# Lexical Evidence for Early Contact between Indonesian Languages and Japanese

Ann Kumar, Phil Rose

Oceanic Linguistics, Volume 39, Number 2, December 2000, pp. 219-255 (Article)

Published by University of Hawai'i Press
*DOI: 10.1353/ol.2000.0018*

# Lexical Evidence for Early Contact between Indonesian Languages and Japanese

Ann Kumar and Phil Rose

AUSTRALIAN NATIONAL UNIVERSITY

Forty-one pairs of words with CVCVC structure selected from Old Japanese and Old Javanese dictionaries are presented. It is claimed that these are the result of borrowing into an antecedent of Old Japanese from an Indonesian source. Semantic relationships are discussed, and sound correspondences are specified within a discussion of the segmental phonology and phonotactics of the two languages. The agreement in phonological form is shown to be extensive, applying in some cases to up to five segments in each word pair, and to also make sense, given the phonotactic restrictions of the recipient language. The semantic agreement is often also of comparably high specificity, showing moreover a further level of structure in its partial resolution into semantic fields, including some that resonate with nonlinguistic findings related to ritual and rice cultivation in the Yayoi period of early Japanese history. The amount of agreement in semantic and phonological form is shown statistically to be greater than chance. The argument is further strengthened by several additional independent levels of agreement in the data that are discussed within a Bayesian framework. The phonological correspondences map unidirectionally from Old Javanese to Old Japanese, and a search for cognates in all Austronesian languages covered by the major comparative dictionaries reveals that the lexical items are localized to the Indonesian subarea of Malayo-Polynesian. This points to one or more Indonesian languages as the source of the borrowings. The agreement between semantic and archaeological evidence on material and spiritual culture dates the contact to the Yayoi period. The semantic evidence further suggests that, contrary to the received view, important Yayoi innovations are likely to have been introduced into Japan from the south, and not from China or Korea as usually supposed.

## 1. INTRODUCTION[1]

**1.1 AIM.** We wish in this paper to present some linguistic data that are indicative of an early interaction between Japanese and one or more of the so-called "Western Malayo-Polynesian" languages, most probably of Indonesian origin.

**1.2 BACKGROUND.** Nonlinguistic evidence from rice genetics and rituals relating to kingship (Kumar 1992), and from biological anthropology and mito-

*Oceanic Linguistics,* Volume 39, no. 2 (December 2000)

chondrial DNA (Kumar 1998), has provided considerable support for a Western Malayo-Polynesian, and most probably specifically Indonesian, influence on early Japan. The first author has suggested that this took place in the Yayoi period (early centuries B.C. and A.D.). An attempt was therefore made to find out whether any linguistic evidence also existed to support this hypothesis. It was assumed that if such evidence existed, it would, of course, be locatable in historical varieties of Japanese, and in some Western Malayo-Polynesian (WMP) languages.

The results of the investigation showed that there was indeed lexical influence, and this is described in detail below. We first present background information on the two languages we compared—Old Japanese and Old Javanese—and give reasons why the latter was selected as our WMP language. Next we describe the procedure we used to compare lexical material in Old Japanese and Old Javanese. We then present and discuss the results. The second author also addresses the question of whether the high degree of phonological and semantic similarity demonstrated can be shown probabilistically to be indicative of nonfortuitous, bona-fide borrowing.

Attempts to link Japanese genetically with other languages abound. We would like to emphasize that we make no claim in this paper for genetic relationship.

## 2. LANGUAGES COMPARED

**2.1 OLD JAPANESE.** The history of Japanese is divided into a number of different periods. The first, Proto-Japanese, may be dated to the period 300–400 B.C. to 300–400 A.D., a time span corresponding to the Yayoi culture, possibly continuing into the succeeding Kofun period. It is generally agreed that Old Japanese is the language of the oldest extant written records and of eighth-century Nara literary works such as the Kojiki and Man'yōshū, but scholars differ as to when it ends. Some scholars consider Old Japanese to go up to the tenth century at the latest, while Miller (1967:37–39) distinguishes a "late Old Japanese" that includes the twelfth-century *Tale of Genji*. Shibatani (1990:118) considers Old Japanese to be the direct antecedent of modern Japanese, being replaced by the latter somewhere between the twelfth and sixteenth centuries, while others posit an intermediate Middle Japanese. For this paper, Old Japanese is the language recorded in the dictionaries described in 3.1.1.

For this paper, it was decided to use the attested language of Old Japanese rather than Proto-Japanese, because as Campbell (1997:197) points out, undetected borrowings have not infrequently distorted reconstructions, and major unde-

1. A preliminary account of this subject was given in Kumar (1996) and presented as a paper at the Fourth International Symposium on Languages and Linguistics. We gratefully acknowledge much valuable advice from the following colleagues, who are naturally absolved from any of our errors of fact or interpretation: Peter Bellwood, Tony Diller, Geoff Haig, Peter Hendriks, Harold Koch, James Matisoff, Bernd Nothofer, Andrew Pawley, Malcolm Ross, Brian Scott, Royall Tyler, and Alexander Vovin. The authors would also like to thank Megumi Edison, Shun Ishihara, Masa Onishi, and Phichit Roinil for processing data, and our four anonymous reviewers for their helpful comments and critique. The work was supported by an A.N.U. Small Faculties Research Fund Grant of $7000 Australian.

tected borrowings are what are hypothesized here. Further, as pointed out by Hübschmann (1967), such undetected borrowings can also lead to an erroneous view of a language's genetic affiliation.

**2.2 OLD JAVANESE.** It was also decided to use an attested WMP language, rather than a reconstruction. This is because WMP languages do not constitute a genetic unity, as the putative member languages do not share any innovations in either lexicon or morphology. Even attempts to group some of these WMP languages into smaller subgroups and to reconstruct protolanguages for them, such as Nothofer's Proto-Malayo-Javanic, have proved less than convincing (Blust 1981), and, although a search for more satisfactory ones is still under way (Ross 1995:79), it is more likely that the so-called WMP languages constitute the remnants of a loose network of dialects (Pawley 1999:127).

We selected Old Javanese for the following reasons. The first, and most cogent, is that the genetic and biological-anthropological findings mentioned above link Japan to the Indonesian area. Kumar (1992:258–282) describes a specific and detailed set of mythological correspondences in kingship rituals between Japan and Indonesia. For example, in both cases the central myth and ritual of kingship revolves around the moon maiden, associated with rice, whose robe the ruler must assume upon accession, and who, when she has returned to heaven, can be summoned by a burnt offering on the part of the ruler. In both cases, there is a sacred dance dedicated to the goddess, performed at accession, and associated with the ritual numbers 4/5 and 8/9. It is important to emphasize the high degree of specificity involved in the correspondences: the rituals do not refer, for example, to vague and general *topoi* such as "ancestor worship," "sacred marriage," or "triumph of order over chaos."

Second, it has been demonstrated (Morinaga 1968; Oka 1988:145–148) that Japanese rice and Javanese (*javanica*) rice—the third major subdivision alongside the Chinese (*sinica*) and Indian (*indica*)—are more closely related to each other than either is to any other type of rice. *Javanica* rice originates in Indonesia.

Third, those linguists who have suggested that there was an Austronesian element in Japanese seem to derive a significant percentage of the lexemes supporting their theory from Javanese (e.g., Izui's material quoted in Shibatani 1990:104–105).

Lastly, Javanese has the oldest manuscript tradition of any Indonesian (indeed, any Austronesian) language, the only earlier written records being inscriptions such as a small number of seventh-century inscriptions in Old Malay (Kumar and McGlynn 1996:6) and possibly some in Old Cham.

No reconstructions exist for Proto-Javanese. Faute de mieux, therefore, Old Javanese has been used.

Javanese has historically been the politically and culturally dominant language of the region (although in post-Islamic times Malay became the lingua franca), and Javanese loanwords are found not only in other Indonesian languages, but also in mainland languages, in the Philippines, and even in Taiwan (Adelaar 1994). This dominance is reflected in the survival of unusually rich and ancient

written sources in Old Javanese going back to the eighth century—indeed, Old Javanese is the only ancient Southeast Asian literature to be so preserved. A large lexicon is thus available for analysis. Unfortunately, due to the lack of knowledge of earlier levels of WMP languages described above, the date at which Old Javanese emerged as a separate language is not known.

**3. METHODOLOGY.** Having decided to look for evidence of lexical influence between Old Javanese and Old Japanese, the procedure basically involved the arduous task of scanning dictionaries of both languages for possible matches. This was then followed by a multistage culling. We describe below first the dictionaries used, and then the culling procedures.

Like most major South East Asian languages, Old Javanese is heavily impregnated with Sanskrit-derived words, which may constitute as much as half the vocabulary in literary works. These were identified and excluded from the investigation. This is because the hypothesized period of influence, perhaps going back several centuries B.C., is conventionally regarded as predating significant sanskritization of the language. It may be that there *was* in fact already sanskritization in this period, in which case Sanskrit words could have been included in the borrowings. Given the lack of written records for Indonesia 2,000 years ago, we have left this question aside for the moment.

## 3.1 REFERENCE DICTIONARIES

**3.1.1 Old Japanese dictionaries.** We used two references for Old Japanese. Our initial reference work was Ohno, Akihiro, and Maeda (1991) *Iwanami Kogo Jiten*. This dictionary has 43,000 entries. Each entry is transliterated in hiragana, followed by kanji where these exist. Verbs are listed in their "conjunctive" or "infinitive" *(-i)* form. Some linguistic information is provided on word class, subclasses (e.g., transitive and intransitive verbs), and other morphological details (e.g., conjugation class).

The items in this dictionary are drawn from literary works spanning the Nara period (710–785) to the (early) Edo period (1603–1868). It is possible that this literary bias may have led to the exclusion of some lexical items relating to more mundane spheres of life (it does not contain, for instance, any word for 'chisel'). The dictionary's hiragana transliterations were romanized, and its modern Japanese glosses translated into English.

After the first author had identified a corpus of likely cognates in this relatively short dictionary, to obtain a fuller specification of the meanings of the putative borrowings we consulted the more comprehensive *Jidaibetsu Kokugo Daijiten— Joudai Hen* (1967). Each entry is transliterated in hiragana, followed by kanji where these exist. Verbs are listed in the "conclusive" *(–u)* form. The dictionary also provides a small amount of linguistic information on the entries, such as word class and conjugation class. It also provides examples of usage from Old Japanese texts.

Both these dictionaries were used to specify the meaning of the Japanese lexemes.

**3.1.2 Old Javanese dictionary.** Our source for Old Javanese was Zoetmulder's (1982) *Old Javanese-English Dictionary.* Lexemes are listed in their standard romanized form (Old Javanese was written with an Indic Script called Kawi [Kumar and McGlynn 1996:7–10]), and appear in the order of the Roman alphabet. Lexemes are listed under their base (unaffixed) form, followed by derived forms where these exist. Zoetmulder does not usually specify the word class of the lexemes, and in the case of the base form, this is often ambiguous. For example, *kělir* 'screen', and *kukus* 'smoke' can be nouns or verbs, as is indeed also the case with their English glosses. Derived forms with prefixes, infixes, or suffixes, however, are unambiguous with respect to word class. Zoetmulder also includes quotations from Old Javanese texts to illustrate usage.

Although this dictionary has more than 25,500 entries, these all come from literary sources, most of them East Javanese. Because of this regional and literary bias, it cannot be assumed to completely cover the Old Javanese lexicon. It does not include, for instance, vocabulary found in inscriptions rather than in literary works. Finally, a number of words in common usage in modern Javanese and in other Indonesian languages are not found, even though it is highly probable that they existed in the Old Javanese colloquial. For these reasons, there is a case for systematically analyzing words found in Modern Javanese but not in Old Javanese, because the former gives coverage of central Javanese vocabulary. And indeed, even without making a systematic search, one can think of a number of modern Javanese words of CVCVC length that display the same sound correspondences with Old Japanese words as the data discussed here. For example, Mod. Javanese *tukul* 'to grow': Old Japanese *tukuri* 'to cultivate a rice field'; Mod. Javanese *kawah* 'flooding': Old Japanese *kaFa* 'river'; and perhaps most strikingly Mod. Javanese *kěbul* 'smoke, vapor': Old Japanese *keburi,* 'smoke, incense'. Despite the existence of such forms, we have not, for the sake of consistency, included them in the paper.

**3.2 CULLING PROCEDURES.** The romanized and translated entries from the Old Japanese dictionaries were scanned by the first author, who has a knowledge of Old Javanese, for words of generally similar form and meaning to Old Javanese and Kawi. This gross sampling turned up many—ca. 150—prima-facie candidates.

As Ringe (1992:1) points out, any pair of languages can be expected to exhibit a nonnegligible number of fortuitous similarities in the phonological shape of words of similar meaning. Dawkins (1999:151) has a useful acronym for these fortuitous similarities: the *petwhac* (Population of Events That Would Have Appeared Coincidental). Given the phonotactic similarity between Old Javanese and Old Japanese, the size of the petwhac might be expected to be rather large. Therefore the initial candidate list was subjected to a multistage culling designed to constrain it. This culling resulted in the removal of some items consistent with the semantic fields of the borrowings, but with irregularities in form: thus, for example, Old Javanese *kukusan* 'rice steamer' and Old Japanese *ko₂siki* (same

gloss), were excluded on the grounds of metathesis. Nevertheless, the authors remain conscious that different preferences for overall word shape (as opposed to restrictions at the level of the syllable) in the donor and borrowing language may have produced more cases of such metatheses. This first cull resulted in the 82 pairs of items given in Kumar (1996:531–540).

These 82 pairs were then further culled on the basis of word length and closeness of both phonological and semantic correspondences. This resulted in the removal from the data presented here of prima-facie plausible pairs such as Old Javanese *barĕsih* 'to purify ritually': Old Japanese *Farapi* (with the same gloss) (the *s:p* correspondence is irregular).

Nichols (1996:50) argues that while lexical evidence is generally not probative of genetic relatedness, an occasional word is long enough or otherwise sufficiently structured that it can be a good diagnostic—an example being the words for 'widow' in Indo-European languages. Therefore, we have selected only words with five segments (CVCVC) in OJav. The CVCVC linear structure is a canonical one for Javanese and some other Austronesian languages, but is unusual in Japanese (A. Vovin, pers. comm.). This makes it worthwhile examining such forms for evidence of borrowing. However, it has the effect of considerably reducing the number of potential comparisons. At an estimate, there were only between 700 and 800 words of this specified length in Ohno, Satake, and Maeda's (1991) dictionary.

Finally, we have divided the data presented here into a primary and a secondary list. The primary list is characterized by an isomorphism of extremely close phonological and semantic correspondence, and constitutes our "known" data in terms of which we have drawn up a list of sound correspondences. The secondary list does not approach this requirement quite so closely, but nevertheless contains some additional plausible sound correspondences.

It will be appreciated that this whole procedure is very similar to that described in Campbell (in press), who sets out a number of desiderata for those claiming the existence of remote language families such as Amerind, Nostratic, and even Proto-World. He writes that it is desirable that proposals of distant genetic relationships should go through two stages: "Where the intention is to call attention to a possible but as-yet untested connection, one often casts a wide net in order to haul in as much potential evidence as possible. When the intention is to test a proposal that is already on the table, those forms admitted initially as possible evidence are submitted to more careful scrutiny."

While we are not, of course, attempting to establish a distant genetic relationship, we have nevertheless aimed at a very careful and cautious approach for three reasons. First, like the newer distant genetic relationships, our proposal is new in historical linguistics. Second, it relates to the history of the Japanese language, a highly controversial topic in historical linguistics. It is well known how many attempts have been made to link Japanese genetically to some other language—if not from Arabic to Zulu, at least from Basque to Tamil—and the lack of adequate evidence that has made all of them undemonstrable (thus Nichols 1990:515 still lists Japanese as an isolate). Third, our proposal is also revisionary in terms of the

ruling myth (in the nonpejorative sense of Lass 1997:4–5) of Indonesian history. According to this myth, whereas Indonesia has been influenced significantly by other civilizations (both Indic and Islamic), it is not considered to have exerted comparable influence on a similar distant cultural area.

We have also followed Campbell's (in press) other desiderata for establishing distant genetic relationships, namely: eliminating any "semantic promiscuity"; using, despite their frequency in Japanese, no monosyllables, and only words of CVCVC length; and requiring sound-meaning isomorphism. These requirements are highly congruent with those of Ringe, whose statistical work is utilized in the section on probability.

## 4. PHONOLOGY

**4.1 OLD JAPANESE PHONOLOGY.** The phonotactics of Old Japanese were simple, the syllable consisting of an obligatory vowel, preceded by an optional consonant and possibly an optional glide (Vance 1987:56). Thus no syllable-coda consonants were allowed. Syllables consisting of a single vowel were restricted almost exclusively to word-initial position, thus making word-internal vowel sequences very rare (Vance 1987:16). No *r* was allowed word-initially (Itoo and Mester 1995:836 n. 6), neither were voiced obstruents in native words (Vance 1987:134). The combination *wu* was disallowed phonotactically (Matsumoto 1995). Only one voiced obstruent was allowed morpheme-internally—so-called Lyman's Law—(Itoo and Mester 1995:819; Vance 1987:136). Compound-internal morphophonemic alternation between voiceless and voiced obstruents—so-called *rendaku*—was also present (Itoo and Mester 1985:819).

The OJAP consonantal phonemic system was also quite simple. The 13 consonantal phonemes are shown in table 1, compiled on the basis of Ohno (1980:142–152) and Hashimoto (1980:140–143). Vance (1987:134–135, 147–148) notes that OJAP *F* reflects an earlier *p, and that there is some disagreement as to whether it had already become a fricative [ɸ] in OJAP. He also notes (Vance 1987:108) that the voiced obstruents probably had prenasalized allophones word-internally, and that *m* and *b* commonly alternated (Vance 1987:147). The fricatives *s* and *z* had palatal allophones before *i;* alternatively, two affricate phonemes *ts* and *dz* are posited, also with palatal allophones before i.

**TABLE 1. OLD JAPANESE CONSONANTAL PHONEMES**

| | LABIAL | DENTAL /ALVEOLAR | PALATAL | VELAR |
|---|---|---|---|---|
| STOPS | b | t d | | k g |
| FRICATIVES | F | s z | | |
| APPROXIMANTS | | r | j | w |
| NASALS | m | n | | |

Orthographic evidence suggests that there were two contrasting series of vowels for *i, e,* and *o:* $i_1$, $i_2$, $e_1$, $e_2$, $o_1$, and $o_2$, as well as an undifferentiated *a* and *u.* There is disagreement both on the phonetic values of these and on the number of phonemes they represent (Shibatani 1990:131–139). Reconstructed values from Ohno (1980:154) are shown in table 2. In some cases, Ohno's transliteration does not distinguish the vowels $i_1$ and $i_2$, and $o_1$ and $o_2$. In the majority of cases, this was no problem, because in the introductory part of the dictionary, general rules are given for predicting the form from the (morpho)phonological environment. In the rare cases where there was still uncertainty, Matsumoto (1995) was consulted.

**4.2 OLD JAVANESE PHONOLOGY.** Data on Old Javanese phonology comes primarily from Nothofer's reconstruction of Proto-Malayo-Javanic (1975:16–20), and forms in Zoetmulder's Old Javanese dictionary.

As explained above, Sanskrit words were eliminated from our OJAV data for important methodological reasons, and we therefore consider the phonological structure of indigenous OJAV words only. It is difficult to establish precisely the details of the indigenous OJAV phoneme system, especially the vowel phonemes,[2] because both Nothofer (1975:16–19) and Zoetmulder (1982:xi–xv) use transliterated orthography when referring to Old Javanese segments, and do not phonemicize the Old Javanese graphemes. In the tables below, therefore, we give a likely phonemic array. We use Zoetmulder's symbols, because we quote forms in this system in what follows. Where Nothofer differs from Zoetmulder, the former's transcription is given in brackets.

It is likely that OJAV had a five-vowel triangular system, with schwa augmentation, like Modern Javanese. These vowels are given in table 3.

The indigenous OJAV consonant symbols are given in table 4. As can be seen, the OJAV consonantal array is typologically unremarkable, with five contrasting places for stops and nasals, and two for fricatives. We have analyzed sequences of nasal and stop as a single phonemic unit, as Horne (1963:xv) and Nothofer (1975:8) have done for Modern Javanese. The addition of a set of prenasalized stops then makes three contrasting series of stops.

There is to our knowledge no published phonotactic statement on Old Javanese, so we have based our phonotactic generalizations on the data in Zoetmul-

**TABLE 2. RECONSTRUCTED VALUES FOR OLD JAPANESE VOWELS**

| | FRONT | CENTRAL | BACK |
|---|---|---|---|
| HIGH | i | | u |
| | | ï | |
| MID | e | ë ö | o |
| LOW | | | a |

2. Nothofer (1975:202) gives five "Old Javanese" vowels as reflexes of PAN; he also lists (p. 8) eight vowel phonemes for Modern Javanese, but it is unlikely that é and è, and ó and ò contrast.

der's dictionary. A base word can consist of one to four syllables. The syllable structure appears to be $(C_1)$ $(C_2)$ V $(C_3)$, where $C_1$ is any of the phonemes in table 4, and $C_2$ is either a liquid or a semivowel. There are heavy restrictions on $C_3$ exponents: nonoccurring $C_3$ segments are shaded in table 4. Some cooccurrence restrictions exist for $C_1$ and $C_2$.

**4.3 PHONOLOGICAL CONSEQUENCES OF BORROWING.** Given the inventories and phonotactic structure of the two languages, a putative borrowing of items from an antecedent of OJAV into an antecedent of OJAP will encounter the following problems due to incompatible phonological structure.

- Because OJAP did not allow syllable-final consonants, OJAV syllable-final consonants could either be lost or resyllabified, as for example, OJAV *am. > OJAP *a, or *a.mV. But because there was no *ŋ in OJAP, the only options would be to lose it, or change to a phonetically similar phoneme (a velar stop, perhaps, or an alveolar nasal).
- Lyman's Law will require a change when the OJAV morpheme has more than one voiced obstruent. It is possible that the first will be devoiced because of the restriction on word-initial voiced obstruents in OJAP.
- *Rendaku* may result in the voicing of word-initial obstruents in borrowed OJAV words, if they occur as a nonfinal element in a compound.
- OJAP does not permit word-initial *r,* so word-initial OJAV *r* will be affected.
- OJAV phonemes without counterparts in OJAP (*h, l,* retroflexes, and palatals, for example) will be affected. It is possible that OJAV palatals might be interpreted as palatal allophones of OJAP *s, z ~ ts, dz* followed by a palatal vowel.
- Because OJAP does not allow the sequence *wu,* OJAV *wu* will be affected.

**TABLE 3. INDIGENOUS OLD JAVANESE VOWELS**

| | FRONT | CENTRAL | BACK |
|---|---|---|---|
| HIGH | i | | u |
| MID | e (é, è) | ĕ (ə̃) | o (ó ò) |
| LOW | | a | |

**TABLE 4. INDIGENOUS OLD JAVANESE CONSONANTS**

| | BILABIAL | DENTAL | RETROFLEX | PALATAL | VELAR/ GLOTTAL |
|---|---|---|---|---|---|
| STOPS | p | t | ṭ | c | k |
| | b | d | ḍ | j | g |
| PRENASALIZED STOPS | mb | nd | nḍ | nj | ŋg |
| NASALS | m | n | | ny | ŋ |
| FRICATIVES | | s | | | h |
| LIQUIDS/ GLIDES | w | l | r | y | |

**5. DATA AND ANALYSIS.** In tables 6 and 7 are listed our two sets of data. For convenience, we give the sound correspondences occurring in the data in table 5. Table 6 gives our primary list: 26 pairs of Old Javanese and Old Japanese words that show very close agreement both in phonological form and meaning. These examples we claim represent bona-fide borrowings into an early form of Japanese from a source closely related to antecedents of Old Javanese. The data are divided for convenience into two groups: group A contains the verbs, and group B contains other pairs that display the same high agreement in form and meaning.

Table 7 presents our secondary data. These do not show as high a degree of agreement as the items in our primary list. This is usually either because of small discrepancies in form, or because the meaning of the items is too general, or both. The entries are arranged as follows. For each entry, the word on the left is Old Javanese; on the right, Old Japanese. The OJAV data are cited in Zoetmulder's romanization. In the case of the Old Japanese glosses, for the relatively infrequent cases where there was a difference in the glosses given by Ohno, Akihiro, and Maeda (1991) (O) and the *Jidaibetsu Kokugo Daijiten— Joudai Hen* (1967) (JKD), these are cited separately. Glosses that appear unrelated have been included, but in square brackets. The names of Austronesian languages with putative cognates are given in abbreviated form after the OJAV entry. The unabbreviated Austronesian language names are given in the appendix.

We discuss the data in table 6 first.

**TABLE 5. SOUND CORRESPONDENCES IN PRIMARY AND SECONDARY DATA BETWEEN OLD JAVANESE (OJAV) AND OLD JAPANESE (OJAP)**

| OJAV | OJAP | ENVIRONMENT | OJAV | OJAP | ENVIRONMENT |
|---|---|---|---|---|---|
| p | m | / __+ i $_{\text{verbal suffix}}$ | r | t | / #__ |
| | F | / elsewhere | | Ø | / __# |
| | | | | r | / elsewhere |
| t | t | | l | Ø | / #__ |
| | | | | r | / elsewhere |
| k | k | | s | s | |
| b | F | / #__ | h | Ø | / __# |
| | b | / elsewhere | | r | / + i $_{\text{verbal suffix}}$ |
| d | t | / #__ | w | Ø | / __ u |
| | r | / elsewhere | | w ~ F | / elsewhere |
| ḍ | r | / intervocalically | i | i ~ $e_2$ ~ a | |
| g | k | / #__ , + i $_{\text{verbal suffix}}$ | a | a | |
| | g | / elsewhere | | | |
| m | Ø | / syllable-final | o | o ~ $o_2$ ~ u | |
| | m | / elsewhere | | | |
| n | n ~ Ø | / syllable-final | u | o | / w_ |
| | n | / elsewhere | | u ~ o | / elsewhere |
| ŋ | Ø | / syllable-final | ě | e | / __bilabial C |
| | | | | a ~ u | / elsewhere |

## TABLE 6. PRIMARY LIST OF BORROWINGS

LIST A: VERBS

| | |
|---|---|
| **sawak** (Z1715)<br>'to call out' | **sawak-**<br>O: 'to gather and make a noise in disorder; to gather and cry in tumult'<br>JKD: 'to make noise, to rustle'; '[to be in a disordered or jumbled state]'<br>[JKD notes that the former sense is a derivative of an onomatopoeic verb *sawa.* This appears to be incorrect, given the OJAV form, and the fact that JKD does not list *sawa* separately.] |
| **sosok** (Z1808); AMIS.<br>'to pour out' | **so$_2$so$_2$k-**<br>'to pour, sprinkle, splash water; to rinse, wash off; to gush out' |
| **matur** (Z603)<br>'to present, offer, tell or report to person of higher rank' | **matur-**<br>'to give or present s.t. to a person of high rank/God; to offer prayers; to honor the memory of God by making offerings at a shrine, etc.; to eat and drink (of a person of high rank)'; (aux vb.) nonsubject exalting action of humble person to person of higher rank |
| **kěkěr** (Z839)<br>'to put away or hide, to cover' | **kakar-**<br>of clouds, etc., 'to cover (the sky); to hang over; [to be possessed by a spirit; (of a God) to draw close to a human; to lean; rely; depend'] |
| **pupur** (Z1448)<br>'blow, battle, cock fight with natural spurs' | **FoFur-**<br>O: 'to cut/tear (the body of a bird/an animal) into pieces; to massacre'<br>JKD: 'to cut up; to rip to pieces; to kill s.o. with sword' |
| **mawuwus** (muwus, mowos; Z2339)<br>'to utter words' | **mawos-**<br>O: 'to reveal (to a god, emperor), to report, to tell (to superior); to ask a favor'<br>JKD: 'to say (s.t. to s.o. of high rank); (aux vb.) nonsubject exalting' |
| **(a)namar** (Z1625)<br>'to hide, disguise' | **namar-** (nabar-)<br>'to hide, be hidden' |
| **kukuh** (Z912)<br>'to strengthen, support' | **kukur-**<br>'to tie up, bind, fasten'<br>[*kukuh* means 'tie up' in Modern Javanese. It is therefore possible that this is the original OJAV meaning, and that a semantic shift to 'strengthen' occurred in East Javanese.] |
| **wuwuh** (Z2338); NGA, SAW<br>'increase, growth; to grow, increase' | **wowor-**<br>O: 'to grow in profusion'<br>JKD: 'to bend' [under the weight of something], e.g., a branch bending under the weight of plentiful blossoms |
| **tutup** (Z2084); IND SUN WOL MAD DAA BUG BAT SAN<br>'cover, lid'; **tumutupi** (with um infix), 'to cover' | **tutum-**<br>'to wrap (s.t.); to cover; to conceal (s.t.) inside (s.t.); [to dam up a stream by heaping up soil]' |
| **tatap** (Z1960)<br>'to get on top of, above s.t.' | **tatam-**<br>'to fold up'. Also used of rocks in layers *(iFa tatami)* and of folded cloth of paper mulberry fibre used for Shinto ceremonies. |
| **susup** (Z1875); SUN, KAU<br>'to enter, go in; to penetrate' | **susum-**<br>'to go forward, advance rapidly, proceed' |
| **kukup** (Z913)<br>'to gather up' | **kukum-**<br>'to wrap up, hold, cover; to make invisible' |
| **tutug** (Z2083)<br>'to the end, completely, as far as, to extend, reach up to' | **tutuk-**<br>'to last, continue, follow, go on' |
| **sěkar** (Z1728)<br>'flower'; [in verbal forms]<br>'to bloom' | **sakar-**<br>'to flourish'<br>[This may be morphologically complex, from *sak-* 'to bloom', and *-ar* 'unspecified derivational suffix'(?)] |

**TABLE 6. PRIMARY LIST OF BORROWINGS (CONTINUED)**

| LIST B: OTHER WORD PAIRS | |
|---|---|
| **parĕm** (Z1294)<br>'to clear (of weather, after rain)' | **Fare**<br>'to clear up, after rain, cloud, fog, mist, snow' |
| **sawah** (Z1715)<br>'paddy' | **saFa**<br>'swamp, marsh' |
| **wawar** (Z2228)<br>'strips of leaf hung on a string [e.g., for ceremonial usage]' | **wawara-ba**<br>O: 'torn leaf'. JKD 'frayed or withered leaf' |
| **duduk** (Z422); MAG<br>'bamboo spear, any weapon for stabbing or piercing' | **turuki**<br>O: 'general term for swords'. JKD: 'a sword' |
| **paḍaŋ** (Z1225)<br>'cleared[field]' | **Fara**<br>'a plain, a field; a flat wide area' |
| **guḍaŋ** (Z507)<br>'building to keep grain, treasure, furniture' | **kura**<br>'storehouse for grain, treasure, furniture'[ e.g., the three kura in which the divine regalia of the Yamato court were kept; a raised platform, esp. a board on the top of pillars where offerings were made to Gods; 'a horse-saddle'] |
| **kadaŋ** (Z761); SAS<br>'kinsman, relative' | **kara**<br>'kinsman, (blood) relative/ relationship' |
| **lĕsuŋ** (Z1016); KAL TAG AKL PAL MOL KAG BLA SAB ACE IND SUN BAL<br>'rice mortar, pounder' | **usu**<br>'mortar for grain, rice cakes and making purified alcohol, and also used to accompany incantation for fertility' |
| **tapih** (Z1948)<br>'(waist) cloth' | **taFe$_2$**<br>'fibre from bark of trees like the Paper Mulberry; cloth made from such fibre; generic term for all types of cloth' |
| **gilaŋ-gilaŋ** (Z525); AKL BUR CEB MAL(?)<br>'shine, glitter' | **kiragira+si** (also **kirakira+si**)<br>O: 'well made-up and beautiful' (of one's appearance) |
| **wakul** (Z2176); TAG MOL(?) SAB(?) KON (?)<br>'a type of basket' | **Fako**<br>O: 'a container with a lid, in which one keeps s.t. important or secret. In ancient times, people believed that it could contain a soul, or a power to bring people fortune'. |

## 5.1 PRIMARY DATA

**5.1.1 Phonological correspondences.** The claimed high level of phonological agreement in both the primary and secondary data rests, of course, on identity/similarity in the inferred phonological content of the orthographic symbols, which themselves are already second- or possibly even third-order abstractions. (Recall the original data is in Javanese script, and hiragana or kanji.) The actual degree of phonological agreement between the OJAV and OJAP words therefore depends ultimately on the methods of reconstruction used to arrive at such orthographic representations, and the status of the representation—whether it is relatively more phonemic or more phonetic. Although these things are not made explicit in the texts, we have assumed that the phonological content of the symbols is relatively uncontroversial. Thus, for example, we assume it is unproblematic to say that there is good agreement in the initial consonants of OJAV *sawak* and OJAP *sawak-*, because we can assume that both OJAV and OJAP *s* can be shown to represent a

voiceless alveolar fricative phoneme segment of some kind. Likewise *a* can be taken to represent a low vowel of some kind.

The generally high level of phonological correspondence in the primary data is obvious, and hardly requires comment. In many cases, as with *sawak/sawak-* and *matur/matur-*, the two forms are identical. Nevertheless, because not all correspondences are identity correspondences, it is important to specify the details, which are given in table 5 and discussed below (consonants are dealt with first, then vowels). We assume the OJAV segments as basic and indicate what they correspond to in OJAP. This is required by the directionality of the correspondences, because it is clear that the OJAP data can be predicted from the OJAV, but not *vice versa*. Phonological environments are also specified in terms of OJAV, but it is often necessary to state grammatical environments in OJAP. We do not expect borrowings to always be uniform, because, inter alia, in some cases differences are due to borrowing from different dialects or sociolinguistic variants (Campbell 1996:99), or to subsequent change. Nevertheless, it is necessary to point out the apparent irregularities, and this is also done below.

**Stops (p t k, b d g, ḍ).** The voiceless OJAV *k* and *t* show identity correspondences with OJAP *k* and *t*. Of their voiced counterparts, OJAV *g* corresponds to OJAP *k* word-initially, and also before the verbal *-i* suffix (*dudu**g**/tudu**k**i*). Otherwise, there is an identity correspondence word-internally. OJAV *d* corresponds to OJAP *t* word-initially. Word-medially, OJAV *d* corresponds to *r*. OJAV *b* corresponds to a voiceless OJAP *F* word-initially. OJAV *p* corresponds to OJAP *F* word-initially and word-internally. Before the OJAP verbal suffix *-i*, OJAV *p* corresponds to OJAP *m* (*tutu**p**/tutu**m**i*.) The OJAV retroflex *ḍ* corresponds, not surprisingly, to OJAP *r*.

Two clear generalizations from the above are that OJAV voiced stops correspond to OJAP voiceless segments word-initially, and that occurrence before the OJAP verbal suffix *-i* can evince changes in voicing ($g > k$), or nasality ($p > m$). The former is predicted from OJAP phonotactics; the latter from OJAP morphology (there is no conjugational form with *p*).

**Nasals (m n ŋ).** All occurrences of OJAV *ŋ* are syllable-final in our data, and in this position, the OJAV velar nasal corresponds to zero in OJAP. The putative WMP-OJAP pair *bu**ŋ**a/pa**n**a* 'flower' (Shibatani 1990:105), which was too short to make it into our final list, suggests a differential with respect to position in syllable-structure, with syllable-initial *ŋ* corresponding to OJAP *n*. The OJAV bilabial and alveolar nasals *m* and *n* are retained in all nonsyllable-final occurrences. The final *m* in *parĕm* is lost; there are no examples of final OJAV *n*.

**Liquids (r l).** OJAV *r* corresponds to OJAP *r* medially and finally. OJAV *l* appears to be lost word-initially (*l**ĕ**suŋ/usu*). Our list of culled shorter words also shows loss of word-initial *l*, as in *l**u**we/uwe* 'hungry'. OJAV *l* corresponds to OJAP *r* medially (*gi**l**aŋ-/ki**r**a*). In final position, it is lost (*waku**l**/Fako*).

**Fricatives (h, s).** OJAV *s* corresponds to OJAP *s* in all cases. Occurrence of OJAV *h* is restricted to word-final. Before the OJAP verbal suffix *-i,* it corresponds to OJAP *r* (*kuku**h**/kuku**r**i*).

**Glide (w).** OJAV *w* corresponds unpredictably to either OJAP *w* or *F* as in *sa**w**ak/ sa**w**ak-*, but *sa**w**ah/saFa*.

**Vowels. i:** OJAV *i* corresponds mostly (2/3 cases) to OJAP *i*. There is one unpredictable correspondence with $e_2$ (*tap**i**h/taF**e**₂*).

**e:** Interestingly, there are no examples in our data with OJAV *e*. It is not clear why this is so. It seems unlikely to be due to sampling error, given the overall large number of vowels used.

**a:** OJAV *a* corresponds to OJAP *a* in all word-internal cases. The one occurrence of OJAV word-initial *a (**a**namar/namar-)* seems to have been lost, though possibly it was not borrowed, as many Javanese verbs have the nasalized form without the *a-* prefix.

**o:** OJAV *o* corresponds fairly equally to OJAP *o* (1/4 cases), $o_2$ (2/4), and *u* (1/4).

**u:** OJAV *u* corresponds mostly (16/21 cases) to OJAP *u.* The remaining four correspondences are with OJAP *o,* three of which are after OJAV *w* (e.g., *wuwuh/ wowor-),* and are therefore explicable by the OJAP constraint on *wu.*

**ĕ:** OJAV *ĕ* corresponds mostly (3/5 cases) to OJAP *a*. One of the remaining three correspondences, before a bilabial, is with OJAP *e (parĕm/Far**e**)*, and the last is unpredictably with OJAP *u (lĕsuŋ/**u**su).* The situation is complicated by many instances of Javanese doublets with *ĕ* replacing *a* or *u.*

The same correspondences as listed above are also found in further items listed in Kumar (1996:531–40), which are not included here because of their shorter length.

**5.1.2 Morphology.** One obvious morphological fact in the primary list data is the relatively high number of verbs borrowed: 17, compared to ten nouns and one adjective. Verbs are also the largest category in the secondary list in table 7. It is necessary to ask whether this is a methodological artefact: perhaps the majority of OJAP lexical items of the required segmental length are verbs. Otherwise it is of interest because, of course, it is usually claimed that verbs are rarely borrowed**.**

**5.1.3 Semantic correspondences.** It is clear that the OJAV and OJAP data are closely related semantically. For example, out of 27 pairs, 14 have the same word in the gloss, and four have synonyms, such as 'cry' for 'call out' in the pair *sawak /sawak-*, 'gather up' for 'hold' in the pair *kukup/kukum-*, or 'bloom' for 'flourish' in the pair *sĕkar/sakar-*. There are also examples of hyponymy, as in *duduk/turuki* ('weapon' and 'sword') and *wakul/Fako* ('basket' and 'container').

Ideally, in order to demonstrate the degree of semantic closeness between the OJAV and OJAP data, the semantic relationships should be characterizable in the same systematic and quantifiable way as the phonological correspondences. This is not a problem where identity correspondences are concerned, of course, but in cases of nonidentity, the establishment of a systematic basis for quantifying

semantic distance is elusive. How does one systematically represent the common and the divergent elements in those items in which the glosses are not identical?

One way to do this would be to state the meaning of each lexeme using the Natural Semantic Metalanguage (NSM) method devised by Wierzbicka (1996). This would show precisely what semantic components are shared between the putative cognates and enable quantified analysis. We give here an illustration of how this works for some pairs.

(1) OJAV *pupur,* 'blow, battle, cock fight with natural spurs' and OJAP *FoFuri,* O: 'to cut/tear (the body of a bird/an animal) into pieces; to massacre'; JKD: 'to cut up; to rip to pieces; to kill s.o. with sword'.

In this case, both lexemes cover two meanings, expressed in semantic primitives as follows:

a. some people make some other people die
b. people make something [a bird] die

(2) OJAV *mawuwus,* 'to utter words', and OJAP *mawosi,* O: 'to reveal (to a god, emperor), to report, to tell (to superior); to ask a favor'; JKD: 'to say (s.t. to s.o. of high rank)'; (aux) nonsubject exalting.

Both lexemes include the semantic primitive 'to say', but the second has the additional element: ' because you want someone else to know something'.

NSM representation also highlights the amount of semantic agreement involved in lexemes with highly specific meaning. Consider the pair OJAV *guḍaŋ* **'**building to keep grain, treasure, furniture …' and OJAP *kura* 'storehouse for grain, treasure, furniture; location on a higher than ground level for putting things on …'. These share at least the following components:

> people do something to some thing [material]. because people want many good things to be in one place; people think that nothing bad can happen to these things in this place.

**5.1.4 Semantic fields.** A further factor of importance in the semantics of the data is that some of the items belong to two main semantic fields. One relates to land clearing and rice cultivation and preparation; the other to divinity and ceremony. Exemplifying the former are word pairs referring to an *open or cleared field,* a *rice field,* a *rice mortar,* and (in the second list) a *boundary of a rice field.* Exemplifying the latter are words that include reference to *presenting something to a person of high rank or god,* and to *making offerings at a shrine; reporting to emperors and gods; the three* kura *in which the regalia of the Yamato court were kept; the approach of a god; the fences of shrines, temples, and imperial courts; use in Shinto ceremonies; fertility incantations; a container for the soul.*

These groupings are significant because they provide two additional levels of coincidence that have to be explained. The first coincidence is that many of the words that show linguistic correspondences between Old Japanese and Old Javanese do not constitute a random set, but are thematically related. The second coincidence is, of course, that these two themes correspond very well to two areas

of demonstrated early cultural interaction between Japan and Indonesia noted above, namely kingship/divinity and rice.

Perhaps the greatest significance of the semantics of the data, however, has to do with the fact that they show striking agreement with the most often quoted innovations of the Yayoi period (see Saitoo 1983:321–322). Thus our data contain word pairs referring to rice cultivation (e.g., *sawah/saFa*: 'rice paddy'); introduction of metallurgy (*duduk/turuki:* 'sword, weapon'); weaving of cloth (*tapih/taFe*$_2$: 'cloth'); development of communal storehouses (*guḍaŋ/kura:* 'storehouse'); and introduction of new rituals and beliefs. The received view is that these innovations entered Japan from China and Korea: "... the growing stream of mongoloid settlers from Korea and China ... brought with them important new knowledge and technology. They introduced rice, and the technique of growing it in flooded paddies. They knew how to make bronze and smelt iron for tools and weapons. The women could weave cloth. These invaders and their ideas were absorbed without too much upheaval" (Thorne and Raymond 1989:191–192). However, our data seem to point rather convincingly to a southern source for these Yayoi innovations.

**5.2 SECONDARY DATA.** Table 7 presents the secondary data. These are pairs that for various reasons do not show quite as good an agreement in form or meaning or both.

**5.2.1 Phonological correspondences in secondary list.** In addition to the sound correspondences listed in 5.1.1 for the primary list, the secondary-list data contain the following additional correspondences that make sense in terms of phonotactic mismatch. OJAV *r* corresponds to OJAP *t* word-initially (***r****awuh/****t****aFur-*), and is lost word-finally in nonverbal forms (*wuki****r****/woka*), as is also OJAV *s* (*kiki****s****/kaki*), and *t* (*nura****t****/nuri*). Word-initial *u* acquires a palatal onglide (***u****ndur/****yu****dur-*). A second way of resolving the OJAV *wu* sequence is by dropping the *w* (***w****utuh/utu*). Syllable-coda *n* is either dropped or resyllabified (*tu****n****tu****n****/to*$_2$*to*$_2$*no*$_2$*F-*). There is also the intervocalic identity correspondence OJAV b : OJAP b (*baru****b****uh/Fo*$_2$*ro*$_2$*****b****-*).

As with the primary list, there are several problematic cases with irregular correspondences. The apparent loss of syllable-initial *ṭ* in *(pasṭi/besi)* is irregular, because the syllable-final *s* would be expected to be lost instead. Perhaps also a rhotic would be expected to occur because of the retroflex consonant.

Of particular interest is the *i* : *a* correspondence (*wukir/woka*), which occurs three times. The presence of adjacent velar consonants makes the correspondence implausible, because the change would then represent a height dissimilation. (It is unlikely that it is a backness assimilation, otherwise an *i* : *u* correspondence would be expected.) It is more possible that some kind of height dissimilation is being triggered by a sequence of two high vowels. In this connection, the pair *isi/asi* 'to fill' is worth noting (culled on grounds of segmental length).

Examples of vowel correspondences that do not occur in the primary data are: OJAV *a* corresponding to *e*, *u*, and *o*$_2$; and OJAV *u* corresponding to *o*$_2$ and *i*. The latter possibly involves reanalysis, with the *i* as verbal suffix.

## TABLE 7. SECONDARY LIST OF BORROWINGS

| | |
|---|---|
| **rawuh** (Z1525)<br>'to fall, come down; to arrive, visit' | **taFur-**<br>'to fall, come down, collapse; [to die]' |
| **(a)nusup** (Z1875)<br>'to penetrate or enter into a place where one is unseen/hidden' | **nusum-**<br>1. 'to conduct/carry out/do something in secret'<br>2. 'to steal [something]; to make [something] one's own in secret' |
| **bataŋ** (Z224)<br>'guess, interpretation, probability' | **Fata**<br>O: 'by any chance, by any stretch of the imagination'<br>JKD 1. 'perhaps [...on one hand... on the other hand], used to make assumption'; 2. 'perhaps' (used like a conjunctive: second clause is "deduced" from first clause) |
| **undur** (Z2121); MIN, IND, SUN, ACE<br>'to retreat, retire' | **yudur-**<br>O: 'to give way, hand over'<br>JKD: 1. 'to give [somebody something one owns]'; 2. 'to move away [from something]; to step aside in order to pay due respect to somebody' |
| **wukir** (Z2322); AKL PAL MOL KAG PAI IBA PGS TAG BKL HAN HLG TBWK BTK CEB NGD MAL ACE TB REJ PS SAN RAT MAK<br>'mountain' | **woka**<br>O: 'mountain'; JKD: 'hill, low mountain' |
| **kikis** (Z866)<br>'border, boundary fence' | **kaki**<br>O: 'fence, boundary (of shrine, temple, imperial court, rice-field)'<br>JKD: 'a hedge, a fence, any form of barrier which makes a boundary/partition' |
| **piriŋ** (Z1367); SUN, MAD, WOL<br>'a plate' | **Fira+ka**<br>'a plate' [ka: 'vessel or container'] |
| **wutuh** (Z2338); (? IBA KB TB DPB SAS)<br>'whole, intact' | **utu**<br>'utterly, totally, entirely, completely; empty, hollow' |
| **sasar** (Z1705)<br>'to go astray, to stray from the right way' | **sasuraF-**<br>(V4) 'to wander, roam, go out as one pleases'<br>(V2) 'to lose one's way'<br>[The morphological status of the OJAP -aF- is not clear.] |
| **tuntun** (Z2066); MOL SAB BAL<br>'to lead, command' | **to$_2$to$_2$no$_2$F-**<br>'to lead, command'<br>[The morphological status of the OJAP -*o$_2$F*- is not clear; note, however, its similarity to the previous example.] |
| **pasṭi** (Z1314); IND SUN<br>'fixed, decided' | **F/be$_2$si**<br>O: modal auxiliary indicating certainty, inevitability<br>JKD 1. (judging from the situation) 'must be, should be'; 2. 'ought to; might at least'; 3. 'to be able to'; 4. 'had better, must' |
| **tutug** (Z2083); PAI PAL MOL SAB BAT MAG RTI<br>'to strike or hit' | **tutuk-**<br>'to poke, to tap, to knock' |
| **tětěg** (Z1997); PAI BON KAN IFG IFGBT ILK PGS MAR DLT<br>'signal block' | **tatak-**<br>O: 'to strike in succession'; JKD: 'to hit, to knock, to tap' |
| **barubuh** (Z219)<br>'to make a thundering noise especially when crashing down' | **Fo$_2$ro$_2$b-**<br>'to fall into pieces and disappear; to be destroyed' |
| **nurat** [derivational form of **surat,** (Z1864)] ISN PAL MOL KAG BLA SAB BAT MIN RTI BUR<br>'to draw or write' | **nur-**<br>'to paint; [to become wet, damp, or stained (by blood, etc.)]' |

**5.2.2 Semantics of secondary list.** Out of 15 items, five have the same word in the gloss, and five have synonyms (e.g., *sasar,* 'to go astray' and *sasuraF-,* 'to lose one's way'). Apart from *kikis* 'fence' and *piriŋ* 'plate', the other items fall mostly into the category of basic vocabulary. There are several items that, despite their phonological similarity, have been relegated to the secondary list because they may be onomatopoeic: *barubuh/Fo₂ro₂bi, tětuk/tatak-, tutug/tutuk-.*

**5.3 PUTATIVE COGNATES OF THE OLD JAVANESE WORDS IN OTHER AUSTRONESIAN LANGUAGES.** In order to test whether the hypothesized influence can be narrowed down to a particular part of the Austronesian world, the distribution of the items in the list was determined by using Tryon's (1995) *Comparative Austronesian Dictionary* and Blust's (1995) on-line *Austronesian Comparative Dictionary*, and has been indicated in tables 6 and 7.

It was found that half or more of the items in both lists (17 out of 26 in the first list and 7 out of 15 in the second) occur only in Old and Modern Javanese. Of those items that do have cognates, the overwhelming majority, in both lists, are found in WMP, mainly Indonesian, languages. Furthermore, the Old Japanese tokens are phonetically closer to the Javanese and WMP forms than to the handful of non-WMP cognates. Indeed, in some instances the Old Japanese reflects a specifically Javanese innovation such as the replacement of PAN *b by *w* as in the case of *wuwuh*. For this item, Blust reconstructs PMP *bubuq, 'increase, growth' (this item is only attested in WMP languages and no PAN form is reconstructed). It is also true that the Old Japanese tokens are phonetically closer to the Javanese and WMP forms than to the reconstructed Proto-Austronesian forms. Compare, for example, OJAP *tatak-* with OJAV *tětěg* 'signal block', for which Blust reconstructs PAN *CegCeg and PMP *tegteg.

The above phenomena point rather clearly to a *localized Indonesian*, not general Austronesian influence. Indeed, this linguistic distribution evidence fits very neatly with the evidence of mitochondrial DNA. Kumar's (1998) study of d-loop sequences showed that Japanese and Indonesians share lineages not present in other Asian populations sampled, such as Mongolians, Chinese, and Taiwanese. Because these lineages are not shared with Taiwan, we must conclude that the genetic link between the Japanese and Indonesians does not go back to the original Austronesian migration, but dates from a later period when the Austronesians had already migrated to Indonesia.

**6. PROBABILISTIC EVALUATION OF OJAV/OJAP SIMILARITIES.** In the sections above, we have listed OJAV and OJAP material that shows considerable agreement in both phonology and semantics. As already mentioned, it is generally accepted that "... any pair of languages can be expected to exhibit a nonnegligible number of fortuitous similarities" (Ringe 1992:1). According to Trask (1996:221), this is because the large numbers of meanings in a language have to be encoded with a limited number of phonological segments, and the fortuitous similarities are therefore the consequence of the structural property of double articulation. Many examples of such fortuitous similarities are quoted in the literature: *man* in Korean

and English; *bad* in Persian and English; 'eye' in Modern Greek and Malay *(mati/mata)* (Hock 1991:557). We add a less well-known example: the word for 'dog' in the Australian language Mbabaram (Dixon 1983:107). It is very likely that Mbabaram has not borrowed this word from English, because Mbabaram *dog* derives regularly, if somewhat complicatedly, from Proto-Australian *gudaga. (The low central vowel between the *d* and *g* is changed to *o* by segments in the preceding syllable [raised by the high vowel and backed by the velar consonant], which then disappears, along with the final vowel [Dixon 1991:362–363].)

Of course, how easy it actually is to find fortuitous similarities depends on the similarity criteria in both semantics and phonology. With strict criteria—identity correspondences, for example—it is perhaps not so easy to find uncontentious examples of a *nonnegligible number* of such fortuitous similarities. For example, Trask's (1996:220) list of 20 pairs of words from Ancient Greek and Hawaiian is designed to demonstrate how easy it is to find sound-meaning correspondences due to chance. However, this is at the expense of some phonological liberties taken with both the Greek and with the Hawaiian that would have made the examples less similar: Hawaiian glottal stop is omitted; long and short vowels in both Hawaiian and Greek are ignored.[3] (Because both languages have long and short vowels, this is a serious omission.) Moreover, putative segmental correspondences are often not between contiguous items. This means that the length of the words is deceptive, in that the amount of agreement is not actually greater than Mbabaram/English *dog*. Most significantly, it also turns out that at least one pair—the best one, *meli-meli* 'honey', the only example with identity correspondences in all four segments—is in fact a borrowing after all (Clark 1998; Campbell 1999).

**6.1 PREVIOUS APPROACHES.** Most of the discussion on fortuitous similarities occurs in the context of determining genetic relatedness (e.g., Campbell 1998:316–317; Hock 1988:556–580; Lass 1997:105, 123–130; Ringe 1992, 1998, 1999). Interestingly, studies of borrowing such as Adelaar (1994), O'Grady and Tryon (1990), and Evans (1992) do not address the problem of nonnegligible fortuitous similarities. They appear to assume that borrowing is indicated by (unquestionably present but unquantified) resemblances in form and meaning. The authors wholeheartedly concur with Ringe's "nonnegotiable" condition for scientific linguistics that "any claim that a similarity between two human languages is significant must be supported by a demonstration that it could not be the result of sheer chance; ..." (1999:213). In this section, then, we address the question of whether it can be shown that the similarities in form and meaning demonstrated above are due to nonrandom effects.

Probably the best known attempts to address problems of this type are Nichols (1996) and Ringe (1992, 1998, 1999). The former is concerned with characterizing the classical comparativist approach. The latter aims to provide objective tests of the

3. *Noonoo* is *no'ono'o*; *manao* is *mana'o*. *Lahui* is *laahui; manthano* is *manthanoo; men* is *meen; kion* is *kioon; hikano* is *hikaanoo*.

validity of claims of remote relationship from long range comparativists. Below we look briefly at those parts of these works that are relevant for the present purposes.

**6.1.1 Individual identifying evidence.** An approach that appears to deal directly with the probability of combinations of sounds was proposed by Nichols (1993:24–25, 1996:48ff.), who characterizes the kind of evidence, called *individual-identifying*, that was accepted as probative of genetic relationship by classical comparativists. Saying that a word is individual-identifying means that "all the languages having it have acquired it, ultimately, from a single source" (Nichols 1996:50).

The implication is this. If the probability associated with the sounds in a given word is low relative to an estimate of the number of languages in the world, the same conjunction of sounds in different languages with similar sound inventories must be due to some factor other than chance.[4] To be confident of the results, a threshold of less than $p = .00001$ (i.e., one in a hundred thousand) is proposed, on an estimate of "about a few thousand languages" for the world, and a statistical confidence limit of 0.05 (Nichols 1996:49).

The inventory and phonotactics of OJAV and OJAP phonemes have been described in sections 4.1 and 4.2. According to this, an estimate of the average probability associated with an OJAV CVCVC root like *sosok* would be $5 * 10^{-6}$, or 5 in a million,[5] and that for the corresponding OJAP root would be either $2 * 10^{-5}$ (2 in 100,000), or $7 * 10^{-6}$ (7 in a million), depending on whether five or eight vowels are recognized for OJAP.[6] One of these, the estimate for OJAV, clearly exceeds Nichols's "one in a hundred thousand" threshold by at least an order of magnitude; the other misses it by a whisker. It is clear, then, that by Nichols's criteria the OJAP/OJAV evidence for one word, if not totally overwhelmingly individual-identifying, is exceedingly near, given the exiguous difference between one in a hundred thousand and two in a hundred thousand.

Evidence can turn out to be wrong, however. Blood group evidence might point toward the suspect as the criminal, but it can be reversed by much stronger evidence from, say, a DNA analysis. Nichols (1996:49) is careful to point out that the individual-identifying notion is *diagnostic*, not proof, and carefully distinguishes it from approaches like Ringe (1992).[7] She also adds in the context of establishing genetic relatedness (1993:24) that not just one word but many are needed.

---

4. Nichols asks, for example: what are the chances of finding a word meaning *widow* with an initial *w* or *v* ? Assuming an inventory of 20 consonants, both *w* and *v* have a probability of 0.05, so the probability of either *w* or *v* is their sum, or 0.10. That is, we would expect to find *w* or *v* in the word for *widow* in 1 out of 10 languages purely by chance (or, with an estimate of 4000 languages in the world, 400 of the world's present languages).

5. This figure is based on an assumption of probability of 0.04 (one in 25) for C1, 0.067 (one in 15) for C2 and C3, and 0.167 (one in six) for the vowels. $0.04 * (0.067)^2 * (0.167)^2 = 0.000005008$.

6. This figure is based on assumptions of 0.077 (one in 13) for the consonants, and 0.2 (one in five) for the vowels. $(0.077)^3 * (0.2)^2 = 0.00001826$.

7. This distinction does not seem to have been appreciated by Hoenigswald (1999:167).

**6.1.2 Whole root comparison.** Ringe (1999) bases on Nichols's insight and contains an approach that can be directly applied to the OJAP/OJAV data. He quantifies the ease with which chance matches between languages can be found in roots with a CVC phonotactic structure. He demonstrates, for example, that with data from 25 unrelated languages possessing a typically sized phonemic inventory, an acceptable semantic and phonological match will in the long run occur by chance with a probability greater than one in four (Ringe 1999:218). This means that in the long run, a little more than one root out of every four examined will show a match somewhere in the 25 languages.

Although he does point out that the chances of demonstrating significant effects are better with a root longer or more complicated than CVC, Ringe is primarily concerned with the *number* of languages necessary for a given effect. In this paper, however, we are concerned with probabilities within a *single pair* of languages, and want to know what the probability is that we will find a match by chance in CVCVC forms if we look at just two languages. This is very similar to the "shared birthdays" problem. Our case is analogous to asking what the probability is, given one person's birthday, of finding the same birthday in a second person. What is the probability, given our OJAV forms, of finding the same forms in OJAP, or given our OJAP forms, of finding the same forms in OJAV?

Before we estimate the probability of the observed correspondences in the OJAP/OJAV data, however, it is necessary to address the problem of what constitutes "the same birthday," that is, a phonological-semantic match. The reason why this is important is that the degree of confidence in the nonrandomness of the patterns is ultimately a function of the number of matches: the higher the number, the greater the likelihood that the patterns are significant.

**6.2 WHAT IS A MATCH?** In the probabilistic evaluation of data like ours, the most important thing is to insist on minimal latitude, both phonologically and semantically, between putative matches. This is because the probability of finding matches purely by chance increases dramatically as a function of the degree of semantic and phonological latitude in the criteria for a match (Ringe 1992:64–70, 1999:231–233), and therefore it gets harder, with increasingly lax correspondences, to demonstrate greater-than-chance similarity.

Semantically, for example, given a probability *p* for a particular phonological shape, if a match is allowed between two items with this shape that are even closely related semantically, like bamboo spear/sword, or rice paddy/swamp, this is equivalent to increasing the number of possible allowable matchings from one to four ('paddy': 'swamp'; 'swamp': 'paddy'; 'swamp': 'swamp'; and 'paddy': 'paddy'). Consequently the expected incidence of the forms will increase by a factor of four, and the probability must then become 4*p*.

Phonologically, too, if a disjunctive match is allowed between, say *p* and either *p* or *b,* in a system where consonants have a probability of occurrence of .05, the probability of observing either a *p:p* or a *p:b* match will be .1, not .05.

Ideally, then, it is clear that acceptable matches should exemplify identity correspondences in sound and meaning. Because semantic matches can exert greater influence on the outcome, the criteria for defining them are discussed first.

**6.2.1 Semantic matches.** Pace Ringe (1999:216), it is not entirely unproblematic to define criteria for a semantic match. Semantically, the nature of our observation data imposes certain difficulties in assuring the strictest identity correspondences. For one thing, because a certain amount of semantic change is to be expected, we would not even expect to find a large number of semantic equivalents in our data, however equivalence is defined.

The strictest requirement for a semantic match between two forms should presumably be that "they mean the same thing." Semantic identity is established on distributional equivalence and demonstration of reduction to the same semantic invariant (Wierzbicka 1996:242–244). Such a requirement is not adhered to in practice, however, because either it would simply take too long to do a proper semantic analysis of the forms in question, or because the available material is inadequate (an absence of native speakers for acceptability judgments, for example). Thus Ringe's (1998, 1999) criteria for semantic match are couched in terms of translation and reference: semantically acceptable forms are those that "translate each other, in the sense that they are used to refer to the same narrowly defined real-world entities" (1999:216). Although we have, of course, no way of knowing whether any of our OJav and OJap forms were translation equivalents, it is reasonable to assume the existence of the same English gloss in both OJav and OJap dictionary entries, (or obvious synonymous expressions like '*building for keeping* grain, treasure, furniture', versus a '*storehouse for* grain, treasure, furniture') would entitle us to infer that they could have been translational equivalents thus construed.

In practice, however, Ringe (1992) uses Hawaiian, Turkish, and Navajo dictionaries to furnish equivalents for the English terms in the Swadesh lists. Therefore his translation criterion reduces further to dictionary equivalence: two items match semantically if their dictionary glosses appear to refer to the same "narrowly defined" thing. While this seems both a sensible and practical option, and is the one we have in fact used in this paper, it is important to be aware of its limitations (not the least of which is that dictionary glosses do not necessarily represent the linguistic meaning).

As can be appreciated from some of the glosses in tables 6 and 7, it is in the nature of dictionaries to attempt exhaustiveness in the specification of meanings for a given phonological form. This clearly militates against establishing identities, because two phonologically similar forms would have to agree in all aspects of sense, reference, and polysemy of their entries for them to be equivalent. (The specificity of dictionary entries also reflects the knowledge and interpretation of lexicographers and the material available to them.) Take the pair *guḍaŋ/kura,* the glosses for which share the highly specific 'storehouse for grain, treasure, furniture'. The OJap entry, however, includes also 'a raised platform', which means that the entries are not totally equivalent (although in some cultures, especially Austronesian ones, a

raised platform would be the usual form for a storehouse). Or consider the pair *matur/matur-,* which share again a highly specific: 'to present to a person of higher rank'. The OJAV and OJAP glosses differ essentially in two points: the OJAP meaning does not include 'to tell' (to a person of high rank), and the OJAV meaning does not include reference to God (in addition to a person of high rank).

It seems to us that, although semantic identity remains ultimately indeterminate with this approach, it is nonsensical that such pairs as *matur/matur-* or *guḍaŋ /kura* should be excluded from candidacy because their dictionary glosses are not totally equivalent. We thus adopt Ringe's dictionary equivalence criterion and take to be a semantic match cases where there is a substantial overlap in the dictionary glosses.

It is possible to find eight pairs in our first data set for which we think that the degree of semantic closeness will not be controversial (note that closely related items like 'paddy/swamp' or 'bamboo spear/sword' are not included). These words are given again, with glosses, in table 8, and it is these pairs that we shall use in subsequent probabilistic evaluation. Constraining the semantics in this way means that the probability of the phonological shape will not be increased.

**6.2.2 Phonological matches.** As mentioned above, the maximally small degree of latitude required for probabilistic evaluation needs to be guaranteed not only by exact semantic match but also by "rigorous sound correspondences." In the following discussion, although we are solely concerned with evaluating the phonological correspondences in the semantically "good" data in table 8, it should be remembered that these correspondences obtain not just for the seven pairs in table 8, but hold in general for *all* our primary data (the correspondences for which have been summarized in table 5). The data in our primary list, it will be seen, also enable us to evaluate the correspondences in the semantically good list.

As criterion for phonological matching, Ringe (1999:216) advocates "match[ing] the phonemes of the two languages in some way that is guided, but not narrowly

**TABLE 8. OJAV/OJAP WORD PAIRS WITH NEAR IDENTITY SEMANTIC/ PHONOLOGICAL CORRESPONDENCES (OJAV ON TOP)**

| | | |
|---|---|---|
| 1. | **'building to keep grain, treasure, furniture'**<br>**'storehouse for grain, treasure, furniture'** | guḍaŋ<br>kura |
| 2. | 'to **present**, offer, tell or report **to a person of higher rank'**<br>'to give or **present** s.t. **to a person of high rank/**God; to offer prayers'; ... | matur<br>matur- |
| 3. | **'to pour** out'<br>**'to pour,** sprinkle, splash water; to rinse, wash off; to gush out' | sosok<br>$so_2so_2k$- |
| 4. | **'to hide,** to disguise'<br>**'to hide'** | (a)namar<br>namar- |
| 5. | '**strips of leaf** hung on a string'<br>'**torn frayed or withered leaf'** | wawar<br>wawara-ba |
| 6. | '**to clear** (of weather, **after rain)'**<br>**'to clear** up, **after rain**, cloud, fog, mist, snow' | parěm<br>Fare |
| 7. | **'rice mortar**, pounder'<br>**'mortar** for grain, **rice** cakes and purified alcohol' ... | lěsuŋ<br>usu |
| 8. | **'kinsman, relative'**<br>**'kinsman**, (blood) **relative**, relationship' | kadang<br>kara |

restricted, by their phonetics." This is designed to allow not only identity correspondences like OJAV *k* : OJAP *k,* but also to sanction desirable matchings of the complementary distribution type actually found in our data between OJAV *p* and OJAP *F.* (Tables 1 and 4 show that OJAV lacks *F* and OJAP *p.*)

Application of this idea is not problematic, but some discussion is required of typical complexities where disparities obviously result from between-language differences in phonotactics and phonemic inventory. For example, OJAP phonotactic restrictions are responsible for the more complicated, but regular mapping of OJAV *m, n, ŋ* to both OJAP Ø word-finally and *m, n,* – elsewhere.

Such cases are totally analogous to those between genetically related languages caused by sound change discussed in Ringe (1999:223), and must be considered as phonological matches because the correspondences are rigorous. Thus, for example, *parěm : Fare* is considered a phonological match, as is *kadaŋ : kara.* Similarly, OJAP *F* and *b* are a match for OJAV *b,* and OJAP *t* and *r* are a match for OJAV *d,* because they occur in complementary distribution. The lack of *l* in OJAP is responsible for the mapping of OJAV *l* to OJAP *r.*

Finally, we have to point out cases where the mapping is unpredictable, but where phonetically similar segments are still involved. For example, although in the data in table 8 OJAV *o* corresponds to OJAP $o_2$, in the larger data set, OJAV *o* corresponds in general unpredictably to the three phonetically similar OJAP vowels *o,* $o_2$, and *u.* Another example is OJAV *ě,* which corresponds to the three OJAP vowels *a, e,* and *u,* with only the correspondence with the last being unpredictable. These cases, too, are considered matches, so that, for example, a pair like *sosok :* $so_2so_2k$- is not rejected on the grounds of unpredictability of *o* : $o_2$.

The probability of the matches of the items in table 8 is assessed using a method similar to Ringe (1999). This involves first estimating the probability of a match in phonological shape for two words with the same meaning, for example OJAV sosok : OJAP $so_2so_2k$- {pour}. Then a binomial distribution is consulted to find what the probability is of observing a match with this probability by chance in a single pair of unrelated languages. The null hypothesis is that this probability is due to chance, and that there has therefore been no interaction. A conventional significance level of 0.05 is assumed.

**6.3 PROBABILITY OF PHONOLOGICAL MATCH.** It is possible to estimate the probability of a match in phonological shape for the same meaning in three ways, of increasing reliability. Firstly, one might use Ringe's (1999:220) approximation of .00054 for CVCVC roots, which is based on the most "lax criteria." It would also be possible to estimate probabilities from the OJAV and OJAP quasi-phonemic inventories and phonotactic restrictions in section 4 above. For example, given the OJAV initial consonant inventory size of 25, the probability associated with a single word-initial consonant would be 0.04. Or given the phonotactic restrictions on the 13 OJAP consonants (no voiced obstruents; no *r*), the probability associated with word-initial position would be 0.125 (one in eight). It would also be necessary to allow for the restrictions on morpheme-internal voiced obstruents in OJAP.

However, neither of these approaches can take into account the actual lexical incidence of segments in our data, and this is potentially dangerous. Suppose, for example, that an OJAV consonantal segment was so restricted in the OJAV dictionary that one would expect it to be sampled only very rarely by chance; or that an OJAP consonantal segment were particularly common. Using approximations from Ringe or the phonemic inventories/phonotactics would in both cases result in the probability estimate for a phonological shape being less than it really is in our data, and this would bias the results towards significance. We therefore eschewed both these approaches in favor of working with estimates of the actual incidence of consonantal segments in the OJAV and OJAP dictionaries.

In order to estimate the actual incidence of consonantal segments, a program was written to generate random numbers within a range corresponding to the number of pages in each dictionary. Pages corresponding to these random numbers were then consulted, the first CVCV(C) form identified, and the consonantal segments entered separately for $C_1$, $C_2$, and $C_3$ in a spread sheet. If there was no CVCV(C) form on the randomly selected page, the next page was used. If the page had already been consulted and its first CVCV(C) form entered, the next word was taken. A total of 339 words were sampled for OJAV, and 200 for OJAP. This method thus gave us a profile of the distribution of individual consonantal segments in the OJAP and OJAV dictionaries.

Tables 9 and 10 present the incidence and associated probability for the OJAV/OJAP data. Table 9 shows that our fears concerning differences in the lexical incidence of some OJAV segments are well-founded. For example, word-initially the voiceless stops *p, t, k, s, w* are common, whereas *y, ḍ, c* are rare, and there is no example of the velar nasal or palatal nasal. (There are, to be sure, examples in the dictionary, but not enough to ensure that one or more is sampled at random.) Missing also is the set of prestopped nasals. There is less variability in the incidence of the middle consonant, although *w* is still common. There are, however, some large differences in $C_3$ probability: *-k, -ŋ,* and *-h* are common, accounting for nearly 50% of the occurrences. It is worth pointing out that the overall values do not differ much from those in the literature: 17 $C_1$ consonants gives a probability of 0.056; 20 $C_2$ consonants gives 0.05, and 14 $C_3$ gives 0.07.

The OJAP data in table 10 also show considerable between-segment differences in incidence, depending on position in word. The voiceless obstruents *F, t, k, s* are relatively common in $C_1$ position, no doubt partly because of the phonotactic restrictions on the voiced obstruents (which it is nice to see confirmed in the data). In $C_2$ position, the voiced obstruents *b* and *z* and the glides *w* and *y* are rare. In $C_3$ position, *s* and *r* clearly predominate, accounting for nearly 50% of all occurences. This was due to the high incidence of forms with the *-si* adjectival suffix, and verbal forms in *r* + *i* (e.g., *matur*+*i*).

Table 5 shows that the mapping between OJAV and OJAP vowels is far from haphazard. But because it is not as good as the consonants, and contains some examples of nonpredictability, we decided to adopt Ringe's (1999:217) approach, which allows for "less stringent" requirements on vowel matching between two vowels in a triangular

**TABLE 9. INCIDENCE AND PROBABILITY OF CONSONANTS IN 339 OLD JAVANESE CVCVC WORDS RANDOMLY SELECTED FROM ZOETMULDER (1982)**

| | C1 | | C2 | | C3 | |
|---|---|---|---|---|---|---|
| | COUNT | PROB | COUNT | PROB | COUNT | PROB |
| p | 34 | .1 | 24 | .071 | 8 | .024 |
| t | 39 | .115 | 14 | .041 | 35 | .103 |
| ṭ | — | | 4 | .012 | * | |
| c | 8 | .024 | 11 | .032 | | |
| k | 32 | .094 | 13 | .038 | 55 | .162 |
| b | 20 | .059 | 22 | .065 | 16 | .047 |
| d | 10 | .029 | 9 | .027 | 9 | .027 |
| ḍ | 4 | .012 | 14 | .041 | | |
| j | 14 | .041 | 4 | .012 | | |
| g | 21 | .062 | 6 | .018 | 7 | .021 |
| m | 15 | .044 | 19 | .056 | 11 | .032 |
| n | 8 | .024 | 23 | .068 | 27 | .08 |
| ny | — | | 2 | .006 | | |
| ŋ | — | | 32 | .094 | 48 | .142 |
| s | 45 | .13 | 20 | .059 | 16 | .047 |
| h | 20 | .059 | 21 | .062 | 55 | .162 |
| r | 15 | .044 | 22 | .065 | 23 | .068 |
| l | 16 | .047 | 40 | .118 | 25 | .074 |
| w | 35 | .103 | 34 | .1 | — | |
| y | 3 | .009 | 5 | .015 | 4 | .012 |
| mb | — | | | | | |
| nd | — | | | | | |
| nḍ | — | | | | | |
| nj | — | | | | | |
| ŋg | — | | | | | |
| TOTAL | 339 | .996 | 339 | 1.0 | 339 | 1.001 |

* Shading indicates phonotactically excluded positions.

five-vowel system, and therefore assigns vowels a probability of .4. (Actually, assuming a six-vowel system for OJAV and a five-vowel system for OJAP, the mean of the probabilities should be .35, but it is silly to worry about differences of this magnitude, and we have taken the value that biases the outcome in favor of the null hypothesis.)

With these figures, it is now possible to estimate the probabilities for the semantically "good" data in table 8. These are given in table 11. Thus the probability of the OJAV phonological shape for {pour} *(sosok)* is .13 * .4 * .059 * .4 * .162 = .0002, and the probability of OJAP {pour} *($so_2so_2k$-)* is .14 * .4 * .145 * .4 * .12 = .00039. It can be seen that the probabilities are different for the two languages, but variations in the overall probability due to between-language differences can be accounted for by

**TABLE 10. INCIDENCE AND PROBABILITY OF CONSONANTS IN 200 OLD JAPANESE CVCVC WORDS RANDOMLY SELECTED FROM OHNO, AKIHIRO, AND MAEDA (1991)**

| | C1 | | C2 | | C3 | |
|---|---|---|---|---|---|---|
| | COUNT | PROB | COUNT | PROB | COUNT | PROB |
| F | 36 | .18 | 14 | .07 | 26 | .13 |
| t | 29 | .145 | 22 | .11 | 10 | .05 |
| k | 39 | .195 | 35 | .175 | 24 | .12 |
| b | * | | 5 | .025 | 6 | .03 |
| d | | | 16 | .08 | 3 | .015 |
| g | | | 18 | .09 | 5 | .025 |
| s | 28 | .14 | 29 | .145 | 49 | .245 |
| z | | | 5 | .025 | ? | |
| m | 17 | .085 | 14 | .07 | 22 | .11 |
| n | 24 | .12 | 12 | .06 | 5 | .025 |
| r | | | 18 | .09 | 46 | .23 |
| y | 15 | .075 | 6 | .03 | 3 | .015 |
| w | 12 | .06 | 6 | .03 | 1 | .005 |
| TOTAL | 200 | .995 | 200 | 1.0 | 200 | 1.0 |

* Shading indicates phonotactically excluded positions.

**TABLE 11. PROBABILITIES FOR SEMANTICALLY EXACT MATCHING PAIRS**

| | | p OJAV | p OJAP | p Combined |
|---|---|---|---|---|
| 1. | (a)namar / namar- | 0.00001 | 0.00031 | 0.00016 |
| 2. | matur / matur- | 0.00001 | 0.00034 | 0.00018 |
| 3. | sosok / $so_2so_2k$- | 0.00020 | 0.00039 | 0.00030 |
| 4. | wawar / wawara-ba | 0.00011 | 0.00007 | 0.00009 |
| 5. | parĕm / Fare | 0.00104 | 0.00259 | 0.00182 |
| 6. | guḍaŋ / kura | 0.00012 | 0.00281 | 0.00146 |
| 7. | kadaŋ / kara | 0.00044 | 0.00281 | 0.00162 |
| 8. | lĕsuŋ / usu | 0.00006 | 0.02320 | 0.01163 |

simply taking their mean (Ringe 1992:222), because the probability of finding a phonological match in OJAV for OJAP {pour} is therefore .00039; and the probability of finding a match for OJAV {pour} in OJAP is .0002.

As can be seen, the combined probabilities range, in the four CVCVC words, from 0.00016 (for *sosok* / *$so_2so_2k$-*) to 0.00009 (for *wawar* / *wawara-ba*). Now an event with such a low probability as even the highest value of 0.0003 would only be expected to occur by chance in the long run in one trial comparison between two languages if many words were available for inspection. The binomial distribution can be used to estimate the number of times one would expect to find by chance, with a given vocabulary size, exact phonological/semantic matches of the type listed in table 11. We take for *a-fortiori* reasons the probability of the highest word pair (0.0003) as the probability

of the event, and the approximate number of CVCVC words in the OJAV dictionary (3,050) as the number of trials. The real number of trials will actually be the number of semantically identical CVCVC pairs in the two dictionaries, and because the two dictionaries are far from identical in word content, this will be considerably smaller than 3,050. However, given the problems of establishing semantic identity mentioned above, this is a simpler (and far less time-consuming!) option.

Because the binomial distribution of $m$ matches for an event of probability $p$ in $n$ trials is given by $C(n,m) * p^m * (1-p)^{n-m}$, (where C is the binomial coefficient), the cumulative binomial probability for zero to three matches is 98.57%. This means that the probability of finding four or more matches due to chance is about 1.4%, or 14 in 1,000. One can therefore be between 95% and 99% sure of a nonchance effect with four CVCVC matches. As we in fact find four, we therefore reject the null hypothesis of no interaction, and conclude that the observed agreement in the OJAP/OJAV forms is most probably not due to chance.

Although it is not necessary, in fact, to look further than these four CVCVC pairs for indication that the degree of agreement is greater than chance, it is also instructive to examine the three CVCV(C) pairs ('to clear up'; 'storehouse'; 'kinsman') (given its low probability, examining the remaining word pair *lěsuŋ / usu* is otiose).

These three pairs have CVCVC structure in OJAV, but due to the lack of a final consonant in OJAP, the final OJAV consonant must be ignored (because any consonant would have been accepted, it must be assigned a probability of 1). The highest combined probability for these data is 0.00182. A binomial distribution for p = 0.00182 over 3,050 trials shows that one would need 10 or more matches to be 95% sure that the effect was not due to chance. There are only three, so on their own, these three forms in table 11 would not be enough to demonstrate a nonrandom similarity. This demonstrates the advisability of using longer strings than CVCV(C) in work of this kind.

Finally, we examine some semantically related pairs, like *duduk* 'bamboo spear, stabbing weapon': *turuki* 'term for swords'. As explained above, the combined probability of pairs like these must be multiplied by four to give an estimate for expected incidence. The data are given in table 12. As can be seen, the estimated incidence for these forms ranges from 0.0027 to 0.00028, with a mean of 0.00108. A binomial distribution for p = 0.00108 over 3,050 trials gives a cumulative probability of 94.9% for zero to six trials, so at least seven matches are required to be 95% sure of nonrandom effects. The six forms in table 12 thus miss being significant at 95% by a very thin whisker.

**6.4 ADDITIONAL LEVELS OF EVIDENCE.** We have demonstrated in the preceding section that evaluation of the probability of the phonological shape of OJAV/OJAP semantically matched items is sufficient to demonstrate nonrandom effects. But there is in fact more evidence than this. It does not require much thought to see that the coincidences that would have to be shown as fortuitous for our hypothesis to be rejected are not restricted to sound-meaning correspondences, but exist on several other independent levels.

First, there is the existence of sound-meaning correspondences (in themselves agreement on two levels). Then there is the obvious fact, already mentioned, that the sound-meaning correspondences obtain for many items. Of these items, there are two independent, semantically-related facts indicating a high probability of nonfortuitous occurrence. We have pointed out above that many of the paired items are culturally relevant monomorphemic lexicalizations with a high concentration of semantic components. Also, there is the incidence of items with semantic components related to ceremony and rice. Add to this the fact that several independent areas of nonlinguistic evidence—from DNA, teeth, rice genetics, and ritual—also point to an early Indonesian influence on Japan. Finally, there is the fact that the last two of these areas of nonlinguistic evidence—rice and ritual—coincide rather well with the linguistic evidence.

**6.4.1 Bayesian probability.** The contribution of the different levels of agreement to the strength of the evidence can perhaps best be understood within a nonfrequentist (i.e., Bayesian) conceptual framework (Robertson and Vignaux 1995; Rose 1998:420). A Bayesian approach is designed specifically to evaluate the probability of a hypothesis, or, equivalently, the strength of evidence. This is quantified in a *Likelihood Ratio* (LR), a simple formula for which is given at (1).

$$(1)\quad LR = \frac{P(E/H_1)}{P(E/H_2)}$$

The LR is the ratio of the probability of the evidence assuming the hypothesis to be correct ($H_1$), to the probability of the evidence assuming that it is false ($H_2$). Applied to the present data, what we would need to know is the probability of observing the evidence—that is, the degree of phonological and semantic correspondences present in the OJAV-OJAP data, and their agreement with the nonlinguistic evidence—assuming that there has been borrowing, and the probability of observing this evidence assuming no interaction. This would result in a statement of the following kind: "One would be *n* times more likely to observe these phonological and semantic correspondences if there had been borrowing than if there had not been borrowing" (where *n* is the Likelihood Ratio). Thus, rather than inferring the probability of a hypothesis from the evidence (sometimes called the "prosecutor's fallacy"), we need to assess the probability of the evidence, given the hypothesis (Sjerps and Biesheuvel 1999:215, 216; Broeders 1999:230).

**TABLE 12. PROBABILITIES FOR SEMANTICALLY RELATED PAIRS**

| | | p OJAV | p OJAP | p Combined | estimated incidence ( p Combined* 4) |
|---|---|---|---|---|---|
| 1. | duduk/turuk- | 0.00002 | 0.00063 | 0.00032 | 0.00129 |
| 2. | sawak/sawak- | 0.00034 | 0.00008 | 0.00021 | 0.00084 |
| 3. | popor/FoFur- | 0.00008 | 0.00026 | 0.00017 | 0.00068 |
| 4. | mawos/mawos- | 0.00003 | 0.0001 | 0.00007 | 0.00028 |
| 5. | kukuh/kukur- | 0.00009 | 0.00126 | 0.00067 | 0.0027 |
| 6. | wuwuh/wowor- | 0.00027 | 0.00007 | 0.00017 | 0.00067 |
| | | | | | x1 = 0.00108 |

One of the strengths of the approach is that, unlike frequentist analyses, it allows the combination of evidence from separate, independent sources. This is particularly useful in the present case, with its several levels of agreement. (This is, incidentally, how an individual piece of evidence can turn out to be wrong: the Likelihood Ratio from, for example, blood group analysis might be strong, but it can be reversed by much stronger counter-evidence from a highly negative LR from the DNA analysis.)

In order to evaluate the denominator of the LR for the OJAP/OJAV data, then, one would first need to compare pairs of words in pairs of languages that can be assumed not to have interacted, either directly or via intermediate languages, and determine the probability of observing phonological and semantic matches. The numerator would be supplied from pairs of languages that are known to have borrowed. This would yield a LR for phonological-semantic matches.

In order to accommodate the second level of agreement, however, there would need to be a semantic constraint, and this is another aspect in which both Ringe's and Nichols's approaches require modification in the context of the present discussion. They both concentrate exclusively on phonology and hold meaning constant, by treating it as random. That is, they are concerned with the probability of observing a phonological match between a pair of lexical items the meaning of which has been chosen at random.

This is, of course, totally justified from the point of view of conducting a controlled experiment. However, words are combinations of both phonological and semantic form, and differences in semantic structure are also relevant to our data. For example, a pair of words with matching phonology would surely be considered more likely to indicate a nonfortuitous event—borrowing or genetic relationship—if their meaning involved the monomorphemic lexicalization of something highly specific and culturally relevant, such as 'storehouse for treasure or grain', or 'present to someone of a higher status', than if it was general and part of core vocabulary, such as 'ear'. (Such high specificity of meaning is no doubt the main reason why studies of borrowing between individual languages, such as those mentioned above, do not concern themselves with the problem of fortuitous similarities.)

A second LR would have to be determined, therefore, that quantified the probability of observing phonological-semantic matches for specific culturally relevant lexicalized items in languages that have borrowed and those that have not. In other words, what is the probability that two languages that have not borrowed have the same phonological shape for a monomorphemic word like 'building for storehouse or treasure'?

Additional LRs would need to be estimated for all the remaining levels, the product of all of which would quantify the overall strength of the evidence.

The tasks of calculating all these LRs obviously fall well outside the scope of the present paper. They are in any case otiose, because it is possible to get an idea of how low the LR denominator will be from previous studies, such as Hock (1986:chap. 18); Trask (1996:366–368); Campbell (1998:316–317); Lass (1997:105, 123–130), and Jones (1989), which are concerned with demonstrating

phonological-semantic similarities between languages. (This may be in order to demonstrate genetic relationships or chance similarities.) In such studies, we have simply not been able to find any occurrences of CVCVC matches. For example, in Jones's (1989) attempt to demonstrate a genetic relationship between Warlbiri, Finnish, and Tamil—three languages that can reasonably be presumed to have borrowed minimally—the greatest degree of matching is for two, not necessarily regularly corresponding, segments (CV). This shows that the value of the LR denominator for cases with CVCVC matches can be expected to be very small indeed. Because the LR numerator can also be expected to be big, the LR itself, and the strength of the evidence for borrowing, will be very large.

**7. DIRECTIONALITY.** We mentioned in 5.1.1 that the correspondences point to a directionality from OJAV to OJAP. However, because this directionality could be an artefactual function of the difference in size between the two consonantal inventories, it is necessary to point to evidence from correspondences that show it is not exclusively so. Such evidence exists in the word-initial consonantal correspondences. We find OJAP *k* corresponding both to OJAV *k* and *g* in, for example, *kadaŋ/kara* 'kinsman' and *guḍaŋ/kura* 'storehouse', OJAP *t* corresponding to OJAV *t* and *d* in *tapih/taFe*$_2$'cloth' and *duduki/turuki 'sword',* and OJAP *F* corresponding to OJAV *p* and *b* in ***p****arěm /****F****are* 'clear up' and *baru****b****uh/Fo*$_2$*ro*$_2$***b****-,* and so forth. These are, of course, only unidirectionally predictable, and are not related to inventory size.

In addition, the data also contain examples like OJAV *lěsuŋ* : OJAP *usu* 'mortar', and so on, where it would be counter-intuitive, to say the least, to assume default epenthesis of *l* word-initially and *ŋ* word-finally in OJAV.

**8. COMPETING ETYMOLOGIES/DERIVATIONS.** The authors are well aware of competing etymologies for two of the items listed, and it is worth mentioning them in the light of the quantified evidence presented above.

Ohno (1980) connects OJAP *kara* 'kinsman' with Tungusic *kala, 'clan', and a form *xala was also proposed by one of our reviewers. Using Ringe's approximations of .05 for consonants and .4 for vowels, the probability associated with a Tungusic *x/kala is .0008, which means the combined probability with OJAP *kara* is (.00281 * .0008 =) .00181. We do not know the size of the Tungusic corpus available for comparison. Neither, in the apparent dearth of other Tungusic-OJAP etymologies, do we know to what extent the sound correspondences between *x/kala and *kara* are recurrent. Even granting, generously, a small corpus size comparable to ours of 3050, *and* regular correspondences, an event of this probability will occur by chance six times. A single word correspondence like this one, therefore, can easily be random.

One of our reviewers noted that OJAP *Fako* 'basket' can be related to a modern Korean form *pakwoni* (with the same gloss). This hypothesis founders on the same grounds as the previous example: lack of sound correspondences, and infraliminal

probability. It is also worth noting that, in emphasizing the former reason, Sakiyama (1996:281, 285, 286) explicitly attributes an Austronesian origin to OJap *Fako.*

**9. SUMMARY AND CONCLUSION.** This linguistic investigation has been offered in the spirit of Nichols (1996:65), who observes that historical linguistics "is likely to be improved if it can take into account the evidence offered by archeology, human biology, and other fields."

We set out in this paper to investigate whether existing nonlinguistic evidence for early contact between Japan and Indonesia could be buttressed by linguistic evidence. We presented many word pairs from Old Japanese and Old Javanese with substantial phonological and semantic agreement. The agreement in phonological form was shown to be extensive, applying in some cases to up to five segments in each word pair. Moreover, it also made sense given the phonotactic restrictions of the recipient language. The semantic agreement was often also of comparably high specificity, and, moreover, showed a further level of structure in its partial resolution into semantic fields, including some that resonate with the nonlinguistic findings related to ritual and rice cultivation in the early Japanese Yayoi period.

It is worth pointing out that the degree of phonological similarity demonstrated exceeds by far the agreement claimed to exist for putative genetic relationships between Japanese and other languages (or language families). Several scholars have remarked on the lack of systematic sound correspondences between Japanese and putative related varieties, even for the most plausible hypothesis, namely Korean (Shibatani 1990:112–117; Sakiyama 1996:349).

Our findings lend plausibility to other resemblances that could not have been countenanced because of their brevity. Some shorter forms that show the same correspondences are, for example, *luwe/tuwe* 'hungry'; *iwak/iwo* 'fish'; *ikan* 'fish'/ *ika* 'squid'; *buŋa/Fana* 'flower'; *nyai/nai* 'lady-in-waiting'; *cuki/tuki*$_2$ 'tribute'; *suku/so*$_2$*ko*$_2$ 'bottom, base'; *baru* 'new planting' / *Faru* 'Spring'; *isi/asi* 'to fill'; *(a)tahu/atahu* 'skilled' (Kumar 1996:527–540).

The findings also suggest that it will be worth sifting other lexicons for more pairs. One obvious Old Javanese source is the so-called Kawi ('old') vocabulary, a lexicon of considerable antiquity that was preserved as a literary vocabulary into the period of modern Javanese (Dirdjasiswaja 1956; Gericke and Roorda 1901). Kawi is thought to represent an ancient central Javanese dialect, and therefore is likely to provide some correction to the East Javanese bias of Zoetmulder (1982).

Because we have shown that the phonological correspondences map unidirectionally from Old Javanese to Old Japanese, and that the lexical items appear to be localized to the Indonesian subarea of Malayo-Polynesian, we conclude that there is indeed also linguistic evidence of a significant early contact from Indonesia to Japan. The agreement between semantic and archaeological evidence on material and spiritual culture dates the influence to the Yayoi period, which postdates by a millennium the dispersal of Malayo-Polynesian in the Indo-Malaysian area. Fur-

thermore, the linguistic evidence also suggests that, contrary to the received view, Yayoi innovations such as swords, warehouses, fences, rice-mortars, plates, cloth, and baskets—as well as ideas of royalty and divinity, and of the divinity of royalty —are likely to have been introduced into Japan from the south, and not from China and/or Korea as usually supposed.

## APPENDIX: ABBREVIATIONS FOR AUSTRONESIAN LANGUAGES AND SUBGROUPINGS

Membership in major Austronesian subgroups is given following the names of Austronesian languages. Because of the uncertainty surrounding the linguistic grouping of languages geographically located in Indonesia, these have been identified simply as WMP, followed by a reference to their geographic location, as for example, "Malay (WMP, Indonesia/Malaysia)."

ACE Acehnese (WMP, Indonesia)
AKL Aklanon Bisayan (WMP, Meso Philippine)
AMIS Amis (F)
BAL Balinese (WMP, Indonesia)
BAT Batak (WMP, Indonesia)
BKL Bikol (WMP, Meso Philippine)
BLA Sarangani Blaan (WMP, South Mindanao)
BON Bontok (WMP, Northern Philippine)
BTK Palawan Batak (WMP, Southern Philippine)
BUG Buginese (WMP, Indonesia)
BUR Buru (CMP)
CEB Cebuano (WMP, Meso Philippine)
CEOC Central–Eastern Oceanic
CMP Central Malayo-Polynesian
DAA Da'a (WMP, Indonesia)
DLT Dalat, Kampung Teh (WMP, Malaysia)
DPB Dairi Pakpak Batak (WMP, Indonesia)
F Formosan
HAN Hanunóo (WMP, Meso Philippine)
HLG Hiligaynon (WMP, Meso Philippine)
IBA Ibanag (WMP, Northern Philippine)
IFG Ifugao (WMP, Northern Philippine)
IFGBT Batad Ifugao (WMP, Northern Philippine)
ILK Ilokano (WMP, Northern Philippine)
IND Bahasa Indonesia (WMP, Indonesia)
ISN Isneg (WMP, Northern Philippine)
KAG Kagayanen (WMP, Southern Philippine)
KAL L(in)imos Kalinga (WMP, Northern Philippine)
KAN Kanakanavu (F)
KAU Kaulong (OC/WOC)
KB Karo Batak (WMP, Indonesia)
KON Konjo (Coastal) (WMP, Indonesia)
MAD Madurese (WMP, Indonesia)
MAG Manggarai (CMP)
MAK Makasarese (WMP, Indonesia)
MAL Malay (WMP, Indonesia/Malaysia)
MAR Maranao(WMP, Southern Philippine)

| | |
|---|---|
| MIN | Minangkabau (WMP, Indonesia) |
| MOL | Molbog (WMP, Meso Philippine) |
| NGA | Ngadha (CMP) |
| NGD | Ngaju Dayak (WMP, Indonesia) |
| OC | Oceanic |
| OJAP | Old Japanese |
| OJAV | Old Javanese (WMP, Indonesia) |
| PAI | Paiwan (F) |
| PAL | Palawan (WMP, Meso Philippine) |
| PAN | Proto-Austronesian |
| PGS | Pangasinan (WMP, Northern Philippine) |
| PS | Proto-Sangiric (WMP, Indonesia) |
| RAT | Ratahan (WMP, Indonesia) |
| REJ | Rejang (WMP, Indonesia) |
| RTI | Rotinese (WMP, Indonesia) |
| SAB | Bangingi [also Balangingi] Sama (WMP, Sulu-Borneo) |
| SAN | Sama North (WMP, Sulu-Borneo) |
| SAS | Sasak (WMP, Indonesia) |
| SAW | Sawu (CMP) |
| SUN | Sundanese (WMP, Indonesia) |
| TAG | Tagalog (WMP, Meso Philippine) |
| TB | Toba Batak (WMP, Indonesia) |
| TBWK | Kalamian Tagbanwa (WMP, Meso Philippine) |
| WMP | Western Malayo-Polynesian |
| WOC | Western Oceanic |
| WOL | Wolio (WMP, Indonesia) |

## REFERENCES


Adelaar, K. A. 1994. Malay and Javanese loanwords in Malagasy, Tagalog, and Siraya (Formosa). *Bijdragen tot de taal- land- en volkenkunde van het Koninklijk Instituut* 150:50–65.

Blust, Robert A. 1981. The reconstruction of Proto-Malayo-Javanic: An appreciation. *Bijdragen tot de taal- land- en volkenkunde van het Koninklijk Instituut* 138/2-3:231–250.

———. 1984–1985. The Austronesian homeland: A linguistic perspective. *Asian Perspectives* 26:45–67.

———. 1995. *Austronesian comparative dictionary.* Computer files, University of Hawai'i at Mānoa, Honolulu.

Broeders, A. P. A. 1999. Some observations on the use of probability scales in forensic identification. *Forensic Linguistics* 6(2):228–241.

Campbell, Lyle. 1996. Phonetics and phonology. In *Kontaktlinguistik/Contact Linguistics/Linguistique de contact*, vol.1, ed. by Hans Goebl, Peter H. Helde, Zdeněk Starÿ, and Wolfgang Wölck, 98–103. Berlin: Mouton de Gruyter.

———. 1997. Genetic classification, typology, areal linguistics, language endangerment, and languages of the North Pacific Rim. *Languages of the North Pacific Rim,* vol. 2 (June). Kyoto: Graduate School of Letters, Kyoto University.

———. 1998. *Historical linguistics: An introduction*. Edinburgh: Edinburgh University Press.

———. 1999. HISTLING posting, Sept 7.

———. In press. How to show languages are related: Methods for distant genetic relationships. In *Handbook of historical linguistics,* ed. by Richard D. Janda and Brian D. Joseph. London: Blackwell.

Clark, Ross. 1998. HISTLING posting, Nov 5.

Dawkins, Richard. 1999. *Unweaving the rainbow: Science, delusion, and the appetite for wonder.* London: Penguin.

Dirdjasiswaja, R. Ng. [1956?]. *Bausastra Kawi–Djawa.* Semarang: Toko buku Pak Roes.

Dixon, Bob. [1983]. *Searching for aboriginal languages: Memoirs of a field worker.* St. Lucia [Brisbane, Qld.]; New York: University of Queensland Press.

Dixon, R. M. W. 1991. Mbabaram. In *The handbook of Australian languages,* vol. 4, ed. by R. M. W. Dixon and Barry J. Blake, 349–402. Canberra: Australian National University Press.

Durie, Mark, and Malcolm Ross, eds. 1996. *The comparative method reviewed: Regularity and irregularity in language change.* New York: Oxford University Press.

Evans, Nicholas. 1992. Macassan loanwords in top end languages. *Australian Journal of Linguistics* 12:45–91.

Gericke, J. F. C., and T. Roorda. 1901. *Javaansch-Nederlandsch Handwoordenboek, vermeerderd en verbeterd door Dr. A.C. Vreede met medewerking van Dr. J.G.H. Gunning*. Amsterdam and Leiden: Johannes Müller and E. J. Brill.

Hashimoto, Shinkichi. 1980. *Kodaikokugo no onin ni tsuite* (On the phonology of Old Japanese). Tokyo: Iwanami Shoten.

Hock, Hans Heinrich. 1986. *Principles of historical linguistics.* Trends in Linguistics: Studies and Monographs 34. Berlin: Mouton de Gruyter.

Hoenigswald, Henry M. 1999. Review of *The comparative method reviewed: Regularity and irregularity in language change*, ed. by Mark Durie and Malcolm Ross, 1996. *Diachronica* 16:165–177.

Horne, Elinor C., with Kustiani and Koentjaraningrat. 1963. *Intermediate Javanese.* Yale Linguistic Series 4, New Haven: Yale University Press.

Hübschmann, Heinrich. 1967. On the position of Armenian in the sphere of Indo-European Languages. In *A Reader in Nineteenth-Century Historical Indo-European Lin-*

*guistics,* ed. and trans. by Winfred P. Lehman, 164–189. Indiana Studies in the History and Theory of Linguistics. Bloomington, Ind.: Indiana University Press.
Itoo, Junko, and R. Armin Mester. 1996. Japanese phonology. In *The handbook of phonological theory,* ed. by John A. Goldsmith, 817–838. Cambridge, Mass.: Blackwell.
*Jidaibetsu Kokugo Daijiten—Jodai Hen* (A periodized dictionary of Japanese—Old Japanese) 1967. Tokyo: Sanseido.
Jones, Alex. 1989. Australian and the Mana languages. *Oceanic Linguistics* 28:181–196.
Kishida, Takeo. 1984. *Kokugo onin henka no kenkyu* (A study of phonological changes in Japanese). Tokyo: Musashino shoin.
Kumar, Ann. 1992. Kingship and religion: Looking from Java to Japan. In *Looking in odd mirrors: The Java Sea,* ed. by V. J. H. Houben, H. M. J. Maier, and W. van der Molen, 258–288. Semaian 5. Leiden: Rijksuniversiteit.
———. 1996. Does Japanese have an Austronesian stratum? In *Pan-Asiatic Linguistics: Proceedings of the Fourth International Symposium on Languages and Linguistics.* vol. 2:522–543. Bangkok: Mahidol University.
———. 1998. An Indonesian element in the Yayoi?: The evidence of biological anthropology. *Anthropological Science,* 106:265–276.
Kumar, Ann, and John McGlynn. 1996. *Illuminations: Writing traditions of Indonesia.* New York: Weatherhill.
Lass, Roger. 1997. *Historical linguistics and language change.* Cambridge Studies in Linguistics 81. Cambridge: Cambridge University Press.
Martin, Samuel E. 1987. *The Japanese language through time.* New Haven: Yale University Press.
Matsumoto, Katsumi. 1995. *Kodai Nihongo boin ron: Jodai tokushu kanazukai no saikaishaku* (A study of Old Japanese vowels: A reinterpretation of special orthography). Tokyo: Hitsuji Shobo.
Miller, Roy Andrew. 1967. *The Japanese language.* Chicago: University of Chicago Press.
Morinaga, T. 1968. Origin and geographical distribution of Japanese rice. *Japan Agricultural Research Quarterly* 3(2):1–5.
Murayama, Shichiroo. 1978. *Nihongo-keitoo no Tankyuu* (Investigations into the Origins/Genealogy of the Japanese Language). Tokyo: Taishookan Shoten.
Nichols, Johanna. 1990. Linguistic diversity and the first settlement of the New World. *Language* 66:475–521.
———. 1993. East of Europe: Languages and peoples of Northern Eurasia. Unpublished reading materials for Slavic 37. University of California, Berkeley.
———. 1996. The comparative method as heuristic. In *The comparative method reviewed: Regularity and irregularity in language change*, ed. by Mark Durie and Malcolm Ross, 39–71. New York: Oxford University Press.
Nothofer, Bernd, 1975. *The reconstruction of Proto-Malayo-Javanic*. Verhandeling van het Koninklijk Instituut voor taal- land- en volkenkunde 73. 's-Gravenhage: Martinus Nijhoff.
O'Grady, G. N., and D. T. Tryon. 1990. Early Austronesian loans in Pama-Nyungan? In *Studies in Comparative Pama-Nyungan,* ed. by G. N. O'Grady and D. T. Tryon, 105–116. Pacific Linguistics Series C-111. Canberra: Australian National University Press.
Ohno, Susumu. 1980. *Nihongo no sekai 1: Nihongo no seiritsu* (The world of Japanese 1: The establishment of Japanese). Tokyo: Chuokoronsha.
———. 1982 Correspondences between Tamil and Japanese—A reply to Murayama. *Kokugogaku* 130:117–134.
———. 1989. The genealogy of the Japanese language: Tamil and Japanese. *Gengo-Kenkyu: Journal of the Linguistic Society of Japan* 32–63.
———. 1994. *Nihongo no kigen* (The origin of Japanese). Tokyo: Iwanami Shoten.
Ohno, Susumu, Akihiro Satake, and Kingoro Maeda. 1974. *Iwanami kogo jiten* (Iwanami's Dictionary of Old Japanese). Tokyo: Iwanami Shoten.

Oka, H. I. 1988. *Origin of Cultivated Rice.* Tokyo: Japan Scientific Societies Press, and Amsterdam: Elsevier.

Pawley, Andrew. 1999. Chasing Rainbows: Implications of the rapid dispersal of Austronesian languages for subgrouping and reconstruction. In *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, ed. by Elizabeth Zeitoun and Paul Jen-kuei Li, 95–138. Symposium Series of the Institute of Linguistics, Academica Sinica. Taipei: Academica Sinica.

Ringe, Donald A., Jr. 1992. On calculating the factor of chance in language comparison. *Transactions of the American Philosophical Society* 82:1–110.

———. 1998. Nostratic: Sifting the evidence. In *Nostratic: Sifting the evidence*, ed. by Joseph C. Salmons and Brian D. Joseph, 153–197. Current Issues in Linguistic Theory 42. Amsterdam: John Benjamins

———. 1999. How hard is it to match CVC-roots? *Transactions of the Philological Society* 97:213–244.

Robertson, B., and T. Vignaux. 1995. *Interpreting evidence.* Chichester: Wiley.

Rose, Phil. 1998. A forensic phonetic investigation into non-contemporaneous variation in the F-pattern of similar-sounding speakers. In *Proceedings of the 5th International Conference on Spoken Language Processing*, ed. by R. H. Mannell and J. Robert-Ribes, 217–220. Canberra: Australian Speech Science and Technology Association.

Ross, Malcolm. 1995. Some current issues in Austronesian linguistics. In *Comparative Austronesian dictionary,* pt. 1, fasc. 1, ed. by Darrell T. Tryon, 45–120. Berlin: Mouton de Gruyter.

Saitoo, Tadashi. 1983. Yayoi culture. In *Kodansha encyclopaedia of Japan*. 1st ed, vol. 8:321–322. Tokyo: Kodansha.

Sakiyama, Osamu. 1996. Formation of the Japanese language in connection with Austronesian languages. In *Prehistoric Mongoloid dispersals*, ed. by Takeru Akazawa and J. E. Emoke Szathmary, 349–358. Oxford: Oxford University Press.

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

Sjerps, Marjan, and Dirk B. Biesheuvel. 1999. The interpretation of conventional and 'Bayesian' verbal scales for expressing expert opinion: A small experiment among jurists. *Forensic Linguistics* 6:214–227.

Thorne, Alan, and Robert Raymond. 1989. *Man on the rim: The peopling of the Pacific*. North Ryde, N.S.W.: Angus and Robertson.

Trask, R. L. 1996. *Historical linguistics*. London: Arnold.

Tryon, Darrell, ed. 1995. *Comparative Austronesian dictionary.* Trends in Linguistics Documentation 10. Berlin: Mouton de Gruyter.

Vance, Timothy J. 1987. *An introduction to Japanese phonology.* New York: State University of New York Press.

Vovin, Alexander. 1994. Is Japanese related to Austronesian? *Oceanic Linguistics* 33:369–390.

Wierzbicka, Anna. 1996. *Semantic primes and universals*. Oxford: Oxford University Press.

Zoetmulder, P. J. (with the collaboration of S. O. Robson). 1982. *Old Javanese-English dictionary.* Koninklijk Instituut voor taal- land- en volkenkunde. 's-Gravenhage: Martinus Nijhoff.

———. 1983. *De Taal van het Adiparwa: Een Grammaticale Studie van het Oudjavaans*, KITLV Indonesische Herdrukken [first published 1950]. Leiden: Foris.

Asian History Centre
Australian National University
Canberra 2600, AUSTRALIA
ann.kumar@anu.edu.au

Arts Faculty Linguistics Department
Australian National University
Canberra 2600, AUSTRALIA
philip.rose@anu.edu.au