

## Japanese and the Other Languages of the World

journal or publication title	NINJAL Project Review
number	1
page range	29-45
year	2010-05
URL	<a href="http://doi.org/10.15084/00000554">http://doi.org/10.15084/00000554</a>

# JAPANESE AND THE OTHER LANGUAGES OF THE WORLD

## 日本語と世界の諸言語

**Bernard Comrie** (バーナード・コムリー)

Max Planck Institute for Evolutionary Anthropology, Leipzig and  
University of California, Santa Barbara  
(マックスプランク進化人類学研究所／カリフォルニア大学サンタバーバラ校)  
comrie@eva.mpg.de

**Abstract:** Linguistic typology provides a means of locating individual languages, such as Japanese, against the background of cross-linguistic variation. The body of the article addresses this issue using the World Atlas of Language Structures (WALS) as a research tool, for testing both the frequency of individual features and the strength of correlations between features. It is shown that while Japanese and English are typologically very different from one another, they are about equally typical in terms of overall cross-linguistic variation. The consistent head-final nature of Japanese is argued to be a contingent feature of Japanese rather than a reflection of strong universal correlations across different constituent order features. Moving beyond the material presented in WALS, the article concludes by situating both the Japanese unified noun-modifying construction and the language's unusually rich system of verbs of giving with respect to other languages of the world.

《要旨》言語類型論は日本語等の個別言語を通言語的変異に照らして位置づけるための1つの方法を提供してくれる。本論では個々の特徴の生起頻度と複数の特徴の相関関係の強さの両方を検証するために、WALS (『言語構造の世界地図』) を研究手段に用いて言語間変動の問題を考察する。日本語と英語は言語類型論的に非常に異なるものの、通言語的変異を総合的に見ると、どちらの言語も同じ程度に典型的であることが明らかになる。また、日本語が一貫して主要部後続型の語順を取ることは、異なる構成素の語順に見られる強い普遍的相関性の反映であるというよりむしろ、日本語の偶発的な性質であると主張できる。最後に、WALSの守備範囲を超えた現象として、多様な意味関係を一様に表す日本語の名詞修飾構造、および類例がないほど豊かな日本語授与動詞の体系に触れ、それらを世界の他の言語との関係で位置づけることで本稿を締めくくる。

## 1. Introduction

The basic questions that this paper addresses are the following: To what extent is Japanese a “typical” language in terms of worldwide cross-linguistic variation? To what extent and in which cases does Japanese go together with the majority of the world's languages? In which cases does it behave like a minority? Are there cases in which Japanese is unique or at least follows a very rare type? In other words, we are asking about the typological – not the genetic (genealogical) – position of Japanese among the languages of the world.

There are various ways in which one might set about answering questions of this kind, such as the typological profile of Japanese in comparison with other languages of the world

presented in Tsunoda (2009). For the bulk of the present article, I will adopt a different and very specific methodology in an attempt to give an admittedly partial but hopefully methodologically consistent answer to these questions. This methodology consists in examining databases of cross-linguistic diversity across the world's languages in an attempt to assess the position of Japanese within that diversity. More specifically, the methodology involves examining the materials presented in Haspelmath et al. (2005; 2008), hereafter referred to as *WALS*; see further section 2. In the final sections 5 and 6, a very preliminary look is taken at material not covered by *WALS* and where considerable further research is required in order to answer the question of the typological position of Japanese among the world's languages. Throughout, emphasis will be placed on examining particular areas in which claims have been made with regard to Japanese, testing them against what we know about cross-linguistic structural diversity.

## 2. The *World Atlas of Language Structures* (*WALS*)

*WALS* was published in 2005 in book form with an accompanying CD-ROM (Haspelmath et al. 2005), and is now available on-line (Haspelmath et al. 2008). The on-line version is updated periodically, and all references in the present article are to the data as presented in the 2005 version. The URL links given for individual maps will, of course, take the reader to the latest version of that map.

### 2.1. What is *WALS*?

*WALS* is in the first instance an atlas with 142 maps, each map showing the geographical distribution of the feature values for a particular structural feature of language; see section 2.2 for examples. On the maps, each language is represented as a dot in a particular geographical location. In addition to the book version, which provides a visualization of what we know about the structural diversity of the world's languages, comprehensible to anyone who can read maps, there is also an accompanying CD-ROM, which includes the database underlying the maps plus an interactive reference tool (developed by Hans-Jörg Bibiko) which can be used to generate new maps, to test hypotheses about correlations across features, etc., as discussed briefly in section 2.3 and in more detail in section 4. Since April of 2008 all *WALS* material has also been available on-line, thus making the material accessible in interactive form to all researchers.

*WALS* covers 142 typological features, broken down into the following rubrics:

Phonology	19
Morphology	10
Nominal Categories	28
Nominal Syntax	7
Verbal Categories	16
Word Order	17
Simple Clauses	24
Complex Sentences	7
Lexicon	10
Others	4

In total, 2,560 languages, or slightly more than one third of the total number of languages in the world, are mentioned on at least one WALS map, with the average number of languages per map being 398. The number of languages per map ranges from a high of 1,370 to a low of 35, but this minimum is misleading, since it relates to one of two maps devoted to deaf sign languages, for which available cross-linguistic information is much more restricted than in the case of spoken languages, even if recent years have seen a huge expansion in work on deaf sign languages from different parts of the world. The total number of data points in the WALS database is about 58,000. As a spin-off the five-year project that led to WALS also gave rise to a list of 6,700 bibliographical references, in effect the catalog for a library of someone wanting to undertake extensive cross-linguistic research.

Although the set of languages covered does vary from map to map, the editors nonetheless tried to attain some measure of comparability by establishing two samples of languages. The basic sample of 100 languages, which is by and large genealogically balanced but with some concessions to geographical balance and occasional other factors, comprises languages for which contributors were urged to include relevant data wherever possible. The extended sample of 200 languages (including the 100 from the basic sample) comprises a further 100 languages for which inclusion of data was recommended.

## 2.2. Individual features

As an illustration of an individual feature that is mapped in WALS, we may consider the relative order of object and verb in the clause (Dryer 2005d). For this map, Dryer identifies three types: languages with dominant OV order, such as Japanese; languages with dominant VO order, such as English; and languages in which there is no dominant order, such as German, which has VO order in main clauses but OV order in subordinate clauses. These possibilities are illustrated in (1).

### (1) Order of Object and Verb

Object–Verb (OV)	Japanese	<i>(gakusei ga) hon o katta</i>
Verb–Object (VO)	English	<i>(the student) bought the book</i>
No dominant order	German	<i>(der Student) kaufte das Buch</i> (main clause) <i>(... dass der Student) das Buch kaufte</i> (subordinate clause)

The numbers of languages of each type in Dryer’s sample are set out in Table 1.

Table 1: Order of Object and Verb (feature 83)

Type	Languages	Genera	Families
OV	640	237	120
VO	639	162	63
No dominant order	91	60	26

The figures given here actually go somewhat beyond what is shown directly on the map and make use of the further possibilities afforded by the interactive reference tool. In particular, table 1 shows not only the number of languages belonging to each type, but also the number of genera including languages of each type, and the number of families including languages belonging to each type. A “family” is to be understood in the traditional sense of the largest

genealogical unit generally accepted by specialists and exemplified, for instance, by Indo-European; the classification used by WALS is rather conservative, so that Japanese (-Ryukyuan) is considered a separate family. A “genus” (plural: genera) is a group of languages whose genealogical relatedness is rather obvious even without detailed comparative research, usually corresponding to a time depth not greater than 3,500 to 4,000 years; illustrations would be the major branches of Indo-European such as Germanic.

It is important to note that the figures for languages, genera, and families can suggest rather different pictures. In table 1, for instance, the language figures suggest that OV and VO are about equally frequent across the world. The figures for genera suggest rather a ratio of about 3:2 in favor of OV, while those for families raise the ratio almost to 2:1. Especially in maps with a large number of languages, going well beyond even the extended 200-language sample, it is inevitable that some languages families will be over-represented, and in the case of this particular map it happens that the world’s two largest language families, Niger-Congo and Austronesian, consist overwhelmingly of VO languages. As a heuristic, it seems that the figures for genera probably give the best single guide to the statistical distribution of feature values, since they abstract away from discrepancies by having a large number of closely related languages sharing the same feature value, while also avoiding controversies between “clumpers” and “splitters” that inevitably arise in establishing family-level units. In other words, despite the near-identical numbers of OV and VO languages in the sample of languages for this map, an argument can nonetheless be made for a skewing towards OV on the basis of the figures for genera.

### 2.3. Correlating features

At least since the pioneering research of Greenberg (1966), one of the main interests of typologists has been testing whether there exist correlations between logically independent feature values, and indeed this search for correlations has extended well beyond typologically oriented linguistics into, for instance, mainstream generative grammar. One of the main contributions of WALS is that it permits testing claims about correlations against a large database. For instance, a slight reformulation of the universals 3 and 4 proposed by Greenberg (1966: 110) suggests that, while it is logically possible for either value OV or VO to correlate with either postpositions or prepositions, there is nonetheless a correlation of OV order with postpositions and VO order with prepositions. WALS enables us to test this claim.

First, we may look at the stand-alone map showing the distribution of postpositions versus prepositions (Dryer 2005b). For present purposes, we may simplify the map slightly by combining under the general rubric “Other” the three relatively small classes of languages identified by Dryer in addition to postpositional and prepositional, namely “inpositions”, “no dominant order”, and “no adpositions”. Prepositions and postpositions are illustrated in (2).

(2) Order of Adposition and Noun phrase

Postposition	Japanese	<i>Tookyoo kara</i>
Preposition	English	<i>from Tokyo</i>

Table 2 shows the number of languages, genera, and families representing each of these three types in Dryer’s sample.

Table 2: Order of Adposition and Noun phrase (feature 85)

<i>Type</i>	<i>Languages</i>	<i>Genera</i>	<i>Families</i>
Postposition	520	212	115
Preposition	467	122	45
Other	87	53	26

It will be noted that there is a slight skewing towards postpositions even at the level of languages, and that this skewing increases as one progresses to genera and families.

But what is of immediate interest to us is whether there is any correlation between the logically independent features represented in tables 1 and 2. By using the interactive reference tool, it is possible to create a combined map, giving the numbers of languages, genera, and families with each of the four values obtained by combining OV/VO with postposition/preposition (Dryer 2005g).

Table 3: Combining features 83 and 85

<i>Type</i>	<i>Languages</i>	<i>Genera</i>	<i>Families</i>
OV & Postposition	427	180	98
VO & Preposition	417	115	44
VO & Postposition	38	22	16
OV & Preposition	10	7	5

(This particular combination of maps is one of three included in WALS to illustrate the possibility of this feature of the overall project, which means that it is possible to access the map directly on-line rather than downloading the interactive reference tool and creating the combined map anew. The map includes a fifth category, 141 “languages not falling into one of the preceding four types”, which have been excluded from table 3 as not directly relevant.)

In this case, the hypothesis of a positive correlation between OV and postposition, between VO and preposition is borne out: 844 languages are consistent with the hypothesis (the first two lines), only 48 go against it (the bottom two lines); at the level of genera, the figures are 295 to 29; and at the level of families, 142 to 21. Of OV languages, postpositional languages outnumber prepositional by 427 to 10; for VO languages, prepositional languages outnumber postpositional by 417 to 38. Postpositional languages prefer OV order by 427 to 38; prepositional languages prefer VO by 417 to 10. Similar skewing applies at the genus and family levels. The correlation thus represents a bidirectional implication. In section 4, further use is made of the possibility of correlating logically independent feature values.

### 3. English versus Japanese

The material provided by WALS can also be used to answer the following question: Which of the two languages, English and Japanese, is more “typical” typologically, i.e. shares the greatest typological similarity with the world average? This question is interesting at least from the perspective of the history of linguistics, given that much of twentieth-century linguistics was based on English, and one might therefore wonder whether choosing a different starting point, say Japanese, might have given not only a different picture but also one that would have been more representative of the world’s languages overall.

The basic methodology used in order to answer this question on the basis of WALS ma-

materials is the following. First, for each feature, identify the percentage of the world's languages that have the same feature value as Japanese, and the percentage that have the same feature value as English. Then, for each of the two languages, calculate the average of the percentages that were arrived at in this way. This gives an overall index of typicality for each language, with a higher index representing greater typicality.

Two examples will illustrate the general methodology. The first relates to the size of inventory of phonemic vowel qualities in a language (Maddieson 2005), for which relevant figures are given in table 4. Japanese falls, in Maddieson's classification, into the set of languages with 5–6 phonemically distinct vowel qualities, and this covers 288 or 51.15% of the 563 languages in Maddieson's sample, so Japanese received the value 51.15. English falls into the set of languages with 7–14 vowels, to which belong 32.50%, so English receives the value 32.50. For this feature, Japanese is more typical than is English.

Table 4: Vowel quality inventories (feature 2)

	<i>Japanese</i>	<i>English</i>
Feature value	5–6 vowels	7–14 vowels
# of languages	288	183
Total # of languages	563	563
Percentage	51.15	32.50

Consider now the order of relative clause and head noun within the noun phrase (Dryer 2005e). As shown in (3), Japanese illustrates the type where the relative clause precedes the noun, while English represents the opposite type where the relative clause follows.

(3) Order of Relative Clause and Noun

Relative clause–Noun	Japanese	<i>[gakusei ga katta] hon</i>
Noun–Relative clause	English	<i>the book [that the student bought]</i>

Table 5 shows the number of languages, genera, and families belonging to each of these two types of languages, plus a group of "Others", again combining a number of different types identified by Dryer and containing smaller numbers of languages.

Table 5: Order of Relative Clause and Noun (feature 90)

<i>Type</i>	<i>Languages</i>	<i>Genera</i>	<i>Families</i>
Noun–Relative clause	507	165	63
Relative clause–Noun	117	52	29
Other	81	51	37

Table 6 takes the relevant data from table 5 and presents it in a way comparable to table 4, showing that in this case Japanese represents a minority type, 16.60% of the world's languages, while English belongs to the majority type, 71.91%.

Table 6: Order of Relative Clause and Noun

	<i>Japanese</i>	<i>English</i>
Feature value	Rel–N	N–Rel
# of languages	117	507
Total # of languages	705	705
Percentage	16.60	71.91

By averaging through all the percentages like 51.15 and 16.60 for Japanese, like 32.50 and 71.91 for English, one arrives at an index of typicality for each of the two languages. For the purposes of the present paper, and in contrast to the methodology used in Comrie (forthcoming), I restricted the calculation to all and only those features for which WALS gives data for both Japanese and English; this turns out to be 127 features. The average degree of typicality of each language is then as shown in table 7.

Table 7: Average percentage (degree of typicality) of Japanese and English

<i>Japanese</i>	<i>English</i>
40.74	40.84

Needless to say, this small difference is not statistically significant. Indeed, in experimenting with different precise measurements following the general methodology outlined above, it turns out that Japanese and English always end up with around the same measure, with Japanese sometimes slightly more typical than English, English sometimes slightly more typical than Japanese. Typologically, Japanese and English are clearly very different from one another, but on a global scale neither can be considered significantly more typical than the other.

There are, incidentally, various ways in which the methodology outlined above could be refined. For instance, instead of taking the percentage of languages that behave like Japanese or English, one could take the number of genera, as a more reliable overall measure of the world's genealogical linguistic diversity – this should be easy to implement, although it remains a task for future work. Much less easy to implement, because of controversies concerning the details, would be attempts to weight some features more than others, as being more significant. At an intermediate level, one might make allowances for feature values that are not logically independent of one another. For instance, since Japanese lacks verbal person marking (Siewierska 2005a; feature 102), it necessarily lacks verbal marking for third person (Siewierska 2005b; feature 103); the fact that English differs from Japanese on feature 103 is logically dependent on the fact that it differs in regard to feature 102. Less obvious is how one would make modifications because of correlations between logically independent feature values, for instance the fact that the OV/VO opposition is not empirically independent of the postposition/preposition opposition. All of this must remain for future research.

#### 4. Japanese as a consistent head-final language

One of the most frequently iterated typological claims about Japanese is that it is a head-final language, and the data in WALS not surprisingly bear this claim out. In the present section, however, I want to address not so much the correctness of the claim, but rather its implications for general linguistic theory, in particular against the background of the discussion in

section 3 about the different degree of typicality of different languages.

For the purposes of this section five canonical headedness oppositions covered by WALS are taken. Table 7 indicates the constituent orders found in Japanese, a consistently head-final language; Italian, a consistently head-initial language (at least on these features); and English, a language that is notoriously mixed with regard to its headedness. The five features examined are the order of verb and object (where the verb is the head), the order of adposition within the adpositional phrase (where the adposition is head), and the order of possessor (“genitive”), adjective, and relative clause within the noun phrase (where the noun is head). In table 8, and likewise in the examples (4)–(8), the head is underlined.

Table 8: Head position in Japanese, English, and Italian

	<i>Japanese</i>	<i>English</i>	<i>Italian</i>
OV/VO	<u>OV</u>	<u>VO</u>	<u>VO</u>
NP–Po/Pr–NP	NP– <u>Po</u>	<u>Pr</u> –NP	<u>Pr</u> –NP
Gen–N/N–Gen	Gen– <u>N</u>	Both	<u>N</u> –Gen
Adj–N/N–Adj	Adj– <u>N</u>	Adj– <u>N</u>	<u>N</u> –Adj
Rel–N/N–Rel	Rel– <u>N</u>	<u>N</u> –Rel	<u>N</u> –Rel

Illustrative examples are provided in (4)–(8); in each case, the examples within a given number are translation equivalents.

(4) OV/VO

Japanese           (*gakusei ga*) *hon o katta*  
 English            (*the student*) *bought the book*  
 Italian             (*lo studente*) *comprò il libro*

(5) NP–Po/Pr–NP

Japanese           *Tookyoo kara*  
 English            *from Tokyo*  
 Italian             *da Tokyo*

(6) Gen–N/N–Gen

Japanese           *gakusei no hon*  
 English            *the student’s book/the book of the student*  
 Italian             *il libro dello studente*

(7) Adj–N/N–Adj

Japanese           *siroi hon*  
 English            *the white book*  
 Italian             *il libro bianco*

(8) Rel–N/N–Rel

Japanese           *[gakusei ga katta] hon*  
 English            *the book [that the student bought]*  
 Italian             *il libro [che lo studente comprò]*

There are thus five logically independent features, and the investigation will proceed by examining whether each of the other four features shows a correlation of feature values with the feature values for OV/VO (for which see (1) and table 1). In other words, is there cross-linguistically a predominance of consistently head-final (OV, postposition, Gen–N, Adj–N, Rel–N) and consistently head-initial (VO, preposition, N–Gen, N–Adj, N–Rel) languages

relative to languages that are inconsistent with regard to head position? For OV/VO and post-position/preposition (for the latter, see (2) and table 2), table 3 in section 2 has already shown that there is indeed a good correlation. The other three features are now examined in turn.

For the order of genitive and noun within the noun phrase, the number of languages, genera, and families in each of the three types identified in Dryer (2005c) is shown in table 9.

Table 9: Order of Genitive and Noun (feature 86)

<i>Type</i>	<i>Languages</i>	<i>Genera</i>	<i>Families</i>
Gen–N	608	233	120
N–Gen	415	107	36
No dominant order	82	55	27

If we now combine features 83 (OV/VO) with feature 86 (Gen–N/N–Gen) – this has to be done using the interactive reference tool, as this is not one of the combinations illustrated by a published WALS map – then the figures generated are as in table 10, where the first two lines are consistently head-final and consistently head-initial respectively, the last two lines inconsistent.

Table 10: Combining features 83 and 86

<i>Type</i>	<i>Languages</i>	<i>Genera</i>	<i>Families</i>
OV & Gen–N	434	178	98
VO & N–Gen	352	92	32
VO & Gen–N	113	45	30
OV & N–Gen	30	18	7

Indeed, there are substantially more languages of the first two lines combined than of the second two lines combined, whether at the level of languages, genera, or families. Each of the first two lines taken individually is also noticeably more frequent than each of the last two lines at the level of languages and genera, although at the level of families VO languages with preposed and postposed genitives almost meet. But overall, the correlation is borne out.

The order of adjective and noun within the noun phrase is treated by Dryer (2005a), with the figures given in table 11 (where “Other” includes not only languages with no dominant order but also a handful of languages where attributive adjectives are expressed as internally headed relative clauses). It will be noted that there are substantially more N–Adj than Adj–N languages.

Table 11: Order of Adjective and Noun (feature 87)

<i>Type</i>	<i>Languages</i>	<i>Genera</i>	<i>Families</i>
Adj–N	340	124	63
N–Adj	768	236	89
Other	105	67	39

If we combine the relevant feature values for OV/VO and Adj–N/N–Adj, we obtain the distribution shown in table 12 (Dryer 2005f).

Table 12: Combining features 83 and 87

<i>Type</i>	<i>Languages</i>	<i>Genera</i>	<i>Families</i>
OV & Adj–N	201	70	43
VO & N–Adj	404	95	33
VO & Adj–N	100	51	28
OV & N–Adj	287	136	61

While there are more languages belonging to the two consistent types (the top two lines) than to the inconsistent types (the bottom two types), it is clear that OV languages in general do not conform to the prediction of consistent head-finality. In particular, at each of the levels of language, genus, and family, more OV languages have the inconsistent order N–Adj than the consistent order. Indeed, one might argue that table 11 if anything tends to reinforce the skewing already noted in table 10: languages are much more likely to be N–Adj than Adj–N. In any event, there is emphatically not a good correlation between OV/VO and Adj–N/N–Adj, in particular between OV and Adj–N. (See further Dryer (1988) and, more generally, Dryer (1992).)

The cross-linguistic distribution of Rel–N and N–Rel in isolation has already been presented, in table 5, and as that table shows, there is a skewing in favor of N–Rel even greater than the skewing shown in table 11 in favor of N–Adj. It will be recalled from table 6 that the Japanese order Rel–N is shared by only 16.6% of the languages in Dryer’s sample. The correlation of OV/VO with Rel–N/N–Rel is shown in table 13 (Dryer 2005h). Although again the top two lines combined are more frequent than the bottom two lines combined, this is largely because of the dominance of line 2, consistently head-final by having both VO and N–Rel. Indeed, in Dryer’s sample of 756 languages only 5 combine VO and Rel–N order: 3 are Sinitic languages (Mandarin, Hakka, Cantonese), 1 of the others, Bai, is either a Sinitic language that has undergone intense contact with Tibeto-Burman or vice versa, i.e. there are either close genealogical or areal links to Sinitic, while the fifth, Amis, is an Austronesian language of Taiwan. Amis seems, incidentally, to be represent an independent instance of this very rare combination; see further Comrie (2008). For OV languages, at the level of languages the consistent combination with Rel–N is only marginally more frequent than the inconsistent one with N–Rel, at the level of genera the relation flips, while at the level of families they almost meet in the middle. In any event, there is no evidence for a bi-directional correlation of the feature values; the overall higher frequency of N–Rel shows itself also when the features are combined.

Table 13: Combining features 83 and 90

<i>Type</i>	<i>Languages</i>	<i>Genera</i>	<i>Families</i>
OV & Rel–N	109	48	28
VO & N–Rel	370	107	38
VO & Rel–N	5	3	2
OV & N–Rel	96	63	31

One might wonder whether there is a correlation between the position of the adjective in the noun phrase and that of the relative clause within the noun phrase, especially given that both are attributes. Table 14 suggests that even this is not the case. Consistent N–Adj with N–Rel is by far the most frequent type, but this is not surprising given that each of N–Adj and N–Rel is more frequent than its opposite. At the language and family levels Adj–N languages are roughly evenly split between consistent Rel–N and inconsistent N–Rel languages, while

at the level of genus the inconsistent type predominates. In the opposite direction, Rel–N does go more frequently with consistent Adj–N, N–Rel with consistent N–Adj. So the results are at best mixed.

Table 14: Combining features 87 and 90

<i>Type</i>	<i>Languages</i>	<i>Genera</i>	<i>Families</i>
Adj–N & Rel–N	81	35	23
N–Adj & N–Rel	365	117	46
N–Adj & Rel–N	24	16	7
Adj–N & N–Rel	81	47	22

Summarizing the data presented in this section, we can say that:

- a) OV/VO, NP–Po/Pr–N, and Gen–N/N–Gen correlate well with one another.
- b) OV/VO does not correlate well with Adj–N/N–Adj or Rel–N/N–Rel; at best there is a one-way implicational tendency: if VO, then N–Adj; if VO, then N–Rel (in this case, almost exceptionless).
- c) Even Adj–N/N–Adj and Rel–N/N–Rel do not correlate well with one another; at best there is a tendency for the combination N–Adj and N–Rel to predominate, but each of these is anyway more frequent than its head-final counterpart.

Returning to Japanese: Japanese is indeed a consistently head-final language, but given the exceptions to consistent headedness noted above, it remains unclear how theoretically significant this fact is; in other words, it might just be a contingent fact about Japanese, in this respect representing a minority of the world’s languages. In particular given the cross-linguistic behavior of Adjective and Relative clause order, where Japanese belongs to a decided minority, it may simply be that consistent headedness, in particular head-finality, is not a manifestation of a deeper principle.

One final clarification point is worth making. The kinds of correlations that have come to be subsumed under the opposition of head-final versus head-initial are often viewed as “Greenbergian” correlations, as if emerging from, for instance, Greenberg (1966). In fact, Greenberg is much more careful in his formulations, and nothing presented above seriously contradicts the correlations actually claimed by Greenberg (as opposed to further levels of generalization created by others). Appendix II in Greenberg (1966: 108–110) shows, with respect to the order of Subject–Object–Verb, Adjective–Noun, Genitive–Noun, and appositions (relative clauses are not considered) that there are 4 combinations that are substantially more frequent than of the other logically possible combinations. (Since Greenberg treats SVO as ternary and each of the other features as binary, there are 24 logical possibilities.) These are the 4 set out in table 15.

Table 15: The four most frequent types in Greenberg (1966: 108–110)

Type 1	VSO	Prep	N–Gen	N–Adj
Type 2	SVO	Prep	N–Gen	N–Adj
Type 3	SOV	Post	Gen–N	Adj–N
Type 4	SOV	Post	Gen–N	N–Adj

The position of the subject has not been considered in this section; for the record, it may be noted that the frequency of SV dwarfs that of VS. Thus Greenberg's types 1 and 2 combine in our terms to give a single consistently head-initial type. Types 3 and 4 are both consistently head-final, except that type 4 has the adjective after the noun. In outline, and with the omission of relative clauses, this is essentially the picture given above on the basis of WALS materials: The features OV/VO, postposition/preposition, and Gen-N/N-Gen do correlate in terms of consistent headedness; the feature Adj-N/N-Adj does not. While we now know much about the details, including geographical distribution, at a general level Greenberg's claim is validated.

## 5. Noun-modifying clauses

In this and the following section, there is a shift from the "macro" perspective of the preceding sections, turning to a "micro" examination of two phenomena of Japanese that stand in marked contrast to English but where global typological studies such as WALS have not yet reached the level of detail to provide comparative data on a world-wide basis.

In this section, we return to relative clauses. As already noted in section 3, Japanese relative clauses are different from English relative clauses in terms of constituent order, with relative clauses being prenominal in Japanese, but postnominal in English. However, there is another significant difference between the Japanese and English constructions. In Japanese, the same construction is used not only as the translation equivalent of English relative clauses, but also as the translation equivalent of so-called *fact-S* clauses in English (clauses serving as the complement of a head noun) and of a number of other English clauses that do not fit into any single well-defined construction internally to English. In other words, Japanese has a single noun-modifying clause construction, where English has a range of different constructions, including relative clauses, as illustrated in (9)–(11); in each example, the modifying/relative clause is set off by means of square brackets.

- (9) [gakusei ga katta] hon  
the book [which/that the student bought]
- (10) [gakusei ga hon o katta] zizitu  
the fact [that/\*which the student bought the book]
- (11) [dareka ga doa o tataku] oto  
the noise of [someone knocking at the door]  
\*the noise [that/which someone knocked at the door]

In English, relative clauses seem to be a distinct construction not only because of the different translations required for these three Japanese examples, but also in that English, in contrast to Japanese, has clear syntactic constraints on "extraction", i.e. the position that can be relativized, as illustrated in the contrast between the English examples in (12) and the Japanese examples in (13), in particular between (12b) and (13b). In each pair of examples, the (b) example is an attempt – successful in Japanese, but unsuccessful in English – to relativize the noun phrase that is underlined in the (a) example.

- (12) a. The person who kept the dog died.  
 b. \*The dog [that the person who kept died] came to the station every evening to greet his master.
- (13) a. Inu o kawaiatte kureta hito ga nakunatta.  
 b. [Kawaiatte kureta hito ga nakunatta] inu ga maiban eki made kainusi o mukae ni kita.

Building on the analysis of this noun-modifying clause construction in Japanese by Hideo Teramura in the 1970s, Matsumoto (1997) argues that these differences between Japanese and English form a single “package”, a set of logically interrelated properties. Japanese has a single noun-modifying clause construction, and thus no syntactic constraints on the position that can be relativized in translation equivalents of English relative clauses (though there are semantic and pragmatic constraints). English has a distinct relative clause construction where the clause-initial relative pronoun represents “extraction” from a position within the clause, with the result that there are constraints on this extraction, e.g. one cannot extract a noun phrase that is already within a relative clause, as in the ungrammatical example (12b). Typological implications of the distinction are pursued, for instance, in Comrie (1998).

Although the distinctions between the “Japanese-type” and “English-type” constructions are now rather well documented, it remains unclear to what extent other languages of the world follow either the English or the Japanese, or conceivably some other pattern with regard to the unity of noun-modifying clauses. In an ongoing project “Relative Clauses and Noun-modifying Clauses: A Cross-Linguistics Investigation” led by Yoshiko Matsumoto and myself and funded in its current initial stages by the Stanford Humanities Center, we are trying to answer this question, initially with regard to a selection of languages of Asia. (See further <https://www.stanford.edu/dept/asianlang/cgi-bin/?q=node/209>.)

We may note first of all that there are other languages similar to Japanese in having a prenominal relative clause that has no overt pronominal reference to the head noun within the relative clause, such as Turkish, which nonetheless do not share with Japanese the property of having a single noun-modifying clause construction and therefore no constraints on extraction. Turkish (Kornfilt 1997: 64) has clear constraints on extraction, e.g. one cannot relativize a noun phrase that is already inside a relative clause, as seen in the ungrammatical relative clause (14), whose intended meaning is ‘the man such that the woman who loves him committed suicide’.

- (14) \*[*sev-en*                *kadın*    *intihar*    *ed-en*]                *adam*  
 love-PRS.PTCP    woman    suicide    do-PRS.PTCP    man

So despite superficial appearances of constituent order and lack of relative pronouns, Turkish is actually more like English in terms of the opposition between a unified noun-modifying clause and a distinct relative clause.

Tamil, as described by Lehmann (1989: 284–294, 310–311), provides a more intriguing set of data, since in general it seems to behave largely like Japanese, so that examples (15)–(17) exactly parallel Japanese examples (9)–(11).

- (15) *[kumaar-aik kaTi-tt-a] naay*  
Kumar-ACC bite-PST-PTCP dog  
'the dog which bit Kumar'
- (16) *[oru mantiri varu-kir-a] anta vatanti*  
a minister come-PRS-PTCP that rumor  
'the rumor that a minister would come'
- (17) *[naan kiiZee viZu-nt-a] kaayam*  
I down fall-PST-PTCP wound  
'the wound from my falling down'

However, there is at least one twist to the data, since the verb 'to be, exist' in a noun-modifying construction admits only the relative clause interpretation, as in (18), but excludes the *fact-S* interpretation, as in (19).

- (18) *[inta(k) kiraama-tt-il uLLa] iraNTu koovil*  
this village-OBL-LOC exist-PTCP two temple  
'the two temples which are in this village'
- (19) *\*[kaTavu uLLa] nampikkai*  
God exist-PTCP belief  
'the belief that God exists'

At the moment, this seems to be a completely idiosyncratic fact about Tamil, and it remains for future work to see how it can be integrated into a general account of noun-modifying clauses versus relative clauses, including the possibility of intermediate types rather than a clear bifurcation of all the world's languages into one or other of the "pure" types. Preliminary work by Shin-Sook Kim and Peter Sells within the Stanford project suggests that while Korean shares many of the properties identified by Matsumoto for Japanese, its noun-modifying clause construction is nonetheless somewhat more restricted than that of Japanese, for reasons that remain to be clarified.

And at the opposite extreme, even English seems sometimes willing to forego its strict constraints on relative clause formation, in particular the constraint against relativizing a noun phrase that is already within a relative clause, as in example (20b), generally judged acceptable (if not necessarily stylistically impeccable) by native speakers.

- (20) a. You have friends who know some languages.  
b. You choose some languages that you know and some languages that you have friends who know.

Why might Japanese have broader use of noun-modifying clause constructions than other languages, or at least most other languages, including at least some languages that seem otherwise to share many of the relevant properties with Japanese? One possibility, which was initially suggested to me by ongoing research of Nobutaka Takara (UCSB), is that this may be linked to the widespread use of conventionalized noun-modifying constructions with head nouns of low semantic content, such as *koto*, *wake*, *hazu*, *yoo*, *hoo*, *tame*, and even *no*. But the details remain to be worked out.

## 6. Verbs of giving

In contrast to the preceding section, with its discussion of a widespread syntactic construction with ramifications across much of the syntax of the language, the present section is concerned with a very specific lexical phenomenon of Japanese, namely verbs of giving – a phenomenon that is a well-known source of difficulties for non-native learners of Japanese. Simplifying somewhat, we can say that corresponding to the single lexical item *give* in English (and indeed its translation equivalents in most other languages), Japanese has a four-way distinction as in (21), with two relevant parameters.

- (21) *ageru*      inferior gives to superior in out-group  
       *yaru*        superior gives to inferior in out-group  
       *kudasaru*   superior gives to inferior in in-group  
       *kureru*    inferior gives to superior in in-group

In terms of social status, *ageru* and *kureru* group together in that they involve someone of lower status giving to someone of higher status, thus contrasting with *yaru* and *kudasaru*, which involve someone of higher status giving to someone of lower status. In terms of in-group versus out-group, *ageru* and *yaru* group together in that the gift is directed away from the speaker, the speaker's in-group, contrasting with *kudasaru* and *kureru*, for both of which the gift is directed towards the speaker, the speaker's in-group. With regard to this latter parameter, we might say that with *ageru/yaru* the gift **goes** away from the speaker, while with *kudasaru/kureru* it **comes** towards the speaker.

The first parameter, that of social status, is found in several other languages that tend to lexicalize or grammaticalize social status, such as Korean, which distinguishes between unmarked *cwuta* and marked *tulita*, the latter specifically indicating that a giver of lower social status is giving something to a recipient of higher social status. But what about the second, deictic parameter, whereby the gift either goes away or comes towards the speaker?

While I am not aware of any other language that exactly parallels the Japanese system as set out in (20), a number of languages do use different roots or stems to express the notion of giving depending on the grammatical person of the recipient, i.e. on deixis (Comrie 2003). Such languages are scattered across the world, nearly always a tiny minority, though there are some concentrations of such languages, e.g. in the Madang area of Papua New Guinea and among Zapotec and Otomí languages in Mexico. Closely related languages, and even dialects of the same language, may differ in that one has the distinction while the other lacks it, e.g. the Dravidian language Malayalam has the distinction, while the closely related Tamil does not (at least in its standard variety).

Among languages with a deictic contrast in the lexical equivalents of 'give' according to the person of the recipient, the usual distinction is to have one form for first or second person recipient, another for third person recipient, as in Malayalam *taruka* (variant: *tarika*) 'give to 1/2' versus *koTukkuka* 'give to 3'. This pattern is also found in: Kolyma Yukaghir (a Yukaghir language of northeastern Siberia), Tsez (a Nakh-Daghestanian language of the North Caucasus), Lepcha (a Tibeto-Burman language of the Himalayas), Enga, Kewa, and Hamtai (three Trans-New Guinea languages of Papua New Guinea), Manambu (a Sepik language of Papua New Guinea), Saliba (an Austronesian language of Papua New Guinea), Nandi (an East Sudanic language of Kenya), various Zapotecan and Otomí varieties (Oto-Manguean languages of Mexico), and Mískitu (a Misumalpan language of Nicaragua). Much less com-

monly, the opposition is between first person versus second or third person recipient; this 1 vs. 2/3 system is found in Kenuzi-Dongola (an East Sudanic language of southern Egypt and northern Sudan).

Comrie (2003) argues that such distinctions probably reflect grammaticalization of a lexical deictic opposition. In Japanese, the distinction between *ageru/yaru* on the one hand and *kudasaru/kureru* on the other is a lexical deictic distinction, which as such often has clear implications for the grammatical person of the recipient, but is not in one-one correspondence with this distinction; thus, either verb can be used for a third-person recipient, depending on whether or not that person is part of the speaker's closer in-group than the giver. Languages like Malayalam seem to have grammaticalized a distinction of this kind in terms of the opposition between 1/2 and 3 person (or 1 and 2/3 person) (Comrie 2003). While the phenomenon is not identical to the deictic distinction found in Japanese, the two can be rather readily related to one another. The complex system of Japanese can then be seen as the intersection of two parameters that are more widely represented cross-linguistically in translation equivalents of 'give', namely social status and deixis. The combination found in Japanese may be unique, but its component parts or their close analogs can be found elsewhere.

## 7. Conclusions

There are many ways in which one can throw light on the question of the typological position of Japanese among the world's languages. The material from a world-wide typological database project like WALS enables one to sift rapidly through masses of data in order to place Japanese, as a language that is certainly no less typical overall than English, but whose much heralded strict head-finality may well be more of a contingent property than a deep-seated generalization. Nonetheless, more detailed studies of specific phenomena also play a significant role in understanding how Japanese fits in among the typological diversity of the world's languages, whether one is examining phenomena with wide-ranging implications for the structure of the language, such as noun-modifying clauses, or details like Japanese verbs of giving.

### References

- Comrie, Bernard (1998) Rethinking the typology of relative clauses. *Language Design* 1: 59–86.
- Comrie, Bernard (2003) Recipient person suppletion in the verb "give". In: Mary Ruth Wise, Thomas N. Headland and Ruth M. Brend (eds.) *Language and life: Essays in memory of Kenneth L. Pike*, 265–281. Dallas: SIL International and The University of Texas at Arlington.
- Comrie, Bernard (2008) Prenominal relative clauses in verb-object languages. *Language and Linguistics* 9: 723–733.
- Comrie, Bernard (forthcoming) Exploiting the *World Atlas of Language Structures*: New directions in areal typology. *Studies in Language Sciences* 10.
- Dryer, Matthew S. (1988) Object-Verb order and Adjective-Noun order: Dispelling a myth. *Lingua* 74: 185–217.
- Dryer, Matthew S. (1992) The Greenbergian word order correlations. *Language* 68: 81–138.
- Dryer, Matthew S. (2005a) Order of adjective and noun. In: Martin Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie (eds.), 354–357. <http://wals.info/feature/87>.
- Dryer, Matthew S. (2005b) Order of adposition and noun phrase. In: Martin Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie (eds.), 346–349. <http://wals.info/feature/85>.
- Dryer, Matthew S. (2005c) Order of genitive and noun. In: Martin Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie (eds.), 350–353. <http://wals.info/feature/86>.

- Dryer, Matthew S. (2005d) Order of object and verb. In: Martin Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie (eds.), 338–341. <http://wals.info/feature/83>.
- Dryer, Matthew S. (2005e) Order of relative clause and noun. In: Martin Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie (eds.), 366–369. <http://wals.info/feature/90>.
- Dryer, Matthew S. (2005f) Relationship between the order of object and verb and the order of adjective and noun. In: Martin Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie (eds.), 394–397. <http://wals.info/feature/97>.
- Dryer, Matthew S. (2005g) Relationship between the order of object and verb and the order of adposition and noun phrase. In: Martin Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie (eds.), 386–389. <http://wals.info/feature/95>.
- Dryer, Matthew S. (2005h) Relationship between the order of object and verb and the order of relative clause and noun. In: Martin Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie (eds.), 390–393. <http://wals.info/feature/96>.
- Greenberg, Joseph H (1966) Some universals of grammar with particular reference to the order of meaningful elements: In Joseph H. Greenberg (ed.) *Universals of Language*, Second edition, 73–113. Cambridge, MA: MIT Press.
- Haspelmath, Martin, Matthew S. Dryer, David Gil and Bernard Comrie (eds.) (2005) *The world atlas of language structures*. Oxford: Oxford University Press. (Book and CD-ROM.)
- Haspelmath, Martin, Matthew S. Dryer, David Gil and Bernard Comrie (eds.) (2008) *The world atlas of language structures online*. Munich: Max Planck Digital Library. <http://wals.info/>.
- Kornfilt, Jaklin (1997) *Turkish*. London: Routledge.
- Lehmann, Thomas (1989) *A grammar of modern Tamil*. Pondicherry: Pondicherry Institute of Linguistics and Culture.
- Maddieson, Ian (2005) Vowel quality inventories. In: Martin Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie (eds.), 14–17. <http://wals.info/feature/2>.
- Matsumoto, Yoshiko (1997) *Noun-modifying constructions in Japanese: A frame semantic approach*. Amsterdam: John Benjamins.
- Siewierska, Anna (2005a) Verbal person marking. In: Martin Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie (eds.), 414–417. <http://wals.info/feature/102>.
- Siewierska, Anna (2005b) Third-person zero of verbal person marking. In: Martin Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie (eds.), 418–421. <http://wals.info/feature/103>.
- Tsunoda, Tasaku (2009) *Sekai no gengo to Nihongo*. Second edition. Tokyo: Kurosio.

## Bernard Comrie

### Present Positions:

- Director, Department of Linguistics, Max Planck Institute for Evolutionary Anthropology, Leipzig (since 1997)
- Distinguished Professor of Linguistics, University of California Santa Barbara (since 2002)
- Honorary Professor of Linguistics, University of Leipzig (since 1999)

### University Education:

- BA 1968 University of Cambridge: Modern and Medieval Languages
- Ph.D. 1972 University of Cambridge: Linguistics

### Major Publications and Papers:

- 1976. *Aspect*. Cambridge: Cambridge University Press.
- 1985. *Tense*. Cambridge: Cambridge University Press.
- 1989. *Language Universals and Linguistic Typology*. Oxford: Blackwell.
- 2005 (ed., with M Haspelmath, M S Dryer, and D Gil). *The World Atlas of Language Structures*. Oxford: Oxford University Press.
- 2007. Areal typology of mainland Southeast Asia: what we learn from the WALS maps. *Manusya, Special Issue 13*: 18–47.