

## **Is borrowability borrowable?**

Uri Tadmor

Max Planck Institute for Evolutionary Anthropology

Paper to be presented at the symposium

*Language Contact and the Dynamics of Language: Theory and Implications*

Leipzig, 10-13 May 2007

“The words of any language can be divided into two broad categories, closed and open... The closed categories are the function words: pronouns like *you* and *she*; conjunctions like *and*, *if*, and *because*; determiners like *a* and *the*; and a few others. Newly coined or borrowed words cannot be added to these categories, which is why we say they are closed.”

(O’Grady, Dobrovsky, and Aronoff 1989:89)

“[T]hough it is true that some kinds of features are more easily transferred than other... social factors can and very often do overcome structural resistance to interference at all levels.”

(Thomason and Kaufman 1988:15)

Many claims have been made regarding which classes of words are more (or less) borrowable than others. In particular, it has been claimed that members of closed sets, such as numerals and pronouns, are less susceptible to borrowing than members of open sets; that function words in general are borrowed less than content words; and that core vocabulary items are less likely to be replaced by loanwords than specialized ones. These ideas were first formulated based on the analysis of ‘Old World’ languages. In this paper, I will explore borrowed vocabulary in some languages of Southeast Asia that cast doubt on these generalizations. In these languages, members of closed sets (such as numerals and pronouns), function words in general, and core vocabulary items (like body parts and kinship terms) are at least as borrowable as members of open sets, content words, and specialized

vocabulary, respectively. Moreover, it appears that borrowability itself is a borrowable feature. In other words, the fact that, for example, pronouns and body part terms are commonly borrowed throughout Southeast Asia, is due to contact.

In a previous conference paper (Tadmor 2006), I explored a very specific sociolinguistic setting which may result in greater borrowing of high frequency words (basic vocabulary, function words) than other vocabulary. In particular, I discussed the case of a few languages spoken by small Dayak (indigenous non-Malay) groups in western Borneo. For probably over a millennium, Dayaks in this part of Borneo have been under the political and economic dominance of Malay(ic)-speakers, and have been under pressure to assimilate to their language and culture. Many individuals, sometimes entire groups, shifted to Malay(ic) as their first language, and quite a few also converted to Islam. Some other groups, while also coming under Malay dominance, had only limited contacts with the dominant Malayic speakers. They did not acquire sufficient proficiency of Malayic to shift to it from their indigenous Land Dayak languages. Rather, they borrowed extensively from Malay while maintaining their own languages. They could not borrow specialized Malay vocabulary, because they were not familiar with it; so they borrowed high frequency words with which they were familiar, such body parts, basic verbs, numerals, color terms, and occasionally some function words. From a sociolinguistic point of view this process may be viewed as partial relexification; perhaps, given more exposure to Malay, their languages would have been completely relexified, or perhaps even perfect acquisition of Malay would have taken place. So it is not surprising that mostly basic vocabulary was affected by this process. A similar situation has been observed on mainland Southeast Asia, where speakers of Aslian languages borrowed much of their basic vocabulary, but little else, from Malay (Gérard Diffloth, personal communication).

In this paper, I would like to discuss a very different and more widespread sociolinguistic setting, which has also led to substantial borrowing of high frequency

vocabulary in languages of Southeast Asia. This process affected languages on the other side of the spectrum, spoken by large, socio-politically dominant groups.

One category of words that has long been considered as especially resistant to borrowing is pronouns. Indeed, looking at the pronoun inventory of most Indo-European and Semitic languages, which have been studied extensively by linguists since the early 19<sup>th</sup> century, it appears that pronouns are among the most stable and conservative parts of the lexicon<sup>1</sup>. For example, all the pronouns of modern Hebrew, Amharic, and standard Arabic reflect Proto Semitic forms. A similar situation exists in most major Indo-European languages. English has one exception that highlights the rule: it borrowed the 3PL pronoun *they* from Scandinavian. Because pronouns are unusually stable, conservative, and resistant to borrowing in Old World languages, they are included in Swadesh lists, which are supposed to be of universal application.

Yet the situation in languages of Southeast Asia is very different from that which we normally encounter in the Old World. To start off with just a few examples, the most neutral 1SG pronoun in standard Indonesian is *saya*, a loanword derived from Sanskrit. The only 3PL pronoun *meréka*, was borrowed from Old Javanese, and the 2PL pronoun is *kalian*, was borrowed from Minangkabau, a closely related yet distinct language. Indeed, even the base of *kalian*, *kali* ‘time’, is a loanword from Sanskrit.

A complete inventory of pronouns used in Malay-Indonesian, especially if colloquial varieties are considered, runs into the dozens, and includes many other loanwords. And this is not a unique situation in Southeast Asia. A comparative study of pronominal reference in Southeast Asia lists several dozen pronouns *each* for Thai, Burmese, and Vietnamese (Cooke 1968). Although Cooke’s study was concerned

---

<sup>1</sup> This is not to say that pronouns in these languages do not change over time. Pronominal paradigms are often subject to change by analogy, leveling, and syncretism (in addition to phonological change). However, borrowing into the pronominal systems of Indo European and Semitic languages is very rare.

with semantics and syntax rather than with etymology, it is possible to identify many loanwords among these pronouns. Some scholars have already remarked on the unusually large inventories of pronouns in Southeast Asian languages, and one study (Thomason and Everett 2001) specifically mentions Southeast Asia as an area where pronouns are prone to borrowing.

Interestingly, this profusion of pronouns in Southeast Asia seems to be a relatively recent phenomenon. In most language families spoken in the region, it is possible to reconstruct limited, well structured inventories of pronouns, that look like the ‘normal’ pronominal systems of Old World language. The pronouns of Proto Tai are one example (the forms are from Li 1977)<sup>2</sup>:

Table 1: Reconstructed pronouns in Proto Tai

Proto Tai meaning	Proto Tai reconstruction	Thai reflex
1SG	<i>*ku</i>	<i>kuu</i>
2SG	<i>*su</i>	<i>sǔǔ</i>
3SG	<i>*man</i>	<i>man</i>
1PL	<i>*rəu</i>	<i>raw</i>
2PL	<i>*mɯŋ</i>	<i>mɯŋ</i>
3PL	<i>*khlau</i>	<i>kháw</i>

In the same vein, one might argue that the pronominal system of Malay-Indonesian is extremely conservative. All six pronouns that are reconstructible to Proto Austronesian have reflexes in modern Malay-Indonesian. Like the Thai forms, their phonological forms are almost identical to the reconstructed ones, despite the much

---

<sup>2</sup> Li did not distinguish between singular and plural 2<sup>nd</sup> person pronouns; both are simply glossed as ‘you’. However, comparative evidence from Austro-Tai suggests that *\*su* was singular and *\*mɯŋ* was plural. The Proto Tai reconstructions were accessed via the website of the Center for Research in Computational Linguistics, Bangkok (<http://crcl.th.net/>).

longer time span (5,500 years). Malay-Indonesian pronouns that reflect Proto Austronesian reconstructions are listed in table 2.

Table 2: Malay-Indonesian pronouns derived directly from Proto Austronesian

Proto Austronesian meaning	Proto Austronesian form <sup>3</sup>	Malay-Indonesian reflex
1SG	<i>aku</i>	<i>*aku</i>
2SG	<i>(eng)kau</i>	<i>*kaSu</i>
3SG	<i>ia</i>	<i>*ia</i>
1+2	<i>kita</i>	<i>*kita</i>
1PL	<i>kami</i>	<i>*kami</i>
2PL	<i>kamu</i>	<i>*kamu</i>

Yet, as we've seen above, the picture presented in tables 1 and 2 is misleading, because in addition to reflexes of the reconstructed pronouns, both Thai and Malay-Indonesian have many other pronouns, some of which are actually more commonly used than the modern reflexes of the reconstructed forms. We can therefore conclude that the old axiom about pronouns being particularly conservative is true, but does not in anyway preclude the possibility of pronoun borrowing.

Lower numerals are another class of words which have hardly been affected by borrowing in Old World languages. Again, the situation in Southeast Asia is rather different. Malay-Indonesian borrowed its word for 'three', *tiga*, from Indo-Aryan; it completely replaced the Old Malay form, *tlu*. High Javanese then borrowed the word from Malay, as the counterpart of Low Javanese *telu*. Another Indo-Aryan numeral borrowed into High Javanese is *dasa* 'ten'. Malay borrowed the element *-belas* 'teen' from Javanese, to form numerals between 11 and 19, as in *dua belas* 'twelve'

---

<sup>3</sup> Proto Austronesian pronouns occurred in various allomorphs. Those listed here are the ones which gave rise to the modern Malay-Indonesian forms. For a full discussion, see Ross 2006.

(literally ‘two-teen’), where Old Malay had *sa-puluh dua* (literally ‘one-ten two’). Classical Malay also borrowed from Javanese the similar element *–likur*, which forms numerals from 21 to 29, but these have become obsolete in modern Malay-Indonesian. As for Tai languages, they borrowed all their numerals other than ‘one’ and ‘two’ from Chinese (and in some contexts, even those are used). Khmer in turn borrowed the terms for 30, 40, 50, 60, 70, 80, 90, and 100 from Thai.

Terms for body parts have also been borrowed. The Malay-Indonesian word for ‘head’, *kepala*, was borrowed from Sanskrit. It completely replaced the indigenous word (*h*)*ulu*, which now survives only in a few compounds and in the figurative sense of ‘head of a river’. The same Sanskrit word also replaced the original Khmer word for ‘head’; its form in modern Khmer is *kbaal*. In Thai, the indigenous term for ‘head’, *hŭa*, persists side by side with the more polite term, *sĭsà*, derived from Sanskrit. Several other basic body parts in Thai are said to derive from Chinese: *khǎǎn* ‘arm’ (Middle Chinese \**kien*), *khǎǎn* ‘shin’ (Middle Chinese \**γien*), and *khǎǎ* ‘leg’ (Middle Chinese \**khau* ‘leg bone’) (Titima Suthiwan, personal communication). The High Javanese words *asta* ‘hand’, *sirah* ‘head’, and *grana* ‘nose’, are all derived from Sanskrit.

Kinship terms in southeast Asian languages are similar to pronouns in that they are particularly prone to borrowing. The standard Malay-Indonesian terms for ‘father’ (*bapak*) and ‘mother’ (*ibu*) were both borrowed from Old Javanese. The polite terms for ‘son’ and ‘daughter’, *putra* and *putri*, both derive from Sanskrit, as are the polite terms for ‘brother, sibling’ (*suadara*) as well as ‘husband’ and ‘wife’ (*suami, istri*).

Thai has also borrowed kinship terms extensively, although the original Tai terms are all still commonly used as the less polite and less formal counterparts. Thus *phô* ‘father’ is used side by side its polite counterpart *bĭdaa*; *mǎǎ* ‘mother’ is used alongside *maandaa*; and *lŭk* ‘child’ is used alongside *bŭt*. A similar phenomenon can be seen in Khmer. Vietnamese borrowed terms for husband and wife, parents’ siblings, and grandparents from Old Chinese.

Why have Southeast Asian languages borrowed extensively words that tend to resist borrowing in other languages? The Thai and Javanese terms discussed above offer some clear clues. In cases where loanwords coexist with earlier indigenous terms, they are usually used as their polite or formal counterparts. In other words, we are dealing with honorifics.

Honorifics occur in various parts of the world outside Southeast Asia; perhaps the best known are those of Japanese. In Southeast Asia, the honorifics of Javanese are particularly well developed (and well studied), although they are usually known as ‘speech levels’: High Javanese or Krama, and Low Javanese or Ngoko.

However, they occur to one extent or another in many other languages of Southeast Asia, including Balinese, Sundanese, Thai, and Khmer. Honorifics can be derived by various means, for example by attaching honorific affixes or clitics to ordinary words. Ordinary words can be phonologically manipulated to create honorifics, for example by replacing the coda with a designated honorific one, or by changing certain vowels in a predictable way. Sometimes synonyms of certain words are selected and ‘elevated’ to the role of honorifics. Older words that survive in the literature but have disappeared from ordinary speech are particularly favored. All these strategies for deriving honorifics are internal: they utilize a language’s own resources. However, drawing on external sources—i.e. other languages—is also a common way of deriving honorifics in Southeast Asia.

While resisting the urge to fall into cultural stereotyping, it is hardly deniable that the major civilizations of Southeast Asia—such as the Malay, Javanese, Thai, Cambodian, and Vietnamese civilizations—have been exceptionally open and receptive to foreign influence, be it from China, from India, from the Middle East, and more recently from the West. This openness is reflected in their religions, art, music, food, dress, and also language. Lexical borrowing should therefore be viewed in its wider socio-cultural context. The combination of a tendency for formalized expressions of respect (including linguistic ones), and on the other openness to

borrowing from foreign cultures (including from their languages), has been a major factor in the introduction of loanwords into the basic vocabulary of their languages. This process can be summarized as follows:

1. Many of the major languages of Southeast Asia have formalized linguistic expression of respect, in the form of honorifics.
2. By their nature, honorifics affect high frequency words (function words and basic content words) more than low-frequency words (specialized vocabulary).
3. Honorific words can be derived by internal or external means. Southeast Asian cultures are particularly receptive to outside influence, and have used many loanwords as honorifics.
4. Over time, honorifics tend to gradually lose their honorific value. They then become ordinary, unmarked words.
5. This may lead to two consequences. First, there is a need to create new honorifics, to replace the 'bleached' ones. Second, the original (indigenous) words, originally unmarked, now get demoted, and come to be perceived as overly familiar or even impolite. Eventually, they may be avoided and become obsolete, their place taken by the bleached honorifics.
6. Since many of these 'demoted' honorifics were originally borrowed from other languages, this leads to the replacement of basic vocabulary by loanwords.

Let us look at some concrete examples. The Malay terms for 'husband' and 'wife' are *laki* and *bini*, both which are reconstructible to Proto Malayic. Some time in the history of Malay, *suami* and *istri* were borrowed from Sanskrit, as the honorific counterparts of the original Malay words. Over time, these loanwords became honorifically bleached, until they became unmarked, ordinary words. This pushed down the original words *laki* and *bini*, which were now perceived as impolite. A very similar process can be seen in Thai, where the Sanskrit loanwords *sǎǎmii* and *pʰanráyaa* were originally honorific counterparts of the Tai words *pʰǔa* and *mia*. Eventually, the loanwords *sǎǎmii* and *pʰanráyaa* became honorifically bleached, thereby turning the indigenous *pʰǔa* and *mia* into impolite terms. In both languages,



the result was the replacement of native basic vocabulary by loanwords. It is also interesting to note that the Vietnamese terms for ‘husband’ and ‘wife’, *chồng* and *vợ*, are both derived from Chinese (Alves 2007).

The Indonesian word for ‘marry, married’ shows a further development. We do not know the original Malay term for ‘marry’, but since there are good reconstructions for ‘husband’ and ‘wife’ in Proto Malayic, we know that the institution of marriage existed, and may assume that Proto Malayic had a word for it. The earliest recorded Malay term for ‘marry’ is *kahwin/kawin*, a loanword from Persian. It was probably borrowed as the polite counterpart of a Malay term (or may even of a Sanskrit-derived term which may have replaced it) which has since become obsolete through demotion by honorific bleaching. After *kawin* became bleached itself, there was a need for a new polite term for ‘marry’. The gap was filled by borrowing the word *nikah* from Arabic. This demoted the earlier loanword *kawin*, which was now perceived as impolite; indeed, some Indonesians claim that it is only suitable for referring to copulating animals. Thus modern Indonesian has two terms for ‘marry’, both marked: *kawin* ([-polite]) and *nikah* ([+polite]). When Indonesians needed a neutral term, for example to refer to their own marriage or to that of siblings or close friends, they turned to the current dominant source of borrowing: English. Thus the most commonly used term for ‘marry, married’ in colloquial Jakarta Indonesian is *mérit*<sup>4</sup>.

A similar example is the borrowing of the English pronouns *I* and *you* into Southeast Asian languages, when the existing 1<sup>st</sup> and 2<sup>nd</sup> person pronouns all became marked, either as respectful or as disrespectful. The use of *ai* and *yuu* as neutral pronouns in Thai was already observed four decades ago by Cooke (1968:38-41). Exactly the same pronouns were also borrowed into urban Peninsular Malay, probably even earlier. For many Malaysian speakers, these have become the default pronouns,

---

<sup>4</sup> Note that the English word *marry* is also a loanword, again showing that while this phenomenon is typical of Southeast Asia, it is not restricted to it.

used most frequently in everyday conversations. More recently, *you* has also been making inroads into Indonesian, although not *ai*.

Of course, the creation of polite terms (or politeness-neutral terms) is not the only way loanwords can replace nonborrowed high-frequency vocabulary. A completely different scenario which may lead to similar results in the languages of small oppressed minority groups was briefly mentioned at the beginning of this paper. For dominant languages, too, other scenarios are possible. One of these is transfer from a substratum (or superstratum) language. One example is the use of Chinese pronouns by ethnic Chinese in Southeast Asia, including those who no longer speak any Chinese language. Ethnic Chinese in Indonesia often use the pronouns *owé/gua* '1SG' and *lu* '2SG', both derived from Hokkien, when speaking Indonesian. Ethnic Chinese in Thailand often use the very same pronouns when speaking Thai, in slightly different forms: *uá* and *lúúú*.

In both Thailand and Indonesia, these originally Chinese pronouns are now also used by non-Chinese as well. In Indonesia, *gua/gué* and *lu* are the unmarked 1<sup>st</sup> and 2<sup>nd</sup> person pronouns in Betawi, the Malay variant used in Batavia, as well as in Jakarta Indonesian, which developed out of it. In Thailand, non-Chinese sometimes use *uá* and *lúúú* as a sign of intimacy and equality, since the indigenous 1<sup>st</sup> and 2<sup>nd</sup> person pronouns in Thai are all marked, either as respectful or as disrespectful.

Substratum influence can also result in the introduction of other types of high frequency items. Jakarta Indonesian, for example, has borrowed many of its body part and kinship terms from substrate languages. These languages included principally Hokkien, Balinese, and Creole Portuguese (in the 17<sup>th</sup>-18<sup>th</sup> centuries), and Javanese and Sundanese (in the 18<sup>th</sup>-20<sup>th</sup> centuries). Borrowed body part terms in Jakarta Indonesian include *kuping* 'ear', *jidat* 'forehead', *dengkul* 'knee', *sikut* 'elbow', *bréwok* 'beard', *kontol* 'penis', and *téték* 'breasts', and *puser* 'belly button'. These are in addition to standard Indonesian *kepala* 'head', *kerongkongan* 'throat', and *bahu* 'shoulder', all of which are loanwords, and may be used in Jakarta Indonesian

as well. Borrowed kinship terms include *nyak* ‘mother’, *babé* ‘father’, *engkong* ‘grandfather’, *nenék* ‘grandmother’, *oom* ‘uncle’, and *tante* ‘aunt’. Again, these are in addition to standard Indonesian *bapak* ‘father’, *ibu* ‘mother’, *paman* ‘uncle’, *bibi* ‘aunt’, and other borrowed kinship terms. The very common epithets/vocatives *mas* (for males) and *mbak* (for females) are also borrowed.

Finally, I would like mention what David Gil has termed ‘the chameleon-like nature’ of Malay-Indonesian and other languages of Southeast Asian. In Malay-Indonesian, it is considered a mark of friendliness to assimilate linguistically to one’s interlocutor. Despite the fact that I speak fluent Indonesian, some speakers, when talking to me, like to mix in as many English words as they know, and even try to speak in with an English accent. At first, I found this disconcerting, and even offensive, before I realized it was in fact a sign of friendliness. To give an example from Malaysia, if an ethnic Malay is speaking to an ethnic Chinese, he might use the Chinese-derived pronouns *gua* and *lu*, already mentioned above. Similarly, when addressing a friend of Javanese ethnic background, an Indonesian may use the object-focus 1<sup>st</sup> person pronoun *tak-*. This phenomenon is most apparent in pronoun use, but extends well beyond that. The same Indonesian, when speaking to a Javanese, may well use other Javanese terms, for example for ‘eat’, ‘sleep’, or ‘bathe’. But he is unlikely to use any specialized Javanese vocabulary, simply because he does not know it. This assimilatory tendency is basically performance-based. But when it occurs frequently and regularly, it may become part of one’s competence: basic Chinese vocabulary may get incorporated into general Malaysian Malay, and basic Javanese vocabulary may become part of general Indonesian.

We have thus seen that high frequency vocabulary—including function words and basic content words—are quite borrowable in Southeast Asia. This is rather different from the situation in other parts of the world, for example in Europe and the Middle East. The borrowability of high frequency items in Southeast Asia is thus areal feature, which spread by contact among speakers of different languages. In other words, borrowability itself is a borrowable feature.

## References

- Alves, Mark, 2007. Loanword Typology database for Vietnamese. Loanword Typology Project.
- Cooke, Joseph R., 1968. Pronominal reference in Thai, Burmese, and Vietnamese. Berkeley and Los Angeles: University of California Press.
- Li, Fang Kuei, 1977. A Handbook of Comparative Tai. University Press of Hawaii.
- Marsden, William, 1812. A Grammar of the Malayan Language. London: Printed for the author by Cox and Baylis.
- O'Grady, W., Dobrovsky, M., & Aronoff, M., 1989). Contemporary linguistics: An introduction. New York: St. Martin's Press.
- Ross, Malcolm, 2006. Reconstructing the Case-marking and Personal Pronoun Systems of Proto Austronesian. In: Streams Converging Into an Ocean: Festschrift in Honor of Professor Paul Jen-Kuei Li on His 70<sup>th</sup> Birthday. Edited by Henry Y. Chang, Lillian M. Huang, and Dah-an Ho. Taipei: Academic Sinica, 521-563.
- Tadmor, Uri, 2006. The Linguistic Consequences of Long Term Patron-Client Relationship: Dayaks and Malays in Western Borneo. Paper presented at the workshop "Historical linguistics and hunter-gatherer populations in global perspective". Leipzig, 10-12 August 2006.
- Thomason, Sarah G., and Daniel L. Everett, 2001. Pronoun borrowing. Berkeley Linguistic Society 27:301-315.