

The Japanese *Pithecanthropus*: Interpreting Japanese Unfamiliar Compounds Takao Suzuki's Model of Conceptual Combination Through *kun*-Glossing

Simone dalla Chiesa

Università degli Studi di Milano, Italia

Abstract With 'kun-glossing', I refer to the method of interpreting unfamiliar compounds proposed by Takao Suzuki (1926-2021), a well-known Japanese scholar. Readers assign a semantic gloss to each constituent *kanji*, combine those meanings into a new, complex concept and attempt to validate the result(s) encyclopedically. In this essay I verify *kun*-glossing's processing power by applying it to eleven compounds, beginning with *enjin* 猿人 (*Pithecanthropus*). I conclude that this method is only effective in interpreting words that are structured as descriptions, as the neoclassical compounds of science often are, and cannot be considered a universal method to infer the meaning of novel compounds.

Keywords Japanese compounds. Kanji. Compounding. Neoclassical compounds. Appositive compounds. Binomens. Conceptual combination. Onomasiological approach. Takao Suzuki. Ernst Haeckel.

Summary 1 Introduction. – 2 Enter the *Pithecanthropus*. – 3 A Proper Name: *Katsuo no Eboshi* 鯉の烏帽子 (Portuguese Man o' War). – 4 A Natural-Kind Name: *Yōryokuso* 葉緑素 (Chlorophyll). – 5 More Kind Names: *Hishishokubutsu* 被子植物 (Angiosperm), *Rashishokubutsu* 裸子植物 (Gymnosperm). – 6 A Class Noun: *Tanpakushitsu* 蛋白質 (Protein). – 7 The Term *Hōshokusei* 蜂食性 (Apivory) and the Function of *Sei* 性. – 8 A Common Class Noun: *Jidōsha* 自動車 (Automobile). – 9 A Common Class Noun: The Neoclassical Compound *Sekiyū* 石油 (Petroleum). – 10 Discussion. – 11 Conclusions. – 12 A Final Assessment.



Edizioni
Ca' Foscari

Peer review

Submitted 2023-05-19
Accepted 2023-06-20
Published 2023-08-29

Open access

© 2023 dalla Chiesa | © 4.0



Citation dalla Chiesa, S. (2023). "The Japanese *Pithecanthropus*: Interpreting Japanese Unfamiliar Compounds. Takao Suzuki's Model of Conceptual Combination Through *kun*-Glossing". *Annali di Ca' Foscari. Serie orientale*, 59, 609-702.

DOI 10.36687/AnnOr/2385-3042/2023/01/023

609

1 Introduction

Takao Suzuki (1926-2021) was a respected and influential sociologist of language known for his popular essays on the peculiarities of the Japanese language, mind and culture and, quite openly in his later years, on the alleged superiority of Japanese civilisation over the rest of the world. In a short section of his 2014 book, *Nihon no kankaku ga sekai o kaeru* (Japanese Sensibility Will Change the World), Suzuki proposes a subtle interpretation of the function of the so-called ‘Japanese readings’ (*kun’yomi* 訓読み) attached to Chinese ideographic characters (*kanji* 漢字). In Suzuki’s (2014, 208-13) view, *kun* readings are not the ‘Japanese pronunciations’ of *kanji*, but rather represent their metalinguistic semantic glosses, conveniently expressed in the background language itself. Since almost every *kanji* has a *kun* reading, readers can infer the meaning of any unfamiliar *kanji* compound by simply glossing its constituents. This processing method, Suzuki argues, is crucial to Japanese, whose *kanji* lexicon is entirely of foreign origin, and is particularly effective in decoding those infrequent terms known only to specialists.

Suzuki (2014, 108) compares several English “stratospheric words” (*kumo no ue no kotoba* 雲の上のことば) with Sino-Japanese semantic equivalents. One such pair is represented by ‘*Pithecanthropus*’, a neoclassical compound derived from two Ancient Greek roots, and its Japanese translation, *enjin* 猿人, a compound similarly formed by borrowing the roots and graphic signs of the language locally considered of the highest scientific status, i.e. Chinese. Whereas the two Ancient Greek roots are opaque to most English readers, to an ordinary Japanese person 猿 and 人 are quite intelligible, thanks to their glosses, which are, respectively, *saru* (ape) and *hito* (human being). Combining these two concepts easily returns ‘ape-man’.

At first glance, Suzuki’s (2014) model, which I will henceforth call ‘*kun*-glossing’, does not seem particularly original. For one thing, the system of assigning Japanese readings to *kanji* was created just for glossing purposes – as pointed out by Suzuki (1963, 28) – and, indeed, the term *kun* 訓 means ‘gloss’ and ‘translation’ (Tōdō 1980; K aizuka, Fujino, Ono 1984).¹ Moreover, other scholars have held similar views, the most quoted being Nomura (1978; 1979). Finally, it is quite intuitive that the meaning of an unknown compound is to be inferred compositionally, by combining the concepts represented by the constituents and thus creating a new, more complex concept. But since the *kun* readings of *kanji* are nothing but ordinary vernacular

¹ See also Matsumura (1988, 88): “the Japanese reading of a *kanji*, corresponding to the meaning expressed by that *kanji*” (*kanji ni, sore ga arawasu imi ni sōtō suru nihongo o ateta yomikata* 漢字に、それが表す意味に相当する日本語を当てた読み方).

words, *kun*-glossing implies that, in Japanese, semantics is accessed through phonetic codes. Whereas in early works Suzuki seemed to adhere to the traditional semiotic view that Chinese characters are translated into meaning directly, without any phonological activity (cf. Suzuki 1963, 24), an interpretive mechanism like *kun*-glossing strongly diverges from that view.

It is not surprising then that, over the years, many scholars have downplayed or bent *kun*-glossing in favour of the traditional *kanji*-to-meaning approach, whilst acknowledging Suzuki by referencing his works. This can be observed when Suzuki's assertions on the efficiency of *kanji* encoding are merely mentioned (cf. e.g. Kageyama 2011; Namiki, Kageyama 2016), quite openly shared (Shibatani 1990, 147) or are subjected to experimental confirmation (most notably by Hatano, Kuhara, Akiyama 1981, but see § 11.2.3). Perhaps the most complete attempt to integrate *kun*-glossing into an organic theory of orthography and *kanji* semiosis is that by Nagano, Shimada (2014).

The importance of Suzuki's model comes from the fact that, despite the lack of systematic generalisations, the author proposes *kun*-glossing as a model with which Japanese speakers might infer the meaning of any unknown *kanji* compound. As Suzuki (1978, 8) explains, *kun*-glossing arises from the basic act of "kun-reading the general meaning of a given Sino-Japanese word" (*aru kango no daitai no imi o, kun'yomi suru* ある漢語の大体の意味を、訓読みする) and is used in all registers (Suzuki 1978, 2), for plain words like *gyūnyū* 牛乳 (cow-milk), *nyūgyū* 乳牛 (milk-cow) (cf. Hatano, Kuhara, Akiyama 1981, 32), *hantō* 半島 (peninsula) and *jidōsha* 自動車 (automobile) (cf. Suzuki 1975, 189) and for technical, scientific words. As such, *kun*-glossing is a mechanism which is constantly at work in present-day reading practice and is endowed with a power extending far beyond the comparatively few instances of it discussed by Suzuki (see § 1.2 for a list of Suzuki's relevant works on the topic).

1.1 The Aim of This Essay

I will challenge the effectiveness of Suzuki's model by analyzing the *kun*-glossing process required to derive meaning from some of the compounds he discusses and from a few more taken from the lexicon. I will focus on the following issues:

- Multiple construals. Multiple construals occur when a constituent *kanji* has more than one *kun* reading, a *kun* reading has more than one meaning or when the meanings of the constituents can interact in more than one way. How can we differentiate the several interpretations?
- Exocentricity and headedness. By definition, semantic exocentricity occurs when a compound or one of the complex

sub-constituents of a compound does not belong to the same semantic class of its morphologic head. In *kun*-glossing terms, the effect is that the parent item is not a hyponym of the particular lexeme chosen to gloss its head. This phenomenon is most often caused by a metaphoric, as in the right-headed compound, 石綿, glossable as *ishi-wata* (stone-cotton), which denotes a mineral, asbestos, not a vegetable fiber. The compound's interpretation is apparently blocked. Can *kun*-glossing provide sufficient material to allow readers to detect exocentricity and overcome this block?

- Suffixes. A problem related to semantic exocentricity occurs when a compound has a suffix. In a *kanji* used as an affix, the original meaning expressed by the *kun* reading has usually undergone some bleaching. As an effect, the efficiency of interpreting suffixed compounds by means of *kun*-glossing each individual constituent is reduced. Could the meaning and role of a *kanji* in suffixal function be derived from its *kun* reading(s) alone?
- Referentiality. The outcome of *kun*-glossing is always a description and corresponds to an intensional meaning. Most of the compounds discussed by Suzuki, however, are kind names which, as names, are commonly considered to only have extension. The very term 'Pithecanthropus' is one of them. Does the descriptive content yielded by *kun*-glossing carry enough information to enable readers to identify the actual denotata of novel compounds?
- World knowledge. A closely related issue concerns the role of world knowledge. Suzuki circles around the issue but never quite addresses it. Once a literal meaning is obtained, how much encyclopedic knowledge, and of what depth and specialisation, is required to give that meaning some denotational value? This is an issue of informativity, for the richer the knowledge required, the smaller the epistemic import of *kun*-glossing.
- Context. The compounds discussed by Suzuki to underscore the merits of *kun*-glossing are almost invariably introduced out of context. This is probably because context, if rich enough, might make the entire *kun*-glossing operation superfluous. In this essay I will play by Suzuki's rules and mostly ignore the role of context (see brief discussion in § 2.2).

Focusing on the above issues means that, in this essay, I will not deal with the history and use of *kun'yomi* - with how the concept of *kun*-glossing evolved under Suzuki, under other Japanese experts and under the intuition of ordinary speakers - and the differences between the various conceptions of it. Neither will I discuss Suzuki's whole view of languages, particularly of Japanese, and of Japanese culture in general. Consequently, I will not examine Suzuki's instrumental use of *kun*-glossing in showing the superiority of Japanese

over English. Suzuki's place in the vast but quite monotonous landscape of Japanese cultural nationalism is a topic of great interest, which falls, however, outside the scope of this essay (on the topic see Vitali 2023). Keeping in mind Suzuki's overall ideological aim, though, does help explain some of his choices, and also the superficiality he sometimes shows in the analysis of his own examples. I will remark on this issue when needed.

Despite my criticism of Suzuki, I am not an advocate of the traditional belief that *kanji* orthography is directly translated into meaning. Seeing if, and how, Suzuki's *kun*-glossing can be reconciled with the traditional view is also beyond the scope of this work but may become the object of future research.

1.2 The Sample

For the purpose of this essay there is no need to take Suzuki's entire body of work into account, which is too vast and mostly irrelevant to the topic. Only those works that specifically deal with *kun*-glossing, even if only indirectly, need be considered. They are Suzuki 1963; 1969; 1975; 1977; 1978; 1990; 2014 and 2017. Their content is quite repetitive, but each of them yields subtly different information on how Suzuki views *kun*-glossing, to the extent that they all have to be taken into account in order to acquire a complete picture of it. Of particular interest are Suzuki 1978 and 1990. Suzuki 1978 provides an explanation as to why Suzuki only exemplifies *kun*-glossing by decomposing Japanese scientific terms that correspond to English neoclassical compounds. Suzuki 1990 is the scholar's most extensive work on the matter, devoting a whole chapter (128-64) to *kun*-glossing and related topics. My main source, however, is Suzuki 2014, because it contains the most finely tuned illustration of Suzuki's model (the discussion of 'Pithecanthropus'), because it is part of his most recent original book and therefore it is implicitly proposed as the sum of the author's lifetime thinking.

1.2.1 Suzuki's Examples

In the works I considered, Suzuki discusses the merits of *kun*-glossing at different levels of detail and from slightly different perspectives. To this end, he decomposes, mentions or merely lists a total of 151 Japanese compounds. I have divided them in the following four groups, according to the depth of Suzuki's analysis.²

² I only list the lexemes of the first three groups. Those English expressions without brackets are the ones which Suzuki himself associates or contrasts with the Japanese

- Group A: 17 items. Compounds that Suzuki submits to actual *kun*-glossing or about which he holds that *kun* readings have a meaning-clarifying role: *kagaku* 科学 ‘science’ and *kagaku* 化学 ‘chemistry’, *kōshaku* 侯爵 ‘prince’ and *kōshaku* 公爵 ‘marquis’, *shiritsu* 私立 (private) and *shiritsu* 市立 (municipal) (Suzuki 1963, 24-5);³ *Tatōkai* 多島海 ‘Aegean Sea’; *hantō* 半島 ‘peninsula’, *tantō* 短頭 ‘brachycephalic’, *chōtō* 長頭 ‘dolichocephalic’ (Suzuki 1975, 189-90); *jidōsha* ジドウシャ, 自動車 (automobile) (Suzuki 1978, 3),⁴ *hōshokusei* 蜂食性 ‘apivorous’; *tanpakushitsu* 蛋白質, **ranpakushitsu* *卵白質 (protein), *gesshirui* 齧齒類 (rodents) (Suzuki 1990, 132-5); *enjin* 猿人 ‘pithecanthrope’, ‘Pithecanthropus’ (Suzuki 2014, 208-10); *yōryokuso* 葉綠素 ‘chlorophyll’ (212-13); *jinruigaku* 人類 ‘anthropology’ (Suzuki 2017, 92-3).
- Group B: 19 items. Compounds for which Suzuki suggests that *kun* readings provide transparency, but which he does not analyse: *samidare* 五月雨 ‘April shower’, *tabako* 煙草 ‘tobacco’, *shigure* 時雨 ‘autumn occasional rain’, *tsuyu* 梅雨 ‘plum, or ripening season, rain’, *sumō* 相撲 and 角力 (Suzuki 1963, 34-5). (Suzuki 1963, 34-5);⁵ *shikeien* 齒齦炎 ‘gingivitis’, *suitōshō* 水頭症 ‘hydrocephalus’, *mueitō* 無影灯 ‘scialytic lamp’, *kussetsu* 屈折 ‘refraction’ (Suzuki 1990, 134, 136-7, 144); *rashishokubutsu* 裸子植物 ‘gymnosperm’, *hishishokubutsu* 被子植物 ‘angiosperm’, *kogaran* 胡蛾蘭 ‘phalaenopsis’, *hen'onsei* 恒溫性 ‘homiothermal’, *kōonsei* 變溫性 ‘poikilothermal’ (Suzuki 2014, 210-11); *suiso* 水素 ‘hydrogen’, *kyōsuishō* 恐水病 ‘hydrophobia’, *suisei no* 水棲の ‘aquatic’, *suizokukan* 水族館 ‘aquarium’ (Suzuki 2017, 90-2).
- Group C: 16 items. Compounds mentioned as part of a relevant discussion on *kun*-glossing: *shintōatsu* 浸透圧 ‘osmotic pressure’, *shinshutsu* 浸出 ‘exudation’, *kōshokyōfushō* 高所恐怖症 ‘acrophobia’ (Suzuki 1990, 128, 132); *sasensei* 左旋性 ‘laevorotatory’, *usenisei* 右旋性 ‘dextrorotatory’, *kōjitsusei* 向日性 ‘heliotropism’, *kōgetsusei* 向月性 ‘selenotropism’ (149-50);⁶ *hakkekkyū* 白血球 ‘leucocyte’, *sekkkekkyū* 赤血球 ‘erythrocyte’, *hakketsubyō* 白血病 ‘leukemia’, *kōketsuatsu* 高血圧 (hypertension), *kesshōban* 血小板 (platelet), *jōmyakuketsu* 靜脈血 (arterial blood), *zōketsu*

terms (the large majority), whereas those in brackets are my own translations, provided for clarity.

3 The lexemes of each couple are homophonous and Suzuki considers *kun*-glossing as a tool to disambiguate their meanings.

4 Here Suzuki (1978, 3) considers how the *kun* reading of a *kanji* assigns meaning to the otherwise meaningless *on* reading of it.

5 These lexemes are written with *ateji* 当て字, characters whose readings, either *kun* or *on* 音, bear no relation to the phonological word they represent in writing.

6 Suzuki (1990, 149-50) discusses the possibility of ordinary speakers freely *producing* these compounds.

貧血 ‘anemia’, *shukketsu* 出血 ‘hemorrhage’, *hinketsu* 造血 ‘hematogenesis’ (159-61).⁷

- Group D: 99 items. Compounds only listed in tables (Suzuki 1978, 7; 1990, 133, 144, 159, 161).

It should be noticed that, at an early stage, Suzuki (1963) pondered the function of *kun*-glossing in clarifying the meanings of homophones and *ateji*. It is only later (from 1975 on) that Suzuki shifted his focus to scientific terms, mostly selected from life sciences and in correspondence with English neoclassical compounds, often as calques. Due to a distributional trend in this type of lexeme and to categorial restrictions in the formation of new multiple-*kanji* words, such a choice actually creates a bias in Suzuki’s sample: most of the terms he considers are nouns. Even the terms that he renders with adjectives are actually abstract nouns denoting conditions and only become adjectivals if used preminally. Two examples are *tantō* 短頭 (brachycephalic) and *hōshokusei* 蜂食性, a term which Suzuki (1990, 132) translates as ‘apivorous’ but then paraphrases as “the natural disposition of feeding on bees” (*hachi o taberu seishitsu* 蜂を食べる性質), that is, ‘apivory’. The only items he mentions that are exclusively used in an adjectival function are *shiritsu* 私立 (private) and *shiritsu* 市立 (municipal). Verbal nouns are also under-represented in Suzuki’s choice of examples. Groups A to C only contain 6 verbal nouns: *susei* 水棲 (sea dwelling), *shigure* 時雨 (drizzle), *shinshutsu* 浸出 (exude, percolate), *shukketsu* 出血 (bleed), *zōketsu* 造血 (blood making) and *kussetsu* 屈折 (refract). This bias is unimportant, though, since the *kun*-glossing of compound adjectives and verbs is based on the same principles as that of nouns. Analyzing it, though, is more complicated, so limiting the discussion to nominals keeps things simple and was a wise decision by Suzuki.

Also of note is the fact that out of the 52 compounds in Groups A to C, 29 are formed of more than two characters. Among these, 27 are, or contain as complex constituents,⁸ three-*kanji* compounds. All are right-headed. In approximately one half of them, the head maintains the original lexical meaning, whereas in the rest, the head has undergone some semantic bleaching as a suffixal.

⁷ Suzuki mentions these terms in the main text of 1990. He accompanies them with a long list of 62 more items using the Sinogram 血 (blood). Because they are merely listed, I consider the compounds of the latter set to be part of Group D.

⁸ The expression *kōsho kyōfushō* 高所恐怖症 (acrophobia) is a complex compound actually formed of a two-*kanji* compound followed by a three-*kanji* one.

1.2.2 Compounds to Be Examined

The items I will submit to *kun*-glossing are the following, all taken from Groups A to C:

- *enjin* 猿人 (pithecanthrope, *Pithecanthropus*) (Suzuki 2014, 208-10);
- *yōryoku-so* 葉緑素 (chlorophyll) (212-13);
- *sui-so* 水素 (hydrogen) (Suzuki 2017, 89-92);
- *hishi-shokubutsu* 被子植物 (angiosperm) (Suzuki 2014, 210);
- *rashi-shokubutsu* 裸子植物 (gymnosperm) (210);
- *tanpaku-shitsu* 蛋白質 and the hypothetical *ranpaku-shitsu* *卵白質 (protein) (Suzuki 1990, 135);
- *hōshoku-sei* 蜂食性 (apivory) (132-3);
- *jidō-sha* 自動車 (automobile) (Suzuki 1978, 3).

The compounds I will add to the discussion, not found in Suzuki's works, are:

- *katsuo no eboshi* 鯉の烏帽子 (*Physalia physalis*, 'Portuguese man o' war');
- *kesshoku-so* 血色素 (hemoglobin);
- *sekiyu* 石油 (petroleum).

I wanted the set of case studies to be representative of the various issues generated by *kun*-glossing. Some items I chose because, before analysis, they appeared problematic to processing through *kun*-glossing. Others I picked for the opposite reason. Of great interest are also Tatōkai 多島海 (Aegean Sea) and *hantō* 半島 (peninsula) (Suzuki 1975, 189), but I discuss them in separate essays (cf. e.g. dalla Chiesa 2023).

My discussion will commence with *enjin* 猿人 (pithecanthrope/*Pithecanthropus*). This term represents my main case study, and I will reconstruct the operations required to obtain a satisfactory interpretation of it at the highest level of detail. I will then proceed by applying *kun*-glossing to the other compounds of the set, from more specialised terms to plainer ones, from kind names to class nouns. In most cases I will do so in a coarser way, making constant reference to the distinct operations individuated and discussed in relation to *enjin* 猿人 (pithecanthrope). The case of *rashishokubutsu* 被子植物 (angiosperm), however, stands out since this is the most complicated compound to interpret from among the ones I chose. I will, therefore, analyse it in depth as well.

1.3 Method of Analysis and Sources

In order to claim the merits of *kun*-glossing, Suzuki introduces a *kanji* compound and provides the glosses of its constituents. He never explains how he and his hypothetical readers choose those glosses and assemble their meanings. He then produces the final interpretation of the parent word, which is always close to the actual denotation of the source compound. To test the validity of *kun*-glossing, I will reconstruct all the missing intermediate steps that should necessarily take place in order to obtain the successful outcome envisaged by Suzuki. To do so, I will follow an unsophisticated computational approach that presupposes the three stages below. Each stage is necessary and should be considered an integral part of the process.

1.3.1 *Kun* Assignment

In the first stage, a *kun* reading is assigned to each constituent *kanji*. *Kun* assignment returns a hypothetical compound which might exist in the lexicon but is unknown to the interpreters. This is, of course, the basic assumption of the whole theory of *kun*-glossing, otherwise the process would simply coincide with the act of reading. In order to individuate the correct *kun* glosses, possibly from among several alternatives, and the actual meaning of them, readers rely on their knowledge of the *kanji*-lexicon, that is, of words written in *kanji*-notation, and on the related knowledge of *kanji* as ideograms (see § 1.4 below).

As for the source of such knowledge, I will proceed under the assumption that published dictionaries, with their tens of thousands of entries, form the largest theoretical repository available to readers engaged in *kun*-glossing. To identify roots, to navigate *kun* readings and their related meanings, and to extract those smaller sets of compounds in which a given character appears in the same position, I will rely on Tōdō (1980), Kaizuka, Fujino, Ono (1984), Spahn, Hadamitzky (1989), Ueda et al. (1993), Nelson, Haig (1997) and Guerra (2015). These are dictionaries which list lexemes by *kanji*. Those of them that contain the largest number of compounds are *kanji*-English or *kanji*-Italian dictionaries, because they list items of any frequency, not only the rarest ones. Spahn, Hadamitzky (1989) also list compounds under the constituent characters after the first. In contrast, a database like Tamaoka et al. (2015) proved nearly useless since it contains far fewer compounds than paper dictionaries. Among the many works that list lexemes phonetically, I mainly used the *Kōjien* (1976), the *Gensen* (1986) and the *Daijirin* (Matsumura 1988). My source of Ancient Greek roots and of their meanings is Montanari (2004).

The above-mentioned works are the sources I will implicitly refer to whenever I report on the results of a dictionary search. I will only

quote those works again if strictly necessary. When quoting a dictionary entry I will not indicate page numbers.

1.3.2 Conceptual Combination

The aim of the second stage is to formulate a tentative denotation of the hypothetical compound obtained from the *kun* assignment. To determine the morphological and semantic characteristics of compounds I will base my analysis on Štekauer's (2005; 2006) onomasiological model, integrating it with the classification of compounds by Bauer (2006; 2008; 2010; 2017) and Scalise, Bisetto (2011). As an application of such a theoretical framework to Japanese, the works I found most noteworthy are Kageyama (1982; 2011), Namiki, Kageyama (2016) and Kobayashi, Yamashita, Kageyama (2016).

In reconstructing the interpretive process of N+N compounds I will follow a conceptual combination model based on property mapping (extending the property of one concept to another) and relation linking – determining the thematic relation that may exist between the entities represented by a compound's constituents (Wisniewski 1996; Gagné, Shoben 1997; Shoben, Gagné 1997; Wisniewski, Love 1998). I will also consider the effects on the interpretive process done by distributional factors, namely family size and availability. I take the former notion from Schreuder, Baayen (1997, 121) and Gagné (2011, 267) and will adapt it here to correspond to the number of different compounds formed with any given *kanji* constituent. Its facilitating effect derives from the fact that the more productive a *kanji* is in a particular usage, the easier a construal based on that usage will be when the *kanji* occurs in an unfamiliar compound. For instance, the character 子 carries the general meaning 'offspring' and the special meanings 'child', 'fruit' and 'seed'. When interpreting a plant name, familiarity with the constituents' frequency in botany-related compounds will encourage readers to discard the construal 'child' and help them differentiate between the two remaining meanings.

Availability influences relation linking (Gagné, Shoben 1997; Gagné, Spalding 2004; 2006a; 2006b). I will consider it here as the frequency with which the nominal referent of a constituent *kanji* occurs in any given thematic relation across all the *kanji* compounds in the lexicon. Again, the higher the availability of a specific thematic relation for a *kanji* constituent, the easier it is to interpret the *kanji* as carrying that relation in unfamiliar compounds (Gagné, Shoben 1997, 74).

1.3.3 Encyclopedic Validation

The third and last stage of *kun*-glossing is finding an encyclopedia entry that satisfies the tentative denotation produced by the previous stage (Cohen, Murphy 1984; Murphy 1988; Allan 2006a; Gagné, Spalding 2006b). Its aim is to validate the entire process. This final search cannot be dispensed with because the ultimate purpose of *kun*-glossing is identifying the real-world extension of unknown *kanji* words. Semantic consistency and conceptual plausibility, assessed at the end of the second stage, do not suffice. In some of the cases studied in this essay, therefore, in order to show whether and how encyclopedic validation can bridge the gap between description and reference, I found it necessary to extensively discuss the encyclopedic knowledge available. This will mostly be the case with *enjin* 猿人 (*Pithecanthrope*) but consider, for instance, how crucial a role the specialised knowledge associated with *tanpakushitsu* 蛋白質 (protein) or *yōryokuso* 葉緑素 (chlorophyll) plays in the interpretive process.

In discussing the involvement of encyclopedic knowledge, I was influenced by the mental-file framework proposed by François Recanati (most relevant for me being 2012; 2013; 2017). However, Recanati's model is mainly focused on file dynamics and has the function of supporting his theory of direct reference. It deals with how singular concepts are established, updated, linked, merged, upgraded, etc., not with how entries are searched. Moreover, encyclopedia entries contain several types of information, differentiated on the basis of the rewarding epistemic relation that links them to their objects. Encyclopedic validation does delve, although to various degrees, into the semantic knowledge obtained testimonially from experts, into the information acquired visually from pictures, into that acquired by acquaintance and so on. But these distinctions too have no bearing on *kun*-glossing and are unnecessary for my analysis. I will, therefore, content myself with evoking the mental-file framework, if only loosely, and mostly to exploit the power of its metaphors.

1.3.4 The 'Best-Scenario Hypothesis'

To reach the successful outcome claimed by Suzuki, readers must be able to discard those seemingly valid interpretations that are actually erroneous. During the first stages, this pruning out relies on the frequency effects mentioned above. In each particular instance, the combinations best documented in the lexicon will suggest themselves as the most probable. Consequently, in order for readers to be positively affected by distributional factors, I will assume that they have a thorough knowledge of all the readings, meanings and uses of the

compound's constituents and that they are familiar with all the other compounds in the lexicon using those constituents.

This is an ideal state of affairs, for a speaker's actual mental lexicon is only a fraction of the contents of paper-grade dictionaries. Knowledge of the entire lexicon with the only exception of the particular item to be *kun*-glossed is also highly improbable. This is why I will often refer to the readers committed to *kun*-glossing as 'hypothetical'. Setting up such a scenario is necessary for two reasons. First, testing the validity of Suzuki's model means that a path to the successful outcome of the interpretive process must be found at all costs, no matter how difficult it may be. Readers should be equipped with the best possible tools for their endeavour. The 'best-scenario hypothesis' provides them with those tools. Second, to postulate a more natural kind of actor would force me to differentiate between levels of lexical knowledge in each particular instance, something not only impractical but never done by Suzuki himself. I will address the shortcomings of the Hypothesis in the conclusions.

1.4 *Kanji* in the Japanese Writing System

Suzuki's *kun*-glossing concerns present-day written Japanese. In order to understand it there is no need to know the history of *kun* reading or of the Japanese writing system in general. The history of individual lexemes is also irrelevant, for even when a lexeme is borrowed from Chinese as a spoken word, with an adapted Chinese pronunciation, it is only the graphic representation of it which is involved in *kun*-glossing. In this brief section, therefore, I will only synthesise the main features of the Japanese writing system in a fashion consistent with Suzuki's line of reasoning and with the notions he uses.

Taking the view of the Japanese writing system as a notation for representing the spoken language, Japanese content words are commonly written using only one or two Chinese characters (*kanji* or 'Sinograms'). Since this happens regardless of the phonological length and etymology of the roots, *kanji* is often considered to be 'logograms'. Functional morphemes are expressed phonetically, with a syllabic script. Intermediate between the two categories are affixes, written in *kanji* but partially bleached of lexical meaning.

Kun-glossing, though, works at the level of individual *kanji*, now conceived as graphemes carrying discrete meanings. This requires a change of perspective. It is *kanji*, not individual words, which is now taken as the unit of meaning (and, then, as 'ideograms'). Each *kanji* has one or more 'readings' or 'pronunciations', corresponding to the ancient Chinese spoken words of that meaning. These pronunciations

are labeled *on'yomi* 音読み (tonal readings), or ‘Chinese readings’.⁹ Lexemes made of one or more *kanji* pronounced by *on'yomi* are called *kango* 漢語 (Chinese words). Symmetrically, each *kanji* is also used to represent one or more lexemes of indigenous origin – *wago* 和語 (Japanese words) – having the same meaning as the Chinese word originally written with that character, or a meaning related to it. Those Japanese spoken words are thus considered to be the Japanese ‘pronunciations’ or ‘readings’ (the *kun'yomi*) of the associated *kanji*.¹⁰ What Suzuki considers to be the gloss of a character is exactly one of the *kun* readings of it. *Kanji* affixes are problematic for *kun*-glossing because their *kun* readings do not necessarily correspond to the meanings and functions they bear.

1.5 Notation and Abbreviations

For consistency with the names and titles listed in the references, in the Romanisation of Japanese I followed the Hepburn system, not the Kunrei one used in the literature of linguistics.

I will mark morpheme boundaries in glosses with the conventional symbols (mostly hyphens) and boundaries between the constituents of hypothetical compounds with ‘plus’ signs. The possible phonological realisations of hypothetical compounds I will write in *Italics*, preceded by an asterisk, with no morpheme boundary shown.

The only abbreviations I will use are N for ‘noun’ and V for ‘verb’.

⁹ When a *kanji* has more than one Chinese reading it means that lexemes using that Sinogram were imported at different times in different phonological realisations, depending on the Chinese dialect that was dominant among the Chinese cultural *élite* at that time. This point is of no importance to Suzuki’s argument.

¹⁰ *On'yomi* and *kun'yomi* 訓読み are phonological properties possessed by individual *kanji* (Morton, Sasanuma 1984, 39). Suzuki (1975, 191) considers this fact to represent a “Janus-like duality” of *kanji*. The notion of *kun* reading is, therefore, distinct from that of *kundoku* 訓読, which is a way of voice-reading Classic Chinese texts (*kanbun* 漢文) using Japanese words and grammar (that is, of practically voice-translating them) with the guidance of a system of online glosses and diacritics. *Kundoku* is also irrelevant to Suzuki’s reasoning.

2 Enter the *Pithecanthropus*

In his 2014 book, Suzuki expounds the merits of *kun*-glossing by means of several anecdotes, mostly taken from his experience in the USA. The episode he recounts first and in greatest detail involves the term ‘*pithecanthrope*’ (Suzuki 2014, 208-10).¹¹ Suzuki was at Yale giving a lecture on “Why Japanese *kanji* have two different readings, the *on* and the *kun* reading” in front of tens of students, researchers and fellow lecturers, all from the humanities faculty. He wrote the word <*pithecanthrope*> on the blackboard and, prompted by a puzzled student, asked his audience the meaning (*imi* 意味) of it. To his great astonishment, no one could answer, or even formulate a rough guess based on the form of the word.

This is because, Suzuki reasons in his book, the term ‘*pithecanthrope*’ or ‘*Pithecanthropus*’ is rare and unfamiliar to English speakers. It is made of two single-root stems, *pithec-* and *anthrop-*, meaning ‘monkey/ape’ and ‘man’ respectively. In theory, this would make it possible to decode the compound compositionally. The problem is that, as in most scientific terms, those roots are from Ancient Greek, not English. They are so uncommon that they have no meaning associations in the ‘linguistic consciousness’ of even the most learned English speakers (Suzuki 2017, 90; see also Shibatani 1990, 147), unless the latter knows them as part of their specialisation.

The Japanese equivalent of ‘*pithecanthrope*’ is 猿人, a two-*kanji* word. When the term was added to the Japanese lexicon, it acquired the pronunciation *enjin*, on the basis of the Chinese readings associated with the graphic constituents.¹² Thus, ‘*pithecanthrope*’ and *enjin* 猿人 are homologous in that they are both neoclassical compounds, terms formed with non-native roots imported from whichever prestigious foreign language is locally used as the metalanguage of science. Being made of two Chinese graphemes and adapted morphemes, though, 猿人 is even more alien to Japanese people than the Latin-script ‘*pithecanthrope*’ is exotic to English native speakers.

But now consider the “average Japanese” person (Suzuki 1975, 190), ignorant enough as to have never seen the “big word” (189) 猿人 or heard the term *enjin*. Ordinary people like this do, nevertheless, have knowledge of the *kun* readings of the compound’s constituent *kanji*, because they learned them at school as part of their basic education. Since each reading is an indigenous lexeme representing the

¹¹ Suzuki touches on the issue of the *pithecanthrope* in 1975, 179 and 1990, 138-9 but only addresses it extensively in 2014, 208-10.

¹² I could not ascertain whether this lexeme was an original Japanese calque of ‘*Pithecanthropus*’ or was taken from Chinese together with the spoken word, *enjin*. This point is irrelevant to this discussion, though.

concept its *kanji* stands for, knowing it amounts to knowing the meaning of that *kanji*. The *kun* readings of 猿 and 人 are respectively *saru*, meaning ‘ape’ or ‘monkey’, and *hito* ‘person’, ‘human being’, ‘people’.

Here the glossing function of *kun’yomi* kicks in. By assigning the two *kun* readings to their respective *kanji* as semantic glosses and joining them in the hypothetical compound *saru+hito*, any Japanese reader can easily derive a tentative compositional meaning of 猿人, namely (Suzuki 2014, 209) “a human being resembling an ape, an ape resembling a human being” (*saru mitai na hito, hito mitai na saru* さるみたいいな人、人みたいなさる). With this clue – Suzuki (2014, 210) says – “by using their heads”, a Japanese person will be able “to land close to the thing” (*atama o megurase, seikai (ni chikai mono) ni tōtatsu suru* 頭を巡らせ、正解(に近いもの)に到達できる) for which the unfamiliar lexeme actually stands. According to Suzuki (2014, 208), that reader will easily retrieve the concept of “the ancestor of humanity” (*jinrui no sosen* 人類の祖先), which even junior high school students possess. Thus, *kun* reading provides Japanese people with a powerful tool for accessing the meaning of difficult scientific terms even when they do not know how to read them.

2.1 Unearthing the *Pithecanthropus*: The Process of *Kun*-Glossing

As outlined above, the processing of ‘pithecanthrope’ is based on several presuppositions and is sustained by a series of inferences. In order to determine the features of those presuppositions, the preconditions and key points of the inferential process, and to apply the same interpretive method to other lexemes, I will, in this section, reconstruct the decoding of <猿人>, up to the output ‘ape-man’ and the final, combined concept of ‘ancestor of humanity’.

2.1.1 A Word

Suzuki (2014, 209) sets a stage on which a hypothetical Japanese reader faces a previously “unknown” combination of two *kanji* (*enjin to iu kotoba o sore made shiranakatta hito* 猿人という言葉をそれまで知らなかった人). For Suzuki’s argument to unfold, the term ‘unknown’ must be taken to denote the lack of *enjin*, a *kango* corresponding to the pronunciation of <猿人>, in the mental lexicon. I will use the terms ‘unknown’, ‘unfamiliar’ and ‘novel’ interchangeably.

Confronted with <猿人>, the reader must first determine that the two characters form a single lexeme.

Suzuki introduces 猿人 out of context, associates it with the English nominal ‘pithecanthrope’ and so implicitly declares it a lexeme.

But, in real life, when *kun*-glossing is supposed to be actually used, a reader must acquire the information necessary to identify 猿人 as an independent lexeme from the context. This information might be, for example, that syllabic *hiragana* characters expressing functional morphemes surround 猿人 and, therefore, isolate it as a carrier of semantic meaning. Or, alternatively, that 猿人 is joined to other *known* lexemes, so that it stands out as a short string of undecoded symbols. Awareness of the fact that many Japanese lexemes are formed of two *kanji* also plays a role.¹³

Assigning the status of ‘word’ to 猿人, either over the testimony of an expert like Suzuki or by inference, has an important implication. The communicative act of merely using the term, or of openly instigating a *kun*-glossing process of it, as Suzuki (2014, 208) does, makes it clear that 猿人 represents a public concept, a well-structured term with a specific intension and a life in a community of experts, not the product of the random coupling of two elements from which some meaning might somehow be extracted. Hence, if there is a word, a denotatum thereof must also exist whose cultural salience is high enough to grant it a name and an entry in the lexicon. At the conceptual combination stage, this understanding will assist the reader in dismissing those interpretations of the link between the two constituents which they may deem too rare or improbable to be represented by a fully-fledged lexeme, albeit unknown to them at the moment.

2.1.2 A Noun

Orthographic information also facilitates determining the word class of 猿人. The lexeme might be easily categorised as a nominal because, when in context, it is followed by noun-following particles and not by verbal, adjectival or adverbial morphemes (Suzuki 1963, 16). Suzuki takes for granted, therefore, the fact that his hypothetical reader will be able to identify 猿人 as a noun on syntactic grounds only, regardless of any possible preliminary semantic decomposition of it. This runs in parallel with the fact that he associates 猿人 with ‘*pithecanthrope*’, which he introduces as a noun to his English-speaking audience.

¹³ This is common knowledge but figures are rarely given. Yokosawa, Umeda (1988, 377) report that two-*kanji* words form 70% of the Japanese lexicon; more than 50% according to Kess, Miyamoto (2000, 68). See also Tollini 1992.

2.1.3 An Unknown Noun

Next, the readers have to consult their mental lexicons, try to retrieve an entry having the graphic address <猿人>, fail, and so recognise this two-*kanji* word as an ‘unknown’ noun. In Suzuki’s argument this is the most basic presupposition of all, for it is exactly to reverse such a state that *kun*-glossing is activated. This negative constraint implies that *kun*-glossing should work for any unfamiliar lexemes, regardless of their structure, usage and frequency in the language. It represents an implicit generalisation and gives *kun*-glossing its epistemic significance.

2.1.4 Two Known Sinograms and the Respective *Kun* Readings

For *kun*-glossing to work, readers are required to possess mental entries for 猿 and 人. Each entry must contain an association of graphological (the *kanji*) and phonological *kun* forms: 猿 = *saru* and 人 = *hito*. Moreover, unless the meanings of *kanji* are retrieved directly from the graphic form, without any need for *kun*-glossing, a separate lexical entry is also necessary wherein the phonological form is associated with a meaning. Using English semantic glosses, the two relevant entries consist of the associations *saru* = ‘monkey/ape’ and *hito* = ‘human being’. In other words, Suzuki’s model implies that a concept is not stored along with its graphological specification in a single lexical entry. Rather, it is accessed via the Japanese phonological form associated with it, as if there were two distinct lexicons. This view on the organisation of lexical knowledge (compare it with Allan 2006b) is consistent with Suzuki’s general view of Chinese characters as alien graphemes (cf. Suzuki 1990, 130, 139). Hence, the meaningfulness of *kanji* only stems from an arbitrary association of graphic forms and Japanese sounds. In Suzuki’s view of 2014, nothing in the shape of a *kanji* may help retrieve its meaning.¹⁴

It should be added that, according to Suzuki’s model, the reader might have mental entries for 猿 and 人 in which the graphological forms are associated with *on* phonological forms: 猿 to *en* and 人 to *jin* or *nin*. This prior knowledge is useless, though, because *on* readings do not give access to the lexicon, given that any particular form is associated with too many Sinograms (Suzuki briefly addresses this topic in 1963, 14). For instance, in a medium-to-large size *kanji* dictionary, such as Tōdō (1980), with about eleven thousand *kanji* entries, 79 items are listed under the *on* reading *en*, 72 under *jin* and

¹⁴ Earlier on, however, Suzuki (1990, 157-60) had proposed that the *kanji* components or ‘radicals’ (*bushu* 部首) are meaning-carrying graphemes expressing a semantic classification of the world. As such, he affirmed, they are even glossable by themselves.

28 under *nin*. Without any prior knowledge of the specific meaning and usage of each *kanji* to help discard inconsistent combinations, the forms *enjin* and *ennin* could correspond to any one of nearly eight thousand graphic words. Therefore, although the *on* readings of *kanji* did use to carry meaning, because they represent ancient Chinese spoken words, in present-day Japanese they have no semantic content (Suzuki 1963, 31-2; 1990, 145; Nagano, Shimada 2014, 355 quoting Takashima 2001, 242-3). In summary, in Suzuki's view, the *on'yomi* of a *kanji* merely allows the reader to articulate some sequence of sounds in conventional correspondence with it. Meaning is assigned to a *kanji* and related *on* reading through the *kun* gloss only – this is the core tenet of *kun*-glossing.

Knowledge of the two *on* readings will, of course, allow the retrieval of a phonological word, *enjin* or *ennin*, the former indeed corresponding to an entry in the lexicon. But, again, this is irrelevant, because the very first assumption in Suzuki's argument is that the reader does not know the meaning of *enjin* to begin with.

2.1.5 No Noun **Saruhito* or **Sarubito* in the Lexicon

Japanese has a word-formation mechanism whereby two nouns can be combined in a compound. The *kun* readings *saru* and *hito* would therefore suffice to retrieve an N+N compound directly from the lexicon having the actual phonological realisation **saruhito* or **sarubito*.¹⁵ But now a further constraint is required in Suzuki's line of reasoning: in the lexicon there is no lemma of either phonological form, or, alternatively, the reader does not know the lexeme even if it exists in the language. This is one more constitutive condition, otherwise the lexicon could just be accessed directly and *kun*-glossing would not be needed. The scanning of the mental lexicon for **saruhito* or **sarubito*, then, produces a null output.

2.1.6 *Kun*-glossing 猿人

Having retrieved the *kun* reading of each *kanji*, the reader now assigns it to that Sinogram as the semantic gloss of it. This amounts, of course, to using the readers' target language (i.e. Japanese) as the

¹⁵ For the end of this essay, distinguishing between the two alternate phonological realisations is irrelevant, since the *wago* compound does not exist. Suffice it to say that, if it did exist, the voicing (or *rendaku* 連濁) of the /h/ phoneme, producing **sarubito*, would not occur if the compound were of the coordinate type (Kageyama 1982, 238 quoting Akinaga 1981) but, conversely, would most likely occur if the compound were subordinate or attributive (see classification in Scalise, Bisetto 2011).

glossing metalanguage.

It should be noted that a problem would arise here if a *kanji* had more than one *kun* reading but this issue is never addressed by Suzuki.¹⁶ For 猿人 the matter is irrelevant, because the two *kanji* only have one *kun'yomi* each. At this level of simplicity, *kun*-glossing merely consists of naming a *kanji* by retrieving the citation form of it. But for other lexemes the presence of several *kun* readings might weaken Suzuki's model, while that of more *kanji* sharing the same *kun* reading might even invalidate it. I will discuss this problem later. For the time being, suffice it to say that Suzuki's actual choice of *kun* glosses strongly suggests that, for him, each *kanji* has a sort of main, underlying meaning, or *Hauptbedeutung*, phonologically expressed by one of the *kun* readings of that *kanji* and easily produced as the citation form of it.

2.1.7 A Compound

Suzuki's readers now have to treat the two conjoined glosses as a hypothetical true compound and apply an analysis. To derive meaning from it, a grammar must be adopted in which the two components are bound in a syntactic relation. This operation is an integral part of *kun*-glossing, if 'glossing' is intended as the building of a metalanguage with both simplex sign-meaning symbolic associations and rules to make those symbols interact and produce complex meanings.

Before retracing the conceptual combination process in question, though, it should be noted that the performers of *kun*-glossing are decoders, not coiners of a new lexical item. They deal with an existing combination of concepts, which they know is viable and given in the lexicon. In order to infer it, they may only work with the output of *kun* assignment, whatever it is, and however poor. Their job is to force *saru* and *hito* into all possible syntagmatic and semantic relations and judge the success of each attempt by the meaningfulness, contextual coherence and encyclopedic consistency of the result. In other words, they are not engaged in an onomatological act of naming under onomasiologic rules, but in a semasiological act of interpretation. The environment in which readers have to orient themselves whenever they encounter unfamiliar compounds, though, is the result of onomasiologic processes of word formation. *Kun*-glossing might only work if decoders are aware of the differences among those processes, and intuitively retro-engineer them (as suggested by Stækauer 2011). In the next sections I will proceed under such an assumption.

¹⁶ For an acknowledgement of the problem, cf. Suzuki 1990, 143. Here the author conveniently directs the reader to Morioka 1987.

2.1.8 An N+N Compound Noun

Readers have already identified 猿人 as a noun on morphologic and syntactic grounds. Upon glossing the two constituents, they now generate a hypothetical compound the elements of which, *saru+hito*, are also nominals. Since the constituents belong to the same word class, the compound should also belong to that class, i.e. should be categorially endocentric. Moreover, *saru* and *hito* are also both first-order nouns, denoting first-order entities (persons or things, Lyons 1977, 442, 446) or substances.¹⁷ Conceptually, this suggests that 猿人 might be a first-order noun itself, rather than an abstract noun denoting a property or an eventuality. More specifically, it suggests that 猿人 is also semantically endocentric and, as a hyponym of either *saru* or *hito*, denotes a type of monkey/ape or human being.¹⁸ The task is now to determine how these two concepts are associated.

As any two-constituent compound, Japanese N+N compounds fall into two classes with respect to inner structure. Either they show internal dependency, so that one element represents the morphological head of the parent word and the other one functions as a subordinate specifier, or they have no internal structure. In the following sections I will deal with these two kinds of compound in reverse order.

2.1.9 A Coordinate N+N Compound?

A compound in which neither element controls the other(s) is coordinate. The constituents of a coordinate compound are of the same category and designate separate individuals. For Japanese, Kageyama (2011, 513) distinguishes two types of semantically-endocentric coordinate compound:¹⁹

- Separate-reference type. The compound defines a set and the coordinated elements each refer to a member of it. Anaphorically, pronouns or adverbials – like *sorezore* それぞれ (each), *izureka* いずれか

¹⁷ I take the notion of ‘substance’ from the onomasiological approach to word formation (Stekauer 2005; 2006).

¹⁸ The problem of distinguishing monkeys from apes only arises because English, the metalanguage used in both this essay and in the Yale classroom by Suzuki (2014, 208-9), operates a lexical distinction between those classes of primates. Japanese and Ancient Greek, on the contrary, do not see that distinction, and both neutralise apes and monkeys under the roots *saru* and *pithec-* respectively. Whenever the distinction is unimportant for the accuracy of *kun*-glossing I will translate *saru* with ‘ape’.

¹⁹ Kageyama (2011, 513) collectively refers to Japanese coordinate compounds as ‘dvandvas’ and classifies them into three groups, although the third group, which he dubs the ‘holistic type’, is actually semantically exocentric. In addition to the quoted sources, Japanese coordinate compounds are discussed in a fashion relevant to this essay in Tollini (1992, 91) and Kobayashi, Yamashita, Kageyama (2016, 104).

(either), *izuremo* いずれも (both), *isshoni* 一緒に (together): cf. Kageyama 1982, 235; Namiki, Kageyama 2016, 213 – can pick the internal constituents in any combination thereof. It is in this sense that these compounds are considered ‘multi-headed’ while, in fact, they have no internal headed structure. Using this interpretation, *saru+hito* might stand for a specific, referential set formed by both individual ape(s) and human being(s).

However, as Namiki, Kageyama (2016, 213-14) point out, in both English and Japanese the order of the coordinated members is constrained by cognitive and cultural factors. In a two-noun compound, the left-hand constituent tends to be socio-culturally more salient. The most quoted Japanese example of this type of co-compound – *oya-ko* 親子 (parent-child) – well illustrates this principle. Thus, it is improbable that Suzuki’s hypothetical reader would interpret *saru+hito* as a separate-reference compound, because doing so would imply giving apes priority over humans. If it existed, a co-compound so associating humans and apes would rather be *人猿 *hito+saru*, with the element ‘human’ occurring first.

- Co-participant type (Bauer 2008, 6 and 2017, 90-1). The compound designates entities associated in a reciprocal relationship, the type of which is to be further specified by co-occurring expressions like ‘love’, ‘match’, ‘friendship’ etc. I judge this type to be a subtype of the previous one. Kageyama (2011, 514) names these compounds “relational”, while Lieber (2011, 89, 91-2) considers them to be subjected to a “relationship interpretation” (where they are actually considered exocentric). I will avoid such expressions because of the confusion they ingenerate with subordinate compounds, which are subject to a relation-linking interpretation.

Suzuki (2014, 208) discusses 猿人 in an artificial environment and out of context. But in context, being of the co-participant type, 猿人 would occur in complex expressions like **enjinsensō* *猿人戦争 (the ape-human war) or **enjinkankei* *猿人関係 (the relationship of apes to humans), where the right-hand two-*kanji* heads mean ‘war’ and ‘relation’ respectively. Or it would be associated with the suffix *kan* 間 (space) as in **enjinkan* *猿人間 (between apes and humans). In these cases, the readers would have no problem in interpreting 猿人, provided they know the meaning and function of *sensō* 戦争 (war), *kankei* 関係 (relation) or *kan* 間 (space).

The principle of socio-cultural salience can be appreciated in co-participant compounds as well. Consider, for instance, the hypothetical **enjinkōhai* *猿人交配 (ape-man hybridisation), in which apes and humans are associated by means of the notion *kōhai* 交配 (hybridisation). Putting *saru* 猿 (apes) first would violate the principle of salience. This compound does not exist in Japanese, neither does *人猿交配 (man-ape hybridisation), in which the order of ‘apes’ and ‘humans’ is reversed. In English, however, where the same principle of

saliency applies (Cooper, Ross 1975), an expression ‘human-chimpanzee hybrid’ does exist in which the element ‘human’ comes first, in accordance with the higher saliency given to human beings over apes. Hence, the saliency constraint – to the effects of which Suzuki’s hypothetical readers are supposed to be exposed – tends to exclude a viable construal of *saru+hito* as coordinate.

2.1.10 An N+N Compound with the Structure ‘Modifier + Head’?

If a compound has internal structure, so that the two elements bear syntactically and semantically distinct functions in relation to each other, it may be headed by either constituent.

In Japanese, left-headed *kanji* words do exist, but are a reflection of Chinese syntax and are common only among verbal nouns (Kageyama 2011, 514; Namiki, Kageyama 2016, 210, referring to other authors). Moreover, among Sino-Japanese two-*kanji* nominals – to which *enjin* 猿人 actually belongs – and among the compound nouns formed by conjoining two *wago*, Kobayashi, Yamashita, Kageyama (2016, 104) can only identify one left-headed noun. It is *satsujin* 殺人 (manslaughter). In this V+N *kango*, the head 殺, glossed *korosu* (killing), represents an action, a second order entity. A compound like *saru+hito*, wherein both elements represent substances or first-order entities and are represented by first-order nouns, can only be right-headed. The right constituent functions both as the categorial head, determining the word class of the parent term, and as the conceptual or semantic one, specifying the class to which the denotatum of the compound belongs (Tollini 1992, 92).

Accustomed with semantically endocentric structures, Suzuki hypothetical readers will now interpret the component *hito* as the head and *saru* as the modifier. They will assume that, as a hyponym of *hito*, *saru+hito* refers to a human being about which something is predicated in relation to apes. When so analysed, compounds are interpreted under two strategies.

2.1.11 Subordinate or Relation-Linking Interpretation of *Saru+Hito*

One interpretive strategy consists of assuming that a thematic relation holds between the modifier and the head (Wisniewski 1996; Wisniewski, Love 1998; the strategy is called ‘relation-based approach’ by Gagné, Shoben 1997; Shoben, Gagné 1997). When this happens, the compound is sometimes called ‘subordinate’ (as in Bisetto, Scalise 2005, 326), because the constituents receive a thematic role, such as when arguments are governed by a verb. In the case of N+N words

like *saru+hito*, of course, it is not possible to determine exactly in what relation the two components might be, for no predicate specifying that relation is realised. In onomasiological terms (following Štekauer 2005), the head *hito* represents the base and the rest of the word constitutes the specifying mark. *Saru*, however, only represents the determining component (named ‘motive’) of the mark. The other possible component of it, the determiner, which has the function of specifying the action, process or state linking *saru* to *hito*, is unexpressed. This results in ambiguity (Lyons 1977, 538; Bauer 2006, 722; Štekauer 2006, 35) or opacity of intension.

In order to reduce it, the first step is to consider those thematic relations that may occur between any two first-order entities like apes and humans. A comprehensive, close list of this kind is, of course, impossible to build, the task resembling the hunt for an exhaustive list of semantic roles (a short discussion of the problem in Bauer 2011, 353). Nevertheless, using a mixture of *a priori* intuition and *a posteriori* data validation, a classification of the most salient relational categories can, indeed, be made (Gagné, Shoben 1997 after Shoben 1991). Then, a semantic filter is applied to reduce the range of viable thematic relations involving apes and humans. The filter works analytically on the basis of the intrinsic properties of the two entity types, not over the frequency of the most salient thematic relations involving them which are observable in actual compounds. For instance, a locative relation (either apes or humans are located in the other, in the concrete domain of space or in the abstract domain of time) and a causal relation (either entity produces or generates the other one, as a virus causes the flu) can be discarded as not compatible with the semantics of the two constituents. Out of context (in the scenario envisaged by Suzuki), though, this filter cannot be accurate. The relation ‘material’ (an ape or a human physically made of one or more specimens of the other living entity) might be acceptable in a work of fantasy or science fiction. Thus, the range of semantically consistent types of relationship may not be significantly reduced in this way.

Thematic role assignment may be further channelled by a tendency of considering humans as agents in relation to other non-human entities, both animate and inanimate. Thus, *saru+hito* might denote a class of human beings specialised in the hunting, capturing, or keeping of apes, or in taming apes and performing with them in festivals. **Sarubito* could be the name of the ‘horse healers’ (actually named *sarumawashi* 猿廻し) who used to work with monkeys (Hirose 1984). Yet, the range of conceptually plausible interpretations is still very wide. Moreover, inverting the agent role is also possible, leading, for instance, to a construal of *hito* as the typical prey of anthropophagous apes. And there is a large number of viable construals in which apes are neither agents nor prototypical patients, so that the compound could denote actors performing in

ape roles or in ape costumes, tribes having monkeys as totemic animals and so forth.

At this stage, the most efficient tool for reducing the range of viable interpretations consists of provisionally construing each constituent under one of the thematic relations most frequently associated with it in compounds (Shoben 1991; Gagné, Shoben 1997; see also Štekauer 2005, 216 on distributional constraints). The mental lexicon is searched for all the items containing either 人 as a head or 猿 as a modifier, and the relations that link 人 and 猿 to the other co-occurring constituent(s) are noted. The available relations then compete with each other, with the most frequent one(s) being preferred. In such an operation, it is the modifier which is considered more salient, perhaps because it is processed first (Gagné, Shoben 1997, 74-4; Shoben, Gagné 1997, 40; Tamaoka, Hatsuzuka 1998, 294). As a modifier, then, is *saru* commonly associated with any specific thematic relation? A dictionary search returns 24 two-*kanji* compounds and 3 three-*kanji* compounds having 猿 in the left-hand position. Excluding 猿人, the Sinogram 猿 occurs with a head denoting a concrete object or substance in only 10 of them, none of which is subordinate.²⁰

The distribution of 猿, therefore, does not support a relational interpretation. As regards the head, 人, the 223 compounds wherein 人 occurs as a right-hand constituent contain only 19 items in which the left-hand element denotes a substance as it does in 猿人.²¹ Only 7 of them are relational compounds, with 人 ‘person’ carrying the agent role in 5 and that of possessee in 2. But only in one of these 7 compounds does the first *kanji* denote an animate entity: *fujin* 夫人 (husband-person = ‘wife’, under a thematic relation of possession).²² The character for ‘person’, 人, is more often found carrying the agent or theme role when it occurs in a suffixal function as the rightmost element of three-*kanji* words. Indeed, most compounds of this type display the complete three-constituent onomasiological structure of base, determiner and determining component whereby the standing thematic relation is overtly realised.²³ In two-*kanji* compounds, a human agent

20 The 10 items are *endō* 猿狻 (monkey), *enso* 猿狙 (monkey), *sarumata* 猿股 (under-shorts), *sarumen* 猿面 (monkey face), *sarumanako* 猿眼 (large, sunken eyes), *enkō* 猿猴 (monkey), *enpi* 猿臂 (a long arm), *sarugutsuwa* 猿轡 (mouth gag), *enchō* 猿鳥 (monkeys and birds) and *enkaku* 猿鶴 (monkeys and cranes).

21 I consider not to designate substances those *kanji* denoting conditions (‘love’, ‘disease’), actions (‘travel’), immaterial results of actions (‘poetry’), places (‘mountains’) and locations (‘West’, ‘outside’).

22 The remaining 6 items are *dojin* 土人 (soil/ground-person = ‘aborigen’, ‘savage’); *funabito* 舟人 (boat-person = ‘boatman’); *somabito* 杣人 (timber-person = ‘woodcutter’); *funabito* 船人 (boat-person = ‘sailor’); 家人 (house-person = ‘family people’, if read *kajin*, or ‘retainer’/‘vassal’, if *kenin*); *chajin* or *sajin* 茶人 (tea-person = ‘eccentric’).

23 I only provide two examples here: *shukuhakunin* 宿泊人 (lodge-lodging-person = ‘lodger’) and *kenbutsujin* 見物人 (watching-thing-person = ‘spectator’/‘sightseer’).

or theme in the head position is commonly expressed by means of 者, 家 or 手, not of 人, and members of organisations are named with 員.

Thus, a relational interpretation of *saru+hito* finds no support in the relative frequency of the constituents. This is probably due to similarity, intended as the sharing of definitory properties. While a very high degree of similarity between the head and the modifier concepts favours hybridisation or an interpretation as a coordinate compound, a relational interpretation is favoured by low similarity (Wisniewski 1996, 436, 442-6). Apes and humans, albeit dissimilar, are not dissimilar enough to motivate the construal of a relationship in which the entity represented by the head assumes such a strong and obvious controlling role over that represented by the modifier such as not to require explicit expression in the compound. In this sense, were primatologists so salient in society as to be named with a compound of the lowest possible complexity and of high opacity, *saru+hito* could well refer to them.

2.1.12 Appositive or Property Mapping Interpretation of *Saru+Hito*

Suzuki's hypothetical readers are, therefore, left with two concepts which are both too different to be amalgamated into a co-compound and too similar to support a head-modifier thematic relation. This is the 'Goldilocks condition' that favours property mapping. This strategy consists of construing the modifier noun as referring not to the entity it denotes but to a quality thereof, and in mapping that quality onto the head concept (Wisniewski 1996, 442). The compounds of this type are defined 'appositive' (Bisetto, Scalise 2005, 327; Scalise, Bisetto 2011, 51). Paired concepts of the right level of similarity facilitate property mapping because they have homologous attributes, alignable on the same dimensions and with merely different values when scalable (Wisniewski 1996, 443, 445, 449). Under this condition, nouns are particularly versatile as modifiers because they ground situations directly in experience and, unlike adjectivals, allow not only single properties but also composite sets of them to be ascribed to the head (Wisniewski, Love 1998, 199).

Moreover, I suggest that interpretation is channelled toward property mapping by the very absence of the determiner component, whose presence would have made explicit a thematic relation and forced a linking interpretation. The determiner is omitted here because it would be semantically too bleached to be worth being

Notice the opposite morphological headedness within the right-hand two-*kanji* onomasiological marks.

expressed, as if it were a copular element ('be like'). In other words, a zero determiner is a significant feature in itself. It signals that a determiner is absent because it is unnecessary, and that it is unnecessary because it is actually deprived of semantic meaning.

Taking this view, in a simple compound like *saru+hito*, formed by two first-order nouns denoting animate entities, the modifier *saru* is to be construed as a predicate expressing a quality to be incorporated into the head *hito*. The mapping can occur in the 'physical aspect' domain and between the slots (Cohen, Murphy 1984, 45) or attributes 'gait', 'posture', 'hairiness'; or under 'skills' (like 'tree-climbing'), 'strength' and so on. Given the preeminence of the sense of sight in humans, appearance or visually detectable behaviour are indeed the best candidates to be mapped, to the effect that *saru+hito* is most likely interpreted as denoting a human being that looks like an ape.

2.1.13 A Semantically Exocentric Compound of Opaque Meaning?

So far, though, the reader cannot be absolutely sure whether 猿人 truly designates a type of human being. It might be semantically exocentric and refer to some other entity altogether. Compounds of this kind, which are not hyponyms to either of their constituents, are borne of metonymic and metaphoric processes and encode meaning outside the principles of compositionality (Namiki, Kageyama 2016, 214). There is no emergency procedure to infer their denotation.

Exocentricity is, of course, never contemplated by Suzuki, either for 猿人 or for any of the compounds he submits to *kun*-glossing in his various works. This weakens his general argument because two of the compounds he discusses, *tanpakushitsu* 蛋白質 (protein) (cf. Suzuki 1990, 135) and *hantō* 半島 (peninsula) (cf. Suzuki 1975, 189), are actually semantically exocentric. For *saru+hito*, though, rather than be a strategic omission on his part, ignoring exocentricity is justified on several grounds.

In a typological survey involving a large number of languages, including Japanese, Bauer (2010) individuates five recurrent major patterns in the semantics of exocentric compounds, which largely correspond to the types of syntactic relation between the inner constituents that also occur in endocentric compounds. Four are the exocentric patterns licensed by an N+N structure. None is appositive. As Scalise, Bisetto (2011, 50) point out, appositive exocentric compounds are indeed "not easy to track" typologically, to the extent that the authors provide no example of them in their table.

According to Bauer's (2010) classification:

- In the *bahuvrihi* or possessive pattern (Bauer 2010, 169-70), the compound as a whole denotes an entity which possesses the

object represented by the inner, or morphological, head, whereas the other constituent expresses a property of the head. Under this interpretation *saru+hito* could denote a non-human entity characterised by the possession of human beings with ape-like characteristics. The probability that there is an entity of that kind of such a great salience as to be represented by so opaque a compound seems very low.

- In the hybrid pattern (Bauer 2010, 172-3), the compound is coordinate, designates a singular object and does not allow for separate reference. Kageyama (2011, 153) labels this the “holistic *dvandva*” type and considers it endocentric, as Wälchli (2005, 154) does, while for Lieber (2011, 91) a semantic pattern of this kind comes under an exocentric ‘collective’ interpretation. Were *saru+hito* of this type, the compound might denote:

- a. A taxon whose parts are all kinds of monkeys, apes and humans. This higher taxon does actually exist (it is the primate suborder of Anthroidea, also known as Simii), but has nothing to do with the *Pithecanthropi*.
- b. A hybrid entity whose properties are an amalgama of the properties of human beings and apes. This entity would be neither ape nor human, but an object halfway between them, equidistant from both. Hybrids, though, tend to occur when the constituent members of a compound have the highest similarity (Wisniewski 1996). Although Bauer (2010) does not contemplate this possibility, the hybrid entity could also consist of a property metonymically defined by properties extracted from both humans and apes alike, and best (or prototypically) represented by them: for example, ‘intelligence’.

In the cases (a)-(b), no sense relation occurs between the compound itself and either constituent.

- c. A composite class having apes and humans as the most salient members. Under this interpretation, both *saru* and *hito* would be hyponyms of the parent compound, a sense relation opposite to that occurring in semantically endocentric co-compounds. In this case *saru+hito* could well denote the order of primates.

The three subtypes (a) to (c) are still subjected to the cognitive and cultural constraint pointed out by Namiki, Kageyama (2016, 213-14). Having the constituent *saru* in the left-hand position would make the coordinate pattern anomalous from the point of view of cultural salience.

- In the transpositional pattern (Bauer 2010, 171-2), the compound is subordinate and designates a relationship between entities which is stereotypically represented by the relation between the two constituents’ referents. It

is this relation, left unexpressed, which gives meaning to the compound. This pattern seems to correspond to an exocentric interpretation of Kageyama's (2011, 514) "relational dvandva". Under this interpretation, if 'mother-child' is taken to mean 'love' then, with some stretch of the imagination, *saru+hito* could represent the concept 'strange sibling relation', 'disquieting physical resemblance' and so on.

- In the metaphorical pattern (Bauer 2010, 171-2), one constituent (here *hito*) functions as the head and represents something metaphorically associated with its referent. For example, an instrument, as in the English term 'deadman' (a log sunk in the ground used to anchor guy ropes). *Saru+hito* could therefore denote, for instance, a device capable of entrapping apes as efficiently as human hunters do.

Linguistically, none of the above interpretations can be excluded. But there is a strong factor that prevents the reader from interpreting *saru+hito* as exocentric: semantically exocentric compounds with either 猿 or 人 are virtually nonexistent in the lexicon. The dictionaries I searched list a total of 27 compounds having 猿 as the leftmost constituent. Of these only 4 N+V items are categorially exocentric and only one N+N item is semantically exocentric: *sarumata* 猿股 (monkey-crotch = 'underpants'). As regards 人, of the 70 compounds having an N+N structure in which 人 occurs in the second position, only 2 are semantically exocentric: *tenjin* 天人 (heaven-human_being = 'nature') and *sejin* 世人 (world-people = 'public world'). Quite simply, in Japanese neither 猿 nor, surprisingly, 人 seems to be endowed with enough metaphorical glamour to be drawn into the machinery of exocentric compounding. In this, 人 strongly contrasts with 手 (hand) and 家 (house), which are richly used to refer to human agents.

2.1.14 A Human Being Resembling an Ape, an Ape Resembling a Human Being

The processes I have extensively reconstructed above are not expounded by Suzuki as part of his treatment of *kun*-glossing. Nowhere does Suzuki go into so much detail. Rather, these processes model what must be going on in the mind of the reader if they are to carry the *kun*-glossing of 猿人 to the successful end envisaged by Suzuki.

Thus, the most straightforward interpretation of *saru+hito* ends up being property mapping, because the two constituents have the right degree of similarity. This interpretation yields 'a human being resembling an ape (in outer aspect or behaviour)'. Next is relation-linking.

Alas, out of context, the relation cannot be construed more clearly than ‘a human being doing something to apes’ or perhaps ‘a human being somehow involved with apes’. The least viable interpretation is as a hybrid: ‘an entity half-way between human and apes’. What constitutes ‘half-way’ remains obscure.

As it turns out, after having his hypothetical Japanese reader perform *kun* assignment and ponder a little, Suzuki (2014, 209) asserts without further ado that they will interpret 猿人 as “a human being resembling an ape, an ape resembling a human being” (*saru mitai na hito*, *hito mitai na saru* さるみみたいな人、人みみたいなさる).

Evidently, Suzuki has somehow dissipated all uncertainty and identified *saru+hito* as appositive. This is consistent with the fact that, as seen, ‘a human being resembling an ape’ is the most frugal interpretation of the compound. He only remains noncommittal about what feature of apes is to be mapped onto human beings. The puzzling thing is that he considers the head-modifier relationship of *saru+hito* to be reversible, so that *hito* may be equally construed as the head and as the modifier of *saru*. Moreover, he also implies that the two opposite interpretations are not mutually exclusive but are related in such a way as to actually sum up and express a reciprocal exchange or fusion of properties. This amounts to asserting that 猿人 is a hybrid, even if nothing in the processing of the term justifies such a conclusion. In the light of the lengthy discussion in the previous sections, it is clear that Suzuki’s ‘two-way’ or ‘reversed’ interpretation of *saru+hito* makes no sense. Property mapping only works one way. For the moment, though, I am going to treat Suzuki’s (2014, 209) interpretation as if it were valid.

2.2 Encyclopedic Validation and the Lack of Context

As noted, Suzuki builds up an artificial scenario in which 猿人 is to be processed out of context. It is exactly the lack of context that prevents interpreters from inferring, with more precision, which simian property is mapped over human beings or which thematic relation could link humans to apes. In these conditions, once the vague description of ‘a human being resembling an ape, an ape resembling a human being’ is obtained, the conceptual combination is complete and interpreters can only proceed with submitting the description to encyclopedic validation.

Context can be a destabilising factor in Suzuki’s model. Firstly, it might be rich enough to provide a satisfying inferential interpretation of a compound without the need to activate *kun*-glossing. Secondly, context would weaken the importance of lexical semantics, which is what *kun* assignment and conceptual combination rely on. This is why Suzuki evokes context only rarely, as when he fears that mere

kun-glossing could prove too weak. He does so, for instance, in the case of *hen'onsei* 恒温性 (homiothermal) and *kōonsei* 変温性 (poikilothermal) (Suzuki 2014, 211) or in relation with *shintōatsu* 浸透圧 (osmotic pressure) (Suzuki 1990, 29). In this essay I also consider the function of context as little as possible and only to give robustness to Suzuki's argument, as I have done when discussing how the word class of 猿人 can be determined. I will remark later how important the lack of context and the resulting vagueness are in relation to 猿人.

2.2.1 The 'Ancestor of Humanity'

Encyclopedic validation is the final stage of *kun*-glossing. Interpreters must check whether and how the description obtained so far is consistent with prior knowledge and integrates with it. This is the only way they can identify the actual extension of the term they have just encountered and give *kun*-glossing epistemic relevance.

In Suzuki's (2014, 208) narrative, upon searching their encyclopedia, interpreters now discover that the description matches their beliefs about an entity of great importance in human culture: "the ape-man (*Pithecanthropus*), ancestor of humanity" (*jinrui no sosen ni ataru enjin (pitekantropusu)* 人類の祖先にあたる猿人 (ピテカントロプス)). In order for this indexing to activate, the reader must, obviously, have the relevant entry. But, Suzuki (2014, 208) maintains, knowledge of an ape-man ancestor is so basic that even Japanese junior high school students possess it. The final effect of *kun*-glossing is that the reader can now re-name a pre-existing conception 'a creature that resembled both an ape and a human being and was the ancestor of humanity [about which I learned at school / from books / from TV / etc.]' with 猿人. Thus, thanks to *kun*-glossing, 猿人 proves to be informative even to Japanese children, while 'pithecanthrope', like many other semantically opaque "highbrow" (Suzuki 1975, 190) scientific terms, does not mean anything, even to American university professors.

2.2.2 Ernst Haeckel, the *Pithecanthropus erectus* and the *Anthropopithecus*

To explain how *kun*-glossing can help organise knowledge about the entity going under the name of 猿人, we first need to understand what the word 'Pithecanthropus' actually stands for.

'Pithecanthrope', 'ape-man' and 猿人 are not natural words, spontaneously coming into existence and grounded by the repeated use in a community of speakers. Rather, they are artificial compounds whose predecessor, 'Pithecanthropus', has a well-recorded history.

This history reveals that ‘*Pithecanthropus*’ was coined with a certain inner structure and a derived ‘literal meaning’ intentionally and for specific ends.

The term ‘*Pithecanthropus*’ and the German equivalent *Affenmensch* (Ape-like man) were coined by the German evolutionary biologist and embryologist Ernst Haeckel (1834-1919) and first appeared in *Natürliche Schöpfungsgeschichte* (1968, 508; 1880, 270-1).²⁴ Haeckel (1868, 507; 1889, 709-10) coined these terms in order to name a hypothetical primeval human which descended from humanoid apes (*Menschenaffen*), was similarly speechless (*alalus*), thoughtless and without consciousness, but had forelimbs and hindlimbs specialised in different ways and was therefore capable of upright walk.

Haeckel (1880, 293) deemed the existence of such an immediate precursor to “Genuine or Talking Men” to be necessary in light of Darwin’s theory of evolution, but fossil remains had yet to be found. He strongly suggested looking for this ‘missing link’ in Malay. At the time, that same hypothetical precursor was already referred to as *Anthropopithecus* (a term in which the two constituents for ‘ape’ and ‘man’ are inverted), to the effect that, when, in 1892, Eugène Dubois (1858-1940) discovered human fossil material with compatible traits in Java, he first dubbed the new species *Anthropopithecus erectus* – only to rename it *Pithecanthropus erectus* one year later. A similar hominid was discovered in the 1920s in Zhōukǒudiàn, near Beijing. Commonly known as ‘Peking Man’ – in Chinese actually *Běijīng yuánrén* 北京猿人 (ape-human of Peking) – this hominid received the specific name *Sinanthropus pekinensis*: in Chinese *Zhōngguó yuánrén běijīng zhǒng* 中国猿人北京种 (cf. van Oostersee 2001 for a history of the *Sinanthropus*), again with the constituent 猿人. It was only in 1950 that the remains of *P. erectus* and *S. pekinensis* were recognised to belong to one and the same species and reclassified as *Homo erectus* (in Japanese *genjin* 原人, glossable ‘origin-human’). Since then, the term ‘*Pithecanthropus*’ has progressively fallen into disuse. By the time of Suzuki’s (1975) first manifestation of interest in the *Pithecanthropus*, *Australopithecus* had already replaced *Pithecanthropus* as the hominin genus immediately ancestral to *Homo*. The entries for 猿人 in Japanese dictionaries all reflect the new taxonomy and mention *Australopithecus* as a concrete example of the now-desuete taxon.²⁵

²⁴ My main sources on the history of the *Pithecanthropus*-related taxa are Tattersall, Schwartz (2001, 150, 154); Thain, Hickman (2004); Tattersall (2009, 24-34, 58-72) and Wood (2011). The research on Haeckel is mine.

²⁵ In the *Dajirin* (Matsumura 1998) the entry for *enjin* 猿人 begins with: “the most ancient fossil human, which lived between one and three million years ago” (*hyakuman-sanbyakumannen ijō mae ni seisoku shita saiko no kasekijinrui* 一〇〇万〜三〇〇万年以上前に生息した最古の化石人類). Then, more data concerning geographical distribution, brain

I could find no actual occurrence of *Pithecanthropus alalus*, in Haeckel's *Natürliche Schöpfungsgeschichte* (1868) or *Anthropogenie* (1877). Haeckel's use of italics and frequent equiparation of the taxa 'alalus' and 'homo' also hint at the fact that he did not treat the term 'Pithecanthropus' as the name of a genus. However, '*Pithecanthropus primigenius*', a specific name, does occur once in the first edition of the *Natürliche* (1868, 514, Ch. 19), and 'Pithecanthropus' was considered a generic name in later use. Thus, it is justified to say that Suzuki's 'pithecanthrope' is the shortened appellative of a whole genus and a plain-language term designating all the species and living organisms belonging to it.

2.2.3 *Pithecanthropus*: 'Upright As a Human, Speechless As an Ape'

Pithecanthropus erectus and **Pithecanthropus alalus*, the multi-root specific names borne after the adoption of 'Pithecanthropus' as a generic constituent, are structured in accordance with the traditional guidelines of Linnaean nomenclature (Ereshefsky 2001). As such, they are not arbitrary identifiers but carry metadata regarding the object they name. They encode:

- In the shared generic constituent, that, in the line of descent of Man, the genus occupies the intermediate position between apes and true ('genuine') humans. Therefore, the term 'Pithecanthropus' carries, first of all, an indexical function.
- As specific names, that the species in question belong to a genus whose organisms possess both ape-like and human-like features. This is the descriptive function of the term 'Pithecanthropus'.
- As an effect of the headed syntax of the generic constituent, that the genus and all species and organisms belonging to it are part of humanity, not of the realm of apes, but still lack some feature to be classified under the genus *Homo*. This is the taxonomic function of the term 'Pithecanthropus'.
- In the case of **P.alalus*, that, in this species of the genus, the lacking essential feature is speech. The name does not overtly

capacity, other cranial features, stance and tool making follow. At the end, the *Australopithecus* is indicated as an example of 猿人. The entry in the *Kōjien* (1976) dictionary, the standard reference work for many years, is quite similar: "an early human, which existed more than 500 thousand years ago" (*gojūmannen ijō mae seizon shita shokijinrui* 五〇万年以上前生存した初期人類). The entry is completed by the same data as above. The time span reported by the *Dajirin* is consistent with the *Australopithecus* only, while that in the *Kōjien* also applies to *H.erectus* and *P.erectus*. Among the entries in other dictionaries, two provide *pithekanthropus* ビテカントロプス (*Pithecanthropus*) as a scientific name (Tokieda, Yoshida 1978; *Gensen* 1986), four mention a morphologic resemblance with great apes (Tokieda, Yoshida 1978; *Kokugo daijiten* 1981; *Gensen* 1986; Umesao et al. 1995).

encode the features ‘thoughtless’ and ‘consciousnessless’, as Haeckel (1889, 710) considered that they derive from the lack of speech by necessity. This is the explicative, or intensional, function of the term ‘*Pithecanthropus*’.

- In the case of *P. erectus*, that, in this species, an upright stance is present, but that such a necessary feature is not sufficient for the species to belong to the genus *Homo* (cf. Haeckel 1889, 709). Again this is the explicative, or intensional, function of the term ‘*Pithecanthropus*’.

Thus, despite being proper names, ‘*Pithecanthropus*’, a two-root compound, and the binomens *P. erectus* and **P. alalus* do have semantic content. They encode a description of the essential properties of the objects belonging to each taxa, a description that determines, or at least restricts, their extension (Cumming 2019, 5-6). ‘*Pithecanthropus*’ names a hybrid animal which is upright (as the specific name *erectus* specifies) as humans but speechless (*alalus*) as apes. It means ‘upright as a human, speechless as an ape’.

The problem is that such a complex relationship between the constituents cannot be inferred by means of a free, unguided processing of the term. Any uninformed attempt to decode ‘*Pithecanthropus*’ by means of ordinary conceptual combination is doomed to be misleading, as it too easily penetrates the apparent transparency of the term. It can only return the image of hairy, bent, long-armed, flat-nosed humans, or of monkey-tamers, ape-worshippers, actors in monkey costumes and so on. The only correct key to combine the concepts of ‘ape’ and ‘human’ and obtain the meaning of ‘*Pithecanthropus*’ is that expressly provided by the coiner of the term in his treatise, where the relationship between apes and human, encoded in ‘*Pithecanthropus*’, is explained at length. To understand what a *pithecanthrope* actually is, one must know what Haeckel wanted it to be. This knowledge is extrasemantic: after all, ‘*Pithecanthropus*’ is still a proper name.

2.2.4 猿人: A Proper Name of Ambiguous Content

In the taxon names discussed above, information is encoded in Latin, the metalanguage used in Western science, taking the form of Latin or Latinised Greek roots, Latin word order and Latin morphology. When the term ‘*Pithecanthropus*’ is rendered in the scientific metalanguage of eastern Asia, Chinese, the apparent transparency and hidden semantic ambiguity affecting the relationship between the constituents of the source compound carry over to 猿人 and unavoidably to *saru+hito*.

Consequently, only readers who are highly knowledgeable in the history of paleoanthropology and in evolutionary thought can

correctly capture the intensional meaning conveyed by 猿人. As a prerequisite, they must possess an encyclopedia entry ‘*Pithecanthropus*’ which details the essential properties of the object entity according to Haeckel (1868, 507; 1889, 709-10) and which contains the history of the header and semantic information about the constituent stems, *pithec-* and *anthrop-*. When processing 猿人, after *kun* assignment, they must not combine the two concepts *saru* and *hito*, for this would lead them astray. They should, rather, launch an immediate search for an encyclopedic file whose header is formed by roots semantically equivalent to ‘ape’ and ‘human being’. They would thus retrieve the very entry ‘*Pithecanthropus*’, which contains Haeckel’s instructions for correctly combining the concepts of ‘apes’ and ‘humans’. A factor facilitating the search is that the modifier-first syntax of ‘*Pithecanthropus*’ is identical to that of Japanese - *wago* - and Sino-Japanese - *kango*- appositionals, as well as to that of Chinese endocentric attributive compounds (Ceccagno, Scalise 2006, 240). As a consequence, when ‘*Pithecanthropus*’ was translated into Chinese, a straight lexical calque like 猿人 sufficed whose inner structure mirrored that of the source word. In the end, expert readers will establish the identity of the entity named 猿人 with the *Pithecanthropus* and use the newly-learned scientific term 猿人 to relabel (or assign a bilingual header to) the encyclopedia entry.

In the unrealistic scenario in which a Japanese reader is familiar with ‘*Pithecanthropus*’ but not with 猿人, such an outcome is doubtless of great epistemic value. But the process outlined above is out of the reach of readers as ordinary and ignorant as those considered by Suzuki (2014, 208). His prototypical *kun*-glossing user is a junior high school student who cannot have any expert knowledge of Haeckel. When Suzuki (2014, 208) offers the outcome of processing 猿人 as “the ape-man (*Pithecanthropus*), ancestor of humanity” (*jirinui no sosen ni ataru enjin (pitekantropusu)* 人類の祖先にあたる猿人 (ピテカントロプス)), the term ‘*Pithecanthropus*’ ピテカントロプス he adds in brackets is an explanation directed to the readers of his book, not a word miraculously flashing in the mind of his fictional interpreters at the conclusion of their conceptual labor.

2.3 Suzuki’s Own Knowledge and the Success of *Kun*-Glossing

When Suzuki begins his discussion of 猿人, he is already well aware of all of the above. He, therefore, creates a scenario in which encyclopedic knowledge and the outcome of conceptual combination are carefully balanced to give *kun*-glossing a certain utility.

First of all, he introduces 猿人 completely out of context. With no additional information to channel a conceptual combination, any specific quality mapping or relation linking interpretation of *saru+hito* is

unwarranted. Readers can only construe the hypothetical compound in the vaguest possible way. Consequently, and knowing that 猿人 is not a common noun, he is very careful to avoid saying that *saru+hito* captures the ‘meaning’ of 猿人. Suzuki (2014, 209-10) only maintains that *kun*-glossing helps readers “land close to the thing”, giving them a “comprehension” (*rikai* 理解) of 猿人 which “you cannot say is right on target, but does not land far from it either” (*atarazu to iedomo tōkarazu* 当らずと言えども遠からず). What readers obtain, then, is the vague interpretation of ‘a human being resembling an ape, an ape resembling a human being’, as ‘resemblance’ is the criterion most easily applied in lack of further evidence.

The ‘two-way’ interpretation reveals Suzuki’s own beliefs on the meaning of 猿人. The only possible explanation for such a syntactic pastiche is that Suzuki wanted to impose an interpretation of the compound as a hybrid. Since 猿人, being a non-coordinate compound, can only have a single referent, Suzuki implied that it designates neither ‘ape’ nor ‘human’ but an individual that is so similar to both as to be difficult to be distinguished from them. In this, it is probable that knowledge of ‘Anthropopithecus’, the alternate term used to refer to the same evolutionary stage as *Pithecanthropus*, also played a role. Apparently, Suzuki ignored Haeckel’s (1868, 507; 1889, 709-10) core distinction between ‘ape-like humans’ (*Affenmenschen*) and ‘human-like apes’ (*Menschenaffen*) but did know of the *Anthropopithecus*, and thus believed that, having this term the same extension as ‘Pithecanthropus’, the two roots for ‘ape’ and ‘human’ are syntactically interchangeable.

When ‘ordinary Japanese people’ use the literal meaning and ambiguous ‘denotation’ of 猿人 to activate their limited encyclopedia, they retrieve a correspondingly vague entry. The match is based on the knowledge that, in the ancient past, humans and apes were evolutionary closer than they are today and shared many features – body forms, behaviour – so that a lexical item in which *saru* and *hito* are conjoined may well refer to those remote ancestors. This is a level of knowledge consistent with junior high school students. Not only does their conception need not to be complete, but, up to a certain degree, it may also contain false beliefs. The very translation of *saru+hito* proposed by Suzuki (2014, 209) as, “an ape resembling a human being”, is semantically false, historically wrong and anthropologically farfetched, for the pithecanthrope was no ape at all, and no hybrid either.

3 A Proper Name: *Katsuo no Eboshi* 鯉の烏帽子 (Portuguese Man o' War)

The expression 'Portuguese man o' war' is the plain name for a genus of marine organisms, actually consisting of colonies of jellyfish-like zooids. It also designates the only species of the genus, the *Physalia physalis*. The two expressions, therefore, stand in the same relation occurring between 'ape-man' and the Linnaean binomens *P. erectus* and **P. alalus*. Moreover, 'Portuguese man o' war' is iconically motivated by the fact that the unsubmerged, visible part of the floating colonies so named resembles a sail. Like 'ape-man' and 'pithecanthrope', it describes a property of the denotatum. There are several important differences though:

- In *Physalia physalis*, the scientific binomen, the same stem, *physal-*, is used twice and the relationship between the two constituents does not translate into a description.
- 'Portuguese man o' war', the plain term, entails a description, but the described property is accidental, not essential.
- As a denoting expression, 'Portuguese man o' war' is categorially endocentric, because it refers to a first-order entity (a warship of old) just as the floating colonies of organisms are. Semantically, however, it is exocentric, because wooden artifacts are not lumps of tiny organisms.

Interpreting an exocentric expression like 'Portuguese man o' war' with the glossing tools that Suzuki uses on 'big' scientific compounds seems to be pointless, for how can a reader bridge the referential gap and jump from a sailing ship to a small, floating creature? But the strategy adopted in Japanese for naming this genus and species, which runs in parallel with that used in English, is interesting and well illustrates certain problems and limits of *kun*-glossing.

3.1 *Physalia Physalis* in Japanese

The family name Physaliidae, the generic name *Physalia* and the specific name *physalis* are not rendered in Japanese by means of corresponding 'neoclassical' *kanji* calques. The three New Latin terms come from Ancient Greek φουσαλλίς (bubble/bladder) (Scarborough 1992, 40). A calque of *Physalia physalis* of the type of 猿人 would then produce something like *泡泡 (or, in the more common orthography, *泡々), glossable *awa-awa* (bubble-bubble). It is evident that, for this hypothetical compound, *kun* assignment would not provide an output of any semantic utility and nor would it start a follow-up search in the encyclopedic domain.

It is no surprise, then, that the Japanese scientific name of the *Physalia* family, genus and only species in them is not calqued after the New Latin compound. Instead, the Japanese vernacular expression that designates the creature is used. What is interesting is that this expression is also a description, and that it operates by equating the *Physalia* to a hat worn by a fish, along an iconic path analogous with that of the English complex term. The Japanese name is 鯧の烏帽子 or, if written in the *katakana* phonetic script as appropriate for specific names, カツオノエボシ (*katsuonoeboshi*). If a morpheme-by-morpheme, Suzuki-style *kun* assignment and an English glossing of the term are run in parallel, the output is:

Table 1 Structure of *katsuonoeboshi*

Orthographic tier	鯧	の	烏	帽	子
Phonetic tier	<i>katsuo</i>	<i>no</i>	<i>e</i>	<i>bo</i>	<i>shi</i>
<i>Kun</i>-gloss tier	katsuo		karasu	∅	ko
Semantic/English gloss tier	sardine	genitive	crow	cap	child

The name is a noun phrase which is right-headed by *eboshi*, a term denoting a rigid, rimless and peakless headgear worn by the *élite* in the Heian period (AD 794-1185). The genitive modifier proclaims the hat to belong to sardines. The choice of sardines as possessor is motivated by the fact that *Physalia* colonies are often found in the vicinity of schools of that fish (Matsumura 1988). Thus, 鯧の烏帽子 does carry metadata about the named object, just as 猿人 does. The encoded information does not capture essential properties of the referent, though, but only its look and the circumstances of its appearance. Moreover, it is significant and relevant only to fishermen at sea.

3.2 Decoding an Exocentric Compound

The decomposition of 鯧の烏帽子 reveals several issues intrinsic to *kun*-glossing. One is that 帽 is not associated with any *kun* reading, does not allow *kun* assignment and therefore interrupts the whole process. This happens quite often with Japanese *kanji* and Suzuki, as previously pointed out, never addresses the problem. Let us assume, though, that, under the ‘best-scenario hypothesis’, every Japanese reader knows that 帽 stands for ‘hat’. Yet, recovering some componential meaning from 烏帽子 (crow-hat-child) is still problematic. On the basis of the onomasiological and syntactic rules outlined above, the head of 烏帽子 is 子 (child). Let us further take it for granted then that Japanese readers are all familiar with the fact that 子 occasionally occurs as the rightmost, suffixal-like constituent in nouns

naming instruments or small objects. Or, alternatively, that they simply know that 帽子 corresponds to the plain-language lexeme *bōshi* (with a long vowel) and that this word means ‘cap’ or ‘hat’. Japanese speakers use it in their daily speech, know that it is written 帽子 and do not need to *kun*-gloss it. Hence, they are not bothered by the fact that 帽子 is actually semantic exocentric and does not denote a child somehow bound to hats.

Still, these pieces of lexical knowledge do not solve the problem of the presence of *karasu* 烏 (crow). Is 烏帽子 a type of hat in some thematic relation with crows or does it share some properties with them? Birds do not wear hats. Thus, the most plausible interpretation is based on property mapping, whereby ‘crow’ is used to map the quality ‘black’ onto ‘hat’. This is a metonymy quite common with crows. Next, the issue as to whether the black hat in question belongs to sardines needs to be solved. But since fish do not wear hats either, the reader will have no problem inferring that the term must designate something else. In parallel with the whole-part relationship linking owner to hat, it must be a body-part. Thus, the entire parent expression is reckoned to be exocentric too. Iconicity suggests that it designates a class of semi-submerged, fish-related objects, perhaps dorsal fins, probably black. If, as an extension of the prior knowledge of 帽子 = ‘hat’, the denotation of *eboshi* 烏帽子 is also known, the property of being *eboshi*-shaped is additionally ascribed to the body part in question.

Prior lexical knowledge and iconicity are, then, enough to reveal the exocentricity of both the parent compound and one of its constituents. But when world knowledge is searched for an object matching that description, none are found. The readers then have to create a new 鰹の烏帽子 file to house the little information carried by the name itself and the data about similar entities that they will possibly obtain in the future. This fatally misleads them, for the creatures designated by the scientific name *Physalia physalis* are very small – so much so that when at sea their emerged parts are visible only at close quarters – and are pale blue-green.

3.3 Nested Exocentricity

The case of 鰹の烏帽子 is interesting because the expression represents an instance of multiple semantic exocentricity: 帽子 is not a hyponym of ‘child’ just like 鰹の烏帽子 is not a hyponym of ‘cap’ or ‘hat’. When combining concepts in order to derive meaning from the compound, nested exocentricity is to be solved modularly, one element at a time, first at the level of the head and then at the level of the whole expression. In the case of 鰹の烏帽子, *kun* assignment is made ineffective at the first level. It can reach the output ‘sardines’

tall hat', so capturing the iconicity of the entire parent expression, only if the suffix-like use of 子 is already known and the semantic exocentricity of 帽子 is thereby ascertained. The fact that *kun* assignment can only be started after determining the value of certain constituents by other means is a frequent problem in the interpretation of complex compounds.

4 A Natural-Kind Name: *Yōryokuso* 葉緑素 (Chlorophyll)

The successful interpretation of 葉緑素 (chlorophyll), yet another *kanji* calque of a neoclassical compound, relies once more on whether the function of one constituent is known prior to and independently of *kun* assignment. Suzuki (2014, 212-13) discusses it as follows. The string of Sinograms under analysis, 葉緑素, is the written notation of *yōryokuso*. Whereas the latter carries no meaning for most Japanese speakers, the former will activate in their minds the *kun* readings of the three *kanji*, respectively *ha*, *midori* and *moto*. According to Suzuki (2014, 212) these *kanji* "provide an idea of the rough meaning" of the entire compound (大よその意味の見当がつかます). The evoking is so spontaneous that Suzuki feels no need to specify what *ha*, *midori* and *moto* actually mean. He also seems to believe that the assembled meaning of these three components is self-evident, because he does not explain in encyclopedic terms what 葉緑素 is.

In the overwhelming majority of Japanese three-*kanji* compounds, one constituent is either a prefix (or a prefix-like modifier) or a suffixal head (Kobayashi, Yamashita, Kageyama 2016, 107). Because a discussion of the strategy adopted in order to successfully analyse a thoroughly unfamiliar three-*kanji* compound might be quite lengthy, it is more convenient to assume that, at this early stage, lexical knowledge will already make readers exclude a prefixal role of the leftmost element 葉 (leaf).²⁶ They will, therefore, assign the role of affix to the

²⁶ This working assumption is somehow forced. The dictionaries I used list 50 different compounds having 葉 as the leftmost constituent. Of these, 9 are three-*kanji* compound nouns. In 4 of the 9, 葉 is actually a modifier of the right-hand two-*kanji* base. They are *habotan* 葉牡丹 (purple kale), derived by adding 葉 to the lexical item *botan* 牡丹 (peonia); *hachōseki* 葉長石 (petalite) from *chōseki* 長石 (feldspar); *hageitō* 葉鶏頭 (amaranth) from *keitō* 鶏頭 (chicken-head = 'plumed cockscomb', a plant, both 葉鶏頭 and 葉鷄 are semantically exocentric); *hamusha* 葉武者 (common soldier/valueless *samurai*), from *musha* 武者 (warrior). In these compounds, 葉 is not a bleached-out prefixal, yet it functions in a similar way, modifying an existing lexeme. The rightmost element of the two-*kanji* string, therefore, is not a suffix. Only in the 5 remaining compounds is the rightmost character indeed a suffixal head. A ratio of five to four does not represent a solid base for excluding that 葉 might indeed be a modifier of 緑素. It is interesting to note that the 5 compounds include 葉緑素 and *yōryokutai* 葉緑体 (chloroplast). Knowledge of the latter compound would help readers correctly identify as suffixes 素 and 体, the two constituents that alternate at the right of 葉緑.

rightmost constituent, 素. This affects the general strategy adopted in decomposing the compound, as 素 becomes the starting point and is processed first.

Japanese people learn 素 as part of the *kanji* syllabus in the 5th grade of elementary school (Monbukagakushō 2009). They are familiar with it because the *kanji* occurs in the names of the eleven chemical elements most often mentioned in ordinary discourse and early science classes, like *sanso* 酸素 (oxygen) – the *kanji* 酸 is also taught in the 5th grade – and *suiso* 水素 (hydrogen). These eleven names are all two-*kanji* Sino-Japanese compounds in which the first character is unique to each element and the second one, 素, is read *so*. The first constituent is obtained by calquing the New Latin scientific name – as in *tanso* 炭素 (carbon), wherein 炭 represents the word *sumi* (charcoal, coal) – by metonymy – as in *enso* 塩素 (chlorine), which uses *shio* 塩 (salt) from the name of the most common salt, sodium chloride – or other methods.

Suzuki does not analyse the names of chemical elements, neither includes any of them in his long lists of “big words” (Suzuki 1975, 189). But, Suzuki (2017, 89-92) mentions 水素 (hydrogen) as a typical case in which Japanese readers can retrieve the concept of ‘water’ by means of assigning the *kun* reading *mizu* to 水 in a series of water-related scientific terms. When considering 水素, though, he provides no gloss for 素.²⁷

4.1 The Function of 素

No *kun* reading of 素 is taught at school as part of the official *kanji* syllabus (Bunkachō 2010). Dictionaries list *moto* as its second, non-standard *kun* reading, with the meaning ‘base material’, ‘ingredient’ or ‘essential element’. Suzuki (1963, 41) suggests the meaning, “fundamental constituent”. Not only are these special meanings all consistent with the nature of the chemical elements, making their names all hyponyms of *moto*, but they correspond to the scientific suffix *-gen* (origin). Being glossed *mizu-moto* (water-origin), 水素 can thus be considered a straight calque of ‘hydro-gen’ and is decodable as ‘the essential constituent of water’ or ‘the building block of water’.

The role of 素 can be more precisely determined on the basis of lexical frequency. This element is not bound. A dictionary search shows that it occurs in the rightmost position in 28 compounds. Of these, 6 have non-technical meanings related to the concepts of simplicity and frugality. The remaining 22 items are scientific terms and in

²⁷ Matsunaga (1996, 9), which shows an explicit appreciation of *kun*-glossing, erroneously reports that Suzuki does.

all of them 素 functions as a suffix-like head, specifying the semantic class ‘chemical substance’ of the parent word. For instance, *dokuso* 毒素 (poison-base_substance = ‘toxin’) and *shikiso* 色素 (colour-base_substance = ‘pigment’). A comparison of these two terms with atom names suggests that 素 marks the entity – either an atom or a molecule or a family of molecule – which provides the essential properties to a substance and is fundamental in making that substance as we perceive it. Combined with 水 (water), 素 encodes the information that the compound refers to the chemical entity that constitutes the essence of water as we commonly know it. Combined with 色 (colour), 素 similarly denotes a type of molecule which gives objects their distinctive colours. In onomasiological terms, the morpheme represented by 素 is the base, an Agent, while the mark expresses the effected entity (water, poison, salt etc.). In 色素 (pigment), 色 is the mark and realises the effected variable ‘colour’.

In summary, knowledge of the specific suffixal role fulfilled by 素 when occurring as the head of scientific terms facilitates compound interpretation in two ways. First, it allows the compound to be identified as endocentric, both categorially and semantically. Second, it conveniently constrains the possible interpretations of the relationship between the constituents, as it dictates that the compound may only be interpreted relationally, the thematic relation being ‘cause’ and the head functioning as causer.

4.2 *Kun*-Glossing 葉綠素

After processing 素, Japanese readers will proceed by assigning *kun* readings to 葉 and 綠, the two elements of the mark. In this operation they will be guided by the knowledge that this complex mark constitutes the distinctive part of the name of a chemical substance that gives an object its essential property. Still, the interpretive process requires a more complex reasoning than that done for the names of chemical elements and for 色素 (pigment) and 毒素 (toxin), in which the mark is constituted by one character only.

Kun assignment yields *ha-midori* (leaf-green). The term *midori* names a green colour, distinct from *ao* 青, which denotes the blue of the sky and traffic lights. The mark is obviously calqued from ‘chlorophyll’ green-leaf, but the two constituents are rearranged in an order consistent with Japanese syntax. The reason of it will be evident in a moment.

One possible strategy is to consider *ha-midori* to be an N+N appositive compound in which the first constituent, *ha* (leaf), maps a property of leaves onto the head, *midori* (colour green). This leads to a viable interpretation whereby some quality typically associated with leaves is predicated of the colour green. The two-*kanji* element

would denote a type of green defined by the fact of having the same chromatic properties as leaves as distinct from those exhibited by other green objects.

Considering this type of green to be an effected property in a thematic relation with the causer brings the reader to interpret the parent compound *ha+midori+moto* as subordinate, with the meaning ‘substance which causes (an object) to be leaf-green’. The entity made green by such a pigment, though, is not specified in the compound, and this creates the misbelief that the substance represented by 素 may be the cause of the colour of any leaf-green thing, from the oil paint in a tube to a Lego block, while, in fact, chlorophyll is only present in living plant tissue.²⁸ Knowledge of the use of *moto* 素 in forming the designators of natural-kind entities like chemical elements allows readers to discard this common-noun interpretation. The definite description ‘the pigment that causes an object to be leaf-green’, though, is no better, because it leads to the equally false belief that chlorophyll must be present in all leaf-green things as the sole chemical substance capable of giving their colour.

Compare this with the processing of *kesshokuso* 血色素 (hemoglobin), where the same error occurs.²⁹ The two-*kanji* mark, 血色, is glossed *chi-iro* ‘blood-colour’. Under a property-mapping interpretation, *chi-iro* also denotes a red hue. Consequently, *chi-iro+moto* might simply mean ‘a pigment determining a blood-red colour’, implying that there can be hemoglobin in all sorts of blood-red objects. If the compound is, instead, interpreted to mean ‘the pigment that determines the same colour as blood’, the implication is now that the chromatic properties of all blood-red things must be due to hemoglobin. Again, neither interpretation is correct, for hemoglobin is only present in blood and in the roots of some plants and is solely responsible for those tissues’ red hue.

Alternatively, the type of green denoted by *ha+midori* could be linked to the head, *moto* 素, as a quality. The complete three-constituent compound would then be interpreted as ‘a leaf-green basic substance’. Construing 血色素 (hemoglobin) in the same way yields the homologous ‘a blood-coloured substance’. Interestingly, both outputs capture some truth, for chlorophyll and hemoglobin, massively present in plant and blood tissues, have their typical chromatic effects because they themselves reflect the light of certain wavelengths only. Chlorophyll is, indeed, leaf-green to the eye. Yet these interpretations

²⁸ My main sources in biology are the *Iwanami seibutsugaku jiten* (1985) and Thain, Hickman (2004).

²⁹ Suzuki (1990, 160) considers the *kun*-glossing of several lexemes using the Sino-gram 血. None of them is 血色素, though. This might be due to the fact that the Japanese term is not a calque of English ‘hemoglobin’, so that no comparison between the two languages is possible.

are wrong as well, because they imply that chlorophyll and hemoglobin can colour any object. Moreover, in order to treat these three-constituent compounds as appositive and not subordinate, one would have to ignore the thematic role of ‘causer’, which is a distinctive feature of 素. As a matter of fact, a property interpretation is similarly not available for several other two-constituent compounds having 素 as the head: whereas 毒素 (toxins) are *doku* 毒 (poisonous), 水素 (hydrogen) is not *mizu* 水 (water) or water-like, and *tanso* 炭素 (carbon) contains no *sumi* 炭 (coal), just as diamonds are not coal-like either. In the end, faced with the hypothetical compound *ha+midori+moto* and armed with the only weapon they can deploy, i.e. knowledge of the role of 素 as causer, the reader will, in all likelihood, discard an attributive construal of the mark in relation to the base.

4.3 The Correct Processing of 葉綠素

The interpretations outlined above are erroneous because they originate from processing the two-element mark separately from the head. In other words, 葉綠 is mistakenly construed as an N+N appositive sub-compound in the same relation with 素 as the one-element mark in the names of chemical elements or in 色素 (pigment). This is why 素 ends up being interpreted as the agent causing a general property, 色 (colour), or a specific property, 葉綠 (leaf-green), in objects. In fact, as noted, the ternary structure of *ha+midori+moto* is much more complex than that.

An interpretation of 葉綠素 that completely agrees with what we know of chlorophyll is only obtained if *ha-midori* 葉綠 is analysed as if composed of two separate constituents, linked to 素 independently of each other. This allows 素 to maintain the role of ‘causer’ and also provides an explanation as to why the original syntax of ‘green’ and ‘leaf’ in *chloro-phyll* is reversed in *ha-midori*. Under this interpretation, *midori* (green) still expresses the result state put into being by the chemical action of the substance designated by 素. However, *ha* (leaves) now represents the concrete entity in which ‘green’ is effected, not a quality mapped on to this colour. As in a locative alternation, this entity is semantically both a place and an affected object. A second construal is therefore possible whereby ‘leaves’ are a patient and ‘green’ their result state. The three compounds in which 素 is associated with a colour term all share such a structure, as shown in table 2.

Table 2 Structure of pigment terms

Modifier (Mark)		Suffixal head (Base)	Denotation	
Affected entity (substance)	Effected entity (property)	Agent (effector, affector) (substance)	Literal meaning (a description)	Translatant
Place	Result object	Causer		
Patient	Result state	Causer		
	色	素	colour-causing substance	'pigment'
	<i>iro</i>	<i>moto</i>		
血	色	素	substance	'hemoglobin'
<i>chi</i>	<i>iro</i>	<i>moto</i>	causing colour in blood	
葉	緑	素	substance	'chlorophyll'
<i>ha</i>	<i>midori</i>	<i>moto</i>	causing green in leaves	

In onomasiological terms, 素 is the base and represents in simple form both the agent, here a class of molecules, and the action, here an unexpressed verbal element denoting the causing of a property. The modifier represents the determining component of the mark and provides the internal argument(s) of the verbal element. Once a common-noun interpretation is excluded, a definite description, 'the pigment that makes leaves green' or 'that determines green in leaves', is thus obtained which does capture one essential property of chlorophyll and constitutes intensional information. The same approach also gives 'the pigment that causes colour in blood' as the correct interpretation of *chi+iro+moto* (hemoglobin).

4.4 The Limits of *Kun*-Glossing in Interpreting 葉緑素

The first weakness of the process detailed above lies in the very glossing of 素 as *moto*. Contrary to Suzuki's (2014, 212-13) suggestion, there is no way to retrieve the specific role of 素 from the meaning of the gloss. In an early paper Suzuki himself includes 素 in a list of eight characters read *moto* each expressing one particular meaning or *Sonderbedeutung*:

An extreme case is the word *moto* もと. This form can be collated to eight characters, each with a differing shade of meaning. 1) 基 [ki] 'base,' 2) 因 [in] 'cause,' 3) 本 [hon] 'principal, primary,' 4) 原 [gen] 'origin,' 5) 素 [so] 'fundamental constituent,' 6) 下 [ge or ka] 'under, below,' 7) 許 [kyo] 'at, under, Fr. *Chez*,' and 8) 固 [ko] 'origin.' (Suzuki 1963, 41)

Suzuki does not suggest any common semantic denominator for the several instances of *moto* he mentions, but they all seem to share the general meaning of ‘origin’ in the three domains of space, time and process. Six more *kanji* can be added to the list, all falling under that same *Gesamtbedeutung*.³⁰ *Moto* is therefore polysemic and acquires distinct meanings only by virtue of its association with different *kanji*. In *ha+midori+moto*, the semantics of *moto* is hence determined by 素. But this is a complete reversal of the logic of *kun*-glossing, because it implies that it is not the Sinogram’s *kun* reading that assigns meaning to it (the core principle of Suzuki’s theory), but vice-versa. This fact might prove devastating for the whole model.

Suzuki does not address the issue, but he seems to imply that distinguishing the meaning of a particular form from that of its homophones is simply part of *kanji* knowledge. Consistently with the ‘best-scenario hypothesis’, this knowledge includes all relevant usages and special meanings of any given *kanji* and thus effectively substitutes *kun*-assignment. As regards 素, the knowledge of it encompasses the fact that, when heading certain terms, the *kanji* becomes a suffix-like morpheme which encodes the names of important chemical elements and substances. That Suzuki is taking this piece of knowledge for granted is revealed by the fact that, albeit providing the gloss *moto* for 素, nowhere (in 2014) does he explain the character’s function.

Ignoring such a special naming function would make it impossible to recognise a compound headed by 素 as a term referring to a chemical substance. Knowing it is also indispensable in order to choose between the common-noun construal ‘a pigment which...’ and the definite construal ‘the pigment that...’. Readers would have no way to correctly identify 葉綠素 as a three-element name without prior knowledge that, as a suffixal head, 素 typically encodes the names of certain natural kinds. This lexical knowledge is still not resolute, though, because some of the scientific terms headed by 素, like 毒素 (toxin) and 色素 (pigment), are class nouns.

But the main weakness of processing 葉綠素 by means of *kun*-glossing lays elsewhere. It is the fact that a judgment on the validity of an appositive construal of 葉綠 – finally yielding ‘the substance that makes objects leaf-green’ – as opposed to an interpretation of those elements as the arguments of *moto* – resulting in ‘the substance that makes leaves green’ – may only be *a posteriori*, driven by what ‘chlorophyll’ actually denotes. To be able to differentiate between the two interpretations, the reader has to already know that leaves are green because of one particular chemical substance. They only lack the scientific name of it. The informed *kun*-glossing of the unfamiliar term then allows them to assign the correct header to the relevant

30 They are 旧, 氏, 故, 原, 資 and the most relevant, 質.

encyclopedia entry. If such a reader does exist, this outcome certainly represents an improvement in their encyclopedic knowledge, but only in the sense of better organizing previous beliefs and further justifying them, not of creating the basis for expanding them in number and content. Under the conditions necessary to interpret 葉緑素, the informative value of *kun*-glossing is of little worth.

5 More Kind Names: *Hishishokubutsu* 被子植物 (Angiosperm), *Rashishokubutsu* 裸子植物 (Gymnosperm)

While discussing *kun*-glossing, Suzuki briefly mentions *hishishokubutsu* 被子植物 (angiosperm) and *rashishokubutsu* 裸子植物 (gymnosperm) as examples of Japanese technical words that are allegedly easier to understand than English neoclassical compounds (2014, 211). He does not use *kun*-glossing on either term, though, taking their transparency for granted. In this section I will challenge this assumption.

5.1 Four-Character Compounds

Four-character compounds can be analysed in two basic ways. Firstly, they may be formed of two sub-components comprising two *kanji* each. This is by far the most common pattern (Kobayashi, Yamashita, Kageyama 2016, 113). For the terms under discussion, this would require decomposing 被子植物 into 被子 plus 植物 and 裸子植物 into 裸子 plus 植物. The alternative is a structure of affixals built around a unitary two-character lexical unit: Kobayashi, Yamashita, Kageyama (2016, 115) individuate nine possible combinations. For the sake of simplicity, I assume that Japanese readers will identify the rightmost two-character element of both terms as *shokubutsu* 植物 (plant), without the need to decompose the whole compound and tentatively *kun*-gloss each of the four *kanji*. Moreover, I also take it for granted that they will recognise 植物 as the categorial and semantic head, making 被子植物 and 裸子植物 both hyponyms of 'plant'. Being so right-headed is indeed the norm for complex compounds of this type (Kageyama 2011, 514; Kobayashi, Yamashita, Kageyama 2016, 113-22). This leaves the marks, 被子 and 裸子, to be interpreted. The relationship they bear to the base is to be sought in function of their meanings. I will analyse 被子植物 (angiosperm) first.

5.2 Construing 子

As seen, the only *kun* reading of 子 is *ko* (child). As in the case of 帽子 (hat) discussed above, all the attempts at interpreting 被子 based on that literal meaning would fail. Under the ‘best-scenario hypothesis’, however, Japanese readers are supposed to be familiar with the less frequent, more specialised usages of 子. In one of them, certainly motivated by the original meaning ‘child’, 子 is associated with *mi* (fruit) and *tane* (seed). And indeed, excluding 被子植物 and 裸子植物, the *kanji* occurs as ‘fruit’ or ‘seed’ in 15 biology-related compounds. These two meanings are consistent with the language of botany, but only one of them defines what an angiosperm actually is.

In differentiating between the two construals, readers will be influenced by the relative frequency of the two meanings. The *kanji* is only associated with ‘fruit’ in *yuzu* 柚子 (*Citrus junos*) and in *nasu* 茄子 (eggplant), while it occurs as ‘seed’ in *shushi* 種子 (seed), *hōshi* 孢子 (spore), the names of the gametes *seishi* 精子 (spermatozoon) and *ranshi* 卵子 (ovum), *shinō* 子囊 (seed pod), *shiyō* 子葉 (baby leaf/ seed leaf, meaning ‘cotyledon’), in *tamago* 玉子 (the everyday orthography of the term for bird eggs), and in a few more items, for a total of 13 compounds. The much larger size of the set of words with 子 = ‘seed’ will bring the reader to discard the gloss *mi* (fruit) and construe 子 as *tane* (seed) and this is even before attempting an interpretation of the other constituents of the compound.

5.3 Interpreting 被子植物 by *Kun*-glossing 被 as *Kaburu* (Wear on Top)

The only *kun’yomi* officially taught at school in association with 被, the mark’s left-hand constituent, is the abstract transitive verb *kōmuru* (receive/suffer/sustain) (Bunkachō 2010). This verb can thus be considered to express the principal meaning of the Sinogram. However, several more verbs are listed in *kanji* dictionaries under 被 as alternate readings. They are *kaburu*, *kabuseru*, *kabuseru*, *kazuku*, *kazukeru* and *ōu*. All share amongst themselves and with *kōmuru* the underlying meaning or *Gesamtbedeutung* ‘put/receive on top’. I will first assume that, if asked to name the character, a Japanese speaker will pick the form *kaburu* (wear on top) and not *kōmuru*. This is for two reasons. First, *kaburu* better captures the general meaning. It must be noted that *kaburu* assigns the grammatical role of object to the entity displaced on top of something else or, alternatively, to the instrument whereby one covers something up (Koizumi et al. 1989). Second, the abstract meaning of *kōmuru* is incompatible with the concrete noun ‘seed’ in its phrase. This connects to the following.

When a *kango* has a V+N structure, the left-hand verbal constituent may be the head, with the right-hand constituent then representing the verb's internal argument (Kageyama 2011, 514). A compound of this type is an independent lexical item syntactically behaving as a verbal noun. In 被子 only the first one of these conditions is fulfilled: the element is *kango*-like (and is actually read *hishi*), but is not a free-standing lexeme and does not function as a verbal noun (Tamaoka et al. 2015). Suzuki's hypothetical readers, though, are not supposed to know that, because *kun*-glossing is only based on the association of a Sinogram with a *kun* reading and is activated exactly in order to obviate the lack of lexical knowledge about unfamiliar terms.

Thus, despite Kageyama's (2011, 516) bold claim that "left-headed S[ino]-J[apanese] compounds are semantically transparent thanks to the ideographic nature of Chinese characters", readers must painstakingly proceed with glossing 被子 as *kaburu-tane* and with construing the corresponding hypothetical compound as a verbal noun in which the right-hand element, 子 (seeds), represents the object put on top of a third, unmentioned entity. As a result, they will interpret the element as a verbal phrase: 'wear seeds on top'. This interpretation, though, is not strongly supported by frequency. *Kanji* dictionaries list 11 two-*kanji* compounds having the structure [被 + concrete noun]. Only 5 of them are left-headed, have verbal semantics and express an action of wearing.³¹ The remaining 6 compounds are right-headed. They are actually noun phrases in which the left-hand element, 被, functions as modifier.³² This fact has some relevance and will be recalled later.

Finally, 植物 (plant), the nominal base, is to be combined with 被子, the verbal mark. The meronymic relation linking seeds and plants and the fact that in 被子 no overt subject is expressed make 植物 the ideal candidate for the role of wearer. Under this interpretation, the compound is interpreted as 'a plant with seeds on top'.

5.4 Interpreting 被子植物 by *Kun*-Glossing 被 as *Ōu* (Cover/Wrap/Conceal)

Alas, the above interpretation is not consistent with the actual denotation of 'angiosperm'. The error originates from glossing 被 with the verb *kaburu* and construing *tane* as the displaced patient or instrument, in accordance with the participant-role structure of the verb.

³¹ They are *hikō* 披甲 (wearing armor), *hishu* 被酒 (showering in sake = being very drunk), *hiken* 披堅 (wearing [hard] armour: 堅 'hard' = *kenkō* 堅甲 (hard armour), *hikatsu* 被褐 (wearing rough clothes), *hishatsu* 被髮 (wearing hair [on the forehead]).

³² The six compounds are *kazuki* 被衣 (lady's veil), *hifuku* 被服 (clothes), *hikin* 被衾 (bed clothing), *hiseki* 被錫 (wig), *himaku* 被膜 (capsule/film), *hifu* 被布 (overcoat).

Among the *kun* readings subsumed under 被, only the verb *ōu* (cover/wrap/envelop/screen/hide/conceal) has a meaning and role structure compatible with the concept of ‘angiosperm’, as it takes as its object the covered theme entity and demotes the covering instrument to an adjunct (Kokuritsu kokugo kenkyūjo 1972, 695). The gloss *ōu* allows ‘plant’ to be construed as the coverer or hider of ‘seeds’, the object, and thus to interpret the compound as ‘a plant wrapping up/covering/concealing its seeds’. This is indeed the description which captures what angiosperms actually are: flowering plants, the ovules of which are first contained in an ovary and then, once fertilised, are borne as seeds within a protective involucre (which is nothing but the fruit itself). This is the iconic motivation underlying both the English and the Japanese names and the definition of the two terms in dictionaries.³³

But what can make one reader prefer *ōu* over *kaburu* and one role structure over the other, thereby discarding the ‘wear on top’ interpretation in favour of the ‘wrap up’ one? Certainly not frequency effects, because the lexicon does not support the correct interpretation. In all the 11 two-*kanji* [被 + concrete noun] compounds found in dictionaries, 被 always carries one of the meanings ‘wear on the head’, ‘from the head’ or ‘over other clothes’ whilst the other constituent invariably realises the displaced entity.

5.5 *Kun*-Glossing 裸子植物

The only way out is *kun*-glossing the companion compound, 裸子植物 (gymnosperm), first. The need to do so is implicitly suggested by Suzuki (2014, 211) himself because he introduces the two botanical terms as part of one and the same problem, to be addressed in its entirety and with a consistent set of tools.

The process follows the same steps as those performed for 被子植物. It starts with dealing with the mark, 裸子, by means of first glossing 子 as *tane* (seed). The subsequent glossing of 裸 is hardly problematic. This *kanji* has only one *kun* reading, *hadaka*. In origin, this form is a noun meaning ‘naked body’ or the condition thereof but is ordinarily used in an adjectival function as ‘naked’. The *kanji* is indeed at the left of 子 (seed), in the ordinary position that Japanese modifiers take in relation to their heads. One dictionary (Tōdō 1980) also associates the verb *hadanugu* (undress/become naked) with the Sinogram. Construing 裸 verbally would require evaluating the role

³³ The neoclassical compound ‘angiosperm’ literally means ‘seed in a vessel’ (Thain, Hickman 2004; Scarborough 1992, 28) and in its ordinary usage is semantically exocentric, as plants are not seeds. The complete expression is ‘angiosperm plants’.

structure of *hadanugu* (which is intransitive) and to take into account that this verb entails the state ‘naked’ as the result of a process. Straightly glossing 裸 with *hadaka* (naked) is much more frugal. The two-*kanji* mark is therefore interpreted as a noun phrase, an expression translatable as ‘a seed which is naked’.

This concept is then to be combined with that of ‘plant’, the base of the complex compound. To simplify, I will not discuss the process required for interpreting the compound [‘naked seed’ + ‘plant’] as coordinate (which would originate ‘naked seeds and plants’) or by means of property mapping (‘a plant looking like a naked seed’). I will only consider relation linking. Under this strategy, a seed and a plant may be in one of the following three thematic relations:

- Derivative: the parent compound is decoded as ‘a plant growing from a naked seed’.
- Causal: ‘a plant producing naked seeds’.
- Meronymic or possessive/locative (theme-place): ‘a plant with naked seeds’.

These three interpretations are equally robust, although it might be argued that the causal and meronymic construals amount to the same thing in as much as botanic objects are concerned. Given that seeds are parts of plants, a part-whole relation seems the one which will be favoured by readers.

Availability confirms this intuition. Upon searching a biology dictionary (*Iwanami seibutsugakujiten* 1985) and a large number of Internet sites for compounds headed by 植物 and having the mark entirely written in *kanji*, I obtained 49 items, not counting 被子植物 and 裸子植物. In 21 of them the mark denotes a plant part, with the thematic relation linking ‘plant’ to the mark being invariably ‘possession’.³⁴ In 17 of these items the mark is a noun phrase containing a modifier that specifies a property of the part, along the simple decomposition

(plant) HAVE [(property) (part)]

Again, in 3 of the 17 items the modifier internal to the mark is verbal, giving the complex compound the same structure as 被子植物.

³⁴ I hereby list all the 21 compounds grouped by the type of physical feature represented in the mark, with little or no further comment: 輪葉植物, 鱗葉植物, 小葉植物, 大葉植物, 双子葉植物, 單子葉植物, 觀葉植物, 莖葉植物, 楔葉植物 (leaves); 顯花植物, 隱花植物 (flowers); 毬果植物, 球果植物 (fruits); 種子植物 (seeds); 裸莖植物 (stem); 蔓植物 (vines); 齒朶植物 (branches); 首飾植物 (necklace-like part); 羊齒植物 (tooth-like parts); 維管束植物 (vascular bundle); 隔膜形成體植物 (phragmoplast). In two more compounds, *yūhaishokubutsu* 有胚植物 and *yūsetsushokubutsu* 有節植物, the marks are verbal phrases, left-headed by the element *yū* 有 (have) and completed by the object *hai* 胚 (embryo) and *setsu* 節 (joint) respectively. The relation of possession is therefore realised overtly as ‘a plant having embryos/joints’.

The 3 items are 顯花植物, with a mark glossable *arawareru-hana* (appearing-flower), designating Phanerogamae or flowering plants; 隱花植物, with *kakureru-hana* (hiding-flower), designating Cryptogamae or flowerless plants; and 觀葉植物, with *miru-hana* (beholding-leaves), a non-technical term for plants beautiful to behold, i.e. ornamental plants.

The locative or possessive interpretation ('a plant with naked seeds') is hence the one best supported by availability. Not only is the output identical to a plain-language description of the essential property of gymnosperms – plants whose seeds are not concealed in a protective involucre but lay exposed – but duplicates element-by-element the literal meaning of the English neoclassical compound, wherein the root *gymn* means 'naked' and *sperm* means 'seed'.

5.6 Processing 被子植物 After 裸子植物

The successful *kun*-glossing of 裸子植物 leads to the correct interpretation of 被子植物. The two compounds only differ in their leftmost element. This must be the locus where their difference in meaning is expressed. Once 裸子植物 is processed and 裸 construed as 'naked', the several *kun* readings of 被 are examined for those belonging to the same semantic field as 'naked', i.e. 'cloth-wearing'. The special meaning carried by *kōmuru* is dismissed now. Among the remaining *kun* readings, *ōu* is individuated as the only one in direct correspondence with 'naked', because the theme in its role structure is not a displaced entity. This licenses a construal of 被 as an antonym of 裸 (naked) and an interpretation of the compound mark, 被子, as a noun phrase and not a verbal phrase. The fact that 6 of the 11 two-*kanji* listed compounds having the structure [被 + concrete noun] are noun phrases supports this interpretation. The final outcome is the correct one: 'a plant with clothed seeds'.

The only issue is that, in this way, *ōu* is forced into the same construction of adjective *hadaka*, which predicates a property of a plant's seeds and have 'seed' as the subject, as if *ōu* were unaccusative. But *ōu* (like *kaburu*) is transitive. It predicates the action of a plant on its seeds and have 'seed' as the object. This is irrelevant. In the interpretive process I have just briefly outlined, the verb *ōu* does not frame a plant and seeds in the argumental structure it carries when in dictionary form. Rather, it assigns meaning to the *kanji* and determines the compound's internal participant-role structure. Once selected as gloss because of the antinomy with 'naked', *ōu* simply assigns the theme role to 'seed', in accordance with its own role structure and just like 'naked' does. This makes it possible to interpret the entire compound along the line of 'a plant that clothes/wraps up its seeds' or the equivalent 'a plant with wrapped-up seeds'.

5.7 The Limits of *Kun*-Glossing in Interpreting 被子植物 and 裸子植物

The process detailed above is driven by a series of choices:

- processing 裸子植物 (gymnosperm) before 被子植物 (angiosperm);
- assigning the gloss *tane* and the meaning ‘seed’ to 子;
- glossing 裸 with *hadaka* (naked) and not with *hadanugu*;
- selecting the ‘possessive/locative’ relation between 裸子 and 植物 (plant) from three plausible thematic relations;
- inferring a homology between 被子植物 and 裸子植物 from their similar structure;
- contrasting 被 and 裸 (naked) as antonyms;
- individuating in *ōu* the *kun* reading of 被 in antonymy with ‘naked’.

This is the only path that leads to the correct interpretation of both compounds. At each point, if the wrong decision is taken, a mistake occurs that directs the process not so much to a dead end – for a totally meaningless output would only make it clear that the path has to be restarted from an intermediate point – but, more dangerously, to an apparently viable wrong result. Definitions like ‘a plant wearing seeds on top’ and ‘a plant carrying clothed fruits’ are all plausible and, if heuristically taken to truthfully capture the meaning of the source word, would fatally generate false beliefs.

Certainly, lexical knowledge, familiarity with the use and distribution of the constituents, thematic relation frequency and the words’ family size do influence readers in those choices, making some of them quite straightforward. But this only means that *kun* assignment hardly plays any role at all. In itself, *kun* assignment does not even provide a clue whether to interpret 子 as ‘child’, ‘fruit’ or ‘seed’, because the only *kun* reading of 子 is *ko* (child). Choosing *tane* is a matter of *kanji* lexical knowledge, not of an intelligent application of *kun*-glossing. As regards 被, *kun* assignment and conceptual combination yield the misleading *kuburu*, not *ōu*. Only a clever contrast of 被 with 裸 (naked) directs the reader to *ōu*, but this implies that the entire construal process is guided by the knowledge of another, albeit related, word, not built up over a patient sequence of inferences kicked off by *kun* assignment. And it should be clear that, in the above discussion, I could proceed in a successful direction only because my choices were driven by what I already knew of angiosperms. In other words, as I had to model it here, the whole process of *kun* assigning plus conceptual combination is *a posteriori*, channelled by what I knew the endpoint would be. I had to work backwards from the denotation of the terms, in a sort of well-informed retro-engineering – while *kun*-glossing, as a one-to-one mapping of Sinogram to concept, would have taken me nowhere.

6 A Class Noun: *Tanpakushitsu* 蛋白質 (Protein)

Dictionaries list *tanpakushitsu* 蛋白質 as having only one meaning: ‘protein’. The first two constituents alone (*tanpaku* 蛋白) form an independent lexical item meaning ‘albumen’. Suzuki (1990, 135) discusses 蛋白質, without providing an English translation, as an instance in which *kun*-glossing proves impossible. This is because, he argues, one *kanji* in the compound (蛋) has no *kun* reading, is not used in any other current Japanese word and, therefore, has no meaning to present-day Japanese people. Suzuki adds that the same opacity equally afflicts *tan*, the *on* reading of 蛋, and, by extension, the entire parent compound when the first constituent is written phonetically (たん白質).

This obscure *kanji* was imported from Chinese in order to represent the word ‘egg’ (and it is still used in this way in Chinese today). What, then, if 蛋白質 were to be re-written as *卵白質, Suzuki wonders, substituting 蛋 with 卵 (*ran*), the character currently used in Japanese to write the word *tamago* (egg)? According to Suzuki, if faced with the unfamiliar three-*kanji* sequence *卵白質, any Japanese reader would easily interpret it as ‘*tamago*’ no ‘*shiro(mi)*’ no *seibun* (the component of egg-white):

初めて耳にしたときは一瞬ためらった人でも、表記を見れば、「ああ《たまご》の《しろ（み）》の成分かと分かるに違いない。(Suzuki 1990, 135)³⁵

Suzuki’s use of double guillemets (which I have rendered with single quotation marks) suggests that he wants to represent a *kun*-glossing operation whereby his reader assigns, on sight, the glosses *tamago* (egg) to 卵 and *shiro* (white) to 白. *Shiro* can have two meanings: ‘white part’ or ‘albumen’. In fact, what Suzuki is doing in the above paragraph is not retracing the *kun*-glossing of *卵白質, but is directly showing the best possible output of the process thus started, that is, the combined concept which ends up being closest to the word’s true denotation.

In the following sections I will address the issues arising from Suzuki’s (1990, 135) treatment of this imaginary compound. Can *kun*-glossing actually yield the interpretation he proposes? And, more importantly, can such an interpretation be effectively linked to the concept of ‘protein’?

³⁵ “There is no doubt that even those who hear [**ranpakushitsu*] for the first time and are momentarily disoriented, once they see it in written form will understand it and say “Ah! isn’t this the substance that makes up the ‘white part’ of ‘eggs’?”.”

6.1 The Function of 質

Being a three-*kanji* compound, *卵白質 is to be decomposed along the lines outlined above for 葉綠素 (chlorophyll). Either the rightmost or leftmost characters is an affix. It is reasonable to suppose, as Suzuki does, that Japanese readers will easily single out the right-hand character, 質, as a suffixal head. This *kanji*, also learned in the 5th grade of elementary school, has no official, or ‘standard’, *kun* reading (Bunkachō 2010; Monbukagakushō 2009), but is associated with *moto* (content) and *tachi* (disposition/nature). Usually, though, it is named by means of *shitsu*, its main *on* pronunciation. As an independent lexeme, *shitsu* means ‘quality’. The Sinogram is also a well-documented suffix, expressing either the meaning of ‘content’ or ‘quality’. In the sentence above, Suzuki (1990) considers it indeed synonymous to *seibun* 成分 (component part: cf. Tōdō 1980).

A dictionary search returns 11 compounds headed by 質 in a suffixal usage as ‘content’ or ‘component’. They can be classified into two subtypes:

- a. Place-Theme. The compound designates a substance comprised in the entity represented by the mark. This is the less frequent combination, occurring in only 3 two-*kanji* compounds and 1 three-*kanji* compound.³⁶
- b. Quality-Theme. The compound designates a substance having the properties expressed by the mark and contained in some other, unspecified substance. This combination is represented by 2 two-*kanji* compound and 5 three-*kanji* compounds.³⁷

In both subtypes the denotatum of 質 is a theme, linked to a place entity in a locative relation. This contrasts *moto* 質 with *moto* 素 because, in the compounds headed by the latter, the standing relation is ‘cause’ and the base designates the unique causer of the effect described by the mark.

6.2 Kun-Glossing 卵白 and Combining It with 質

What is left after processing 質 is the two-*kanji* element 卵白. These two Sinograms actually form *ranpaku* (albumen). However, this fact and the existence of *ran’ō* 卵黄 – a sister compound meaning

³⁶ The compounds are *kakushitsu* 角質 (keratin), *kosshitsu* 骨質 (bony tissue), *shishitsu* 齒質 (dentin) and *shokubutsushitsu* 植物質 (vegetal matter).

³⁷ The compounds are *dokushitsu* 毒質 (poisonous ingredient), *shishitsu* 脂質 (lipids), *genkeishitsu* 原形質 (protoplasm), *shibōshitsu* 脂肪質 (fat), *hōrōshitsu* 琺瑯質 (enamel), *kōbutsushitsu* 鈹物質 (mineral matter), *denkaishitsu* 電解質 (electrolyte).

‘egg-yellow’ = ‘yolk’ – must be ignored because, as repeatedly pointed out, knowledge of the source lexeme would obviously make *kun*-glossing unnecessary.

The two constituents of the mark, 卵 and 白, are then to receive *kun* glosses. In the resulting N+N hypothetical compound, *tamago+shiro*, the head is *shiro*, an abstract nominal, and the modifier is *tamago*, a first order noun. Since *shiro* (white) denotes a property, the only strategy to construe the mark is property mapping, whereby the substance *tamago* (egg) transfers its chromatic quality to the colour white, in parallel with the interpretation of *ha+midori* in ‘chlorophyll’. *Tamago+shiro* is so decoded as the type of white most typically or uniquely seen in (evidently boiled) eggs. Next, *moto* 質 is to be combined with the mark. Only a property-mapping interpretation – quality-theme (b) above – is possible, leading to decoding *卵白質 as ‘a substance as white as boiled eggs and found in some other substance’.

I consider a relation linking strategy according to subtype (a) to be unacceptable because it would require construing *tamago+shiro* not as a chromatic property but as a substance in which the referent of the parent compound is located. This would make *卵白質 semantically exocentric. I see no reason why readers should contemplate the possibility of exocentricity. Not only is the property-mapping interpretation (b) straightforward and satisficing (Simon 1956), but it is also better represented in the lexicon. Alas, embracing semantic exocentricity is exactly what is needed here, for 卵白, just like the English phrase ‘egg white’, is indeed exocentric.³⁸

6.3 *Shiromi* しろみ (Albumen)

How can hypothetical readers realise the semantic exocentricity of 卵白? Suzuki does not trace the *kun*-glossing operation to any of its possible outputs. Rather, he implies that, at the end of it, perhaps because they are suspicious or unsatisfied, Japanese readers will search their lexicons for items in which the concept of ‘white’ is somehow associated with that of ‘egg’. They will then retrieve two plain-language items. One is *shiromi* (written 白み) ‘white part’, a deadjectival noun obtained by applying a semi-productive morphologic rule compatible with colour terms to *shiroi* 白い (white). The other is *shiromi* (written 白身), lit. ‘white-body’, a homophone referring to fish meat and, indeed, egg white.

³⁸ Thus, in Ueda et al. 1993, the English definition associated with *tanpaku* 蛋白 and *ranpaku* 卵白 is ‘white of an egg’, an obvious attempt to clarify the exocentricity of the two terms.

It is not clear which *shiromi* a Japanese reader would prefer for interpreting the source compound. From the use of brackets in the quote above, which singles out the morpheme *mi* as if it only had a grammatical function, Suzuki seems to imply the first one. However, lexical knowledge also supports the second one, because this second *shiromi* can be contrasted with *kimi* 黄身 (yellow-body) and meaning ‘yolk’. In either case, though, *shiromi* is eventually identified as the object that comprises the substance to which the whole complex compound refers, consistently with subtype (b) above. All the pieces now click into place and the words that Suzuki puts in the mouths of his hypothetical readers must be taken to mean “ah! isn’t this a component of the albumen of eggs?”. This outcome is of great significance because it shows how *kun*-glossing, combined with basic lexical and morphological knowledge, has enabled readers to interpret 卵白 as an expression referring to a white object, a substance (a first order entity), not as a semantically endocentric term denoting a type of white (a second-order entity). *Kun*-glossing has led to overcoming the exocentricity block.

6.4 蛋白質 (Protein)

Once the meaning of 蛋 is clarified, 蛋白質 is effectively decoded as a kind of molecule because (1) molecules of this kind are part of the protective and nutritional substance enclosed in the eggs of birds and some reptiles, and (2) this substance, which appears transparent when raw, solidifies, becomes opaque and takes the visual quality of ‘white’ if exposed to high temperatures (as happens in cooking).

A whole class of molecules is thus designated with an expression denoting one of its many properties, in this case that of being one of the constituents of a type of biological substance. This is a synecdoche, like referring to the human being by means of the term ‘farmer’. From the point of view of sense relations, this metonymy also makes 蛋白質 ‘a component of the albumen of eggs’ a superordinate of all sister terms designating specific types of protein, from ‘hemoglobin’ to ‘enzyme’ – and even of ‘albumin’ itself, because the proteins of this group are not the only ones found in egg white. But since hemoglobin is not a component of the albumen of eggs, this ends up being a second instance of semantic exocentricity. In this case, though, I do not see how a description like ‘a component of the albumen of eggs’ could be expanded to ‘a class of organic molecules forming an important part of all living organisms³⁹ and existing in, among other things, the

³⁹ This is the modern, synthetic meaning of ‘protein’ according to the *Online Etymological Dictionary*.

albumen of eggs' and linked to the general concept of 'protein'. Not even a thorough knowledge of biology, of all types and functions of proteins, of the composition of chicken eggs and ova, of the developmental processes of embryos and, why not, of the chemical processes involved in cooking, can provide the data needed to infer 'protein' from 'a component of egg white'. There is no way to overcome this second exocentricity block. In the end, 蛋白質, or the domesticated version thereof, *卵白質, cannot even be used to correctly (re)name a misnamed or unnamed encyclopedia entry accidentally coinciding with the concept of 'protein'.

7 The Term *Hōshokusei* 蜂食性 (Apivory) and the Function of *Sei* 性

A discussion of 蜂食性 is interesting because this compound contains 性, a suffixal element which recurs quite often in Suzuki's examples. Suzuki (1990, 132-3) discusses the *kun*-glossing of 蜂食性 in some detail, to the extent that I classified this item in Group A. In his works, he explicitly put 6 other items containing 性 in relation to *kun*-glossing (two in Group B and four in Group C, see the list in § 1.2), while he mentioned 4 more without comment (Group D).⁴⁰ Among them, I decided to discuss 蜂食性 because processing this compound is straightforward and unambiguous, to the extent that it represents perhaps the most successful instance of *kun*-glossing in Suzuki's entire corpus of 'big' technical compounds.

As noted earlier, Suzuki (1990, 132) decodes 蜂食性 as "the natural disposition of feeding on bees" (*hachi o taberu seishitsu* 蜂を食べる性質). This description is not the outcome of *kun* assignment but the conception obtained at the end of the entire *kun*-glossing process. As regards the *kun* assignment, the gloss of the first constituent, 蜂, is *hachi* (bee) and that of the second, 食, is *taberu* (eating). In contrast, 性 (learned in the 5th grade, Bunkachō 2010) has no official *kun* reading, but is usually named with *sei*, one of its two *on* readings, meaning 'sex'. This Sinogram has widespread usage as a semi-productive suffix in the formation of abstract nouns, not unlike the English derivative suffixes '-ness', '-ity' and indeed the '-y' of 'apivory'. Two examples are *seisansei* 生産性 (productivity) from *seisan*

⁴⁰ They are *enshinsei* 遠心性 (centrifugal), *kyūshinsei* 求心性 (centripetal), *gyōshokusei* 魚食性 (piscivorous) and *sōshokusei* 草食性 (graminivorous), all represented as adjectives (Suzuki 1978, 7; 1990, 131, 133). The only plausible reason for rendering 草食性 with 'graminivorous' rather than with the more ordinary 'herbivorous' is that Suzuki wants to provide an English translation as detached as possible from daily plain language. 魚食性 (Piscivorous) is mentioned by Hatano, Kuhara, Akiyama (1981, 31) as part of a discussion of Suzuki's approach.

生産 (production) and *jiyūsei* 自由性 (freedom), from *jiyū* 自由 (freedom), constructed as ‘the property of being free’. In this function, 性 is bleached of lexical meaning.

However, 性 also has a non-standard *kun* reading as *saga*, a desuete term whose meaning coincides with Suzuki’s translation, *seishitsu* 性質 (natural disposition/birth character). Consequently, prior knowledge of the abstract, suffixal function of 性 is highly probable. However, to interpret 蜂食性 this knowledge is not necessary, if readers can gloss 性 with *saga* and, of course, know the meaning of this old term. *Kun* assignment then yields *hachi-taberu-saga* (bee-eating-natural_disposition), in which *saga* functions as the head and the left-most two elements constitute the onomasiological compound mark. *Taberu* is the determining component, expressing the category ‘action’, while *hachi* is the determiner, specifying the internal argument of *taberu*. The two elements are also in the ordinary object-verb syntax of Japanese and, hence, are unambiguously interpreted as denoting the act of eating bees. The lexicon contains several compounds with the same structure. One is *panshoku* パン食 (bread-eating), referring to a Western-style kind of meal centred on bread consumption rather than on rice. When the verbal phrase *hachi+taberu* modifies *saga*, eating bees is interpreted as a type of natural disposition, to the effect that the final outcome of *kun*-glossing coincides exactly with the source concept encoded by 蜂食性. *Kun*-glossing yields homologous results for *gyoshokusei* 魚食性 (piscivory), wherein *sakana* 魚 (fish) substitutes *hachi* 蜂 (bee) as the object; and for *sōshokusei* 草食性 (herbivory) in which *kusa* 草 (grass) occurs (Suzuki 1978, 7; 1990, 131, 133).

8 A Common Class Noun: *Jidōsha* 自動車 (Automobile)

‘Automobile’ is a nineteenth-century loanword originating in the French expression *voiture automobile*, in which the element ‘automobile’ is an adjective. In turn, this term stems from *automobilis*, a neoclassical compound (*Online Etymological Dictionary*; CRNTL). The corresponding Sino-Japanese lexeme is *jidōsha* 自動車. Thus, regardless to the fact that the Japanese term might actually be a semantic calque of its English and French counterparts, 自動車 represents yet another instance of a Japanese word coined with high-standing foreign (Chinese) elements, in analogy with English technical terms, and rightly belongs to Suzuki’s favourite choice of examples.

Suzuki (1978, 3) discusses 自動車 in an unusual way. When facing *jidōsha*, a phonological word segmentable in the three morphemes *ji-dō-sha*, he argues that Japanese people are able to interpret it as the three-word noun phrase *mizukara ugoku kuruma* みずから動くくるま (a cart moving by itself). This happens because the elements occupying

the same position in each parallel sequence, namely *ji* and *mizukara* (oneself), *dō* and *ugoku* (move), and *sha* and *kuruma* (cart), are paired (*ketsugō* 結合) through the intermediation of one particular character, 自, 動 and 車 respectively. Thus, rather than applying *kun*-glossing to the constituent *kanji*, Suzuki decomposes *jidōsha* as a phonological word, applies what could be considered a sort of *kanji*-glossing to each of its morphemes, and then implicitly suggests that the *kanji* so assigned receive meaning through an ordinary *kun*-glossing operation. Although not inconsistent with his general model, such a mechanism weakens it because, as Suzuki (1963, 31-2; 1990, 145) himself points out, *on* forms like *ji*, *dō* and *sha* are just as alien and devoid of meaning as the *kanji* they are written with. Whereas to associate 自, 動 and 車 with, respectively, *ji*, *dō* and *sha* is merely part of *kanji* knowledge, to proceed in the opposite direction would require lexical knowledge of both *jidōsha*, the spoken form of the parent compound, and of 自動車, its written form. This would, of course, make *kun*-glossing superfluous. What is interesting, though, is that straight *kun*-glossing works quite well in extracting meaning from 自動車 anyway.

Mizukara, the gloss of 自, is a nominal meaning ‘oneself’ also used adverbially as ‘by oneself’, ‘with one’s own forces’. The *kanji* also carries a non-standard reading, *onozukara* (spontaneously). The second *kanji*, 動, has the *kun* readings *ugoku* (move, intransitive) and *ugokasu* (move, transitive). Suzuki (1978, 3) chooses the former, unaccusative verb as the gloss because it is less marked. Either would work, though, as I will show in a moment. The third *kanji*, 車, is glossed with its only *kun* reading, *kuruma*. This term has the primary meaning of ‘wheel’, as in *kurumaisu* 車椅子 (wheel-chair). By synecdoche, it also denotes any vehicle on wheels. *Kuruma* is now used mostly to refer to cars.

In a three-*kanji* compound, both or either distal element could be affixes and either could be the head. Left-hand adverbial constituents can be the head only if they have a negative function (Kageyama 1982, 227-30; Namiki, Kageyama 2016, 212), though. *Mizukara* is not negative and, hence, cannot head *mizukara+ugoku+kuruma*. The first-order nominal *kuruma*, on the other hand, is an ideal candidate for the role of head. The compound is, therefore, interpreted as semantically endocentric and as denoting a type of wheeled vehicle. In *mizukara+ugoku* the adverbial and verbal constituents are in the ordinary modifier-head syntax. *Mizukara* (by itself) is construed as an adverb of manner which modifies *ugoku* (moving), so yielding ‘self-moving’, a verbal phrase. That 自動 occurs very frequently in Japanese with the meaning ‘automatic’ is irrelevant in this context because *kun*-glossing is used when readers have no lexical knowledge of this kind. In turn, *mizukara ugoku* precedes *kuruma*, in the position of a relative clause. ‘Self-moving’ is therefore construed as a property of ‘car’. It is after such a process that Suzuki’s proposed interpretation obtains: *mizukara+ugoku+kuruma* denotes a ‘self-moving wheeled

vehicle'. Were 動 to be glossed transitively with *ugokasu* (move) then 自 could just be interpreted as 'oneself', yielding the equally viable 'a wheeled vehicle moving itself'.

The descriptions so obtained not only capture the concept of 'car' but are generic enough to even agree with the Japanese classification of vehicles contained in the Road Traffic Act, whereby 自動車 stands for all self-propelled wheeled vehicles whose path of movement is not restrained by tracks or power lines (*Dōrokōtsūhō*, Ch. 1 Art. 2-9; see also Takashima 1994).

9 A Common Class Noun: The Neoclassical Compound *Sekiyu* 石油 (Petroleum)

The *kanji* of 石油 are glossed, respectively, *ishi* (stone) and *abura* (oil/fat/grease). The lexeme is only apparently a semantic calque of 'petroleum', a loanword from Medieval Latin composed of two roots having the same meanings as the Japanese *kanji* (on the history of the Latin term, cf. McDonald 2011, 353-5, 363). Instead, 石油 is borrowed from Chinese, in which it is documented as early as the eleventh century.⁴¹ It is unclear, of course whether the term had the same extension as today's 石油 and as 'petroleum'. To establish the thematic relation at the basis of 石油 and 'petroleum' is also difficult. The relation might be 'origin' or 'place', if the referent is extracted from the ground. Alternatively, it might be 'material' or 'appearance', if the substance is obtained from open tar pits or outcrops.

The similarity between the New Latin and Chinese compounds might, therefore, be an odd case of convergent word formation, so that no direct comparison can be done between the two terms. Furthermore, whereas the usage of 'petroleum' is not common in English, that of *sekiyu* 石油 is widespread across all registers. Because of these two factors, the compound is not an ideal *kun*-glossing candidate for Suzuki, who tends to limit his discussion to those scientific, highly specialised terms that have a homologous structure in English and Japanese. And, indeed, nowhere does he analyse it. In an early article of 1963, though, he does touch on the relationship of *abura* with the *kanji* used to represent it in writing. Moreover, the way in which the concepts of 'stone' and 'oil' are combined in 石油 offers an interesting opportunity to discuss how effective *kun*-glossing is in a situation as simple as this. That is why I will devote this last section to an analysis of this compound.

⁴¹ According to *Science and Civilization in China* (Golas 1999, 201, quoting further Chinese sources) the term was coined by Shěn Kuò 沈括 (1031-1095) and first occurred in vol. 24(1) of the *Mèng xī bitán* 夢溪筆談 (Dream-pool essays) of 1088.

9.1 *Kun*-Glossing 石油 and the Polysemy of *Abura*

The *kun*-reading of the first Sinogram of the compound is *ishi*. This lexeme means ‘stone’ – or ‘rock’, as a synonym of *ganseki* 岩石, to be intended as a mineral formation and not as a landscape object (Matsumura 1988). In contrast, *abura*, the *kun*-reading of the second Sinogram, is highly polysemic:

the word *abura* あぶら ‘oil, fat, grease etc.’ for instance has at least three characters as its semantic equivalents, each having a distinct *On*-form of its own. 1) 油 [yu, yū] ‘oil,’ 2) 脂 [shi] ‘fat,’ 3) 膏 [kō] ‘ointment,’ all of which are read as *abura*. (Suzuki 1963, 40-1, square brackets in the original)

As Suzuki suggests, there are other *kanji* associated with *abura*. One is 膩 (sticky/viscous fat) (ねっとりした脂肪); another is 肪 (animal fat). If *abura* has many meanings, and each of them is represented in writing by a different *kanji*, what clarifies the meaning of the form in any one instance of its use is the character. In other words, the specific meanings of *abura* are determined by their respective ideograms, so that upon seeing the character 油 a Japanese reader will immediately know that the gloss of it, *abura*, means ‘oil’ and not, for instance, ‘animal fat’. It is the same problem I discussed above with regards to *moto*. Unlike 素, though, 油 has no affixal usage which can be registered as part of a reader’s lexical knowledge.

Using *abura* as a semantic gloss, then, causes the problem of which one of the several *kanji* meanings is to be considered relevant in the conceptual combination. In this case, the issue can be solved by assuming that *abura* denotes a greasy substance of unspecified viscosity and origin, going from vegetable oil to a thick fluid like asphalt, and that this is the underlying meaning that the term carries across all usages and different *kanji* instantiations.⁴² This is the only way to process 石油 by means of *kun*-glossing, and I will proceed under this assumption.

The form **ishiabura* does not exist in the lexicon. To interpret 石油, the reader has to combine the concepts expressed by the two first-order nouns that compose it.

⁴² According to Tōdō (1980) and Matsumura (1988), the *kanji* 油 and 脂 do indeed differentiate between two viscosity degrees.

9.2 *Ishi+Abura* As a Coordinate Compound

As a hypothetical compound, *ishi+abura* might be coordinate. If of the separate-reference type, it would denote a referential set formed by stones/rocks (hereafter mostly ‘stones’ for brevity) and greasy substances (‘grease’ for brevity). An encyclopedic search returns no entry matching this interpretation. Yet, readers cannot exclude in principle that such a set exists, is relevant to some obscure field of knowledge and that 石油 is actually established in scientific language.

Here, the reader is influenced by frequency effects. A search among the 150 items having the structure [石 + concrete noun] listed in dictionaries returns only 2 coordinate compounds, both composed of a hyponym and its superordinate (Bauer 2017, 86): *sekishō* 石礁 (stone-sunken_rock), and *sekiso* 石礎 (stone-foundation_stone). Moreover, among the 27 N+N compounds having 油 as the rightmost constituent, none is coordinate. On distributional grounds, then, it looks highly improbable that 石油 is a co-compound of the separate-reference type.

Alternatively, *ishi+abura* might be of the co-participant type. In this case, as noted earlier, the content of the reciprocal relationship between the two conjoined terms should be further specified. But since 石油 is not a co-compound, there can be no contextual specification of this kind. Thus, the reader will exclude a co-participant interpretation as well.

9.3 *Ishi+Abura* as an Appositive Compound

Greasy substances might be mineral entities, just as stones are. Between the two concepts the right amount of similarity exists to support a property-mapping interpretation. *Ishi+abura* would, then, be right-headed and refer to a type of grease, with *ishi* denoting the property of stones that is to be ascribed to it, thus identifying its type.

As regards the relevant feature to be transferred to grease, there are only three attributes whose values may be considered prototypical of stones: colour (grayish), shape (roundish) and hardness (hard). In human experience the most salient attribute is hardness, as it is constant and is assessed after direct physical contact with the object. This concept is also commonly associated with the high viscosity or ‘thickness’ of fluids. The most plausible property interpretation of *ishi+abura*, then, consists of mapping the value ‘high’ of the attribute ‘hardness’ of stones to the corresponding attribute ‘thickness’ of grease. The result is a hyperbole: the compound would denote a greasy substance of such a high viscosity as to make it metaphorically rock-like hard. Alternatively, if colour is mapped, the compound might denote a grayish grease. Mapping shape is also compatible

with grease. Natural asphalt, a highly viscous or semi-solid form of petroleum, can solidify in the shape of rock-like mounds when seeping from cracks in the ground.

A hardness-based interpretation is plausible because it bridges the conceptual gap between stones and grease as mineral entities but is not supported by the lexicon. Of the 150 [石 + concrete noun] compounds listed in dictionaries, only 12 are semantically endocentric appositionals. Of these, 6 can be interpreted as mapping the property 'hard', 6 encode 'shape', 2 encode 'colour' (there is some overlapping).⁴³ As regards the compounds having a *kanji* glossable *abura* (油, 脂, 膏) as the head, a dictionary search returns only 3 appositive compounds, mapping shape, smell and chemical composition.⁴⁴ In conclusion, the distribution of 石 and of 油, 脂 and 膏 in actual compounds does not justify a property-mapping interpretation.

9.4 *Ishi+Abura* As a Subordinate Compound

One more possible interpretive strategy is relation linking. Dictionaries list 104 subordinate N+N compounds in which *ishi* 石 occurs as the modifier, 24 compounds headed by 油, and 5 headed by *abura* 脂. In this set, the thematic relations that may plausibly involve the two types of constituent are:

- Material. Unsurprisingly, the most frequent role carried by 石 (stone) is 'material', with 79 items. In the case of *ishi+abura*, though, the 'material' thematic relation is to be ignored because it would yield the semantically unacceptable 'a grease consisting of stones'.
- Locative. This is the second most frequent thematic relation involving 石 (stone), occurring in 18 items. It does not involve the mark in any of the compounds headed by 油 or *abura* 脂. A 'locative' relation linking produces two possible interpretations. One is place-theme, whereby the modifier, 'stone', carries the role of 'place' while the head, 'greasy substance', is the theme or locatum. This type is represented by 9 of the 18 compounds,

⁴³ I classify them as follows: 石人, 石女, 石友, 石婦, 石塩 (hardness); 石鹼 (hardness and shape); 石花菜, 石龜, 石榴, 石鼓文 (shape); 石斑魚 (shape and colour); 石鱒 (colour).

⁴⁴ Besides 石油 and the variant 石腦油, the relevant two-*kanji* appositionals are *suiyu* 水油 (liquid hair/lamp-oil) and *kōyu* 香油 glossable *kaori-abura* and meaning 'balm'. The third appositional is *sekinōyu* 脂肪油 (fatty oil), in which 油 heads 脂肪 (fat). This mark is in itself a synonymic co-compound (Bauer 2008, 10). As a matter of fact, it is glossable as *abura-abura*, to the effect that *kun*-glossing *shibōyu* 脂肪油 would yield *abura-abura-abura*. As regards the two other *abura* characters, 肪 only occurs in 脂肪, and 膩 in the two co-compounds *shiji* 脂膩, one more *abura-abura*, and *kuni* or *kōji* 垢膩, glossable 'grime-fat'. They are irrelevant for this discussion.

with the theme role realised by either a living being: *sekiso* 石鼠 (mole cricket), *sekisen* 石蘚 (rock moss) -, a structural feature - *sekirai* 石罅 (rock fissure) - or a landscape feature: *sekka* 石瀨 (shallow stream running on rocks). Under this interpretation, *ishi+abura* would mean 'greasy substance contained in rocks' or 'hidden among stones'. The remaining 9 items belong to the theme-place type, in which 石 (stone) is the *locatum* and the 'place' role is carried by a landscape feature: *sekiki* 石磯 (stony beach), *sekiden* 石田 (stony field), *sekiro* 石路 (stony road). If of this type, *ishi+abura* would mean 'stone-containing grease'. Putting stones in grease could be a way of preserving them.

- Instrumental. This relation can be seen in only 5 of the 104 items having *ishi* (stone) as the modifier⁴⁵ and in 5 of the 24 items headed by *abura* 油.⁴⁶ Under this interpretation, *ishi+abura* would denote a type of grease obtained with a stone tool, like stone-pressed vegetable oil.
- Derivative. This relation only occurs in 2 compounds among those having 石 (stone) as the modifier.⁴⁷ As regards *abura* (grease), it occurs in 19 compounds among the 24 headed by 油,⁴⁸ and in all the 5 items headed by 脂.⁴⁹ According to this interpretation, 石 (rock) represents the source substance from which a grease is obtained via some kind of transformation.
- Benefactive. This relation is not established in any compound in which 石 and 油 / 脂 / 膏 occur, respectively, as modifier and head. In this interpretation, the modifier, 'stone', would represent the beneficiary of an effect caused by the head, 'grease', as in the use of the latter as a lubricant to reduce attrition between two stones.

The above data shows a strong polarisation in the availability of the thematic relations. The locative relation may only be observed in compounds having *ishi* 石 as the modifier, while the derivative relation

⁴⁵ They are *ishiyumi* 石弓 (crossbow/catapult), *sekka* 石火 (flint fire), *sekyo* 石渠 (stone-cut groove), *sekishō* 石匠 (stonemason), *ishiku* or *sekkō* 石工 (stonemason).

⁴⁶ They are *tōyu* 灯油 (lamp oil), *kamiabura* 髮油 (hair oil), *hakutōyu* 白灯油 (kerosene), *tōshin'yu* 灯心油 (lamp oil), *kikai abura* 機械油 (machine oil).

⁴⁷ One is *ishioto* 石音 (sound of rocks), in which 'stone' is more properly a cause. The other is *sekiro* 石蠟 (paraffin), glossable 'stone-wax'.

⁴⁸ They are: *komeabura* 米油 (rice-bran oil), *kan'yu* 肝油 (liver oil), *tōyu* 桐油 (tung/ nut/wood oil), *gyōyu* 魚油 (fish oil), *tsubakiabura* 椿油 (camellia oil), *kōyu* 鉱油 (mineral oil), *taneabura* 種油 (rapeseed oil), *watabura* 綿油 (cotton oil), *nukaabura* 糠油 (rice-bran oil), *shōyu* 醬油 (soy sauce), *geiyu* 鯨油 (whale oil); *hennōyu* 片腦油 (camphor oil), *gomaabura* 胡麻油 (sesame-seed oil), *nataneabura* 菜種油 (rapeseed oil), *menjitsuyu* 綿実油 (cottonseed oil), *shōnōyu* 樟腦油 (camphor oil), *hakkayu* 薄荷油 (peppermint oil), *kijōyu* 生醬油 (raw soy sauce) and *kanran'yu* 橄欖油 (olive oil).

⁴⁹ They all denote animal fats or resin: *gyūshi* 牛脂 (beef tallow), *tonshi* 豚脂 (lard), *jūshi* 獸脂 (animal fat), *geishi* 鯨油 (blubber), *jushi* 樹脂 (resin).

almost exclusively occurs when *abura* is the head. *Ishi* and *abura* are in compatible distribution only under the instrumental relation, which is, however, underrepresented in absolute terms. The effect of this polarisation is that in *ishi+abura*, the left-hand constituent calls for a locative interpretation, whereas the right-hand one calls for a derivative interpretation. In order to choose between the two conflicting interpretations, the reader has to resort to encyclopedic validation.

9.5 *Ishi+abura* As an Exocentric Compound

The last issue to be addressed is whether either *ishi* or *abura* have such a frequent use in metaphors as to suggest an exocentric interpretation of the compound. In this case, what would the most plausible meaning of it be?

The *kanji* for *abura*, 油, does not occur as the head or onomasiological base in any N+N semantically exocentric compound. In contrast, of the 150 [石+N] compounds listed in dictionaries, 31 are semantically exocentric and 2 have a secondary exocentric interpretation. This ratio is significant so that the possibility that 石油 is exocentric as well cannot be dismissed. All 33 compounds fall under the metaphorical pattern described by Bauer (2010, 171-2). Among them, 5 stand out: *ishiwata* 石綿, glossable 'stone-cotton', and *sekijū* 石絨, glossable 'stone-wool_cloth', both meaning 'asbestos'; *sekijun* 石筍, glossable 'stone-bamboo_sprout' and meaning 'stalagmite'; *sekishōnyū* 石鍾乳, glossable 'stone-clot-milk' and meaning 'stalactite'; and *sekkō* 石膏, glossable *ishi-abura* 'stone-meat_fat' and meaning 'gypsum'/alabaster'. In these expressions, the base denotes a substance resembling the denotatum of the parent compound in the fields 'shape' (first four items) and 'colour' or 'translucency' (last two items). The mark, 'stone', then maps a quality ('hard') onto the base or expresses the thematic role of 'material' in relation to it. Knowledge of these particular items and the awareness of their compositional principles might direct the reader to interpret the compound as referring to a stone- or a rock-hard formation in the shape of flowing oil. The compound could mean 'flowstone'.

The high number of exocentric compounds having 石 as a mark does indeed give strength to an exocentric interpretation of 石油. The question now arises as to whether this interpretation could be validated encyclopedically.

9.6 Encyclopedic Validation: The Successful Interpretation of *Ishi+Abura*

The reader reaches the stage of encyclopedic validation having discarded the interpretations of *ishi+abura* as coordinate and appositive on the basis of family size, the ‘material’ relational interpretation on semantic grounds and the ‘benefactive’ one because of availability. They are left with four possible relational interpretations – place-theme locative, theme-place locative, instrumental and derivative – which are all conceptually consistent and supported by the lexicon, and with a possible interpretation of the compound as exocentric. They now search their encyclopedic knowledge for entities matching any of the viable interpretations. Assuming they:

1. possess an entry for the actual object (oil, petroleum);
2. know, as part of the knowledge stored thereby, that the object in question is a greasy substance and not simply an undescribed liquid;
3. are ready to switch from the construal of *ishi* as ‘stone’ to that as ‘rock’;
4. are acquainted with the images of oil wells extracting that substance from the ground, where rocks are,

the search will return the very concept of ‘petroleum’ under two alternative descriptions: ‘a grease-like substance found in rocks’ and ‘a grease-like substance derived from rock’. These are indeed the descriptions yielded by the two best available thematic relations, the place-theme locative and the derivative one. Both are conceptually sound. That only the former description is congruent with the denotatum is irrelevant, for a certain confusion about the procedure of oil extraction can only be expected and is even justified historically. In the knowledge and communicative practices of “the man in the street” (Suzuki 1975, 189), the latter conception would work as well without invalidating the rest of the file’s content. Thus, the conflict between the two polarised interpretations, which cannot be solved, is actually neutralised by taking both descriptions as true and making them coexist. This is one more instance of successful *kun*-glossing.

10 Discussion

The full process of *kun*-glossing unfolds in three stages: *kun* assignment, conceptual combination and encyclopedic validation. I considered these activities to occur in steps because each stage provides the raw material used in the next one: no encyclopedic search can commence without a hypothetical concept to feed in it, but no conceptual combination is possible without first assigning a semantic value

to each *kanji* constituent. The process is, of course, far too complex to be so linear, and yet at each stage certain problems arise which, if not solved properly, will negatively affect the operations to be carried out in subsequent stages. In the next sections I will examine each stage, underscoring the specific problems presented and the solutions adopted by Suzuki's hypothetical readers or which are implicitly available to them.

10.1 Stage One: *Kun* Assignment

Recalling the *kun* readings of each Sinogram in a compound, selecting one of them and assigning it to that *kanji* as its unique semantic gloss is the first and definitory operation of *kun*-glossing. The problems to be faced at this stage are the following.

10.1.1 No *Kun* Reading

A Sinogram completely lacks *kun* readings under two concurrent conditions. First, the official *kanji* syllabus determined by the Government (Bunkachō 2010) prescribes that no *wago* is to be taught at school as the character's reading. Second, dictionaries do not associate the Sinogram with any unofficial *kun* reading either. When existing, these non-standard readings are listed in the section denominated *imi* 意味 or *igi* 字義 (meanings) of a *kanji*'s entry. They correspond to less frequent or desuete *kun'yomi* and are, in fact, alternative semantic glosses expressing the several special meanings the Sinogram may have (Tōdō 1980, 1-2).

A zero *kun*-reading situation effectively kills *kun*-glossing at the start. Unsurprisingly, Suzuki (1990, 134-5) considers this a very rare occurrence, never addresses the issue and only includes three instances of it in his well-devised list of examples. One is 齧, occurring in *gesshirui* 齧齒類 (rodents). Its *kun* reading, *kajiru* (munch), is so uncommon, he writes, to be virtually unknown. Then there is 齧, a constituent of *shigin'en* 齧齧炎 (gingivitis). This *kanji*, he maintains, has no *kun* reading at all, while, in fact, dictionaries assign *haguki* (gums) to it. Only 蛋 is truly devoid of *kun* readings so that, for this Sinogram, Suzuki (1990, 135) envisages a substitution with the common 卵.

I briefly considered the problem when dealing with *bōshi* 帽子 (hat). Not only does 帽 have no *kun* reading, but 子 (child) has no gloss consistent with the meaning of the compound. To discuss how 鯉の烏帽子 may be processed through *kun*-glossing, I had to assume that at least one component of it, 帽子, was previously known.

10.1.2 One *Kun* Reading Only

If the Sinogram only has one listed *kun* reading, either official or non-standard, then that reading is used as the gloss. This type of *kanji* forms the large majority of the compounds I analysed. For example, *saru* 猿 (ape), *hito* 人 (person), *ha* 葉 (leaf), *tamago* 卵 (egg), *taberu* 食べる (eat), *kuruma* 車 (car) and *ishi* 石 (stone) all belong to it. Moreover, they have only one *kun* reading each, which is part of the official school's syllabus. This situation represents the most favourable starting point of a *kun*-glossing operation. It is consistent with the prototypical interpretive scenario envisaged by Suzuki, which involves junior high school students. The most outstanding exception is 性, whose singular *kun* reading is non-standard and is listed in the 'Meanings' section of the dictionary entry.

10.1.3 Multiple *Kun* Readings

This problem can take two forms.

(1) A *kanji* is associated with numerous *kun* readings, all semantically related. In this case, several different situations may occur, requiring different approaches.

- *Hauptbedeutung* (principal meaning) approach. If only one of the listed *kun*-readings is part of the officially school syllabus then that reading is considered the most important and adopted to express the main meaning associated with the Sinogram. This glossing strategy works well with, among many, 裸 and 自. It does not work with 子, whose only official reading, *ko*, means 'child' and is too specific to be compatible with plants and marine organisms, making it obvious that another approach is needed. Similarly, the only official reading of 被, the verb *kōmuru* (receive/suffer), is semantically incompatible with 'seed' because it is abstract and its sole argument designates a damaging entity. Again, this directs the reader to use another approach.
- *Gesamtbedeutung* (general meaning) approach. If the Sinogram has more than one official *kun* reading (and possibly other alternate, non-standard readings as well), a common underlying meaning is extracted from the particular meanings expressed by all the listed readings. Then the reading that best captures the underlying meaning, that which is closest to prototype, is used as a gloss. This is indeed the approach proposed by Lehmann (2004, 1841) for interlinear glossing. The approach works well with 裸, whose alternate reading is very rare anyway and is not even listed in every dictionary. In most cases, however, this approach produces too general a concept to be of

any great use. In the case of 子 it returns ‘offspring’ and, again, the reading *ko*, which are too vague and do not allow any differentiation between ‘child’, ‘fruit’ and ‘seed’. For 被, the general meaning approach seems viable. In fact, as I showed, following it would be an error, as it leads to ‘wear (or ‘worn’) on top’, which does not capture neither what actually happens to seeds nor what links seeds to plants under the concept of ‘angiosperms’. In this specific case, the prior processing of 裸子植物 prevents confusion by directing the reader to the right relation between the constituents.

- *Sonderbedeutung* (special meaning) approach. Interpreters take into account all the *kanji*’s *kun* readings and alternate semantic glosses and select one. To do so they do not have to indiscriminately assign all the possible glosses and then test them, one by one, by putting each construal through the next stages of the process, going back and forth until some validation is obtained. There are several tools that can assist them in pruning out unviable glosses. In the case of 子, semantics makes the reader favour *mi* (fruit) and *tane* (seed) in relation to plant names; then, frequency directs them to ‘seed’. In the case of 被, contrasting this *kanji* with 裸 restricts their choice of semantic glosses to *ōu*. In the case of 素 and 質, both glossed *moto*, the valid special meanings are individuated by identifying the suffixal functions of the Sinogram.

(2) All or some of the meanings of the available glosses belong to unrelated semantic fields. This situation is expressly feared by Morton, Sasanuma (1984, 38, 40). But when analyzing Suzuki’s 151 words and during my search for additional compounds to process, I could find no instance of obvious conflict. *Mi* (fruit) and *tane* (seed), the two glosses of 子, are indeed incompatible, a particularly dangerous form of conflict because both concepts belong to botany and yield plausible outputs. However, I showed that frequency directs the reader to the right construal without letting the conflict block the process.

10.1.4 Several *Kanji* With the Same *Kun* Reading

When a glossing form is shared by more than one *kanji* there is a set of homophonous *wago* whose slightly distinct meanings are expressed by different *kanji*. Among the cases discussed in this essay, this problem is best exemplified by *abura*, the term used to gloss the right constituent of 石油 (petroleum). In this instance, a *Gesamtbedeutung* approach is followed whereby the distinct special meanings of *abura* (all related to lipids of different thickness and origin) are neutralised under the general meaning ‘greasy substance’. This

strategy allows the two component concepts to be combined in a large number of ways, without forcibly limiting them to a few sense relations only. Its greatest advantage, though, is that it makes it unnecessary to distinguish the meanings of *abura* by *kanji* – an operation that would entail a reversal of the *kanji*-to-meaning relationship and negates the very logic of *kun*-glossing.

Another example is *moto*, gloss to both 素 and 質 in 葉綠素 (chlorophyll) and 蛋白質 (protein). Twelve more *kanji*, all with related meanings, have *moto* as their listed *kun* reading. Determining the *Gesamtbedeutung* ‘origin’ of *moto* does not help. Instead, a strategy is required that consists of first recognizing the suffixal function of the two characters and then extracting their distinct naming usages from the lexicon. Only in this way can the meanings of the two relevant *moto* and the particular thematic roles they carry in compounds become clear.

10.1.5 Semi-Bleached Suffixes

In certain cases, one key constituent cannot receive a lexical or grammatical meaning on the basis of *kun* assignment alone. To interpret the parent compound, the reader must acknowledge the specific usage of that constituent, assign the special meaning based on that usage to it and only then proceed with *kun*-glossing the remaining component(s). Besides 素 and 質, this is what is done to 子 in 帽子 (hat). This *kanji* has a frequent usage as a suffix to name small objects, tools, garments and microscopic entities. This abstraction exploits a different metaphor than that of 子 (child) → (offspring) → (seed). In the case of suffixal 性, function and meaning are retrievable from *kun* reading only in those compounds that, like 蜂食性 (apivory), denote natural properties. In other instances – as in the mentioned 生産性 (productivity) –, however, the suffix has undergone semantic bleaching and interpreting the compound by means of translating 性 literally as *saga* (natural disposition) would be impossible.

10.2 Stage Two: Conceptual Combination

Conceptual combination is channelled by tentatively assigning the unfamiliar compound to one of four fundamental types: coordinate, appositional (or attributive, if the modifier is not nominal), relational and semantically exocentric. Basically, obtaining a plausible result under one of these four interpretations does not exclude the possibility that the compound might actually belong to one of the other types. However, it may interrupt the process and prevent a reader from looking for a more robust output under another interpretation.

As part of the ‘best-scenario hypothesis’, I have assumed that all the four possible interpretations are attempted in every instance and that they are taken in no specific sequence.⁵⁰ The order in which I will discuss them in the next sections is, therefore, irrelevant, although I will deal with exocentricity last because it is the least probable and the most difficult to overcome.

The following are the problems and the possible solutions for each of the four interpretations.

10.2.1 Coordinate Interpretation

Among the compounds I analysed, only two are constituted by a pair of first-order nouns naming first-order entities or substances. One is *saru+hito* (ape+man), this essay’s main case study. Here, the two named entities have the right degree of similarity to combine into any subtype of co-compound. However, the principle of salience, whereby the left-hand constituent of a co-compound is socio-culturally more important than the right-hand one, makes it improbable that a reader would interpret *saru+hito* as coordinate. *Ishi+abura* (rock+grease), on the other hand, could be coordinate, because the two constituents have the right degree of similarity and there is no way of ruling out that, in some scientific field, rocks might be more salient than greasy substances. A mechanism of lexical validation then kicks in whereby the reader looks for co-compounds having *ishi-* and *-abura* as constituents. They will hardly find any. Unsupported by distribution, a co-compound interpretation is ruled out in this case as well.

10.2.2 Appositive and Attributive Interpretations

Appositive interpretation is based on property mapping, whereby the value of one of the attributes of the mark is ascribed to the base. Since comparable types of attribute can be found in almost any two entities, this strategy is epistemically very powerful, to the extent that, if it yields a plausible match, it might prevent the reader from looking for other interpretations. Thus, for *karasu+bōshi* (crow hat), appositive ‘black hat’ trumps the relational ‘a hat worn by crows’ and does indeed capture an actual feature of the referent. In other cases, though, property mapping proves wrong. In *saru+hito*, the similarity

⁵⁰ Taking this position, after Wisniewski, Love (1998), I distance myself from the so-called “last resort hypothesis”, whereby a property-mapping interpretation of an N+N compound is only attempted if relation linking fails (Downing 1977; Gagné, Shoben 1997; Shoben, Gagné 1997).

between the two entities is just right, so that the erroneous outcome ‘a human being resembling an ape’ (transferring the values of attributes like ‘nose shape’, ‘hairiness’, ‘tree-climbing skills’, ‘intelligence’, etc. of apes to the corresponding slot of humans) would satisfy and make relation linking superfluous. The same holds for *ishi+abura* (rock+grease). Interpreting it as ‘stone-hard grease’ = ‘very thick grease’ would be a mistake. What compels the reader to rule this interpretation out is the very low frequency of 石 and 油 as, respectively, the modifier and head of appositive compounds.

Interpreting *ha+midori* as ‘the green colour of leaves’, *chi+iro* as ‘the colour of blood’ and *tamago+shiro* as ‘the white colour of eggs’ also originates false beliefs, but for different reasons. Here the problem is that, under property mapping, the elements in each pair are associated and locked together, while in fact they are arguments of the third constituent of their respective compound and are related to it independently of each other. This issue may only be prevented at the first stage by means of identifying the rightmost constituent as the affixal base and individuating the specific thematic role it bears in relation to the elements it heads as part of its lexical properties.

Interpretation fares better when the modifier is not nominal but verbal or adjectival and the compound is attributive. This is obvious because, in such instances, the semantic relationship between mark and base is always overtly expressed. Thus, in the compound mark of 裸子植物 (gymnosperm), construing *hadaka* 裸, a noun meaning ‘naked body’, as ‘naked’, an adjective, as most frequently happens, leaves no doubt that *hadaka+tane* 裸子 is to be interpreted literally as ‘naked seed’. The meaning of *hadaka* is clear even if the head constituent, 子, is wrongly construed to mean ‘fruit’ or even ‘child’. The interpretation of 裸子植物 opens the way to the successful decoding of 被子植物 (angiosperm). The evident homology of the two compounds directs the reader to construe 被 as the antonym of 裸 (naked) and 子 (seed) as the theme.

10.2.3 Relational Interpretation

The power of property mapping makes it necessary to postulate that relation-linking is attempted first, or at least that both strategies are pursued in any case regardless of their relative success. This is what I have assumed in this essay.⁵¹ The problem with subordinate N+N compounds is that when the mark is simple and a verbal element is

⁵¹ At the level of my analysis, and *contra* what I asserted in the previous footnote, it is therefore reasonable to postulate that in a reader’s mind relation linking is performed first, as claimed on experimental grounds by the supporters of the last resort hypothesis.

lacking, there is no overt indication about how to link the constituents. This leaves a vast array of theoretically viable thematic relations to be somehow reduced in number.

The main filter is provided by availability, i.e. the frequency with which each constituent occurs in any given relation as part of all the compounds in the lexicon (Gagné, Shoben 1997). Admittedly, the low availability of one particular relation cannot rule out that relation might occur in the very compound under interpretation, for *kun*-glossing is typically applied to technical terms, whose scientific fields are narrow and little known to the average reader. Yet, for any two constituents, certain relations are so underrepresented or even absent in the lexicon that it is justified to assume that interpreters will ignore them. In compounds, for instance, humans are rarely represented as agents by means of 人 and never in relation with an animate mark. This allows any interpretation of *saru+hito* along the line of ‘a human being doing something to/with apes’ to be excluded. Availability effectively reduces the range of plausible interpretations in the processing of *ishi+abura* as well. First, the most frequently occurring thematic relation, ‘material’, is discarded. This exclusion takes place at a semantic level because, in the combined concept ‘a grease made of stone’, the predicate is not consistent with the properties of its two arguments. Next, ‘benefactive’ is discarded because it is unavailable for the two constituents. Four thematic relations survive. When tested against each other, two emerge as highly plausible. They are admitted to the next stage and undergo encyclopedic validation.

In the three-*kanji* compounds analysed here, recognizing the affixal function of the rightmost constituent, an operation performed during the previous stage, leads to a thematic role being assigned to it in relation to the mark. As a result, *moto* 素 is successfully construed as the unique causer of the entity (either concrete or abstract) described by the rest of the compound. Similarly, *moto* 質 is taken as a substance comprised in another substance, either that expressed by the mark or another unspecified one. Obviously, when the mark of a three-constituent compound contains a verbal element the compound is not N+N anymore and the thematic relation is overtly expressed within the mark itself. This occurs in 自動車 (automobile), in which the mark ascribes the property predicated by the verbal phrase ‘moving by itself’ to the base, ‘car’. In 蜂食性 (apivory), the suffix means ‘natural disposition’. The value of it is predicated by the verbal mark 蜂食 (feeding on bees).

The two four-*kanji* compounds I analysed in this essay – 被子植物 and 裸子植物 – share the same head, 植物 (plant). Once 被子 and 裸子 are decoded as denoting types of seed, a meronymic relation between mark and base suggests itself as highly probable. Availability supports this intuition.

10.2.4 Coping with Semantically Exocentric Compounds

Do *kun* assignment and conceptual combination allow readers to identify a compound as semantically exocentric and infer its denotation? In one case, yes. When processing *卵白質, knowledge of the compounds in which eggs and colour terms are associated supports a construal of 白 (white) in *卵白 (egg+white) as a substance and not as a colour. This understanding is reinforced by the meaning ‘component of a substance’ carried by *moto* 質. As regards the expression 鯷の烏帽子 (sardine’s tall hat = *P.physalis*), the exocentricity of 烏帽子 (crow-hat-child) can only be circumvented by prior knowledge of 帽子 for ‘hat’. The iconicity directs the reader to identify the entire parent expression as semantic exocentric – a protruding fish-part and not an ancient piece of headgear. This overcoming of the exocentricity block is a considerable feat. However, out of context, the fact that the parent expression refers to a living being cannot be inferred by any means. Its denotation remains out of the reach of *kun*-glossing.

In other cases, exocentricity cannot even be detected. The correct literal interpretation of *卵白質 (the main component of egg white) suffices in itself and provides no hint that the term is actually a synecdoche for the superordinate concept ‘protein’.

A related but opposite question is whether the outcome of the operations performed during the first two stages might instead exclude exocentricity. In the processing of 裸子植物 and 被子植物, the issue of the actual exocentricity of the complex marks, 裸子 and 被子, is simply prevented from arising once *ko* 子 (child) is construed as *tane* (seed) due to frequency effects. The analysis of 葉綠素 and 血色素 illustrates that reckoning the suffixal role of *moto* 素 as the causer of a property (a second-order entity) prevents an exocentric interpretation of the mark as a substance. Frequency effects also disallow an exocentric interpretation of the compound in the case of *saru+hito* 猿人. In contrast, the very high frequency of exocentric compounds among those having the structure [石+N] – one third of the listed items – suggests the possibility that 石油 is exocentric as well. In this and in all the other cases, the issue of semantic exocentricity can only be solved by means of encyclopedic validation.

10.3 Stage Three: Encyclopedic Validation

Kun-glossing is concluded by an encyclopedic search designed to validate the new complex concept over existing knowledge. An encyclopedic entry is sought whose content matches the information obtained by conceptual combination. In Suzuki’s model, as reconstructed in this essay, this validation cannot be dispensed with. One reason is that the previous stages often yield several coherent interpretations, even

after semantic filtering. Knowledge consistency is the last tool the reader can apply in order to choose just one of them. A second reason is that, for *kun*-glossing to be informative, the correct encyclopedia entry must be found in which to accumulate all the data concerning the compound's referent, both those gathered from context (including the semantic contents carried by its name) and those acquired in the future. The epistemic value of the entire process depends on that.

The validation is not merely satisfactorial, based on a match between the provisional meaning(s) obtained so far and semantic knowledge. As noted, encyclopedia entries pile up information gathered through any type of epistemic relation. Thus, validation also occurs relationally, over the full range of the interpreter's stored experiences with the objects of the world. It is in this sense that a distinction between an 'encyclopedic' knowledge of semantic nature and a more encompassing 'world' knowledge might be drawn.

However, the output of the validation process is still subject to a final assessment based on simple heuristics (Simon 1956, 129; for an approach parallel to mine, see Chow 2011, 167). The final match, if any, will often be the result of some forced simplification. On the one hand, the description submitted for validation is only a tiny fragment of meaningful information. Even the most complex descriptions constructed by Suzuki's hypothetical readers, like 'a component of the white part of eggs' or 'a plant with clothed seeds', do not contain much data. There is a vast plethora of known objects corresponding to such descriptions. On the other hand, the compound under interpretation might belong to a specialised knowledge field. In this case, a match is only obtained if the reader possesses that relevant knowledge, no matter how much time and energy they might be willing to devote to the first stages of compound processing. This means that, most times, interpreters have to content themselves with obtaining a good-enough match. The height of the threshold of acceptance can be set at will but ultimately depends on how rich an interpreter's knowledge actually is in the relevant field(s).

In this paragraph, I will discuss the two opposite outcomes of the validation process: either a link to an existing encyclopedia entry is established or no match is returned.

10.3.1 No Match is Obtained

This situation occurs when interpreters have no stored information about an object corresponding to the description built up through conceptual combination. For instance, when searching for 'a plant with naked seeds', they may know about plants, about the fact that plants might be classified according to the structure or appearance of the seed they produce, and even that 'naked seeds' may be relevant to that

classification, whatever ‘naked’ might mean. But, if they have no object file on a specific entity matching that description, the search fails.

As an effect, they generate a new file (Allan 2006a, 575). This file is graphically headed by the just-analysed *kanji* string and labeled with the compound provisionally constructed by means of *kun* assignment. At creation, it contains the information resulting from conceptual combination, however poor it might be, but is ready to receive new data, starting with contextual information, and is, of course, linked to the particular concepts used to form its descriptive content.

When the reader must do with what they have learned by means of *kun* assignment and concept combination there is the risk that the file’s content is unknowingly false in relation to the properties of the real object denoted by the compound. If undetected and stored as true, these coherent sets of misbeliefs will create misconceptions. The most extreme case of semantic exocentricity discussed in this essay well illustrates the problem.

At the conclusion of the decoding process of 鱧の烏帽子 (Portuguese man o’ war), the reader has successfully determined the semantic exocentricity of the expression. But an encyclopedic search for objects resembling blackish, sardine-related, *eboshi*-like objects still fails to return an entry. A new file is thereby created to store information about similar objects. But actual *Physalia* are quite different from the description encoded by the expression. As an effect, the new file will unavoidably end up storing irrelevant and even untrue information.

This does not imply that the reader may not have an encyclopedia entry on the creature. On the contrary, they could have a large amount of well-organised and accurate world knowledge about it, for example having encountered a beached specimen, or having been stung when swimming. They might even have studied the creature in a laboratory and, in this case, they will surely have a mental name for it, perhaps even identical to the Linnaean one. Yet no epistemically rewarding relation can bridge the gap between description and reference and return the right entry, or even allow the new file to be merged with the old one in the future. Unless a fortunate event of os-ten-sion occurs, the two files will forever remain apart.

A similar case is that of 蛋白質 (protein). By construing *moto* 質 as ‘component’ and assigning the role of theme to it in a locative relation, the reader is able to overcome the exocentricity block and understand that *shiro* 白 (white) refers to a substance – *shiromi* (egg white) – and not to a colour. Notwithstanding, there is no way for them to infer that ‘a component of egg white’ stands, in fact, for the entire class of chemical molecules it belongs to, no matter how rich their biology-related knowledge might be. Rather, a new file is created whose very existence is misleading and which immediately begins to accumulate information in the wrong place, with no possibility of being linked to, or merged into, the correct concept-file, ‘protein’.

The same risk subsists in relation to semantically endocentric compounds. The case of 葉緑素 (leaf-green-substance = ‘chlorophyll’) and 血色素 (blood-colour-substance = ‘hemoglobin’) is particularly revealing. For these compounds, conceptual combination yields two plausible interpretations: ‘the substance determining leaf-green/blood-colour in objects’ and ‘the substance determining the colour in leaves/blood’. As I have repeatedly remarked, the two interpretations are conceptually very dissimilar, and only the latter is correct. But it is not possible to differentiate between them at an earlier stage. Both survive up to encyclopedic validation. Once this stage has been entered, though, the correct interpretation can only be selected if the reader already has an encyclopedia entry with matching contents. Lacking this entry, the reader will probably choose the former interpretation, not only because construing the relationship between colour concepts and concrete objects by means of property mapping is much more intuitive, but also because relational linking is not supported by the lexicon for the relevant constituents. Thus, again, new files will be created for 葉緑素 and 血色素 to house untrue content. At best, if the reader perceives the danger, both conflicting descriptions will be uploaded and retained. The reader will uphold choosing between them in the hope of future disambiguation by other means.

Similarly, if the reader has no notion of paleoanthropology, no file is returned for *saru+hito*, to the effect that a new file concerning a ‘person resembling an ape’ is created to gather information about hairy, ugly, sub-intelligent, clumsy or savage people. At creation, the file is linked to the entries storing the reader’s experiences in the real world. Subsequently, though, the file will have a life by itself under the new label. It will persist in piling up information under a wrong header (for there is no word of that intension in the lexicon) until some new knowledge is revealed in the future.

For other compounds, a failed encyclopedic validation causes no risk of misinformation. This is the case of 蜂食性 (apivory) and 魚食性 (piscivory) – the compounds most transparent and easiest to decode among those analysed in this essay – and of 裸子植物 (gymnosperm) and 被子植物 (angiosperm), once 子 and 被 are correctly construed. For all these words, if the encyclopedic search fails (a likely outcome), a new file is started which is already named with the correct scientific term and readied to house relevant, true information – to begin with, that encoded by the naming compound itself, whose descriptive content matches the actual properties of the denotatum. These files represent novel and valid encyclopedia entries, albeit still poor. In these cases, *kun*-glossing is informative and enriches an interpreter’s knowledge.

10.3.2 A File Is Returned

Encyclopedia entries are returned for those technical terms which are also in ordinary usage. In the case of 石油 (rock-greasy_substance = 'petroleum'), locative and derivative relations both match the ordinary knowledge about oil acquired through the epistemic relations available in ordinary life: that the object is viscous, is found in the depths of the earth, is pumped out of the ground, is obtained from rocks, etc. The case of 自動車 (self-moving-cart = 'automobile') is more complex. Here, a correct interpretation requires the reader to know about engines, devices that convert power into movement and that, when used to move a cart, are either harnessed to it in the form of animals, like horses, or, with more advanced technology, are man-made and installed on the vehicle itself, thus becoming part of it, to the effect that, when they work, the cart seems to spontaneously displace itself. Readers then individuate a corresponding entity among the objects comprised in their knowledge of the world. It should be noted that, in the case of 葉緑素 (chlorophyll), 血色素 (hemoglobin) and 猿人 (*Pithecanthropus*), the relevant encyclopedia entries must exist and drive the interpretation *a posteriori*. For 猿人, this essay's main case study, the lack of encyclopedic knowledge about the referent is so devastating for the entire process of *kun*-glossing that Suzuki (2014, 208) himself has to dismiss it as impossible.

When a file is returned, the interpreter's knowledge is not enriched with new information. By having them rename an existing entry, *kun*-glossing only serves to reorganise their encyclopedias and enlarge their lexicons. This is indeed the simple effect envisaged by Suzuki.

11 Conclusions

Suzuki's *kun*-glossing model can be reduced to the claim that interpreters can correctly infer the denotation of an unfamiliar *kanji* compound by assigning a Japanese word of known meaning to each *kanji* constituent and then by combining those words. I considered *kun*-glossing to be successful when it yielded the retrieving or the creation of a concept file about the actual object named by the novel compound. As Suzuki (2014, 10) himself points out, the concept so obtained need not be identical to the actual denotation. Being 'close to the thing' suffices. If a file already exists, it is surely less rich than the definitions in paper dictionaries and the entries in encyclopedias. It may even contain some misbeliefs. What counts as a 'correct' outcome is that the target concept file should not be uploaded with misbeliefs as an effect of the interpretive process, that is, that the information reconstructed by means of *kun*-glossing is true with respect to the actual intension of the source compound.

In this essay, I tested Suzuki's model by means of retracing the computational process required to reach the successful outcome defined above. To do so, I added two additional stages to the process – conceptual combination and encyclopedic validation – which are almost entirely ignored by Suzuki. I submitted several compounds to the test, making an effort to find out the path that leads to the successful interpretation claimed by Suzuki, no matter how tortuous that path may be.

To enable the readers to win their interpretive battle against unknown compounds I had to equip them with the best weaponry available. I therefore assumed that (1) they know all the *kun* readings and related meanings of each source Sinogram, and (2) they know all the existing compounds in which that character occurs. The latter assumption is necessary to supply them with a maximum level of familiarity with the usage of each *kanji* in compounds, thus enabling frequency effects to influence their choices. I considered such optimal knowledge to be best expressed by the information found in paper dictionaries, both those listing words by *kanji* ideograms and those listing them by sounds. A dictionary is, after all, the outcome of a corpus-based lexical research carried out with traditional methods. The ideal state of affairs I endowed upon interpreters I dubbed the 'best-scenario hypothesis'.

11.1 When *Kun*-Glossing Succeeds

Under the above complex conditions, *kun*-glossing succeeds in capturing the actual denotation of the majority of the compounds discussed in this essay: 石油 (petroleum), 自動車 (automobile), 蜂食性 (apivory), 裸子植物 (gymnosperm), 被子植物 (angiosperm), 葉緑素 (chlorophyll), 血色素 (hemoglobin). The analysis shows that, for all these items, when multiple conceptual combinations are possible, frequency effects are effective in reducing the number of possible literal meanings. In a few cases, the same factors also direct interpreters to the correct construal of semantically exocentric expressions. They even enable readers to correctly identify suffixal heads and find out their relevant meanings.

For the above compounds, reference is successfully retrieved in two opposite ways. If interpreters already possess an encyclopedia entry whose content matches the description yielded by *kun*-glossing, that file is retrieved and renamed. If they have no such file, a new one is started, headed by the new word and containing true information: true because it is consistent with the properties of the compound's actual denotatum.

11.2 A Criticism of the ‘Best-Scenario Hypothesis’

For some of the compounds that are correctly interpreted – 石油 (petroleum), 自動車 (automobile), 蜂食性 (apivory) – *kun* assignment and conceptual combination are straightforward, as there are few or no alternative outputs to be compared and balanced against each other. For the remaining items, however, there are many viable ways to relate the constituents to each other. Only the thorough knowledge of the lexicon assumed under the ‘best-scenario hypothesis’ can ensure that the reader is sufficiently familiar with the different usages of the constituents to be influenced by distributional factors.

But how plausible is the ‘best scenario hypothesis’? Two related forms of criticism can be levelled against it.

11.2.1 Unlimited Lexical Knowledge

The first criticism concerns the vastness of the lexical knowledge the hypothesis requires. Admittedly, family size and availability are factors in play whenever compounds are to be interpreted. As such, they do presuppose a vast access to the lexicon. It is on this assumption that prior knowledge of 裸 (naked) in *rashishokubutsu* 裸子植物 ‘gymnosperm’ can be legitimately considered a necessary precondition to the construal of 被 as ‘clothed’ in *hishishokubutsu* 被子植物 ‘angiosperm’. But exposing readers to 裸子植物 and 被子植物 at the same time – as Suzuki (2014, 211) cleverly does, aware of the facilitating effect that decoding the former would unavoidably have on decoding the latter – is actually a form of *ad hoc* priming. In ordinary communicative situations this priming may not occur at all. This is the extreme form of a problem that occurs when processing any compound. Making interpreters navigate the whole lexicon until they spot the very usage of a given constituent that explains the behaviour of that constituent in an unfamiliar compound is nothing but a hidden, indirect priming operation, artfully repeated in all instances of *kun*-glossing.

Another unreal situation occurs when interpreting the compounds headed by 素 and 質. Here, *kun* assignment leads nowhere, because *moto* is too polysemic. The actual semantics of the two suffixes is encoded by the *kanji*. As seen, construing them as ‘causing substance’ and ‘component’ respectively is the only way to individuate the correct thematic relations linking them to their marks. But, in order to infer that semantics, the reader has to consider all suffixal usages of the two *kanji* in the lexicon, carefully comparing them, assessing their relative weight. Any mistake or negligence would invalidate the whole process. Epistemically, the ‘best scenario hypothesis’ is indispensable for reconstructing the first two stages of the *kun*-glossing

process as Suzuki envisages them. If considered real, though, the hypothesis becomes a clever tool for creating a purely artificial state of affairs in which the complexity of the lexicon and the operations to access it is conveniently reduced to zero. Under such conditions, any interpretive activity is invariably successful. This casts serious doubt on the effectiveness of *kun*-glossing in ordinary circumstances and suggests that its actual rate of success is much lower than that determined in this essay.

11.2.2 Unreal Interpreters

The second criticism concerns the contrast between the ignorance of the compounds to be interpreted and the very rich lexical knowledge necessary to successfully decode some of them. In general, this state of affairs is intuitively impossible. But what about the actors involved in the particular *kun*-glossing situations staged by Suzuki?

Suzuki only describes real-life *kun*-glossing scenarios in his 2014 book (208, 211). The settings are probably fictional. On the one side, he pitches native English-speaking university professors unable to extract meaning from “formidable-looking tongue twister[s]” and “bombastic English words” (Suzuki 1977, 418), like ‘pithecanthrope’, ‘gymnosperm’, ‘angiosperm’ ‘homoiothermal’, ‘poikilothermal’ (cf. Suzuki 2014, 208-11). On the opposite side he deploys average Japanese people, prototypically represented by junior high school students. These students are allegedly capable of retrieving ‘the ancestor of humanity’ from <猿人> (Suzuki 2014, 208), and of beating native English-speaking typists over the word pair 人類学/anthropology (Suzuki 2017, 92-3).⁵² Indeed, one of the ideas that started the present essay – namely that *kun*-glossing might be at work in everyday reading and for all kinds of unfamiliar compounds – also presupposes a demography of ordinary, relatively ignorant users. But these actors are, by definition, no god-like possessors of a mental lexicon as rich as that of paper-grade dictionaries. Quite simply, they do not fit into the hypothetical ‘best scenario’. No real-world language user may actually fit into it. The existence of Suzuki’s hypothetical reader and, consequently, the ‘best scenario’ itself are unrealistic.

The criticism can be extended to the relation between an interpreter’s lexical and encyclopedic knowledge. As noted above, in some cases the success of *kun*-glossing is completely dependent on pre-existing encyclopedic knowledge. Lacking this knowledge only generates misconceptions. Whereas for 猿人 (‘Pithecanthropus’) the possibility

⁵² The typist introduced by Suzuki (1990, 163 note 8) is actually a character in a novel by Agatha Christie.

cannot be dismissed that junior high school students know about a paleoanthropological entity with some property roughly matching the descriptive content of the compound but have no name for it, the encyclopedic knowledge necessary to correctly interpret 葉緑素 and 血色素 is very specialised. It is difficult to conceive of a deeply erudite reader who is, however, devoid of the lexical knowledge of the item under analysis.

11.2.3 No Experimental Validation

This is why, to my knowledge, Suzuki's model has never been tested experimentally. In general, the relevant literature is narrowly focused on proving the facilitating effect of *kanji* orthography over phonetic encoding (*contra* Suzuki), occasionally considering the position and conceptual type of the constituents as variables. Only two-Sinogram familiar compounds are examined or used as stimuli in experiments.⁵³ *Kun*-glossing, however, only concerns novel compounds. In order to test it, participants should be selected who are positively ignorant of the stimuli. But stimuli range from the rare 葉緑素 (chlorophyll) and 被子植物 (angiosperm), to terms in everyday use like 自動車 (car), 石油 (oil) and 半島 (peninsula). I see no way to individuate such a population. A preliminary filter based on lexical knowledge cannot avoid exposing possible participants to the very stimuli to be used in the experiment. Ignorance of certain compounds may only be presupposed intuitively and yet with no certainty withal. Therefore, whereas high frequency compounds like 自動車 and 石油 cannot be tested at all, even though Suzuki holds that *kun*-glossing can be successfully applied to them as well, very low frequency scientific terms can. Knowledge of individual *kanji* components is simply determined by education level.

The only experimental study I could find which is based on this kind of intuitional approach is published by Hatano, Kuhara, Akiyama (1981). But this paper, written in English and with no reference to a Japanese version, is actually off-topic because, in an effort to ground Suzuki's assertions, the authors compare the interpretation of *kanji* compounds by Japanese undergraduates with that of neo-classical compounds by American students – foreigners, foreign students, Americans return time and time again in the literature. Moreover, while mentioning *kun* readings as a key to accessing meanings,

⁵³ Here and hereafter, my assertions are based on the following additional sources and on the other works discussed in them: Nomura 1978, 1979; Saitō 1981; Morton, Sasanuma 1984; Yokosawa, Umeda 1988; Tamaoka, Hatsuzuka 1995, 1998; Kinoshita 1998; Kess, Miyamoto 2000, 33-82; Tamaoka, Taft 2010; Kuwabara 2015.

the authors do not investigate how *kun* assignment actually works.

Other studies, loosely connected to *kun*-glossing, aim to give a quantitative basis to Suzuki's notion of 'semantic transparency', make it scalar and group compounds by transparency class (see, for example, Honda 2017; and the two similar Kuwabara 2013 and Kawakami 2018). Suzuki (1978, 8; 1990, 134) summarily introduced the notion of semantic transparency as a property of the individual constituents of a compound.⁵⁴ However, after the influential Mori, Nagy (1999) - also inspired by Suzuki - semantic transparency has been reinterpreted as the easiness to relate the complex meaning of already-known compounds to the meanings of their individual constituents (Gagné 2011, 265). In the experiments, Japanese native speakers are asked to subjectively evaluate the transparency of each stimulus. The issues of how they retrieve those particular meanings and of the possible effect of *kun* assignment are ignored. When a process akin to conceptual combination in interpreting unfamiliar Japanese compounds is considered, the population of reference is young learners and, unsurprisingly, foreign students (see, for instance, Kuwabara 2012, 2017). Foreigners are indeed the only type of participant that may be supposed to have never been exposed to the stimuli. At the same time, in the perspective of Japanese linguistic nationalism, they are also the complacent victims of the arrogance of Western neoclassical compounding.

The only element of interest that the above mentioned studies bear in relation to Suzuki's original *kun*-glossing model is that they show that Japanese native speakers are indeed conscious of how uncertain the interpretive path can be that joins the meaning of each constituent to that of the parent compound. For instance, in Kuwabara (2013, 12), the stimulus rated as least transparent out of 500 items is *furo* 風呂 (glossable 'wind-vertebra'), an everyday term for 'bath' which, unsurprisingly, is semantically exocentric.

11.2.4 No Phenomenological Bubble

Lacking testability, Suzuki and his epitomes can only operate with a source word, an input into the decoding process - i.e. *kun* assignment - and the target meaning, successfully retrieved at the end of some undescribed mental operation. The risk is to treat the in-between phase as a phenomenological bubble: an obscure place where interpretive and re-encoding processes occur which are fuelled by

⁵⁴ In 1990, when doing so, Suzuki refers to *Semantics* by Stephen Ullmann (1962) and to the Japanese translation thereof by Yoshihiko Ikegami (1968), one of the most distinguished Japanese linguists.

meaning associations too deeply hidden to be reconstructed and explained. Taking this view, only native speakers can intuitively draw those associations and enact those processes. Such a phenomenology of linguistic experience represents a significant part indeed of Japanese cultural nationalism.

To Suzuki's credit, and despite his ideological biases, his frequent use of the expression "linguistic consciousness of the Japanese" (1963; 1990; 2017) and his reference to "unconscious" processes (1978; 2014; 2017), the scholar does not fall into the trap. He does not detail what happens after *kun* assignment, but neither does he suggest that only Japanese native speakers can understand those processes. Rather, he makes an effort to explain them to his readers, even if in a simplified and lacunose fashion. As for myself, this essay is nothing but an attempt to fill that void by determining a coherent sequence of operations bridging the input and the final outcome. I had to reconstruct the process abductively and in purely theoretical terms, relying on Suzuki's promise of success and without experimental validation. The 'best-scenario hypothesis' is one of the tools I deemed necessary to that purpose; its limits are the limits of *kun*-glossing itself.

11.3 When *Kun*-Glossing Fails

The problem is that there are cases in which *kun*-glossing does not work even under the 'best-scenario hypothesis'. Some cases involve semantic exocentricity. In a few instances, exocentricity is so blatant as to be easily recognisable. For example, context - the variable so strategically downplayed by Suzuki - would immediately clarify that 鰯の烏帽子 (*P.physalis*), and 風呂 (bath) do not refer to, respectively, hats and bones. In the case of *Physalia*, the implausible association of sardines, crows and hats also betrays the exocentric nature of the expression. It is safe to admit that interpreters, aware of the exocentricity of these compounds, would not attempt to forcibly extract meaning from them. These instances do not negate the general validity of the model.

In other instances, though, semantic exocentricity is far less evident and may well go unnoticed. *Kun*-glossing fails when, unable to detect and overcome the exocentricity block, it yields a consistent interpretation which only seems to match the actual denotation of the novel compound. The Japanese term for 'protein' 蛋白質 (or the domesticated version thereof *卵白質), best illustrates the issue. The Japanese compound is a synecdoche meaning 'a component of egg white'. This new combined concept is relevant, for proteins are part of the albumen. *Kun*-glossing has indeed returned a small piece of true knowledge. But this knowledge has hardly any informative value

because it does not enable the interpreters to go further upstream to the actual concept, ‘protein’, even if they already possess the right encyclopedic file.

An opposite type of failure is linked to a lack of relevant encyclopedic knowledge. In such a situation, a new file must be created. *Kun*-glossing fails when the optimal outcome of conceptual combination is a new complex concept the content of which is, however, false with respect to the actual denotation of the source compound. This situation is exemplified by 葉緑素 (chlorophyll) and by the sister compound 血色素 (hemoglobin). Here conceptual combination yields two opposite interpretations. The correct one (the substance causing leaves to be green) has little support from the lexicon and may prevail only if validated by the content of an existing encyclopedia file storing the crucial piece of knowledge that one and only one pigment family is responsible for the unique colour of leaves. Lacking this knowledge, interpreters cannot avoid choosing the more intuitive quality-mapping interpretation ‘a leaf-green substance’ or ‘the substance causing objects to be leaf-green’. If stored under the header, 葉緑素, this description constitutes misinformation.

11.4 *Pithecanthropus* Farewell

The interplay of conceptual combination and encyclopedic validation is particularly complex in the case of 猿人 because this term and the original ‘*Pithecanthropus*’ are only indexes to the works of Haeckel (1868, 1877, 1889), the coiner of the neoclassic term, and their apparent semantics only acts as a remainder to the expert reader of the essential properties of the object they identify.

But if the readers have no paleoanthropological knowledge, they can only interpret *saru+hito* literally. *Kun*-glossing, then, yields nothing more than the trivial association ‘a human being somehow related to apes’. Or, in the more probable case of a quality mapping interpretation, ‘an apish person’. Then, the reader might even start looking in the world around them for people like that. The justification of the existence in the lexicon of this term could be, quite simply, that it is a negative epithet for an ugly person or a prankster. This outcome represents a failure of *kun*-glossing.

Only knowledge of Haeckel’s works and of the history of paleoanthropology enables readers to infer the true meaning of 猿人. After *kun* assignment, they have to skip conceptual combination altogether and directly seek an encyclopedia entry whose header is a compound formed by terms meaning ‘ape’ and ‘human being’. They would then retrieve the lemma, ‘*Pithecanthropus*’, whose entry contains the definitory properties determined by Haeckel (1868, 292-4; 1889, 709-10) and clarifying information about the relationship of the object with

Anthropopithecus and with present-day taxa *Australanthropus* and *Homo erectus*. In this scenario, *kun*-glossing fails for two reasons. First, the reader does not have to combine the two concepts. Actually, they must refrain from doing so. Second, an implausible state of affairs is presupposed in which a Japanese reader is expert enough to have a rich knowledge of *Pithecanthropus* but no Japanese term to name it. Only non-Japanese readers might profitably process 猿人 in this way.

If, however, interpreters have some low-grade encyclopedic knowledge of human evolution, *kun* assignment and subsequent conceptual combination will allow them to activate the general conception – absorbed during science classes, perhaps indeed in junior high school, while watching TV or leafing through *manga* – of an early evolutionary stage of humankind when people were somehow ape-like. Because the word says nothing about the specific apish characteristics of those ancestors of humanity, there is still the risk that readers might form their own images of ape-men as hairy, bent at the neck and knee, long-armed and flat-nosed; and possibly, in a more sophisticated representation, grunting inarticulate sounds and incapable of using tools. Of Haeckel, they know nothing, though, so that any correspondence between their mental image of 猿人 and some of Haeckel's taxonomic traits is merely down to luck.

Admittedly, this is an improvement in comparison with any ape-man conceptions built up without the support of encyclopedic knowledge because those concern types of person, while this one corresponds to a class of human beings with some historical grounding. But the informativity of this outcome is still close to nothing. It consists of recording the fact that yes, there is a Sino-Japanese name for that entity, vaguely remembered from school days, and in opening a link for associating a set of (mostly false) beliefs to it. And yet this is the processing depth implicit in Suzuki's scenario, as he honestly admits: while English people only see a puzzling string of letters forming a word, Japanese people understand that the word has something to do with primitive men.

12 A Final Assessment

In Suzuki's (1963; 1969; 1975; 1977; 1978; 1990; 2014; 2017) model, it is individual *kanji* – coupled with their *kun* readings – which are the fuel of *kun*-glossing. They are enough to start and sustain the whole interpretive process. Taking Suzuki's view, to be successful, *kun*-glossing only needs interpreters who were taught the relevant *kanji* and associated readings as part of the school syllabus. When convenient, Suzuki pretends to ignore that *kun*-glossing requires much more than mere *kun* assignment and even that *kun*-reading selection and assignment are themselves often problematic and need vast lexical knowledge to be effective.

Thus, *kun*-glossing only works when done by the ideal interpreters devised by Suzuki, in the artificial circumstances he sets up for them, and for carefully selected items like the calques of neoclassical compounds, whose semantic transparency has been artificially built-in from creation – but actually not even in all the instances he himself proposes as best examples of it. And in order to be epistemically valuable, *kun*-glossing sometimes demands prior knowledge of the object, at other times it concedes there might be none. At times it returns the exact denotation, though often just a vague idea of it. All things considered, it cannot be taken as a reliable method for processing unfamiliar compounds in general and even scientific 'big words' only. It does indeed enable readers to extract meaning from words like 魚食性 (piscivorous) or generate a new encyclopedic file about the concept descriptively expressed by 被子植物 (angiosperm). But what interest will a Japanese person ever have to componentially decode everyday words like 牛乳 (cow milk), 自動車 (car), 石油 (oil)? The need to *kun*-gloss these terms would imply that there are native speakers who have corresponding conceptions but no name to handle them. This is impossible. Yet Suzuki never acknowledges this issue because he never discusses the real-life situations in which *kun*-glossing may prove helpful, with their complex interplay of lexical semantics, encyclopedic knowledge and context. In his discourse, as in reality, *kun*-glossing is not a model of conceptual combination – and indeed, conceptual combination must be skipped entirely if one wants to retrieve the reference of 猿人. Rather, it is a game-like associative process that allows native speakers to playfully discover the hidden compositionality of familiar words they have never thought about, making them feel proud of the superiority of Japanese over English.

Bibliography

- Akinaga Kazuo 秋永一枝 (1981). *Meikai nihongo akusento jiten* 明解アクセント辞典 (Japanese Accent Dictionary). 2nd ed. Tōkyō: Sanseidō.
- Allan, K. (2006a). "Dictionaries and Encyclopedias: Relationship". Brown 2006, 3: 573-7. <https://doi.org/10.1016/b0-08-044854-2/04335-2>.
- Allan, K. (2006b). "Lexicon: Structure". Brown 2006, 7: 148-51. <https://doi.org/10.1016/b0-08-044854-2/01096-8>.
- Bauer, L. (2006). "Compound". Brown 2006, 2: 719-26. <https://doi.org/10.1016/b0-08-044854-2/00111-5>.
- Bauer, L. (2008). "Dvandva". *Word Structure*, 1(1), 1-20. <https://doi.org/10.3366/e1750124508000044>.
- Bauer, L. (2010). "The Typology of Exocentric Compounds". Scalise, S.; Vogel, I. (eds), *Cross-Disciplinary Issues in Compounding*. Amsterdam: John Benjamins, 167-75. <https://doi.org/10.1075/cilt.311.14bau>.
- Bauer, L. (2011). "Typology of Compounds". Lieber, Štekauer 2011, 343-56. doi.org/10.1093/oxfordhb/9780199695720.013.0017.
- Bauer, L. (2017). *Compounds and Compounding*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108235679>.
- Bisetto, A.; Scalise, S. (2005). "The Classification of Compounds". *Lingue e Linguaggio*, 4(2), 319-30. <https://doi.org/10.1093/oxfordhb/9780199695720.013.0003>.
- Brown, K. (ed.) (2006). *Encyclopedia of Language & Linguistics*. 2nd ed. New York: Elsevier.
- Bunkachō 文化庁 (Agency for Cultural Affairs) (2010). *Jōyōkanjihyō* 常用漢字表 (List of *kanji* Characters for Daily Use). https://www.bunka.go.jp/kokugo_nihongo/sisaku/joho/joho/kijun/naikaku/pdf/joyo-kanjihyo_20101130.pdf.
- Ceccagno, A.; Scalise, S. (2006). "Classification, Structure and Headedness of Chinese Compounds". *Lingue e Linguaggio*, 5(2), 233-60.
- Chow, S.J. (2011). *Heuristics, Concepts, and Cognitive Architecture: Toward Understanding How the Mind Works* [PhD dissertation]. London (Ontario): University of Western Ontario.
- CNRTL (Centre National des Ressources Textuelles et Lexicales). s.v. "Automobile". <https://www.cnrtl.fr/etymologie/automobile>.
- Cohen, B.; Murphy, G.L. (1984). "Models of Concepts". *Cognitive Science*, 8, 27-58. https://doi.org/10.1207/s15516709cog0801_2.
- Cooper, W.; Ross, J. (1975). "World Order". Grossman, L.; San, J.; Vance, T.J. (eds), *Papers from the Parasession on Functionalism*. Chicago: Chicago Linguistic Society, 63-111.
- Cumming, S. (2019). s.v. "Names". Zalta, E.N. (ed.), *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/fall2019/entries/names/>.
- dalla Chiesa, S. (2023). "On Interpreting 'Peninsula' and the Japanese 半島 'Half-Island'". *Geography Notebooks*, forthcoming.
- Dōrokōtsūhō 道路交通法 (Road Traffic Act) (Act No. 106 of 1960; Act No. 76 of 2015). <https://elaws.e-gov.go.jp/document?lawid=335AC0000000105>.
- Downing, P. (1977). "On the Creation and Use of English Compound Nouns". *Language*, 53, 810-42. <https://doi.org/10.2307/412913>.

- Ereshefsky, M. (2001). *The Poverty of the Linnaean Hierarchy: A Philosophical Study of Biological Taxonomy*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/cbo9780511498459>.
- Gagné, C.L. (2011). “Psycholinguistic Perspectives”. Lieber, Štekauer 2011, 255-71. <https://doi.org/10.1093/oxfordhb/9780199695720.013.0013>.
- Gagné, C.L.; Shoben, E.L. (1997). “Influence of Thematic Relations on the Comprehension of Modifier-Noun Combinations”. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 71-87. <https://psycnet.apa.org/doi/10.1037/0278-7393.23.1.71>.
- Gagné, C.L.; Spalding, T.L. (2004). “Effect of Relation Availability on the Interpretation and Access to Familiar Noun-Noun Compounds”. *Brain and Language*, 90, 478-86. [https://doi.org/10.1016/s0093-934x\(03\)00459-0](https://doi.org/10.1016/s0093-934x(03)00459-0).
- Gagné, C.L.; Spalding, T.L. (2006a). “Using Conceptual Combination Research to Better Understand Novel Compound Words”. *SKASE Journal of Theoretical Linguistics*, 3(2), 9-16.
- Gagné, C.L.; Spalding, T.L. (2006b). “Conceptual Combination: Implications for the Mental Lexicon”. Libben, G.; Jarema, G. (eds), *The Representation and Processing of Compound Words*. Oxford: Oxford University Press, 145-68. <https://doi.org/10.1093/acprof:oso/9780199228911.003.0007>.
- Gensen (1986). *Kokugodaijiten Gensen 国語大辞典言泉* (Gensen Japanese Dictionary). Tōkyō: Shōgakukan.
- Golas, P.J. (1999). *Science and Civilisation in China*. Vol. 5(13), *Chemistry and Chemical Technology. Mining*. Cambridge: Cambridge University Press.
- Guerra, S. (2015). *Kanji. Grande dizionario giapponese-italiano dei caratteri*. Bologna: Zanichelli.
- Haeckel, E.H.O.A. (1868). *Natürliche Schöpfungsgeschichte*. Berlin: Georg Reimer.
- Haeckel, E.H.O.A. (1877). *Anthropogenie oder Entwicklungsgeschichte des Menschen*. 3 Bde. Leipzig: Wilhelm Engelmann.
- Haeckel, E.H.O.A. (1880). *The History of Creation: On the Development of the Earth and Its Inhabitants by the Action of Natural Causes*. New York: D. Appleton & Company. Transl. of: *Natürliche Schöpfungsgeschichte*. Berlin: Georg Reimer.
- Haeckel, E.H.O.A. (1889). *Natürliche Schöpfungsgeschichte*. 8 Bde. Berlin: Georg Reimer.
- Hatano, G.; Kuhara, K.; Akiyama, M. (1981). “*Kanji* Help Readers of Japanese Infer the Meaning of Unfamiliar Words”. *The Quarterly Newsletter of the Laboratory of Comparative Human Cognition*, 3(2), 30-3.
- Hirose Shizumu 廣瀬鎮 (1984). *Animaruroa no teishō. Hito to saru no minzokugaku アニマルロアの提唱—ヒトとサル民俗学—* (For a Lore of Animals: The Folklore of Humans and Monkeys). Tōkyō: Miraisha.
- Honda Yumiko 本多由美子 (2017). “Nijikango ni okeru go no tōmeisei. Kōpasu o mochiita go to kōseikanji no bunseki” 二字漢語における語の透明性—コーパスを用いた語と構成漢字の分析— (Word Transparency in Two-Character Sino-japanese Compounds: A Corpus-Based Analysis of Words and Constituents Characters). *Keiryō kokugogaku*, 31(1), 1-19.
- Iwanami seibutsugaku jiten* 岩波生物学辞典 (Iwanami's Dictionary of Biology) (1985). Tōkyō: Iwanami.
- Kageyama, T. (1982). “Word Formation in Japanese”. *Lingua*, 57, 215-58. <https://doi.org/10.1093/obo/9780199772810-0194>.

- Kageyama, T. (2011). “Isolate: Japanese”. Lieber, Štekauer 2011, 512-26. <https://doi.org/10.1093/oxfordhb/9780199695720.013.0028>.
- Kageyama, T.; Kishimoto, H. (eds) (2016). *Handbook of Japanese Lexicon and Word Formation*. Berlin: De Gruyter Mouton. <https://doi.org/10.1515/9781614512097>.
- Kaizuka Shigeki 貝塚茂樹; Fujino Iwatomo 藤野岩友; Ono Susumu 小野忍 (eds) (1984). *Kadokawa kanwa chūjiten* 角川漢和中辞典 (Kadokawa's Medium-Size Kanji-Japanese Dictionary). Tōkyō: Kadokawa.
- Kawakami Masahiro 川上正浩 (2018). “Kanji 2ji jukugo no imi to sore o kōsei suru kanji no imi to no kanrensei (tōmeisei) dētabēsu no kōchiku” 漢字二字熟語の意味とそれを構成する漢字の意味との関連性 (透明性) データベースの構築 (Constructing a Database of the Correlation between Constituent Meanings and Whole Word Meaning (Transparency) in Two-Character Compounds). *Ōsakashōin joshidaigaku kenkyūkiyō*, 8, 19-28.
- Kess, J.F.; Miyamoto, T. (2000). “Kanji Processing”. Kess, J.F.; Miyamoto, T. (eds), *The Japanese Mental Lexicon: Psycholinguistic Studies of Kana and Kanji Processing*. Amsterdam: John Benjamins, 33-82. <https://doi.org/10.1075/z.95.04kan>
- Kinoshita, S. (1998). “The Role of Phonology in Reading Japanese: Or Why I Don't Hear Myself When Reading Japanese”. *Reading and Writing: An Interdisciplinary Journal*, 10, 439-55. https://doi.org/10.1007/978-94-015-9161-4_14.
- Kobayashi, H.; Yamashita, K.; Kageyama, T. (2016). “Sino-Japanese Words”. Kageyama, Kishimoto 2016, 93-131. <https://doi.org/10.1515/9781614512097-007>.
- Koizumi Tamotsu 小泉保 et al. (eds) (1989). *Nihongo kihon dōshi yōhō jiten* 日本語基本動詞用法辞典 (Dictionary of the Usage of Japanese Basic Verbs). Tōkyō: Taishūkan.
- Kōjien* 広辞苑 (Large dictionary) (1976). Tōkyō: Iwanami.
- Kokugo daijiten* 国語大辞典 (Large Japanese Dictionary) (1981). Tōkyō: Shōgakukan.
- Kokuritsu kokugo kenkyūjo 国立国語研究所 (National Institute for Japanese Language and Linguistics) (1972). *Dōshi no imi, yōhō no kijutsuteki kenkyū* 動詞の意味・用法の記述的研究 (A Descriptive Study on the Meaning and Usages of Japanese Verbs). Tōkyō: Shūei shuppan.
- Kuwabara Yōko 桑原陽子 (2012). “Kanji 2ji jukugo no imi suisoku ni oyobasu gokōsei ni kansuru chishiki no eikyō. Shuyōbu no ichi to no kakawari kara” 漢字2字熟語の意味推測に及ぼす語構成に関する知識の影響—主要部の位置との関わりから— (The Influence of Knowledge Concerning Word Structure on Interpreting Unknown Two-kanji Compounds: With a Focus on Headedness). *Fukui daigaku ryūgakusei sentā kiyō*, 7, 1-10.
- Kuwabara Yōko (2013). “Kanji 2ji jukugo no imi no tōmeisei no chōsa” 漢字2字熟語の意味の透明性の調査 (An Enquiry on the Semantic Transparency of Two-kanji Compounds). *Fukui daigaku ryūgakusei sentā kiyō*, 8, 1-14.
- Kuwabara Yōko (2015). “Kanji 2ji jukugo no imi no tōmeisei no bunseki” 漢字2字熟語の意味の透明性の分析 (An Analysis of the Semantic Transparency of Two-Kanji Compounds). *Fukui daigaku ryūgakusei sentā kiyō*, 2, 1-9.
- Kuwabara Yōko (2017). “Kanji 2ji jukugo ga kango dōmeishi ka dō ka no handan ni oyobasu kōseibun no eikyō. Hikanjikeichū jōkyūgakushūsha taishō no chōsa no kekka kara” 漢字2字熟語が漢語動名詞かどうかの判断に及ぼす語構成の影響—非漢字系中上級学習者対象の調査の結果から— (The

- Influence of Word Structure on Determining the Nominal or Verbal Status of Two-*Kanji* Compounds: The Results of an Enquiry Targeting Intermediate-To-Advanced Non-*Kanji* Learners). *Kokusai kyōiku kōryū kenkyū*, 1, 27-36.
- Lehmann, C. (2004). "Interlinear Morphemic Glossing". Booij, G.; Lehmann, C.; Mugdan, J. (eds), *Morphology: An International Handbook on Inflection and Word-Formation*, vol. 2. Berlin: De Gruyter, 1834-57. <https://doi.org/10.1515/9783110172782.2.20.1834>.
- Lieber, R. (2011). "A Lexical Semantic Approach to Compounding". Lieber, Štekauer 2011, 78-104. <https://doi.org/10.1093/oxfordhb/9780199695720.013.0005>.
- Lieber, R.; Štekauer, P. (eds) (2011). *The Oxford Handbook of Compounding*. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199695720.001.0001>.
- Lyons, J. (1977). *Semantics*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/cbo9781139165693>; <https://doi.org/10.1017/CBO9780511620614>.
- Matsumura Akira 松村明 (ed.) (1988). *Daijirin* 大辞林 (Large Dictionary). Tōkyō: Sanseidō.
- Matsunaga, S. (1996). "The Linguistic Nature of *Kanji* Reexamined: Do *Kanji* Represent Only Meanings?". *The Journal of the Association of Teachers of Japanese*, 30(2), 1-22. <https://doi.org/10.2307/489563>.
- McDonald, G. (2011). "Georgius Agricola and the Invention of Petroleum". *Bibliothèque d'Humanisme et Renaissance*, 73(2), 351-63.
- Monbukagakushō 文部科学省 (Ministry of Education, Culture, Sport, Science and Technology) (2009). *Gakunenbetsu kanjihaitōhyō* 学年別漢字配当表 (List of *Kanji* Per School Year). https://www.mext.go.jp/a_menu/shoutou/new-cs/youryou/syo/koku/001.htm.
- Montanari, F. (2004). *Vocabolario della lingua greca. Greco-italiano*. Torino: Loescher.
- Mori, Y.; Nagy, W. (1999). "Integration of Information from Context and Word Elements in Interpreting Novel *Kanji* Compounds". *Reading Research Quarterly*, 34(1), 80-101. <https://doi.org/10.1598/rrq.34.1.5>.
- Morioka Kenji 森岡健二 (1987). *Moji no kinō* 文字の機能 (The Function of Characters). Tōkyō: Meiji shoin.
- Morton, J.; Sasanuma, S. (1984). "Lexical Access in Japanese". Henderson, L. (ed.), *Orthographies and Reading: Perspectives from Cognitive Psychology, Neuropsychology, and Linguistics*. Hillsdale: Erlbaum, 25-42. <https://doi.org/10.4324/9781315107448-3>.
- Murphy, G.L. (1988). "Comprehending Complex Concepts". *Cognitive Science*, 12, 529-62. https://doi.org/10.1207/s15516709cog1204_2.
- Nagano, A; Shimada, M. (2014). "Morphological Theory and Orthography: *Kanji* As a Representation of Lexemes". *Journal of Linguistics*, 50, 323-64. <https://doi.org/10.1017/s0022226714000024>.
- Namiki, T.; Kageyama, T. (2016). "Word Structure and Headedness". Kageyama, Kishimoto 2016, 198-235. <https://doi.org/10.1515/9781614512097-010>.
- Nelson, A.; Haig, J.H. (1997). *The New Nelson Japanese-English Character Dictionary*. Tokyo: Tuttle.
- Nomura Yukimasa 野村幸正 (1978). "Kanji no jōhō shori. Ondoku, kundoku to imi no fuyo" 漢字の情報処理—音読・訓読と意味の付与— (*Kanji*

- Informational Processing: *On Readings, Kun Readings and Meaning Assignment*). *Shinrigaku kenkyū*, 49(4), 190-7.
- Nomura Yukimasa (1979). "Kanji no jōhō shori. Ondoku, kundoku no kensaku katei" 漢字の情報処理—音読・訓読の検索過程— (*Kanji Informational Processing: Retrieval via On and Kun Readings*). *Shinrigaku kenkyū*, 50(2), 101-5.
- Recanati, F. (2012). *Mental Files*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199659982.001.0001>.
- Recanati, F. (2013). "Perceptual Concepts: In Defense of the Indexical Model". *Synthese*, 190(4), 1841-55. <https://doi.org/10.1007/s11229-013-0264-6>.
- Recanati, F. (2017). "Cognitive Dynamics: A New Look at an Old Problem". De Ponte, M.; Korta K. (eds), *Reference and Representation in Language and Thought*. Oxford: Oxford University Press, 179-94. <https://doi.org/10.1093/oso/9780198714217.003.0011>.
- Saitō Hirofumi 齋藤洋典 (1981). "Kanji to kana no yomi ni okeru keitaiteki fugōka oyobi on'inteki fugōka no kentō" 漢字と仮名の読みにおける形態的符号化及び音韻的符号化の検討 (A Study of Morphological and Phonemic Encoding in Reading *Kanji* and Syllabic Characters). *Shinrigaku kenkyū*, 52(5), 266-73.
- Scalise, S.; Bisetto, A. (2011). "The Classification of Compounds". Lieber, Štekauer 2011, 34-53. <https://doi.org/10.1093/oxfordhb/9780199695720.013.0003>.
- Scarborough, J. (1992). *Medical and Biological Terminologies: Classical Origins*. Norman: The University of Oklahoma Press.
- Schreuder, R.; Baayen, H.R. (1997). "How Complex Simple Words Can Be". *Journal of Memory and Language*, 37, 118-39. <https://doi.org/10.1006/jmla.1997.2510>.
- Shěn Kuò 沈括 (1088). *Mèng xī bītán* 夢溪筆談 (Dream Pool Essays). <https://www.gutenberg.org/ebooks/author/2419>; <https://zh.wikisource.org/wiki/夢溪筆談?uselang=ja>.
- Shibatani, M. (1990). *The Languages of Japan*. Cambridge: Cambridge University Press.
- Shoben, E.J. (1991). "Predicating and Nonpredicating Combinations". Schwanenflugel, P.J. (ed.), *The Psychology of Word Meanings*. Hillsdale: Erlbaum, 117-35.
- Shoben, E.J.; Gagné, C.L. (1997). "Thematic Relations and the Creation of Combined Concepts". Ward, T.B.; Smith, S.M.; Vaid, J. (eds), *Creative Thought: An Investigation of Conceptual Structures of Structural Alignment in Conceptual Combination and Processes*. Washington, D.C.: American Psychological Association, 31-50. <https://doi.org/10.1037/10227-002>.
- Simon, H.A. (1956). "Rational Choice and the Structure of the Environment". *Psychological Review*, 63(2), 129-38. <https://doi.org/10.1037/h0042769>.
- Spahn, M.; Hadamitzky, W. (1989). *Japanese Character Dictionary. With Compound Lookup via any Kanji*. Tokyo: Nichigai Associates.
- Štekauer, P. (2005). "Onomasiological Approach to Word-Formation". Štekauer, P.; Lieber, R. (eds), *Handbook of Word-Formation*. Dordrecht: Springer, 207-32. https://doi.org/10.1007/1-4020-3596-9_9.
- Štekauer, P. (2006). "Onomasiological Theory of Word Formation". Brown 2006, 9: 34-7. <https://doi.org/10.1016/b0-08-044854-2/00122-x>.

- Štekauer, P. (2011). "Meaning Predictability of Novel Context-Free Compounds". Lieber, Štekauer 2011, 272-97. <https://doi.org/10.1075/sfs1.54>.
- Suzuki Takao 鈴木孝夫 (1963). *A Semantic Analysis of Present-Day Japanese with Particular Reference to the Role of Chinese Characters*. Tokyo: The Keio Institute of Cultural and Linguistic Studies.
- Suzuki Takao (1969). "Hyōki to shite no kanji" 表記としての漢字 (*Kanji as Notation*). *Gengo seikatsu*, 214, 17-25.
- Suzuki, T. (1975). "On the Twofold Phonetic Realization of Basic Concepts: In Defence of Chinese Characters in Japanese". Peng, F.C.C. (ed.), *Language in Japanese Society: Current Issues in Sociolinguistics*. Tokyo: University of Tokyo Press, 175-92.
- Suzuki, T. (1977). "Writing is Not Language, or Is It?". *Journal of Pragmatics*, 1, 407-22.
- Suzuki Takao 鈴木孝夫 (1978). "Kango wa gairaigo ka" 漢語は外来語か (Are Sino-Japanese Words Loans?). *Gekkan gengo*, 7(2), 2-8.
- Suzuki Takao (1990). *Nihongo to gaikokugo* 日本語と外国語 (Japanese and Foreign Languages). Tōkyō: Iwanami.
- Suzuki Takao (2014). *Nihon no kansei ga sekai o kaeru*. *Gengoseitaigaku bunmeiron* 日本の感性が世界を変える 言語生態学文明論 (Japanese Sensibility Will Change the World: An Ecolinguistic Theory of Civilizations). Tōkyō: Shinchōsha.
- Suzuki Takao (2017). *Tozasareta gengo*. *Nihongo no sekai* 閉ざされた言語・日本語の世界 (A Sealed Language: The World of Japanese). Tōkyō: Shinchōsha.
- Takashima Shizuo 高島鎮雄 (1994). "Jidōsha no teigi to shurui" 自動車の定義と種類 (Definition and Classification of *Jidōsha*). *Nihon daihyakkasho (Nipponika)*. Tōkyō: Shōgakukan. <https://kotobank.jp/dictionary/nipponica/>.
- Takashima Toshio 高島俊男 (2001). *Nihonjin to kanji* 日本人と漢字 (*Kanji and Japanese People*). Tōkyō: Bungeishunjū.
- Tamaoka Katsuo 玉岡賀津雄; Hatsuzuka Makiko 初塚真喜子 (1995). "Kanji ni-ji jukugo no shori ni okeru kanji shiyō no hindo no eikyō" 漢字二字熟語の処理における漢字使用の頻度の影響 (The Effects of *kanji*-Printed Frequency on Processing Two-*kanji* Compound Words). *Dokusho*, 39(4), 121-37.
- Tamaoka, K.; Hatsuzuka, M. (1998). "The Effects of Morphological Semantics on the Processing of Japanese Two-*Kanji* Compound Words". *Reading and Writing: An Interdisciplinary Journal*, 10, 293-322. <https://doi.org/10.1023/a:1008012220770>.
- Tamaoka, K.; Taft, M. (2010). "Sensitivity of Japanese Native Speakers to *On* and *Kun Kanji* Readings". *Reading and Writing: An Interdisciplinary Journal*, 23, 957-68. <https://doi.org/10.1007/s11145-009-9184-8>.
- Tamaoka, K. et al. (2015). "*Kanji* Database: A New Interactive Online Database for Psychological and Linguistic Research on Japanese *Kanji* and Their Compound Words". *Psychological Research*, 81, 696-708. <https://kanjidatabase.com/index.php>.
- Tattersall, I. (2009). *The Fossil Trail: How We Know What We Think We Know About Human Evolution*. Oxford: Oxford University Press.
- Tattersall, I.; Schwartz, J. (2001). *Extinct Humans*. Boulder: Westview Press.
- Thain, M.; Hickman, M. (2004). *The Penguin Dictionary of Biology*. London: Penguin Books.
- Tōdō Akiyasu 藤堂明保 (ed.) (1980). *Gakken kanwadaijiten* 学研漢和大字典 (Gakken's Large *Kanji*-Japanese Dictionary). Tōkyō: Gakushū kenkyūsha.

- Tokieda Makoto 時枝誠; Yoshida Seichii 吉田誠一 (eds) (1978). *Kadokawa kokugochūjiten* 角川国語中辞典 (Kadokawa's Medium-Size Japanese Dictionary). Tōkyō: Kadokawa shoten.
- Tollini, A. (1992). *Kanji. Elementi di linguistica degli ideogrammi giapponesi*. Pavia: Università degli Studi di Pavia.
- Ueda Kazutoshi 上田万年 et al. (eds) (1993). *Kōdansha shindaijiten* 講談社新大字典 (Kodansha's New Large Dictionary). Tōkyō: Kōdansha.
- Ullmann, S. (1962). *Semantics: An Introduction to the Science of Meaning*. New York: Barnes & Noble.
- Ullmann, S. (1969). *Gengo to imi* 言語と意味 (Language and Meaning). Transl. by Ikegami, Yoshihiko 池上義彦. Tōkyō: Taishūkan. Transl. of: *Semantics: An Introduction to the Science of Meaning*. New York: Barnes & Noble, 1962.
- Umesao Tadao 梅棹忠夫 et al. (eds) (1995). *Nihongo daijiten* 日本語大辞典 (Large Japanese Dictionary). Tōkyō: Kōdansha.
- van Oosterzee, P. (1999). *The Story of Peking Man*. Crows Nest: Allen & Unwin.
- Vitali, E. (2023). *The Quest for the Primordial: An Inquiry into the Nationalist Rhetoric of Contemporary Japan*. Roma: Sapienza Università Editrice. <https://doi.org/10.13133/9788893772839>.
- Wälchli, B. (2005). *Co-Compounds and Natural Coordination*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199276219.001.0001>.
- Wisniewski, E.J. (1996). "Construal and Similarity in Conceptual Combination". *Journal of Memory and Language*, 35(3), 434-53. <https://doi.org/10.1006/jmla.1996.0024>.
- Wisniewski, E.J.; Love, B.C. (1998). "Relations Versus Properties in Conceptual Combination". *Journal of Memory and Language*, 38(2), 177-202. <https://doi.org/10.1006/jmla.1997.2550>.
- Wood, B. (ed.) (2011). *Wiley-Blackwell Encyclopedia of Human Evolution*. Oxford: Wiley Blackwell. <https://doi.org/10.1002/9781444342499>.
- Yokosawa, K.; Umeda, M. (1988). "Processes in Human Kanji-Word Recognition". *Man, and Cybernetics = Proceedings of the 1988 IEEE International Conference on Systems* (Beijing; Shenyang, August 8-12, 1988), 377-80. <https://doi.org/10.1109/icsmc.1988.754318>.