Lexical strata in Japanese and Korean and the notion of lexeme

Clemens Poppe
Waseda University

1 Introduction: lexical strata and lexemes

Both Japanese and Korean are languages with three different lexical strata of an originally etymological nature. In both languages, the lexicon consists of so-called 'native' words, Chinese loan words, and more recent 'foreign' loan words (Shibatani 1990: Sohn 1999). In Japanese, for instance, both the free native element *hito* and the bound (°) Sino-Japanese elements °*nin* and °*jin* mean 'human, person' and are written with the same Chinese character. In Korean, the native element that expresses the meaning 'human, person' is *salam*, and the Sino-Korean element with the same meaning is °*in*.

In languages with neoclassical elements like French and English, we can find similar situations. Amiot and Dal (2007) discuss three different roots that refer to 'human': a native root *homme*, a Greek root, °*anthrop*, and a Latin root °*homin*, proposing that they belong to the same lexeme. A similar analysis is proposed for native and Sino-Japanese elements with corresponding meanings by Nagano & Shimada (2014), who also argue that *kanji* (Chinese characters) can be seen as representing lexemes.

The goal of this study is to critically discuss a number of issues in what we may call a 'shared lexeme' approach to native and Chinese loan elements in Japanese and Korean, and to look at the matter from the viewpoint of Construction Morphology (Booij 2010) and the closely related model of Relational Morphology (Jackendoff & Audring 2020).

2 Morphology, pragmatics, and orthography

For both Japanese and Korean, it has been pointed out that words belonging to the different strata have different ranges of meaning and stylistic values. Shibatani (1990: 144) writes that native words have broader meanings than Sino-Japanese and loan words, Sino-Japanese words have a more formal character and are used a lot in learned expressions, and foreign loan words have more modern and stylish connotations. A highly similar characterization of the three lexical strata in Korean is given by Sohn (1999: 88-89). In other words, native elements and Chinese elements have similar functions in the two languages, and what appear to be synonyms in reality are words with different pragmatic or sociolinguistic functions.

Given that such differences in abstractness and formality exist, the question arises why the elements that make up Sino-Japanese words should be analyzed as word-forms of the same lexeme as a corresponding native form with the same meaning. The main argument for such an analysis comes from the difference in morphological behavior between elements of the two strata: native elements are either free forms (in the case of nouns) or inflectional stems (in verbs and adjectives), whereas Sino-Japanese elements for which a corresponding native equivalent exists are bound forms. Based on this observation, Nagano & Shimada (2014) propose that nouns and verbs have two different stems: adopting the notion of 'stem space' (Montermini & Boyé 2012), they propose a distinction between a default stem and a compound stem. Their analysis is given in adapted form in (1a), where the default stem of nouns is called $\text{Stem}_{\text{Free}}$, a default stem which is in complementary distribution with a compound stem '$\text{Stem}_{\text{Comp}}$'. The forms in (1b) may serve as an example: the lexeme HITO

'human, person' has one free word-form *hito*, and two bound word-forms °*nin* and °*zin* (the distribution of which need not concern us here).

(1) a.

LEXEME	[−formal]	[+formal]
[−bound]	$\text{Stem}_{\text{Free}}$	-
[+bound]	-	$\text{Stem}_{\text{Comp}}$

b.

HITO	[−formal]	[+formal]
[−bound]	hito	-
[+bound]	-	°nin °zin

The tables in (1) give the impression that implicational relations exist between the two features. From the viewpoint of the language user, the [±formal] feature would seem to be the trigger to select a word based on bound or free elements. We would thus expect at least the implicational relations [+formal]→[+bound] and [−formal]→ [−bound]. Whether the relations also go in the opposite direction is a question that is more difficult to answer and depends on how we analyze the bound allomorphs of free forms which appear in compounds. There are two types of such allomorphs: *rendaku* stems and apophonic stems (see Labrune & Irwin 2021). To start with the latter type of allomorphy, the lexeme AME 'rain' is realized as *ame* in isolation, but either as *ame-* or *ama-* in compounds: *ame-huri* 'rain-fall' vs. *ama-gasa* 'rain-umbrella'. Whether the free stem or the apophonic stem is used is only partially predictable, depending on a whole range of factors (Labrune & Irwin 2021). The phenomenon known as *rendaku* or 'sequential voicing' refers to cases in which a non-initial compound member that in isolation starts with one of the four voiceless obstruents /h/, /s/, /t/, or /k/ is realized with initial /b/, /z/, /d/, or /g/ in non-initial position in a compound, depending on several phonological, morphological, and lexical factors (Vance 2014; Irwin 2005). In the case of *hito* ('human, person'), for instance, the *rendaku*-stem is °*-bito*, as in *mura-bito* ('village-person = villager'), *turi-bito* ('fishing-person = fisherman'), and *koi-bito* ('love-person, i.e. lover'). It may be clear that if we analyze *rendaku* stems as [+bound] forms, it is not possible to derive the value of the [±formal] from the [±bound] feature. To make things more complicated, sometimes even Sino-Japanese elements may undergo sequential voicing (Vance 1996). An example of such a word in *rin-goku* ('next (to)-country = neighboring country'), which can also be pronounced as *rin-koku*, without *rendaku*. Nagano and Shimada (2014: 343, footnote 29) refer to *rendaku* as a "phonological voicing operation", but this is a highly controversial characterization of the phenomenon; *rendaku* is far from regular and therefore should be viewed as morpho-phonological or lexical (van de Weijer et al. 2013; Vance 2014). As the marking of 'compoundhood' can be seen as the function of both rendaku and apophony (Labrune & Irwin 2021), it seems natural to treat them as compound stems. The consequences of such a view are shown in (2), where the *rendaku*-stem is given as Stem_{Ren} in (2a), and a concrete example in the form of the presumed structure of the lexeme KUNI 'country' in (2b). The following words are examples which contain the different allomorphs: *kuni* 'country', *shima-guni* 'island country', *koku-nai* 'domestic', *kan-koku* 'South Korea', and *chū-goku* 'China'.

(2) a.

LEXEME	[−formal]	[+formal]
[−bound]	$\text{Stem}_{\text{Free}}$	-
[+bound]	Stem_{Ren}	Stem_{Ren} $\text{Stem}_{\text{Comp}}$

b.

KUNI	[−formal]	[+formal]
[−bound]	kuni	-
[+bound]	°-guni	°-goku °koku

Evidently, under the analysis in (2) we need the feature [±formal] to distinguish between native and Sino-Japanese bound stems. To deal with cases in which bound stems that are specified with the same value for the formality feature, we could assign a separate

morphological feature [±rendaku] to *rendaku*-stems as in (3). This feature can be seen as an instance of the type of features used by Fradin (2003) to indicate that a stem is 'reserved' for a certain morphological construction (see also Amiot & Dal 2007). In other words, [±rendaku] could be formulated as [res: non-initial].

(3)

KUNI 'country'		[−formal]	[+formal]
[−bound]		kuni	-
[+bound]	[+rendaku]	˚-guni	˚-goku
	[−rendaku]		˚koku

In Korean, a phenomenon similar to *rendaku* exists which involves obstruent tensification and nasal gemination and goes by the name of *sai-sios* (lit. 'between-s') (see Labrune 1999). The Korean name of this phenomenon refers to the letter *sios* ('s') of the native *hankul* script which is inserted between two members of a compound and written in the coda of the initial member of a compound, where it is optionally realized as /t/ if constraints on syllable structure are satisfied (Sohn 1999). The name 'compound tensification' refers to the tensification of the first consonant of the lax onset of the non-initial member of a compound. For instance, in the compound consisting of the native words *cho* ('candle') and *pul* ('fire'), the second member has an initial tense consonant /pp/ rather than the lax /p/ that appears in the isolation form: *cho(t)-ppul* ('candlelight'). The same tensification can be found in the often-cited compound *pom-ppi* 'spring rain', which consists of the lexemes *pom* 'spring' and *pi* 'rain'. The Sino-Korean element with the same meaning 'rain' is *u*, as in the word *u-chen* ('rain-sky = rainy weather'). As *sai-sios* related phenomena are not fully predictable, the *sai-sios* stems can also be thought to be lexically listed, as in the Japanese case in (3) above. The fact that Korean shows the same patterns as Japanese is important evidence for the idea that knowledge of Chinese characters is not necessary to acquire knowledge of the lexical relations that are interpreted as lexemic by Nagano & Shimada (2014). Korean is generally no longer written by means of a combination of Chinese characters (*hanca*), and more crucially, Chinese characters have no native Korean readings. Despite these facts, Song (1986: 493) explicitly states that "[a] Sino-Korean morpheme is free, if there does not exist a native word, denoting the same sense", and that "[i]f there is a native word, the SK morpheme is bound". The Korean case thus strengthens the case made for the morphological relatedness between native elements and Chinese loan elements in Japanese by Nagano & Shimada (2014).

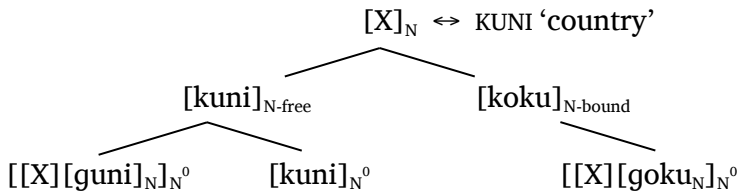
From the above we may conclude that the complementary distribution in terms of the free vs. bound and informal vs. formal distinctions in both Japanese and Korean should be reflected in the morphological analysis. In the final part of this study, an alternative construction-based analysis is proposed in which a formal distinction is made between two types of bound stems: stems which are lexically marked as bound, and stems which are part of a schema that contains a lexical variable.

3 A view from Construction Morphology

In Construction Morphology (Booij 2010), the lexeme can be seen as a set of forms sharing particular semantic features, a morphosyntactic category, and often phonological properties. In a construction-based analysis of the Japanese and Korean data discussed above, native forms and Chinese loan forms can be thought to share their semantics and lexical class feature, but not their phonological form. This more abstract 'lexeme' node in the hierarchical network dominates two types of stems: as shown in (4), a native stem which is specified as a free stem,

and a Chinese loan form which is specified as a bound stem. As a free stem, the native form [kuni] has a word form at the N^0 level. The *rendaku* allomorphs are dealt with by placing them one level lower than the more general schemas with a preceding variable 'X': $[[X][guni]_N]_{N^0}$ and $[[X][goku]_N]_{N^0}$. Although not shown in (4), these two constructions are at the same time instantiations of a more general *rendaku* schema, which captures the morpho-phonological nature of the allomorphy. (Note that both the $[kuni]_{N\text{-free}}$ and $[koku]_{N\text{-bound}}$ schemas also dominate further compound schemas which are not included in (4)) The distinction in formality, finally, can be captured by means of a default construction which expresses the implicational relation $[X]_{S\text{-bound}} \leftrightarrow$ [Pragmatics: formal], although possible alternative analyses will also be considered.

(4)



References

Amiot, Dany & Georgette Dal. 2007. Integrating neoclassical combining forms into a lexeme-based morphology. In Geert Booij, Luca Ducceschi, Bernard Fradin, Emiliano Guevera, Angela Ralli & Sergio Scalise (eds.), *On-line Proceedings of the Fifth Mediterranean Morphology Meeting*, 323–336. Bologna: Università di Bologna.

Booij, Geert. 2010. *Construction Morphology*. Oxford: Oxford University Press.

Fradin, Bernard. 2003. *Nouvelles approaches en morphologie*. Paris: Presses Universitaires de France.

Irwin, Mark. 2005. Rendaku-based lexical hierarchies in Japanese: The behavior of Sino-Japanese mononoms in hybrid noun compounds. *Journal of East Asian Linguistics* 14. 121–153.

Labrune, Laurence. 1999. Variation intra et inter-langue: morpho-phonologie du rendaku en japonais et du sai-sios en coréen. *Cahiers de Grammaire* 24. 117–152.

Labrune, Laurence & Mark Irwin. 2021. Japanese apophonic compounds. *Journal of Japanese Linguistics* 37(1). 25–67.

Montermini, Fabio & Gilles Boyé. 2012. Stem relations and inflectional class assignment in Italian. *Word Structure* 6. 69–87.

Nagano, Akiko & Masaharu Shimada. 2014. Morphological theory and orthography: *Kanji* as a representation of lexemes. *Journal of Linguistics* 50. 323–364.

Shibatani, Masayoshi. 1990. *The languages of Japan*. Cambridge: Cambridge University press.

Sohn, Ho-Min. 1999. *The Korean language*. Cambridge: Cambridge University Press.

Song, Ki Joong. 1986. Remarks on modern Sino-Korean. *Language Research* 22(4). 469–501.

Vance, Timothy J. 1996. Sequential voicing in Sino-Japanese. *The Journal of the Association of Teachers of Japanese* 30(1). 22–43.

Vance, Timothy J. 2014. If rendaku isn't a rule, what in the world is it? In Kaori Kabata & Tsuyoshi Ono (eds.). *Usage-based approaches to Japanese grammar: Towards the understanding of human language*, 137–152. Amsterdam & Philadelphia: John Benjamins.

van de Weijer, Jeroen, Clemens Poppe and Marjoleine Sloos (2013). Family matters: Lexical aspects of Japanese rendaku. In Jeroen van de Weijer & Tetsuo Nishihara (eds.). *Segmental variation in Japanese*, 129–148. Tokyo: Kaitakusha.