

Constructions for measurement and comparison in Japanese and English

Yoko HASEGAWA¹, Kyoko Hirose OHARA², Seiko FUJII³,
Russell LEE-GOLDMAN^{1,4}, Charles J. FILLMORE^{1,4}
¹University of California, Berkeley, ²Keio University,
³University of Tokyo, ⁴International Computer Science Institute

We have been investigating in recent years various constructions in Japanese and English in an *onomasiological*, or encoding, approach. That is, rather than questioning what an expression means (a decoding, or *semasiological*, approach), an onomasiological approach departs from a concept and asks how it can be expressed — radically distinct from organizing expressions by lexical items or phraseological types and then describing their meanings.

Adaptation of the principle of onomasiology in crosslinguistic investigation requires, as the subject matter, clear and universal concepts that can be discussed objectively. In other words, we need to set out in advance the concepts to be expressed and then relate these concepts to the manner of expression in each language. Measurement and comparison are ideal candidates for such a purpose.

With help of several corpora, we have found surprisingly many idiosyncratic rules both in English and Japanese, many of which are not readily available to non-native language users. Not surprisingly, therefore, when checked against popular translation programs, most measurement and comparison sentences we selected are beyond their ability, and they return miserable results. Many of the resultant machine-translations are not only blatantly ungrammatical, but also totally incomprehensible. A serious challenge for us now is how to assemble our accumulated observations about crosslinguistic comparison expressions in an efficient and principled way.

This paper presents our first step toward this goal. As an underpinning, we employ and elaborate the *constructicon* and *lexicon* developed by FrameNet for English comparison constructions. A construction is defined as a grammatical rule that pairs a particular syntactic pattern with the meaning to which it is dedicated. The constructicon is the inventory of constructions in a language.

In English, one of the ways of expressing measurement is by a measurement-value expression followed by some adjectives indicating the *parameter*, e.g. *2 inches **thick***, where the parameter is thickness. To illustrate the idiosyncratic nature of comparison expressions, let us begin with the fact that both the neutral and the marked polar adjectives can be used in a comparative construction, (1a), while the latter is not permitted in non-comparative constructions, (1b).

- (1) a. *2 inches taller, 2 inches shorter, 2 years older, 2 years younger, 2 pounds heavier, 2 degrees colder, 2 dollars more expensive, 20% more likely, 20 IQ points more intelligent*
- b. **2 inches tall, *2 inches short, *2 years old, *2 years young, *2 pounds heavy, *2 degrees cold, *2 dollars expensive, *20% likely, *20 IQ points intelligent*

Also, the measurement adjectivals are straightforward in the predicative use (e.g. *the fence is 4 feet tall*), but eccentric in the attributive use, where a singular-unit noun must be used (e.g. *a 4-foot-tall fence*, vis-à-vis **4-feet-tall fence*).

Furthermore, some expressions are “hidden” comparatives: *ten minutes late* means ten minutes later than some appointed time; *five miles ahead* means five miles ahead of some moving reference point. Such appointed time and reference point must be recoverable from discourse (i.e. definite null instantiation, in FrameNet terms). In fact, all evaluative adjectival expressions involve hidden comparison. For example, in *That building is tall*, the standard of comparison is implicit, e.g. tall for buildings with its function, tall for buildings in the neighborhood, etc. The standard in the latter case, however, is an instance of indefinite instantiation.

Let us now consider Japanese expressions to indicate the idea expressed by:

- (2) This book is 100 pages long.

Japanese uses a noun specifying the parameter, followed by the measurement value: e.g. *nagasa 100 peeji* ‘length 100 pages’, *atsusa 3cm* ‘thickness 3cm’, *ususa 1mm* ‘thinness 1mm’, *haba 2m* ‘width 2m’, *takasa 6 fiito* ‘height 6 feet’.

<i>nagasa</i>	NP	<i>100 peeji</i>	NP
	Parameter		Value

For describing the parameters of an entity and their values in Japanese, the topic-comment construction is used, in which the topic can be distinct from the subject, e.g. (3).

- (3) a. *zoo wa hana ga nagai.*
 elephant TOP nose NOM is-long
 ‘Speaking of elephants, their trunks are long.’
 Lit. ‘Elephant, the nose is long.’
- b. *An wa se ga takai.*
 Ann TOP height NOM is-high
 ‘Ann is tall.’
 Lit. ‘Ann, the height is high.’

The topic-comment construction is used in (4a). When the parameter is inferable from the measurement unit, it can be omitted, e.g. (4b). Alternatively, the existential construction can be used as well, e.g. (4c).

- (4) a. *kono hon wa nagasa hyaku peeji da.*
 this book TOP length 100 page is
 ‘This book is 100 pages long.’
 Lit. ‘This book is length 100 pages.’
- b. *kono hon wa hyaku peeji da.*
 this book TOP 100 page is
 ‘This book is 100 pages.’

- c. *kono hon wa hyaku peeji aru.*
 this book TOP 100 page there-is
 ‘This book has 100 pages.’
 Lit. ‘In this book, there are 100 pages.’

While scalar adjectives behave similarly in English and Japanese, e.g. (5a), significantly different interpretations result when such adjectives are juxtaposed with a measurement expression, e.g. (5b).

- (5) a. *kono hon wa nagai.*
 this book TOP is-long
 ‘This book is long.’
 b. *kono hon wa hyaku peeji nagai.*
 100 page
 ‘This book is 100 pages **longer**.’

While (5a) translates ‘this book is long’, (5b) does not mean ‘this book is 100 pages long’. Rather, it renders only ‘this book is 100 pages *longer*’. That is, Japanese scalar adjectives do not license measurement-value expressions. Additional examples:

- (6) a. *kono hako wa go kiro omoi.*
 this book TOP 5 kg is-heavy
 ‘This box is 5kg **heavier**.’
 b. *kono pen wa nisen-en takai.*
 this book TOP ¥2,000 is-expensive
 ‘This pen is ¥2,000 **more expensive**.’

This obligatory comparison interpretation is neither a property of the measurement expression itself, nor of the scalar adjective itself. Therefore, the comparison reading should be considered a property of the construction, not compositionally derived from the meaning of its component(s). This fact raises a serious question regarding a fundamental assumption of the FrameNet project. Before discussing this problem, we explain how Comparison construction is accounted for in the current development of FrameNet (to be released to the public shortly).

The Comparison construction is defined as:

A general construction that licenses the creation of a complex comparative adjectival predicator and the realization of the arguments of that predicator. The comparative expression indicates the equality or non-equality of two values on a scale.

The Comparison construction has two types of Construction Elements: inner and outer. The inner Construction Elements consist of Marker (e.g. *more*, *less*, *-er*) and Base_expression (e.g. *tall*, *fond of grapes*). The lexical identity of the Marker determines the nature of comparison (greater/less than, equal to).

- (7) *She is 6 inches*

<i>tall</i>	<i>-er</i>
Base_ expression	Marker

than her father

The outer Construction Elements are: Item, normally the external argument, (*that is more interesting than this*), Standard, normally a complement of the Comparison_phrase, (*taller than her*), Difference (*three inches taller than her*), Approximation (*almost taller than her*) and Multiplicative (*four times taller than that one*). Difference, which indicates the difference in values between the Item and Standard, may be a measurement phrase (*three inches*) or a more vague specification (*much*).

(8) [*She* Item] *is* [*almost* Approximation] [*6 inches* Difference] [*taller*] [*than her father* Standard]

The Standard covers all varieties of phrase (*than her, than she is, than expected*) except where it expresses a particular measurement-value (*taller than 2 meters*); in that case Standard_value is used.

This distinction between Standard and Standard_value is significant in English: while the Standard permits a clausal complement, the Standard_value does not.

- (9) a. *John is taller than Bill is.*
 b. **John is taller than 6 feet is.*

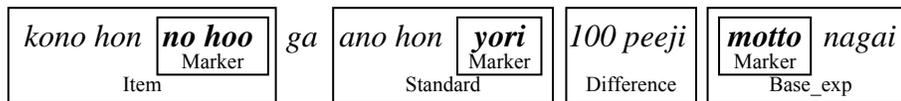
In Japanese, Standard, e.g. (10a), and Standard_value, e.g. (10c), require totally different constructions. With a Standard_value, the existential construction with an NP [measurement-value *ijoo*] must be used.

- (10)a. *taroo wa hanako yori se ga takai.*
 Taro TOP Hanako than height NOM is-high
 ‘Taro is taller than Hanako.’
 b. **taroo wa 6 fiito yori se ga takai.*
 Taro TOP feet than height NOM is-high
 ‘Taro is taller than 6 feet.’
 c. *taroo wa 6 fiit ijoo aru.*
 Taro TOP feet more-than there-is
 ‘Taro is taller than 6 feet.’
 Lit. ‘With Taro, there are more than 6 feet.’”

The two inner Construction Elements, or lexical comparison items, evoke the Comparison frame in FrameNet. The Comparison frame has as its Frame Elements Item, Standard, Scale, and Extent. The Item and Standard are compared with respect to their values on a Scale (e.g. height, intelligence). A particular Construction Element is simply equivalent to the corresponding Frame Element.

Let us now discuss the problem that this analysis raises when it is applied to Japanese Comparison construction. In Japanese, three comparison markers are identified: they can attach to Item, Standard, or Base_expression.

- (11) *kono hon no hoo ga ano hon yori hyaku peeji motto nagai.*
 this book GEN side NOM that book than 100 page more is-long
 ‘This book is 100 pages longer than that book.’



Any one of the markers is sufficient to evoke the Comparison frame in Japanese:

- (12) a. *kono hon **no hoo** ga nagai.*
 ‘This book is longer.’
 b. *ano hon **yorī** nagai.*
 ‘It’s longer than that book.’
 c. ***motto** nagai.*
 ‘It’s longer.’

Furthermore, as discussed earlier as well as shown in (13), the mere juxtaposition of a measurement and a scalar adjective retaining only Difference and Base_exp expression without any comparison marker nonetheless forces a comparison interpretation.

- (13) *100peeji nagai.*
 ‘It’s 100 pages **longer**.’

Non numerical expressions can play a role as Difference and, with a scalar adjective, permit only comparison interpretations:

- (14) a. *kono hon wa **harukani** nagai.*
 ‘This book is longer **by far**.’
 b. *kono hon wa **wazukani** nagai.*
 ‘This book is **a little** longer.’

The question here is how to invoke the Comparison frame. The fact that Japanese comparatives can be recognized even when there is no overt comparative marker makes it clear that we have to be able to describe comparatives without linking them to markers like *more/less/-er*.

Originally, with Fillmore’s *Case Grammar*, FrameNet assumed that the semantic roles assigned to constituents of a clause define **relations to a governing word in the clause**. This would be easiest to express in a dependency notation, where all of the dependents would hang down from the same node — including the subject. The semantic roles, or Frame Elements, could be represented as labels on the branches.

Case Grammar posited a universal set of semantic role types (“cases”) and sorted predicates into semantic types according to the cases that they would select. They were not limited to VPs and did not express linear order. {Exp Cont} would describe psych verbs with experiencer subject, e.g *I fear the apocalypse*; {Agt Pat} fit verbs that could express events in which somebody causes something to undergo change, as in *I broke the bubble*; {Stim Exp} stood for verbs that showed a stimulus of some kind and an experience of a lexically defined emotion, as in *The lightning scared me*; and so on. Verbs classified in this way might have some predictability in their syntactic behavior, but in many cases they can be quite different from verb to verb.

FrameNet has abandoned the idea that there is a fixed set of semantic relations to choose from. It is recognized that a lexical item evokes a frame. The FrameNet annotator finds instances where different lexical items evoke the same frame, describes such a shared frame, and gives frame-specific names to its components and aspects.

A verbal frame-evoker can incorporate information relevant to its frame, so in a sense it is its own Frame Element (e.g. *Let's shelve these books*, where *shelve* indicates Place); or a noun like *remainder* can evoke the frame of some supply of things getting reduced, and this noun, in addition to evoking the frame, also designates what is left (e.g. *There's a remainder of three*). The NP that a noun heads is a Frame Element of the frame it evokes. Studying the combinatorial properties of lexical units has been more or less what the lexicon part of FrameNet is about.

Until now, we could not, strictly speaking, talk about Frame Elements in a construction, because there is no lexical predicator for anything to be Frame Elements of. However, the Japanese Comparative construction led us to consider that a construction itself can evoke a frame. That is, the construction that juxtaposes a measurement expression and a scalar adjective evokes the Comparison frame. We then can say that the components of a construction satisfy the construction's Frame Elements, e.g. the measurement expression satisfies the Frame Element of Difference, and the scalar adjective the Frame Element of Scale.

In this view, instead of thinking of semantic roles as the names of relations to a lexical unit, it is more desirable to think of a frame as something holistic, and the Frame Elements as the components or aspects of situations described independently of mentioning particular lexical units, as can be seen in Fillmore & Atkins' (1992) analysis of *risk*.

Consider the example of Speed. We can say that the concept — the “frame” — that expresses measured speed consists of (i) something that identifies a distance and (ii) something that expresses a time. These are taken as defining a ratio, as in *two miles per hour*. (The “ratio” provision is important: a sentence like *I walked two miles in an hour* is not a speed construction, although it is possible, of course, to calculate the speed from the information it contains.) Each of the phrases that represent these Frame Elements can themselves be constructs licensed by their own constructions.

It would be nice to be able to say for constructions as well as lexical units that certain Frame Elements can be realized with many different syntactic structures, as when we recognize the content of a decision in all the phrases found in *He decided **on applying/to apply/that he should apply***. Similarly the denominator of ratio expressions can be formulated in many different ways: *Six times **a second/per second/every second***.

With this understanding, let us now examine what frames are necessary for interpreting *I am taller than you*. In this case, the Height and Comparison frames are evoked. The Height frame is a subtype of the Dimension frame, which, in turn, is a subtype of the Scalar_property frame. The Height frame has elements Item and Height_value.

Height frame: FE (Item, Height_value)

Comparison frame: FE (Item, Standard, Scale, Extent)

The word *taller* evokes both the Height frame (via *tall*) and the Comparison frame (via the Comparison construction). The question is how to integrate these two frames, or concepts. Is the Height frame subordinated by the Comparison frame? Is the Height_value Frame Element spelled out via some element in the Comparison frame? etc.

Regarding the first question, one could say that Height is made subordinate to Comparison:

GREATER-THAN (MY-HEIGHT, YOUR-HEIGHT)

However, as mentioned earlier, one could also say that all scalar notions are subordinate to comparison. *She is tall*, for example, can be written as:

GREATER-THAN (HER-HEIGHT, AVERAGE-HEIGHT)

Therefore, it might not be all that useful to consider the Height frame to be subordinate to the Comparison frame.

Regarding the second question, we consider the Height_value Frame Element to be specified by the Comparison frame. Consider the sentence *I am tall*. There is a construction, Predicative Scalar Adjective, which renders the interpretation of *am tall* as inherently comparative, again as discussed in the beginning of the paper. Here, the Height_value Frame Element is specified just as “taller than some contextual standard/average”. Now let us bring in the Comparison frame. Instead of some vague value of the height, *taller* says that the Height_value ***exceeds some particular, contextually- or linguistically-retrievable value***. In FrameNet, that implicit value is called Standard_value; the entity that has that value is called Standard. Thus, in *I am taller than you*, the *than you* is the Standard, and from that we figure out the value (it is linguistically-retrievable, because it is expressed by an overt Frame Element). In *I am taller*, the Standard is missing, but ***must*** be retrievable from the discourse or real-world context.

The construction that builds *20 feet tall* also involves the Height frame, but not the Comparison frame; rather the construction supplies the value of the Frame Element of Height_value, which is the value expressed by *20 feet*.

Examining Japanese, we see that the construction that builds (15a) evokes both the Height and Comparison frames.

- (15) a. *50 meetoru takai.*
 meter is-high
 ‘It’s 50 meters higher.’
 b. *ano biru yori takai.*
 that building than is-high
 ‘It’s higher than that building.’

Similarly, (15b) also evokes Height and Comparison. The Comparison construction requires only the presence of an adjective, and it supplies (i) a Comparison frame within which to understand the sentence, and (ii) the syntactic means to express Comparison Frame Elements, e.g. a *yori*-phrase, the adverb *motto*, etc.

Finally, we mention mental-space comparison. The Standard in English comparatives covers varieties of phrase (*than her*, *than she is*, *than expected*). Some Standard expressions evoke comparison between two mental spaces.

- (16) a. *Tony is smarter than anybody thought.*
 b. *Tony thinks he's smarter than he actually is.*

The Standard in (16a) implies a counterpart of the $Item_{\alpha}$ in a subordinate mental space, $Item_{\beta}$, the person that anybody expected him to be. By contrast, the main clause in (16b) creates a mental space, in which the $Item_{\beta}$ is smarter than its counterpart in the reality.

While there is little structural similarity between the two mental space comparisons in English, Japanese employs similar constructions for such a comparison, for example:

- (17) a. *tonii wa omotta yori kashikoi.*
 TOP thought than is-smart
 'Tony is smarter than anybody thought.'
 b. *tonii wa jissai yori kashikoi to omotteiru.*
 TOP reality than is-smart QUOT is-thinking
 'Tony thinks he's smarter than he actually is.'
 Lit. 'Tony thinks he's smarter than the reality.'

Omotta yori 'than thinking' (and its equivalents) indicates that the Standard is in a subordinate mental space, whereas *jissai yori* 'than reality' (and its equivalents) indicates that the Standard is in the reality.

In conclusion, this paper reported several findings of our ongoing investigation of Japanese-English measurement and comparison study as part of the FrameNet project. We have learned that the framework developed by FrameNet for English comparison construction needs to be amended in order to make possible a uniform analysis of Japanese comparison construction. In our modification, not only lexical items but also grammatical constructions can evoke frames. Furthermore, such frames must become more conceptual and holistic than the current notion of frame in the FrameNet project.

Fillmore, Charles J. and B. T. S. Atkins (1992): Towards a frame-based organization of the lexicon: The semantics of RISK and its neighbors. In Lehrer, A and E. Kittay (Eds.) *Frames, Fields, and Contrast: New Essays in Semantics and Lexical Organization*. Hillsdale: Lawrence Erlbaum Associates, 75-102.