologically very significant feature occurs in many I.E. languages:

In Buddhist Sanskrit one finds several passive participles, meaning conveyed by the active suffix *-(a)-nt-(a)* < *-(V)nt-: it is as such words as *dahy-ant-as* (nom. sing. masvc.) "which is burned" semantically analogous to *dahyamānas*; the same happens in the Prakrit languages.

In Iranian, as in Sanskrit, the derivatives of the suffix *-(V)nt- normally have an active, transitive or intransitive meaning, depending on the verb to which they are agglutinated; yet, just as in Sanskrit, one finds visible traces of insensitivity to diathesis of this suffix in Avestic, as evidenced words like Av. *haošy-ant-a* "that ought ot be squeezed" (Solta 1958: 42).

In Greek, just as in Sanskrit and in Avestic, the continuations of the suffix *-(V)nt- > -(V)ντ- form participles with active diathesis (present, future, aorist, like in λύ-ο-ντ-, λύ-σο-ντ-, λύ-σα-ντ-), whereas for the mediopassive (present, future, passive future, middle aorist,

perfect) one finds the continuations of the suffix *-meno-: the types of λυ-ό-μενο-ς, λυ-σό-μενο-ς, λυ-θη-σό-μενο-ς, λυ-σά-μενο-ς, λε-λυ-μένο-ς. Such a division generally applies, and the suffix *-(V)ντ- normally has active, transitive or intransitive meaning, depending on the nature of the verb to which it is attached. However, an original insensitivity towards diathesis emerges from the fact that the suffix involved plays a role *also* in the formation of the aorist passive participle (see the model -(θ)ε-ντ-: λυ-θ-έ-ντ- > λυθεῖς) and in ancient noun forms in which the suffix *-(V)ντ- is clearly used with a middle or even passive meaning: see Chantraine (1990: 218) on Gr. *γέρ-o-ντ- > γέρων, γέροντος "that has grown old, made to grow old > old" comparable to Skr. *járant-*, Ossetian *zärond*.

In Hittite, the continuations of the suffix *-(V)nt- (besides forming plural / collective nouns, which we already mentioned) yield the only participle of the originally diathesis indifferent verb (Benveniste 1984: 126).

In Latin, the morpheme -(V)nt- participates in the formation of present participle forms both of active (such as *dīcēns, dīcentis*) and deponent verbs (*sequēns, sequentis*), with an active-transitive or intransitive meaning. Latin has no present mediopassive participle deriving from the suffix *-m(e)no- for the expression of the aspectual notion "not completed", whereas the passive and deponent past participles are supposedly derived from the ancient verbal adjective *-(C/V)to- (the types represented by *fac-tu-m* and *mon-i-tu-m*). Ancient participle forms of the type *e-vid-ens* "that is seen", *in-fans* "that cannot be told" (translating Gr. ἄρρητον: compare. Acc. 189 *infāns facinus*; Lucil. 140 *facta nefantia*), *gignentia* "plant, growth" which renders Gr. τὰ φυόμενα. The usage of the participle *-(V)nt- with an intransitive-reflexive or, also, passive meaning (e.g. *amantissimus* vs. *amatissimus* and *desiderans* vs. *desideratus* found in funeral inscriptions already during the Republican era and present throughout Medieval Latin epigraphs) continues throughout the entire Latin era and it is also attested in later authors, as one may see from the many examples in Hofmann, Szantyr e Leumann (1972: 387-390).

In Celtic, the category of present participle is absent from the verbal system. However, the diachronic analysis of Celtic data confirms that the continuations of the participle suffix *-(V)nt- could not only combine with active bases, but could also be a part in middle or even passive construction, being insensitive to diathesis: See e.g. the O. Ir. adjective *té/tee* "hot" (< *tep-ent-s; compare. Skr. *tapant-*, Lat. *tepens*), the O. Ir. nouns *care/carae* (masc. sing.) "friend" (< *kar-ā-nt-s, lit. "the one who loves"; compare Welsh *Carantus, Carantius, Carantillus,* etc.), *bragae* "prisoner" (< *brag-a-nt- "the one who is being guarded"; compare Goth. *bairgan* "to guard", Russ. *beregu* "I am guarding"), where a participle *-(V)nt- had an evident passive meaning

In Germanic and Tocharian one also finds strong insensitivity to diathesis in participle forms that continue the suffix *-(V)nt-: in Old Nordic, participles in *-(V)nt- may be found only in popular prose, with intransitive verbs (Solta 1958: 26-27; Evangelisti 1965: 158). In Gothic, one finds a single case of present participle in *-(V)nt-, with mediopassive meaning, which is *aland* < *alan* "to feed, to nourish" (translating Gr. ἐντρεφόμενος of Timoth. I, IV, 6); analogously, we have Goth. *þai daupjandans* (corresponding to Greek οἱ βαπτιζόμενοι of K 15, 29): the verb on which it depends – *daupjan* "to baptize" – is clearly transitive. In both Tocharian dialects the continuation of the suffix *-(V)nt- (Toch. A: *-nt,* Toch. B: *ñca*) form the active present participle, as opposed to the middle present participle in *-meno- (Toch. A: *māṃ*; Toch. B: *-mane*). The semantic values yield complete indifference to diathesis also in Tocharian participles. To be sure, Tocharian has collective nouns attesting the continuation of the suffix *-(V)nt-, marking plural (Toch. A *akär / akrunt* "tear / tears", *tärkär / tärkrunt* "cloud / clouds").

Insensitivity to diathesis in participle formations with the *-(V)nt- morpheme is a typologically relevant feature that hints to the presence of remnants of a grammaticalization process of units in the morphological component in an ancient phase of the Indo-European linguistic tradition, if considered side by side with the morphemes present in some complex words (in Vedic, Greek, Latin) which are the continuation of the root *b^{h}weh$_2$-: it is a very ancient phase, in which the I.E. system yielded agglutinating phenomena, with a high degree of semantic transparency in word formation.

4.2 Summary and (provisional) conclusions

The comparative analysis of complex words in Chinese containing the morpheme 生 *shēng* / *sheng* and of the continuations in I.E. languages of morphs deriving from the roots *h₁es- / *bʰweh₂- (excluding, of course, any genealogic realationship between the two linguistic areas) suggests, thus, three facts:

a) a process of progressive enrichment of word formation in Chinese, which is an isolating system that has, partially, acquired typical agglutinating features. In such a system, however, in complex word formation the elements involved maintain their shape unaltered and, consequently, allow for a high level of semantic transparency;

b) the parallel process of the increasing in morphological complexity in the I.E. system, with its apparently early grammaticalization features and where individual morphemes that acquired a high degree of formal opacity, following the loss of (formal) connection to their original form, have acquired a high degree of formal opacity. This loss of transparency was (is) at such a high level that only the historical linguistic reconstruction yields insights into its underlying mechanisms;

c) from the methodological point of view, the observation of "parallel" morphological strategies documented in typologically and diatopically distant linguistic systems ultimately allows for hypothesizing the presence of semantic universal features operating at the cognitive level, in the speakers' linguistic awareness, leading to other fascinating research topics.

5. Concluding remarks

5.1 Grammaticalizazion and "typological drift" in Chinese

A general assessment of the typological status of the morphology of Chinese is obviously beyond the scope of this talk. Nevertheless, we may take the data presented here as an example of the kind of development which characterizes Modern Chinese.

A "typological drift" toward the agglutinating type, i.e. the acquisition of agglutinating features, may be seen, above all, in the domain of word formation; in the modern language there is a tendency to the development of word formation patterns with "hybrid" properties, which possess both features of isolating morphology (such as the 1:1 ratio between morph and lexical morpheme) and characteristics of the agglutinating "word type" (words made of long, clearly segmentable, morpheme chains).

It is still questionable whether word formation elements such as 生 *shēng/sheng* should be regarded as lexemes or as affixes. In Chinese, the origin of the "suffix-like" 生 *shēng/sheng* is rather peculiar, being born out of the abbreviation of a compound (先生 *xiānsheng*, see above), similarly to Engl. –*gate* or it. –*poli* "scandal related to X", a somehow productive suffixoid which was created by analogy with the word *tangentopoli* "tangent town" (used in reference to Milan in the early nineties, later to indicate the system of corruption in Italian politics and the scandal which followed its unveiling).

The question as to whether 生 *shēng/sheng* has undergone grammaticalization requires a "split judgement": as a suffix-like formative used in nouns for occupations and in its usage with some adverbs (exx, 4 and 6), as well as in its prefix-like identity (with intensifying value) for V and Adj (ex. 8), we may talk about "aborted grammaticalization", as the number of words which are built according to these patterns is negligible, and they are no longer productive; we might speak of grammaticalization / morphologization when 生 *shēng/sheng* is used as a suffix-like formative meaning "student" (ex. 3), where we have both a specialization of meaning and (limited) productivity. The grammaticalization of 生 *shēng/sheng* in a suffix(oid) meaning "student" seems more evident in Japanese, where we have a consistent series of complex words with such morph in suffixal position and with a stable meaning.

5.2 A comparative outlook

In its diachronical evolution, Chinese morphology (specifically, word formation) has shifted from processes modifying the syllable to *word formation schemas* (Booij 2006) based on the agglutination of lexical morphemes; we may also regard some of these lexemes as derivational morphemes, i.e. grammatical morphemes;

The transparency of complex words in Modern Chinese is very high; the main "challenge" to this feature is the production of abbreviated forms which generate "new morphemes" (Packard 2000) →

(17)　　勞動　　　　　　　　　保險　　　　　　　　　勞保
　　　　láodòng　　　　　　　*bǎoxiǎn*　　　　　　　*lǎobǎo*
　　　"labour+act"　　　　　　"protect+risk"　　　　　"labour+protect"
　　　　"labour"　　　　　　　"insurance"　　　　　　"labour insurance"

In the case of 生 *shēng/sheng* used as a suffix(oid), meaning transparency is ensured by the "paradigmatic" relationship with other complex words in the same "word family" (see Booij 2006).

The opposite occurred in the Indo-European system, where, as we have seen, agglutinating features in word formation are not "active" anymore, and early grammatical morphs have become fused as part of the lexical root, thus making the structure of the word opaque, whose original shape is of interest only for etymologcal research.

Having found two morphemes with a common original meaning and similarities in the present semantic value in two unrelated (and typologically distant) language families, one is tempted to hypothesize the action on "meaning universals"; such an explanation, which can hardly hold in the lexical domain, seems to be valid for the evolution of lexical morphemes into grammatical ones (see Heine & Kuteva 2002). In the latter case, one might well put forward the hypothesis that lexemes meaning "to be (born), to be by nature" are a typical source of grammatical morphs meaning "to be in the state / activity of". Of course, this is to be regarded as a "clue", and should be tested on a proper language sample; it is also worth remembering once more that the "student" meaning for 生 *shēng/sheng* has been acquired from the truncation of a compound, and thus it may not easily fit into the comparison with the Indo-European roots investigated here. Although…

Abbrevations used in the glosses

CONG.: conjunction
PREP.: preposition
PTC.: structural particle (possessive, relative)

Selected references

F. Adrados, A. Bernabé and J. Mendoza, Manual de Lingüística indoeuropea. II. Morfología nominal y verbal*, Ediciones Clásicas, Madrid, 1996.

P. Baldi, The Foundations of Latin, Mouton de Gruyter, Berlin-New York, 1999.

W.H. Baxter and L. Sagart, Word formation in Old Chinese, in J.L. Packard (ed.), New approaches to Chinese Word Formation, Mouton de Gruyter, Berlin-New York, 1998, pp. 35-75.

W. Belardi, Linguistica generale, critica e filologia del testo*, Bonacci, Roma, 1990.

É. Benveniste, Origines de la formation des noms en indo-européen*, Maisonneuve, Paris, 1984.

G. Booij, Inflection and derivation, in K. Brown *et alii* (eds.), Encyclopedia of language and linguistics, Elsevier, Oxford, 2006, pp. 654-661 (V vol.).

P. Chantraine, La formation des noms en grec ancien, Klincksieck, Paris, 1979 [1933].

E. Evangelisti, Ricerche sul suffisso -NT- di participio*, Acme 18, 1965, pp. 205-22 [Reprinted in E. Evangelisti, Scritti tocarici e altri studi*, Brescia, 1990, pp. 139-161 (quotations from the reprint)].

T. V. Gamkrelidze and V. Ivanov Vjačeslav, Indo-European and the Indo-Europeans. A Reconstruction and Historical Analysis of a Proto-Language and a Proto-Culture*, Mouton de Gruyter, Berlin-New

York, 1995.

M. Haspelmath, On Directionality in Language Change with particular Refence to Grammaticalization, in O. Fischer, M. Norde and H. Perridon (eds.), Up and down the Cline – The nature of Grammaticalization, Benjamins, Amsterdam-Philadelphia, 2004, pp. 17-44.

B. Heine and T. Kuteva, World Lexicon of Grammaticalization, Cambridge University Press, Cambridge, 2002.

P. J. Hopper Paul, Where do words come from?, in W. Croft, D. Keith and S. Kemmer (eds.), Studies in Typology and Diachrony. Papers presented to Joseph H. Greenberg on his 75th birthday, John Benjamins, Amsterdam-Philadelphia, 1990, pp. 151-160.

P. Hopper and E. Traugott, Grammaticalization, Cambridge University Press, Cambridge, 1993.

C. Lehmann, Thoughts on Grammaticalization, Lincom Europa, München, 1995 [1982].

H. Lin, A Grammar of Mandarin Chinese, Lincom Europa, München, 2001.

F. Mawet, Les participes indo-européens (*-nt-), in G. Bolognesi and V. Pisani (eds.), Linguistica e Filologia. Atti del VII Convegno internazionale di linguisti (Milano, 12-14 IX 1984), Paideia, Brescia, 1987, pp. 357-374.

G. Meiser, Historische Laut- und Formenlehere der lateinische Sprache, Wissenschaftliche Buchgesellschaft, Darmstadt, 1998.

A. Meillet and J. Vendryes, Traité de Grammaire comparée des langues classiques, Paris, 1979[5].

J. L. Packard, The Morphology of Chinese. A Linguistic and Cognitive Approach, Cambridge University Press, Cambridge, 2000.

W. Pan, B. Ye and Y. Han, 汉语的构词法研究 (Hanyu de goucifa yanjiu – Research on word formation in Chinese), Huadong Shifan Daxue Chubanshe, Shanghai, 2004.

E. G. Pulleyblank, From Archaic Chinese to Mandarin, in G. Booij, C. Lehmann and J. Mugdan (eds.), Morphologie-Morphology, Walter de Gruyter, Berlin-New York, 2000, pp. 1730-1740.

R. Solta, Gedanken über das nt-Suffix, Österreische Akademie der Wissenschaften, Phil.-Hist. Kl. 232/1, 1957, pp. 3-47.

C. Watkins, Il proto-indoeuropeo, in A. Giacalone Ramat and P. Ramat (eds.), Le lingue indoeuropee, Il Mulino, Bologna, 1993, pp. 43-93.

Dictionaries and corpora

CCD (2002), The Contemporary Chinese Dictionary - 现 代 汉 语 词 典 (Xiandai Hanyu cidian), Foreign Language Teaching and Research Press, Beijing.

GHKJE (2002), 逆引き広辞苑 (Gyakuhiki Kōjien –Kōjien reverse order Japanese dictionary), Iwanami Shoten, Tokyo (electronic edition).

GHYDCD (2000), 古汉语大词典 (Guhanyu Dacidian – Great dictionary of Classical Chinese), Shanghai Cishu Chubanshe, Shanghai.

HYDCD (1991), 汉语大词典 (Hanyu Dacidian – Great dictionary of Chinese), Hanyu Dacidian Chubanshe, Shanghai.

INKGJT (1994), 岩波国語辞典 (Iwanami Kokugo jiten – Iwanami Japanese dictionary), Iwanami Shoten, Tokyo.

NXCD (2005), 现代汉语逆序词典 (Xiandai Hanyu nixu cidian – Modern Chinese reverse order dictionary), Yanbian Renmin Chubanshe, Yanji.

Academia Sinica corpus of Old Chinese (古漢語語料庫 Gǔhànyǔ yǔliàokù): http://www.sinica.edu.tw/~tibe/2-words/old-words/.

Academia Sinica tagged corpus of early Mandarin Chinese (近代漢語標記語料庫 Jìndài Hànyǔ biāojì yǔliàokù): http://www.sinica.edu.tw/Early_Mandarin/.

Academia Sinica Balanced Corpus of Modern Chinese (現代漢語平衡語料庫 Xiàndài Hànyǔ pínghéng yǔliàokù): http://www.sinica.edu.tw/ftms-bin/kiwi.sh?ukey=344276996&qtype=1.