

Sample of Existing N+V=N Compounds in Japanese

Kumiko Nakamura Akiko Takemura Timothy J. Vance

This small database is the result of an effort to collect a representative sample of existing compounds of the form N+V=N, i.e., compounds that are nouns and consist of a noun followed by the stem form (*ren'yō-kei*) of a verb. A typical example of such a compound is *te+fuki* 手拭き 'hand towel', which is based on the noun *te* 'hand' and the verb *fuk-u* 'wipe'.

The goal was to provide a set of data relevant to a well-known claim about N+V=N compounds in the literature on *rendaku*, namely, that *rendaku* is the norm unless the first element is semantically the direct object of the verb that the second element is based on (Okumura 1955; Kindaichi 1976; Sugioka 2002, 2005; Yamaguchi 2011). For example, in the case of *aji+tsuke* 味付け 'flavoring', the accusative particle *o* in the related phrase *aji o tsuke-ru* 'to add flavor' shows that *aji* is semantically the direct object of *tsuke-ru*. On the other hand, in the case of *kugi+zuke* 釘付け 'attaching with nails', the instrumental particle *de* in the related phrase *kugi de tsuke-ru* 'to attach with nails' shows that *kugi* is not semantically the direct object of *tsuke-ru*.

The first version of this database was compiled in 2002 in order to provide data for a presentation (Nakamura and Vance 2002). Vance did a systematic search using dictionaries, and then Nakamura reviewed the candidate examples and eliminated many for reasons explained below. Ten years later, Takemura reviewed the 2002 version, corrected the errors, and eliminated a few more items.

The first step in the initial search was to make a list of all the non-compound transitive verbs beginning with a voiceless obstruent that appear in Kazama 1979, a reverse dictionary that has a separate section for each part of speech. There is no point in considering intransitive verbs, since an intransitive verb cannot take a direct object. Verbs that do not begin with a voiceless obstruent are also beside the point, since *rendaku* cannot affect a vowel (as in *obo-e-ru* 'to memorize'), a sonorant (as in *nom-u* 'drink'), or an obstruent that is already voiced (as in *das-u* 'to put out'). Kazama 1979 includes many obsolete words, and the search was restricted to modern Japanese by removing from the list every verb that does not appear as a headword in Hasegawa et al. 1986, a small Japanese-English dictionary. Every verb that contains a medial voiced obstruent (e.g., *sage-ru* 'to lower') was also eliminated because of a well-known constraint on *rendaku* called Lyman's Law: *rendaku* almost never affects an element that already contains a voiced obstruent. Consequently, it would not be appropriate to cite N+V=N compounds such as *ne+sage* 値下げ 'price cut' (cf. *ne o sage-ru* 'to lower the price') as support for the claim that compounds of this form resist *rendaku* when the nominal element is semantically the direct object of the verbal element.

The next step in the process was to take the stem form of each verb remaining on the list and find every N+V=N compound containing that continuative form that appears in Kitahara 1990, a small reverse dictionary of Japanese. Then, in an effort to restrict the sample to common words, every such compound that does not appear either as a headword or as a compound under its nominal element in Hasegawa et al. 1986 (the small Japanese-English dictionary mentioned above) was discarded. For example, the verb *kar-u* 'to cut' has the

stem form *kari*, and Kitahara 1990 lists eight N+V=N compounds containing this element, but only three of these appear in Hasegawa et al. 1986: *kaku+gari* 角刈り ‘crew haircut’, *kusa+kari* 草刈り ‘grass mowing’, and *tora+gari* 虎刈り ‘uneven haircut’. Only these three are included in the database.

The next step was to specify two properties for each N+V=N compound in the sample: whether or not the nominal element is semantically the direct object of the verbal element (\pm DO) and whether or not the compound shows *rendaku* (\pm R). To illustrate with the three items containing the stem form of *kar-u* ‘to cut’, *kaku+gari* ‘crew haircut’ is $-$ DO/ $+$ R (cf. *kaku* ‘square’), *kusa+kari* ‘grass mowing’ is $+$ DO/ $-$ R (cf. *kusa* ‘grass’), and *tora+gari* ‘uneven haircut’ is $-$ DO/ $+$ R (cf. *tora* ‘tiger’).

Several compounds had to be eliminated from the sample at this stage for a variety of reasons. First, a few verbs take alternative case frames that make it impossible to decide for sure whether the nominal element in a compound should be considered a direct object. One such verb is *sas-u* ‘to stab’, which allows both *X ni Y o sas-u* ‘to stab Y into X’ and *X o Y de sas-u* ‘to stab X with Y’. The upshot is that a compound like *kushi+zashi* 串刺し ‘skewering’ (cf. *kushi* ‘skewer’) cannot be classified definitely as $+$ DO or $-$ DO, since it can be taken as either ‘stabbing (something) with a skewer’ (cf. *kushi de sas-u*) or as ‘stabbing a skewer (into something)’ (cf. *kushi o sas-u*). Compounds like *kushi+zashi* were simply deleted from the sample.

Also deleted were compounds with a second element that is related to a verb but is not synchronically verbal. For example, *ishi+datami* 石畳 ‘stone pavement’ contains the voiced allomorph of *tatami* ‘mat’, which derives etymologically from the stem form of *tatam-u* ‘fold’, but *tatami* functions as a simple noun in modern Japanese. In keeping with this rationale for eliminating items like *ishi+datami* from the sample, items like *tatami+gae* 畳替え ‘mat changing’ (cf. *kae-ru* ‘to change’) were retained, since the nominal status of *tatami* puts *tatami+gae* in the N+V=N category. Along the same lines, items such as *otoko+buri* 男ぶり ‘manliness’ (cf. *otoko* ‘man’) were eliminated. Although *buri* is etymologically related to the verb *fur-u* ‘to wave’, the semantic connection is tenuous, and the pattern N+*buri* is productive enough that Japanese dictionaries treat it under a separate entry for *buri* as a suffix.

In a small number of cases, an item was eliminated because the apparent nominal element is a bound morph. For example, *ka+barai* 過払い ‘excessive payment’ (cf. *hara-u* ‘to pay’) seems to have the form N+V=N, but Sino-Japanese *ka* ‘excess’ is bound, which means that it is not possible to offer a phrase of the form *ka* PARTICLE *hara-u* in support of categorizing *ka+barai* as $+$ DO or $-$ DO. Also eliminated were two items that are too semantically opaque to allow confident categorization: *mizu+hiki* 水引き ‘paper strings for tying presents’ (cf. *mizu* ‘water’, *hik-u* ‘to pull’) and *heso+kuri* 臍繰り ‘secret savings’ (cf. *heso* ‘navel’, *kur-u* ‘to reel’). (The nominal element in the second of these two examples is historically a folk etymology; the word *heso* ‘spool’ dropped out of use, and speakers reinterpreted the first element in *heso+kuri* as ‘navel’.)

Many vocabulary items, including some N+V=N compounds, vary between a form with *rendaku* and a form without. For example, although Hasegawa et al. 1986 and Kitahara 1990 list *saku+zuke* 作付け ‘crop planting’ (cf. *saku* ‘crop’, *tsuke-ru* ‘to attach’) only with *rendaku*,

the entry in *Kōjien* 『広辞苑』 (a comprehensive dictionary of Japanese) says that both *saku+zuke* and *saku+tsuke* are possible. Since such variability complicates any statistical profile of the database, examples listed both with and without *rendaku* in the 4th edition of *Kōjien* (Shinmura 1991) were simply deleted.

After all these adjustments, the sample contained 446 N+V=N items in which the verbal element is susceptible to *rendaku*. The final step in 2002 was for Nakamura (a native speaker of Japanese) to go through these items and eliminate those that were not in her vocabulary and those that she thought allowed or required a different pronunciation ($\pm R$). An example of the first type is *tonbo+tsuri* 蜻蛉釣り ‘dragonfly catching’ (cf. *tonbo* ‘dragonfly’, *tsur-u* ‘to catch’), which she did not recall ever hearing. An example of the second type is *kasen+shiki* 河川敷 ‘river basin’ (cf. *kasen* ‘river’, *shik-u* ‘to spread’), for which she had the pronunciation *kasen+jiki* (+R). (Shinmura 1991 lists only *kasen+shiki* (–R), but *Nihon kokugo daijiten* 『日本国語大辞典』 [Nihon Kokugo Dai-jiten Dainihan Henshū lin-kai 2000–02] lists *kasen+jiki* (+R) first and gives *kasen+shiki* (–R) as an alternative pronunciation.) These final deletions left a sample of 403 N+V=N compounds in 2002.

In 2012, Takemura (also a native speaker of Japanese) went through the 2002 sample and applied the same criteria that Nakamura used. She deleted 9 more entries, reducing the total number of items to 394. Takemura also added semantic information for each compound, specifying whether it can denote a person, a thing, or an action. These categories are crude, and in a few cases, they are really not appropriate. For example, non-human living creatures were treated as things, so the compound *mizu+sumashi* 水澄まし ‘whirligig beetle’ is categorized as having a “thing” meaning. On the other hand, since it does not denote a tangible, the compound *ko+garashi* 木枯らし ‘cold winter wind’ is categorized as having an “action” meaning but not a “thing” meaning. The upshot is that the semantic information in the database must be used with caution.

The list in its current form is a reasonably representative sample of relevant N+V=N compounds that are likely to be in an ordinary, educated native speaker’s vocabulary.

References

- Hasegawa, Kiyoshi, et al., eds. 1986. *Obunsha’s comprehensive Japanese-English dictionary*. Tōkyō: Ōbunsha.
- Kazama, Rikizō, ed. 1979. *Tsuzuriji gyakujun hairetsu gokōsei ni yoru Daigenkai bunrui goi*. Tōkyō: Fuzanbō.
- Kindaichi, Haruhiko. 1976. *Rendaku no kai*. *Sophia Linguistica* 2: 1–22.
- Kitahara, Yasuo, ed. 1990. *Nihongo gyakubiki jiten*. Tōkyō: Taishūkan.
- Nakamura, Kumiko, and Timothy J. Vance. 2002. *Rendaku in N+V=N compounds: A production task*. Paper presented at LP’2002, Urayasu, Japan.
- Nihon Kokugo Dai-jiten Dainihan Henshū lin-kai, ed. 2000–02. *Nihon kokugo dai-jiten*, 2nd ed. Tōkyō: Shōgakukan.
- Okumura Mitsuo. 1955. *Rendaku*. *Kokugo-gaku jiten*, ed. by Kokugo Gakkai, 961–962. Tōkyō: Tōkyōdō.
- Shinmura, Izuru, ed. 1991. *Kōjien*, 4th ed. Tōkyō: Iwanami.
- Sugioka, Yoko. 2002. Incorporation vs. modification in deverbal compounds. *Japanese/Korean linguistics* 10, ed. by Noriko M. Akatsuka and Susan Strauss, 495–508. Stanford: CSLI.
- Sugioka, Yoko. 2005. Multiple mechanisms underlying morphological productivity. *Polymorphous linguistics: Jim McCawley’s legacy*, ed. by Salikoko S. Mufwene, Elaine J. Francis, and Rebecca S. Wheeler, 204–223. Cambridge: MIT Press.
- Yamaguchi, Kyoko. 2011. Accentedness and *rendaku* in Japanese deverbal compounds. *Gengo Kenkyu* 140: 117–134.