Classifying Correspondence in Japanese and Korean

Francis Bond

NTT Communication Science Laboratories

1-1 Hikari-no-oka, Yokosuka-shi, Kanagawa-ken, JAPAN 239

`bond@cslab.kecl.ntt.co.jp`

Kyonghee Paik

Keio University

6-13-10 Kugahara, Ota-ku, Tokyo, JAPAN 146

`i9724107@mita.cc.keio.ac.jp`

Abstract

In this paper we present an analysis of the ways Japanese and Korean categorize long-distance communication, for example, over the telephone, or through email, fax or letters. In Korean and Japanese, numeral classifiers are used to overtly categorize such communicative acts. Japanese uses more classifiers in this field than Korean, and so there is more possible variation in classifier use. We use the framework of Pustejovsky's (1995) Generative Lexicon to make clear the differences between kind and shape classifiers, and explain the variation in classifier use by showing how different classifiers select for different features of lexical entries.

1 Introduction

In this paper, we examine how Japanese and Korean classifiers are used to pick out certain aspects of the denotations of the nouns they classify.

Classifiers are used to pick out salient features of the noun they classify. These can be **kinds**, that is, semantic categories, or **properties** such as shape or size. Kind classifiers are characteristically used for complex objects and tend to be quite specific Denny (1979). Often several different classifiers can be used to classify the same things.

Earlier approaches to analyzing the meanings of classifiers have concentrated on typological surveys of classifiers Allan (1977); organizing classifiers into a hierarchy Denny (1979); describing classifiers and organizing them into groups according to meaning Unterbeck (1994); or giving semantic analyses for various classifiers, Matsumoto (1986, 1988, 1991).

The typological studies show the importance of classifiers in the worlds languages, and give a broad outline of what needs to be explained, but cannot go into much detail. Detailed studies such as Matsumoto's (1991)'s go further to explain how classifiers are used but are still fairly nebulous.

In this paper, we attempt to give a principled description of the use of classifiers in a small domain using the Generative Lexical framework of Pustejovsky (1995). In particular, we explain the difference between kind and shape classifiers as a difference in the kind of argument they select for. This allows us to provide an explanation for the variance in classifier use: why different classifiers can be used to quantize the same noun. Variance is an instance of selective binding, different classifers select for different parts of the lexical structures of the nouns.

In the Generative Lexicon, words have detailed lexical entries, which can be further extended dynamically, by operations such as type coercion. Each entry can have lexical information in the following categories: ARGUMENT STRUCTURE, EVENT STRUCTURE, LEXICAL INHERITANCE STRUCTURE and QUALIA STRUCTURE. QUALIA STRUCTURE is further broken down into CONSTITUTIVE, FORMAL, TELIC and AGENTIVE qualia which describe an entry's constituent parts; its relation to other things; its purpose; and how it is made respectively.

Japanese has two or three hundred numeral classifiers, although typically individual speakers use far less, between 30 and 80 (Downing, 1995, 346) and Korean has a similar number. In this paper we will concentrate on the use of classifiers to classify long-distance correspondence, such as, letters, telephone conversations and email.

We first give a brief account of the nature of numeral classifiers (2), and then focus on the semantics of classifying communication, first in Japanese (3) and then in Korean (4).

2 Classifiers

2.1 Numeral Classifiers

In many languages, including most South-East Asian languages, Chinese, Japanese and Korean, nouns can not be directly modified by numerals. Instead nouns are modified by a numeral-classifier combination, similarly to the way uncountable nouns are enumerated in English: (1).[1]

(1) 2-tsū-no   denshimēru (Japanese)
    2-tong-ui imeil   (Korean)
    2-CL-ADN   email

    2 pieces of email

In Japanese and Korean, numeral classifiers are a subclass of nouns. The main property distinguishing them from prototypical nouns is that they cannot stand alone. Typically they postfix to numerals, forming a quantifier phrase. Japanese also allows them to combine with the quantifier *sū* "some" or the interrogative *nani* "what" (2); Korean only allows them to postfix to numerals or the interrogative *myech* "what" (3).

(2) Japanese

   a. *2-hiki* "2 animals"   (Numeral)

   b. *sū-hiki* "some animals"   (Quantifier)

   c. *nan-biki* "how many animals"   (Interrogative)

(3) Korean

   a. *2-mari* "2 animals"   (Numeral)

   b. *myech-mari* "how many animals"   (Interrogative)

Both Korean and Japanese have two number systems: a general one based on Chinese, and an alternative native system. In Japanese the native system only exists for the numbers from one to nine, while in Korean it exists for the numbers from one to ninety nine. In Korean, classifiers combine with the native numerals where possible. In Japanese, most classifiers combine with the Chinese forms, however, different classifiers select Sino-Japanese for some numerals, and most classifiers undergo some form of sound change (such as *-hiki* to *-biki* in (2)). We will not be concerned with these morphological changes, we refer interested readers to Backhouse (1993, 118–122) for more discussion.

Quantifier phrases characteristically premodify the noun phrases they quantify, linked by an adnominal case marker, as in (4); or appear 'floating' as adverbial phrases just before the verb: (5). The choice between pre-nominal and floating quantifiers is largely driven by discourse related considerations Kim (1995); Downing (1993). In this paper we concentrate on the semantic contribution of the quantifiers, and ignore the discourse effects.

(4) 2-tsū-no   tegami-o   yonda   (Japanese)
    2-tong-ui pyunji-ru ilkessta (Korean)
    2-CL-ADN   letter-ACC read

    I read two letters

---

[1]We use the following abbreviations: TOP — topic; NOM — nominative; ACC — accusative; CL — classifier; ARGSTR — argument structure; ARG — argument; D-ARG — default argument; CONST — constituative; DIMEN dimensionality

(5) tegami-o 2-tsū yonda (Japanese)
     pyunji-ru 2-tong ilkessta (Korean)
     letter-ACC 2-CL   read

     I read two letters

Quantifier phrases can also function as noun phrases on their own, with anaphoric or deictic reference, when what is being quantified is recoverable from the context. For example (6) is acceptable if the letters have already been referred to, or are clearly visible.

(6) 2-tsū-o   yonda (Japanese)
     2-tong-ul ilkessta (Korean)
     2-CL-ACC read

     I read two letters

We propose the following basic structure (7) for classifiers. They have two variables in the argument structure, the numeral (or equivalent, represented by `numeral+`), and the noun being classified. Because the noun being classified can be omitted in context, it is a default argument, one which participates in the logical expressions in the qualia, but is not necessarily expressed syntactically (Pustejovsky, 1995, 63).[2]

(7)
$$
\text{CL}\begin{bmatrix} \text{ARGSTR} & \begin{bmatrix} \text{ARG1} & \texttt{x:numeral+} \\ \text{D-ARG1} & \texttt{y: ?} \end{bmatrix} \\ \text{QUALIA} & \begin{bmatrix} \text{FORMAL} & \texttt{quantifies(x,y)} \end{bmatrix} \end{bmatrix}
$$

Classifiers differ from each other in the restrictions they place on the quantified variable, y. In the next section, we will discuss the classifiers typically used to quantify correspondence, first in Japanese and then in Korean.

3 Japanese

Classifiers are often divided into two kinds, those that select for the FORMAL `type` of nouns (kind classifiers), and those that select for a certain property of the noun, characteristically its `dimensionality`, a sub value of the FORMAL quale (shape classifiers). We discuss shape classifiers, then kind classifiers, and finish by looking at examples of using two or more classifiers simultaneously.

3.1 Shape Classifiers

We consider first the shape classifiers, which are only used for inanimate nouns. We will only consider the shape classifiers that select for dimensionality. There are three main shape classifiers that select for dimensionality in Japanese (and most classifier languages Allan (1977)): -*hon* "1D" which selects for nouns that extend primarily in one dimension; -*mai* "2D" which selects for nouns that extend primarily in two dimensions; and -*ko* "3D" which selects for nouns in which no one dimension is greater than the others.

The lexical structure for -*ko* "3D" is given in (8), omitting the information about quantification. The structures for -*hon* "1D" and -*mai* "2D" are the same, except selecting for 1 and 2 dimensions respectively.

(8)
$$
\begin{matrix} \text{-}ko \\ \text{"3D"} \end{matrix}\begin{bmatrix} \text{ARGSTR} & \begin{bmatrix} \text{ARG1} & \texttt{x:numeral+} \\ \text{D-ARG1} & \texttt{y:} \begin{bmatrix} \texttt{inanimate} \\ \texttt{DIMEN} & \texttt{3D} \end{bmatrix} \end{bmatrix} \\ \text{QUALIA} & \begin{bmatrix} \text{FORMAL} & \texttt{quantifies(x,y)} \end{bmatrix} \end{bmatrix}
$$

---

[2]We omit the details of the quantification as irrelevant to the current discussion. For all numerical classifiers, the numerical argument quantifies the default argument.

According to the generative lexical framework dimensionality is a role of the FORMAL quale. However, for nouns, the FORMAL quale is used to give its basic category description: the sortal typing of the argument (either a simple or dotted type).

We consider three alternatives to describe the use of shape classifiers in the generative lexical framework:

1. Every object that can be counted by a shape classifier has a type made up of its normal sortal type dotted with it's shape.

2. The FORMAL quale must be allowed to take at least two values: (i) the sortal typing of the argument (possibly a dotted type); and (ii) a feature of dimensionality (DIMEN).

3. Dimensionality is is represented outside the qualia structure.

The first alternative is theoretically possible, but conceptually anomalous. It amounts to saying that any noun that can be classified by a shape classifier has two senses. But in considering something like a piece of paper *kami*, there are not two logically ambiguous senses: `paper` and `2D`, but rather it is `paper` which has `2D` dimensionality.

The second alternative goes against the stipulation of Pustejovsky (1995) that each qualia should have only one value, but fits our intuition best.

There is also a third alternative, that an object's shape is best considered as knowledge from outside the qualia structure. This still leaves us with the problem of how this world knowledge is accessed by the language faculty. Jackendoff (1992; 1996) suggests at least two kinds of mental representations: linguistic and spatial. In the spatial representation, distinctions between extension in primarily one, two or three dimensions can be encoded directly. We speculate that shape classifiers may be accessing this representation directly.

The discussion in the following sections does not need to distinguish between the above three alternatives, as long as it is possible to select for both a noun's sortal type and dimensionality to achieve lexical.

We will tentatively adopt the second alternative in this paper, as it is the simplest to represent. As an example we show the structure of -*kami* "paper" in (9) (omitting irrelevant qualia).

$$
(9) \quad
\begin{array}{l}
\textit{kami} \\
\text{``paper''}
\end{array}
\begin{bmatrix}
\text{ARGSTR} & \begin{bmatrix} \text{ARG1} & \texttt{x:paper} \end{bmatrix} \\
\\
\text{QUALIA} & \begin{bmatrix} \text{FORMAL} & \begin{bmatrix} \texttt{x} \\ \text{DIMEN} & \texttt{2D} \end{bmatrix} \\ \text{TELIC} & \texttt{writes-on(z:agent,x)} \end{bmatrix}
\end{bmatrix}
$$

According to Jackendoff (1996) the structure would be something closer to (10), where CS is the conceptual structure and SR is the spatial representation. The question this raises is, do shape classifiers directly access the spatial representation, or are they mediated by the conceptual structure?

$$
(10) \quad
\begin{bmatrix}
\text{Phono:} & \text{/kami/} \\
\text{Syntax:} & \text{+N, -V} \\
\text{CS:} & \begin{bmatrix} \text{ARGSTR} & \begin{bmatrix} \text{ARG1} & \texttt{x:paper} \end{bmatrix} \\ \text{QUALIA} & \begin{bmatrix} \text{FORMAL} & \texttt{x} \\ \text{TELIC} & \texttt{writes-on(z:agent,x)} \end{bmatrix} \end{bmatrix} \\
\text{SR:} & \begin{bmatrix} \text{DIMEN } \texttt{2D} \end{bmatrix}
\end{bmatrix}
$$

Another possible approach, between the second and third alternatives, is to add a new quale. The AQUILEX II project Copestake and Briscoe (1995), does this, although they are not motivated by the use of classifiers. In their relativised qualia structure they add a new quale, FORM which has a further feature SHAPE, where dimensionality can be encoded.

3.2 Kind Classifiers

In this section we look at kind classifiers, concentrating on those used to classify communication, particularly over the phone, or through fax, email, letters or postcards.

There is a major lexical distinction in Japanese in that *denwa* "phone", and *fakkusu* "fax" (and for some speakers *denshi-mēru* "email") are all verbal nouns, that is, they can be combined with *suru* "do" to become verbs. Note that these three can also become verbs in English, by a process of zero derivation.[3]

We give a partial hierarchy of semantic types in Figure 1. Types are marked with the classifier that is used to classify them: an explanation will follow. The hierarchy shows only the nodes relevant to the current issues, it is obviously not complete. The top nodes follow Pustejovsky (1995, 90), where `nomrqs` is the least upper bound for the types `proposition`, `event` and `entity`.

We show here the default classifiers for counting events and propositions: *-kai* "event" and *-ken* "proposition".[4]
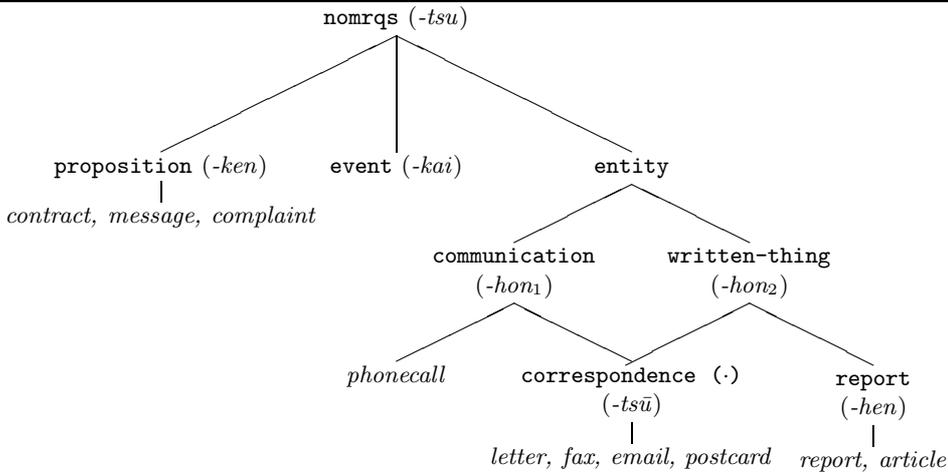

Figure 1: proposed hierarchy for communication in Japanese

Next we will give an example of the lexical structure of *denwa* "phonecall" (11). Because of its verb like properties, it has an explicit event structure. Because it is a subtype of `communication`, its TELIC value, which describes the purpose, includes the fact that someone communicates something to someone else. It consists of someone talking to someone else (the CONSTITUTIVE quale). A telephone call comes into being through someone making a phone call (the AGENTIVE quale).

$$
(11) \quad
\begin{array}{l}
\textit{denwa} \\
\text{``phonecall''}
\end{array}
\begin{bmatrix}
\text{ARGSTR} & \begin{bmatrix} \text{ARG1} & \text{x:communication} \end{bmatrix} \\
\text{EVENTSTR} & \begin{bmatrix} \text{EVENT1} & \text{e:Process} \end{bmatrix} \\
\text{QUALIA} & \begin{bmatrix}
\text{CONST} & \text{talk(z,y:agent)} \\
\text{FORMAL} & \text{x} \\
\text{TELIC} & \text{communicate(z,w:info,y)} \\
\text{AGENTIVE} & \text{phonecall(e,z,y)}
\end{bmatrix}
\end{bmatrix}
$$

tegami "letter" is an example of `correspondence`. `correspondence` is a combination of `communication` and `written-thing`, sharing characteristics of both (technically it is a 'dotted type' (Pustejovsky, 1995, pp 95–97)). Note that, as it is a case of `written-thing`, there is no need to explicitly include `read` and `write` in the TELIC and AGENTIVE roles. A letter is made of sheets of paper, giving the CONSTITUTIVE quale. Faxes, cards and

---

[3]In Japanese, *denwa* "phone" and *fakkusu* "fax" are also ambiguous between the instrument and the communication. In English, *phone* is the instrument and *phonecall* the communication, but *fax* is ambiguous; you can receive a fax using a fax, but you receive a phonecall using a phone. We will ignore this ambiguity in the rest of this paper.

[4]We are unclear on the status of *-ken*. It may well be the case that it does not count propositions, and it is almost certainly the case that it should be further down in the hierarchy, but we place it at the top in the absence of a more detailed analysis.

postcards all have similar lexical structure, although the AGENTIVE for fax will be `faxes(z,x)`; email is different in that it is not made of paper.

(12)
$$\begin{array}{l} \textit{tegami} \\ \text{``letter''} \end{array} \begin{bmatrix} \text{ARGSTR} & \begin{bmatrix} \text{ARG1} & \texttt{x:correspondence} \end{bmatrix} \\ \\ \text{QUALIA} & \begin{bmatrix} \text{CONST} & \texttt{part-of(u:paper,x)} \\ \text{FORMAL} & \texttt{x} \\ \text{TELIC} & \texttt{communicates(z:agent,w:info,y:agent)} \\ \text{AGENTIVE} & \texttt{posts(z,x)} \end{bmatrix} \end{bmatrix}$$

Now we return to the question of classifying communication.

Letters are characteristically counted with the kind classifier *-tsū* "correspondence". This selects for the noun type `correspondence`. The classifier *-tsū* is Sino-Japanese, and has the same use in Chinese. Its use has been extended to such things as faxes, postcards and recently email.[5]

Matsumoto (1991) gives a prototype based analysis, where nouns categorized by *-tsū* should ideally satisfy several conditions such as, a graded (continuosly varying) condition <linguistic message as a main content>, a necessary condition <sent by post>, and a characteristic condition <contained in an envelope>. However, a letter passed on by hand is still counted by *-tsū*, so posting is not a necessary condition. In contrast, a book sent by post in an envelope is not categorized by *-tsū*, although it appears to satisfy all three conditions.

(13)
$$\begin{array}{l} \text{-tsū} \end{array} \begin{bmatrix} \text{ARGSTR} & \begin{bmatrix} \text{ARG1} & \texttt{x:numeral+} \\ \text{D-ARG1} & \texttt{y:correspondence} \end{bmatrix} \end{bmatrix}$$

The classifier *-hon* is characteristically used to count phonecalls, and is less often used to count letters and faxes.[6] It is often argued that cases such as these are extensions of the shape use, for example Lakoff (1986, pp 25–27). We agree that this was probably originally the case, with *-hon* "1D" first used for telegrams, which were originally conceived of as wires even in English (e.g. *I'll send a wire* for *telegram*), however, we argue that the shape classifier *-hon* "1D" has given rise to (at least) two kind classifiers, *-hon₁* "communication" and *-hon₂* "written-thing". They straightforwardly select for noun types, and their representations are given in example (14). hon_1 is the classifier normally used for telephone calls, and hon_2 is the classifier used for counting written reports and articles.

Iida (1996) claims that different classifiers select for different telicities: **activity** is classified by *-kai* "event", **accomplishment** by *hon* and **achievement** by *-ken* "proposition". While it is the case that *-hon₁* "communication" is only used for successful phone calls, those that either reached the receiver or an answering machine, this is not because *-hon* particularly implies accomplishment. It is rather that the classifier puts a bound on the activity measured out by the verb, in this case the phone call. Unsuccessful attempts, or aggregates including unsuccessful events, must be counted using *-kai* "times" the classifier for `events` (or activities).

(14) a.
$$\begin{array}{l} \text{-hon}_1 \end{array} \begin{bmatrix} \text{ARGSTR} & \begin{bmatrix} \text{ARG1} & \texttt{x:numeral+} \\ \text{D-ARG1} & \texttt{y:communication} \end{bmatrix} \end{bmatrix}$$

b.
$$\begin{array}{l} \text{-hon}_2 \end{array} \begin{bmatrix} \text{ARGSTR} & \begin{bmatrix} \text{ARG1} & \texttt{x:numeral+} \\ \text{D-ARG1} & \texttt{y:written-thing} \end{bmatrix} \end{bmatrix}$$

Shape classifiers can also be used to pick out what the thing is made of. In particular faxes, cards and postcards, which generally consist of one sheet are often counted with *-mai* "2D". This is done by a process of type coercion, as they consist of paper, which has the FORMAL DIMENSIONALITY of 2D (a basically flat object that extends in two dimensions). Postcards in particular, which must consist of a single sheet, are often counted

---

[5]Note, however, that the status of email is not universally agreed on. One informant said he counts email with *-hon*, because he considers it to be closer to phone calls than letters. But this is non-standard: Japanese email software, for example, presents the number of email messages using *-tsū*.

[6]There is another classifier used to count phonecalls, *-ko₁* "phonecall", which is the technical term used by telephone engineers. It has a very specific meaning, and only used in technical texts, where it is the only choice. *-ko₁* "phonecall" is written with the Chinese character meaning "call" (as in *yobu* "call"), while *-ko* "3D" is written with the character meaning "individual".

by *-mai*. With letters, which characteristically consist of several sheets of paper, *-mai* is not normally used, and if it does, it refers to the sheets of paper which make up the letter, as in (15). We consider the fact that postcard must be a single sheet to be world knowledge, and the distinction a pragmatic one, although it is a grey area. *-mai* "2D" cannot be used with email, instead the default shape classifier *-ko* "3D", is used.

(15) 3-mai-no tegami
      3-CL-ADN letter

      {a 3 page letter/*3 letters}

The choice of classifier depends on which aspect of the classified noun's meaning is salient. For example, consider the following examples, adapted from Iida's (1996, pp. 132-133) (7a), (7b) and (8b).

(16) watashi-wa tegami-o   3-{tsū/hon$_2$/$^{\#}$kai} kaita
      I-TOP   letter-ACC 3-CL   wrote

      I wrote {3 letters/3 long letters/[a letter/letters] 3 times}

(17) shikashi 2-{tsū/*hon$_{1or2}$/$^{\#}$kai}-shika toukan-shinakatta
      but   2-CL-only   posted

      But I only posted two of them.

(18) Hiromi-wa   1-{tsū/hon$_1$/$^{\#}$kai}-mo tegami-o   yokosanai.
      Hiromi-TOP 1-CL-even   letter-ACC send-NEG

      Hiromi doesn't even send us a letter

      Hiromi hasn't even written once

In example (16), the fact that the letter is being written focuses attention on the writing, so both the classifiers for `correspondence` and for `written-thing` are appropriate. *-kai* can also be used in the same position, but in this case it is the act of writing that is being quantized, not *tegami* "letter". Because the interpretation is so different we mark it with an octothorpe (#).

In example (17), where the letters are being posted, *-tsū* is more appropriate, with *-hon$_{1or2}$* sounding strange. We argue that this is because there is nothing to pick out the letters (referred to anaphorically) as cases of either `written-thing` or `communication`; they are simply letters, so the use of *tsū* is strongly preferred. There is, however, some personal variation here. One informant said they could always replace *-tsū* with *-hon$_1$*, with no change in meaning. We assume that they have internalized a different type hierarchy. Again, *-kai* used here implies that the act of posting was done twice, which is an unlikely scenario.

Finally in example (18), where the emphasis is on lack of communication, *-tsū*, *-hon$_1$* or *-kai* can all be used. Using *-hon$_1$* which includes `communication`, or *-kai*, which counts events, implies quite strongly that there has been no communication by any means, whereas *-tsū* says only that there have been no letters.

The final classifier we consider is *-ken* "proposition", which can be used to count the contents of correspondence, being the classifier of choice for answering machine messages. It is particularly appropriate when the contents of the communication are explicitly referred as in (19), but anomalous if there is no reference to the contents, as in (20). Iida (1996) notes that *-ken* "proposition" is counting "matters, affairs, cases and subjects" in sentences like (19), but then goes on to mistakenly characterize these as achievements.

(19) kujō-no   tegami-ga   3-{tsū/ken} kita
      complaint-ADN letter-NOM 3-CL   came

      3 letters of complaint came

(20) utsukushī tegami-ga   3-{tsū/*ken} kita
      beautiful letter-NOM 3-CL   came

      3 beautiful letters came

When deciding which classifier to use, standard considerations of informativeness suggest that the most specific kind classifier possible will be used. Because each kind classifier selects for a single noun type, they are generally more informative than shape classifiers. Therefore, when there is a choice between a shape and a kind classifier, the kind classifier will be used, unless there is some salient reason for emphasizing the shape. Kind classifiers from higher members of the type hierarchy can be used if there is some motivation, for example, constraints on the controlling verb's argument structure, or adjectival modifiers. Shape classifiers can be used to pick out a certain facet of items. We predict that objects with no special kind classifiers will typically be counted with shape classifiers.

3.3 Multiple Classifiers

Different classifiers can select for different parts of the lexical structure of a noun simultaneously, as in the slightly forced (21).

(21)   tegami-o   5-tsū (awasete)   8-mai moratta
       letter-ACC 5-CL (altogether) 8-CL   received
       I received 5 letters, consisting of 8 sheets (altogether).

Note that each classifier is selecting a different part of *tegami* "letter"'s qualia structure: *-tsū* "correspondence" selects for the FORMAL type and *-mai* "2D" selects for the dimensionality of its constituent sheets of paper.

We make two predictions:

1. There can only ever be one classifier selecting for each variable (noting that you can of course simultaneously select for as many measurable attributes as exist: e.g., weight, height, density ...).

2. Nouns that are of a type associated with a kind classifier, will not be directly selected for by a shape classifier.

This ends our analysis of Japanese classifiers used to count correspondence. We will now turn to Korean.

4   Korean

In Korean, as in Japanese, classifiers are used to pick out salient features of the noun they classify. In Korean, *-kei* "3D", the general classifier for inanimate entities, is the same as the shape classifier used for objects with no salient extension. There are also shape classifiers for one and two salient dimensions: *-charu* "1D" and *-chang* "2D".

In general Korean tends to have fewer classifiers than Japanese, for example *-mari* "animal" vs Japanese *-tō* "large animal", *-hiki* "small animal", *-wa* "bird (& rabbit)", and so on. This tendency is reflected here, where there is no equivalent to *-hon₁* or *-hon₂*.

We present our hierarchy for Korean in Figure 2. It is almost the same as the Japanese hierarchy, but there are fewer kinds that are selected by classifiers.
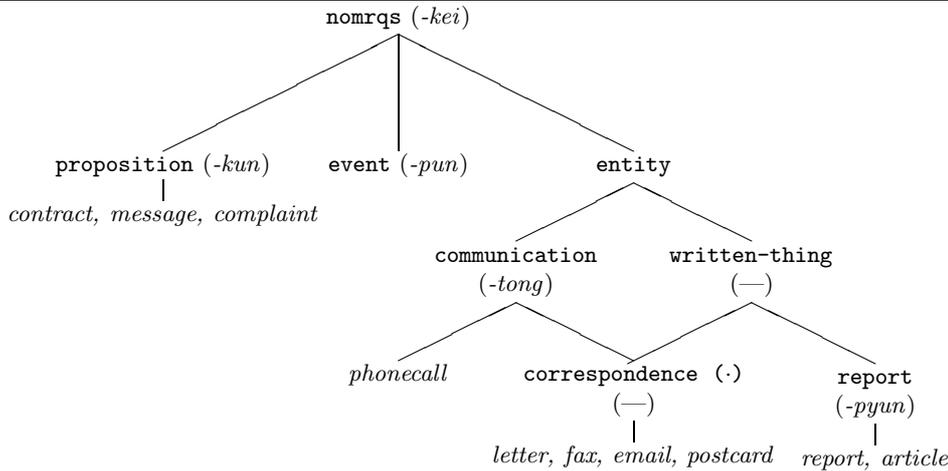

Figure 2: Proposed hierarchy for communication in Korean

Korean does not have a separate classifier for correspondence, thus phonecalls and other correspondence are all classified using *-tong* "communication", whether they are sent by post or made by phone. This classifier has the same origin as Japanese *-tsū* (the character for "pass through, transmit") but selects for a different level of the hierarchy. Because there is no classifier for correspondence itself, and no general classifier for written-things there is less variation than Japanese for counting correspondence. We give the lexical structure of *-tong* in (22).

(22)
-tong
$$\begin{bmatrix} \text{ARGSTR} & \begin{bmatrix} \text{ARG1} & \text{x:numeral+} \\ \text{D-ARG1} & \text{y:communication} \end{bmatrix} \end{bmatrix}$$

In contrast with Japanese, in Korean letter is a verbal-noun: one can say *pyunji-hata* "letter-do", as well as *pyunji-lul ponnaeta* "send a letter". Therefore, *pyunji* "letter", along with *chonhwa* "telephone" and *imeil* "email", must include event structure in its lexical entry, unlike other means of correspondence such as postcards and faxes. The difference can be seen in the following examples, where *pyunji* "letter" can be classified with the event classifier *-bun*, but *yupse* "postcard" cannot.

(23) ku-ekeyso pyunji-ka   1-{tong/bun} issessta
      he-from   letter-NOM 1-CL   had

      There was a letter from him

(24) ku-ekeyso yupse-ka   1-{tong/$^?$*bun} issessta
      he-from   postcard-NOM 1-CL   had

      There was a postcard from him

The shape classifier *-chang* "2D", which is characteristic of two dimensional objects, can also be used if the emphasis is on the sheets of paper in a letter. For example in (25) the emphasis is on not sending even a single page and so *-chang* is interchangeably used with *-tong*. Apart from such uses the kind classifier is generally preferred.

(25) ku-nun {pyunji/yupse}   1-chang-to ponaeji annunta
      he-TOP {letter/postcard} 1-CL-even send   NEG

      He doesn't even send a single {letter/postcard}

(26) ku-nun pyunji 3-{tong/*chang}-lul patassta
      he-TOP letter 1-CL-ACC   sent   NEG

      He sent three letters

5   Conclusion

In this paper we have presented a preliminary analysis of classifiers relating to communication in Japanese and Korean. By using the generative lexical framework of Pustejovsky (1995), we are able to formalize the well known difference between shape and kind classifiers, as well as give explicit motivation for variation in classifier use. The difference between kind and shape classifiers is as follows: kind classifiers select for a noun's FORMAL TYPE; shape classifiers select for a noun's FORMAL DIMENSIONALITY.

The advantage of formalizing the description of classifiers is that it enables us to formulate a number of empirical questions that may other wise be obscured, and thus to advance beyond the important first step of cataloguing phenomena.

This work suggests several further avenues of investigation:

1. Extending the analysis to other classifiers in Japanese and Korean, and classifiers in other languages.

2. In this small semantic domain Japanese makes more distinctions with classifiers than Korean does. Is this the case generally, i.e., does Korean use fewer kinds of classifiers than Japanese?

3. One of the advantages of using explicit lexical structures is that it is possible to implement the grammar and lexicon and test the predictions. We would like to do so.

4. How do the lexical structures interact with the different syntactic positions available: in particular pre-nominal modification and 'floating' quantifiers?

5. Do the shape classifiers access information from spatial memory, and is it separate from linguistic memory?

6. How should the differences in how different people use classifiers be explained? Have people internalised different hieararchies?

7. Does the lexical entry for a noun list its classifier? If so, can it list more than one classifier? We speculate that nouns only list their classifier if it is exceptional, i.e. not predicted by a node in the meaning hieararchy; and that will only ever be one such classifier listed.

Acknowledgments

References

Keith Allan. Classifiers. *Language*, 53:85–311, 1977.

A. E. Backhouse. *The Japanese Language: An Introduction*. Oxford University Press, 1993.

Ann Copestake and Ted Briscoe. Acquision of lexical translation relations from mrds. *Machine Translation*, 9 (3–4):183–219, 1995.

J. Peter Denny. Semantic analysis of selected Japanese numeral classifiers for units. *Linguistics*, pages 317–335, 1979.

Pamela Downing. Pragmatic and semantic constraints on numeral quantifier position in Japanese. *Journal of Linguistics*, 29:65–93, 1993.

Pamela Downing. The anaphoric use of classifiers in Japanese. In Downing and Noonan (1995), pages 345–375.

Pamela Downing and Michael Noonan, editors. *Word Order in Discourse*, volume 30 of *Typological Studies in Language*. John Benjamins, Amsterdam, 1995.

Asako Iida. Aspect and classifiers: A study of Japanese classifiers for counting correspondence. *Keio University Colloquia*, 17:125–137, 1996. Keio University.

Ray Jackendoff. *Languages of the Mind*. MIT Press, 1992.

Ray Jackendoff. The architecture of the linguistic-spatial interface. In Paul Bloom, Mary A. Peterson, Lynn Nadel, and Merril F. Garret, editors, *Language and Space*, pages 1–30. MIT Press, Cambridge, Mass, 1996.

Alan Hyun-Oak Kim. Word order at the noun phrase level in Japanese: quantifier constructions and discourse functions. In Downing and Noonan (1995), pages 199–246.

George Lakoff. Classifiers as a reflection of mind. In Colette Craig, editor, *Noun Classes and Categorization*, volume 7 of *Typological Studies in Language*, pages 13–51. John Benjamins, 1986.

Yō Matsumoto. The Japanese classifier *-hon*: A prototype-semantic analysis. *Sophia Linguistica*, 20/21:73–81, 1986.

Yō Matsumoto. The Japanese classifiers *-ken* and *-mune*: Prototype and background of existence. *Sophia Linguistica*, 22/23:19–29, 1988.

Yō Matsumoto. A semantic structure and system for Japanese classifiers — based on prototype semantics –. *Gengo Kenkyu*, 99:82–106, 1991. (in Japanese).

James Pustejovsky. *The Generative Lexicon*. MIT Press, 1995.

Barbara Unterbeck. Korean classifiers. In Young-Key Kim-Renaud, editor, *Theoretical Issues in Korean Linguistics*, pages 367–385. CSLI, 1994.