

NIKOLAOS VAN DAM

Arabic loanwords in Indonesian revisited

How did Arabic loanwords end up in Indonesian?¹ Various answers to this intriguing question have been put forward in an effort to trace the long road these words may have followed before becoming part of the Malay or Indonesian vocabulary. Some of the arguments used may be theoretically correct, also from a historical point of view, but they rarely provide any solid linguistic evidence. Conclusions have sometimes been drawn on the basis of only a few words, or nonrepresentative or obsolete words. Occasionally, academics have taken over arguments and hypotheses from one another without independent research into their reliability. Some of these arguments and hypotheses seem to have gained the status of 'established fact' as a result of being repeated time and again. In short, some existing theories on Arabic loanwords in Indonesian do not stand up to a reality check.

I shall deal with four questions:

Traces of colloquial Arabic in loanwords in Indonesian. Is there any clear linguistic evidence for a connection with the Arabs of South Arabia or Hadramaut?

Does spelling follow pronunciation or does pronunciation follow spelling?

What is the significance of pseudo-classical endings in Arabic loanwords, such as *nafsu* and *salju*?

Notes on the origin of *-ah* and *-at* endings in Arabic loanwords as described in Stuart Campbell (1996a). Can a Persian connection be demonstrated statistically?

¹ This article is an extended version of a lecture delivered at the Universitas Indonesia, Jakarta, on 20 May 2009, as part of a workshop on loanwords in Bahasa Indonesia, organized by the *Wacana Journal of Humanities* in cooperation with the Department of Linguistics of Universitas Indonesia. I am most grateful for many valuable comments by Rudolf de Jong, which were incorporated when preparing this article.

NIKOLAOS VAN DAM served as Ambassador of the Netherlands to Indonesia (2005-2010), Germany (1999-2005), Turkey (1996-1999), Egypt (1991-1996) and Iraq (1988-1991). He studied Arabic and political and social sciences at the University of Amsterdam, and served most of his academic and diplomatic career in the Arab world, also covering Libya, Lebanon, Jordan and the Palestinian Occupied Territories. Dr. N. van Dam can be reached at nikolaosvandam@gmail.com. Most of his publications can be downloaded from www.nikolaosvandam.com.

Traces of colloquial Arabic

A major problem in studying Arabic loanwords in Indonesian is that some of their original traces were erased through the language reform of Bahasa Indonesia, part of which was an artificial process. During this reform, or standardization or codification process, language committees and others decided what was to be considered correct and what not (Van Dam 2009a, 2009b). *Pesantren* must have played an important role, because those in Indonesia who wish to learn (classical) Arabic have mostly been trained at these schools (Versteegh 1997:238; Steenbrink 2006:338). As a result, some words underwent change, while others disappeared. Nevertheless, colloquial Arabic elements can still be found in some Malay dialects, such as Betawi as spoken in Jakarta. And although these colloquial elements have generally not been incorporated into the official Indonesian language, they can nevertheless provide us with valuable information on the history of Arabic loanwords in Indonesian.

Arabic loanwords that may have come to Indonesian through Austronesian languages from within the archipelago, such as Javanese or Sundanese, or Malay dialects such as Betawi, have usually been transformed into a new and more classical Arabic form, and some may have disappeared.

The paths travelled by Arabic loanwords – or rather by the people carrying them – remain to be studied. Were they traders? Islamic teachers? Arabs, Persians, Tamils? If they were Arab traders, one would expect the Arabic loans in Indonesian or Malay to at least contain a substantial number of colloquialisms, because traders (like most people) do not generally use classical Arabic when communicating with others in their mother tongue. The majority of traders would not even have had a very good command of classical Arabic. The case could be somewhat different if Islamic teachers were involved. As most Arabic loanwords in Indonesian have a classical appearance, it seems reasonable to assume that they reached the Indonesian Archipelago mainly through people who had a command of written Arabic: Islamic teachers and scholars (of Arab, Persian, Indonesian or other origin) at *pesantren*, in mosques, and elsewhere, and perhaps also through compilers of dictionaries.

But as I shall demonstrate below, there is also evidence of direct face-to-face contact, most probably through Arab traders who used colloquial Arabic.

Both Campbell (1996a) and Versteegh (2003) have noted that many Arabic loanwords in Indonesian have undergone a process of re-arabization. Loanwords that had previously taken a form deviant from classical Arabic were at a later stage given a new form more in line with the formal original classical Arabic.

The relinquishing of the alphabets in which Malay was originally written (including Arabic script in an adapted version) may have led to changes in how loanwords came to be written in Malay/Indonesian in the Latin

alphabet, and subsequently in the pronunciation of words, as a result of the transcription chosen.

I shall discuss in particular some of the works of Versteegh and Campbell, because they have dealt extensively with Arabic loanwords in Indonesian (and Malay), and have also noted colloquial Arabic traces in Arabic loanwords in Indonesian. Both scholars deserve high credit for their work, and my remarks should be seen as no more than minor observations, or notes in the margin. This review is certainly not intended to repeat their work, and refers only to parts of it.

Most Arabic loanwords, due to their classical form, do not bear any colloquial traces that would give a clue as to their regional origin. Versteegh (2003:221) notes however that:

Of particular interest are those loanwords that betray an Egyptian source, in which *j* is realized as *g*, as in *gamal* 'camel' (= Arabic *jamal*, Egyptian Arabic *gamal*), *gengsi* 'prestige, status' (if this word actually derives from Arabic *jins*, Egyptian Arabic *gins*, rather than from Chinese), and those having a voiced *g* for Arabic *q*, as in *gamis* 'shirt' (= Arabic *qamis*), *gereba* 'waterskin' (= Arabic *qirba*). The latter category might go back to Persian in which the reflex of Arabic *q* is *g*. With regard to the possibility of Egyptian provenance, it should be remarked that according to some sources [...] traders from Cairo had been active in Java as early as the 11th century.

Campbell (2007:340) notes, on the other hand, that there are 'very few unequivocally colloquial Arabic loanwords in Indonesian and Malay, suggesting that colloquial Arabic was not the medium of discourse with Malays'. Campbell (1996b) further suggests that 'there is very little evidence of regionalisms or colloquialisms' and that there is no trace of high-frequency characteristically urban colloquial verbs such as *shāf* (to look) and *rāh* (to go), which would constitute more extensive and systematic evidence of colloquial Arabic influence (Campbell 2007:344).

Note, however, that the Arabic colloquial *shuf* (imperative of *shāf*) does occur in Betawi, although it has not been absorbed into official Indonesian.

Campbell (1996b) further remarks that:

One pronunciation diagnostic is the Arabic *q*, which always occurs as *k* in Malay. In regional varieties of Arabic, *q* often appears as the glottal stop or as *g*; coffee is usually 'ahwah or gahwah in speech, but qahwah in writing. If colloquial varieties of Arabic were a source of borrowing, we would expect to see, for example, gahwa or ahwa for kahwa, gima or imah for kimah, etc.

Whereas the Arab colloquial forms *gahwe* and *gahwa* indeed do not occur in of-

ficial Indonesian, as noted by Campbell, I did encounter these forms in Betawi and colloquial Jakartan Indonesian, respectively. This is a clear indication of colloquial influence, as shown below. It is difficult, however, to find any clear clue as to the exact period or geographical area in which such colloquial words were absorbed into Malay.

Concerning Versteegh's remarks on the various pronunciations of *jīm*, I would at first sight have considered looking for a simpler explanation. In southern parts of the Arabian Peninsula, including Yemen, from which many Arab immigrants in Indonesia originate, it is quite common to realize *jīm* as *gīm*. That might, therefore, have been a more probable explanation, all the more so since there has been hardly any Egyptian presence in Indonesia. But when looking into this issue in more detail, it turns out to be very difficult to find any solid evidence of a South Arabian, let alone Hadrami, origin. Since Hadramaut happens to be the region within Yemen where the phoneme /y/ is the regular reflex of classical *j (Al-Saqqaf 2009:687; Vanhove 2009: Table 1; Van den Berg 1886:259), one might have expected this aspect to be reflected somewhere in Arabic loanwords in Indonesian, at least if the Hadrami people in Indonesia played any linguistic role in the past in bringing these words to the Indonesian archipelago. This might have provided a unique clue, but I could not find one single Arabic loanword in Indonesian that reflects this rather characteristic Hadrami *j-y* phenomenon, described by Al-Saqqaf (2009:687) as the '*g-yodization, that is, changing *j to /y/ [j]'.² Nor did I find any Arabic loanwords containing a g instead of the classical *jīm*, which would point to southern parts of the Arabian Peninsula as a place of origin (see also Woidich and Zack 2009).

Concerning Versteegh's remarks on a possible Persian influence for the pronunciation of *qāf* as /g/ in words like *gamis* or *gereba*, it should be noted that in Persian the reflex of the Arabic *qāf* (ق) is not /g/ but *ghayn* (غ); that is, it is written *qāf* but pronounced *ghayn*. Most Persian speakers do not differentiate between *ghayn* and *qāf* (Lambton 1963:xviii). It should be added that in various areas of the Arabic-speaking world *qāf* is realized as /g/ (including in Egypt itself, where *jīm* can also be pronounced *jīm*).

In addition, note that the words used as examples by Versteegh are hardly used in contemporary Indonesia, if at all. (This is not to deny that they may have been used in the past.) The word normally used for camel in Indonesian is *unta*, which is of Hindi origin (Jones 2007a). I could not find the word *gamal* in the Malay-Dutch dictionaries of Van der Tuuk or Klinkert, which were both compiled in the nineteenth century, nor in Wilkinson (1903); but it is included in more recent dictionaries. (According to Van den Berg (1886:253) the word *jamal* was rarely used in Hadramaut.)

² Johnstone (1965:235) has noted that the j>y change is not exclusive to a few regions in Hadramaut and Dhofar (Oman), but also occurs elsewhere in the Arabian Peninsula.

The word *gereba* can be found in the list compiled by Russell Jones (1978, 2007a, 2007b), but not in the *Kamus besar bahasa Indonesia* (Pusat Bahasa 2008), nor in the (more extensive) *Kamus lengkap, Indonesia-Inggris* of Alan M. Stevens and E. Ed. Schmidgall-Tellings (2004). In Klinkert's *Nieuw Maleisch-Nederlandsch woordenboek* (1930) it can be found as *kirbat*. The Arabic origin of *gengsi* is considered by Jones (1978) (as well as by Versteegh himself) doubtful. *Gamis* is still in common use today.

The /g/ (classical Arabic /q/) of *gamis* and *gereba* might point in the direction of a colloquial influence, but given the wide geographical spread of /g/ as a reflex for /q/, it is too general to provide any definite regional clue. Similar examples of possible colloquial influence are *gisah* (story: compare classical Arabic *qiṣṣah*; *kisah* in standard Indonesian) in an Ibanic language of Western Kalimantan (Tjia 2007:393); in Jakartan *wagtu* (time: compare classical Arabic *waqt*; standard Indonesian *waktu*), *gomar* (moon: compare classical Arabic *qamar*) and *gahwa* (Abdul Chaer 2009a); and in Betawi *gahwe* (coffee), and *gahar* (anger: compare classical Arabic *qahr*) (Saidi 2007).

Finally, the influence of Hadrami Arabs on the process of Arabic loanwords in Indonesian should not be overestimated, because many among the later generations of the Hadrami community in Indonesia lost their ability to speak Arabic as a result of intermarrying with Indonesians. Most of the offspring adopted the language of their Indonesian mother (Mobini-Kesheh 1999:22; Van Dam 2007a, 2010:212-3).

Coming back to the Arabic *qāf* in a more general sense, note that in Arabic dialects the *qāf* may be pronounced either *qāf*, /g/, *kāf*, *ghayn*, *jīm* or *hamzah* (').

The *qāf* in Arabic loanwords in Indonesian usually becomes /k/.

It would be nonsense, however, to maintain that loanwords starting with /k/ (of which there are not many) have been adopted from Arabic (or other) speakers who pronounced the *qāf* as *kaf* in their dialects. Similarly, it would be far-fetched to conclude that the words written with an initial *qāf* in Arabic and written in Indonesian with *q* (like *qari* or *qiraah*) are due to the dialectal background of speakers pronouncing *qāf* in their dialects (*Qur'ān* could be an exception, but this is in itself a classical word).

Jones's list gives only two examples of *qāf* being reflected by /g/: *gereba* and *gamis* (aside from the obsolete word *geramsut*), and we have just seen that these two words may have a colloquial Arabic origin. All other *qāf*-initial Arabic loanwords in Jones's list start with a *k*. I therefore conclude that *qāf* is pronounced /k/ because that is the normal way for an Indonesian speaker to pronounce *qāf*, simply because it is not part of the phoneme inventory of Indonesian, and the phoneme /k/ in Indonesian is phonetically the nearest alternative. The same applies to various other phonemes, to which I shall come back later.

Words that are unmistakably colloquial

I would like to explore another method by looking for words whose structure or meaning betrays an unmistakable colloquial and, if possible, regional origin.

In modern Indonesian the name for Wednesday (*Hari Rabu*) not only seems to reflect a colloquial background, but also a regional one. In Yemen and parts of Saudi Arabia (not in Oman, however) Tuesday and Wednesday are called *thalūth* and *rabūʿ*. The latter form may have led to the Indonesian *Rabu* (I owe this observation to Peter Behnstedt). One might think of *al-yawm al-rābiʿ* – ‘the fourth day’ – as the origin of *Rabu*, but this would be illogical, considering the forms of the other weekdays. The other weekdays in Indonesian: *Hari Ahad* (classical Arabic: *yawm al-ʾaḥad*), *Senin* (*al-ʾithnayn*), *Selasa* (*al-thalathāʿ*), *Kamis* (*al-khamīs*), *Jumaat* (*al-jumʿah*) and *Sabtu* (*al-sabt*) are all closer to the classical Arabic forms and do not betray any clear dialect influence with the exception of *Senin*, which, in contrast to *Rabu*, has no clear regional origin. The alternative Indonesian form *Isnain*, also found in various dictionaries, is somewhat closer to classical Arabic (*al-ʾithnayn*), but should nevertheless be considered as having a colloquial Arabic background. The classical Arabic diphthong *ay* is usually retained in Indonesian as *ai*, and in this case would probably have resulted in **Isnain* (or **Isnayn*).

I have not yet found other clear examples of colloquial Arabic in official Bahasa Indonesia, except for the words *khalas* or *halas* (finished: compare colloquial Arabic *khalās*) and *fulus* (money: compare colloquial Arabic *fulūs*), all included in the *Kamus Besar Bahasa Indonesia* (Pusat Bahasa 2008; *khalas* was not included in Moeliono 1988).

More examples of colloquial Arabic can, however, be found in Malay dialects.

Betawi has several. In the *Glossari Betawi* of Ridwan Saidi (2007) I found various typically Arabic colloquial loanwords (most of which are not mentioned by Abdul Chaer 2009a, 2009b). For instance: the word *syuf* (look!) is typically colloquial (not classical) Arabic (*shuf* or *shūf*). Other examples of colloquialisms are *harbatè* (making something difficult: compare colloquial Arabic *kharbat*), *kul* (eat!), *taʿal* (come! compare colloquial Arabic *taʿāl*), *yahana* (to pretend or deceive: compare colloquial Arabic *yakhūn* or Hadrami *yakhan-nuh*), *anè* (I: Arabic *ʾanā*) and *ènè* (you [2nd person masculine singular]: Arabic *ʾanta*), both used in polite speech for keeping or creating distance (Grijns 1991:21), *kaʿal* (penis [also a term of abuse]: Hadrami Arabic *kʿāl* (testicles); see also *kiʿl* in Behnstedt 2006:1076), *nus ajus* (middle-aged woman: colloquial Arabic for ‘half’ is *nuṣṣ* [classical Arabic *niṣf*]; *ʾajūz* is ‘old woman’), *regut* (sleep: Arabic *ruqūd*), *tapran* (poor: Arabic *ṭafraṅ*).

Surabayan Malay has several examples of Arabic loanwords that appear to have a colloquial background, like *yukul* (to eat: compare colloquial Arabic

yākul/yōkul), *asrob* (drink!: Arabic *'ashrab*), *zen* (good, nice: compare colloquial Arabic *zēn*), *rejak* (to go home: Arabic *raja'* to return) (Hoogervorst 2008:39, 42).

Jakartan has *emak* (mother: Arabic *'ummak*, 'your mother') (Abdul Chaer 2009a), reflected in Betawi in *emak bumi* (mother earth) (Saidi 2007).

Campbell (2007:340) observes correctly that 'Arabic loanwords mostly function as nouns in Indonesian and Malay, and few Arabic morphological rules have been borrowed'. Various of the Arabic *colloquial* examples cited above indicate however that in those cases the morphology of verbs has been followed or at least imitated (like *syuf*, *kul*, *ta'al*, *asrob*, all being similar to 2nd person singular masculine imperatives, as well as *yahana* and *yukul*, both having a verb structure similar to the 3rd person singular masculine imperfect). Stevens and Schmidgall-Tellings (2004) mention *yalamlam* ('the stage in the pilgrimage where the pilgrims from Yemen assume the pilgrim's garb'), from the Yemeni colloquial Arabic verb *lamlam*, *yilamlim* 'to collect' (Behnstedt 2006:1128), providing an interesting Yemeni context.

All of the above-mentioned colloquial Arabic words are in common use in Hadrami Arabic.³ This does not necessarily imply that all these Arabic colloquial loanwords originate from Hadramaut itself, because most of these words are also in use elsewhere. When taking the history of the Hadrami Arabs in the Indonesian archipelago into account, however, it does not seem to be an unreasonable assumption. Nevertheless, apart from a few specific Hadrami words (which in turn do not occur in standard Indonesian), there is not any linguistic evidence for such a thesis.

The 'classical Arabic loanwords' may not have any clear link with Hadramaut or South Arabia, and should be dealt with as a category different from the 'colloquial loans'.

There are also Arabic loanwords that in Arabic itself are not necessarily colloquial but are nevertheless rejected in standard Indonesian because they are considered to be colloquial in Indonesian. Betawi (Saidi 2007), for instance, has *tajir*⁴ (Arabic for trader: *tājir*) meaning here 'filthy rich', and *zub* (penis [also a term of abuse]: Arabic *zubb*). The *Kamus besar bahasa Indonesia* (*Pusat Bahasa* 2008) does not generally mention such words, presumably because they are colloquial. The more 'classical' *zakar* (penis: Arabic *dhakar*) has however been accepted as standard Indonesian.

³ I owe the above observations on Hadrami Arabic vocabulary to A. Al-Saqqaf.

⁴ Van der Tuuk (1877) also mentions *tajir* as the 'Batavian' pronunciation of (Arabic) *ta'zīr* (punishment; in Bahasa Indonesia *takzīr*), whereas Stevens and Schmidgall-Tellings (2004) interpret *tajir* as a contraction of '*harta banjir*' (very rich). Abdul Chaer (2009a) mentions *tajir* (rich) as Jakarta dialect, but not as Betawi (2009b).

Does spelling follow pronunciation?

Various scholars who have studied Arabic loanwords in Indonesian have given more attention, or attached more importance, to the characteristics of the Arabic language that provided these loanwords, than to the recipient languages and their linguistic particularities, also where their pronunciation is concerned.

Let us have a look at the way Indonesians tend to pronounce various Arabic phonemes, in an effort to establish whether or not Arabic loanwords came via written sources or via people speaking classical or colloquial Arabic (or other languages with loans from Arabic), and whether oral sources can be traced.

I think the pronunciation of various phonemes in Arabic loanwords in Indonesian, such as the *ḍād* (ض), *zā'* (ظ), *zā'* (ز), *'ayn* (ع), *fā'* (ف), *ṣād* (ص), and others, has to do with the way the recipient Indonesian community was and is inclined to pronounce such foreign sounds, and that the pronunciation of the loanword in the source language is less of a factor. An unknown Arabic phoneme was interpreted and pronounced by listeners as the phoneme phonetically nearest to it in their own phoneme inventory (Uhlenbeck 1949:71-82; Adelaar 1985:10-3; Machali 2006a, 2006b).

Even the spelling chosen at a given time may have been decisive for the pronunciation afterwards. Normally, spelling would follow the way a word was pronounced. But Indonesian has various examples that indicate that pronunciation followed spelling. Once a different spelling was introduced, a word came to be pronounced differently as well (Machali 2006a, 2006b; Van Dam 2009c; Vikør 1988).

Let me demonstrate this point with an example of Dutch loanwords. It is clear that the Indonesian names of the months originate from Dutch, and not from English: *Maret* (Dutch: *maart*), *Juni*, *Juli* and *Augustus*. *Januari*, *Juni* and *Juli* were originally pronounced in the Dutch way as *Yanuari*, *Yuni* and *Yuli*. Irrespective of the post-revolutionary spelling rules for Indonesian (according to which former *j* was written *y*), the initial letters for the names of the months remained the same, as a result of which the pronunciation is as it is today (pronounced like the *j* of English *jack*). There was apparently no collective language memory causing people to maintain the original (Dutch) pronunciation. Words without a long-established traditional pronunciation thus appear to be pronounced as they are written, and not necessarily as they were pronounced in the past.

Another example is the name for Friday: at first it was *Juma'at*, *Jum'at* or *Jumaat* (all pronounced more or less the same way), but now the spelling *Jumat* has also become very common, which, although still generally pronounced with a glottal stop as in the past, can after some time be expected to lead to a more general pronunciation in which the original *'ayn* (already in the form of a glottal stop) will be completely lost in writing and subsequently also in pronunciation (resulting in *Jumat*).

Many other Arabic loanwords provide examples of this phenomenon, for instance where the original Arabic *'ayn* has been replaced by /k/. Originally, the /k/ in place of the *'ayn* was meant as a substitute (Indonesians generally could not pronounce the *'ayn*). In this case, /k/ came to be pronounced as a glottal stop, as it is in other Indonesian words (for instance *menunjukkan*, which is usually pronounced *menunju'kan*). Similarly, *menikmati* will originally have been pronounced *meni'mati* (from the Arabic *ni'mah*), particularly by those who knew the Arabic origin of the word. But, perhaps because many Indonesians were not aware of the original Arabic pronunciation, they started to pronounce the word as it was written. Words like *makna* (Arabic *ma'nā*) and *yakni* (Arabic *ya'nī*) are similar cases. The original Arabic pronunciation, or something similar to it, has therefore been lost (if it was ever used). It goes without saying that there are no Arab speakers who have ever pronounced words like *ma'nā* as *makna* in Arabic.

All these examples lead to the conclusion that the sound changes referred to above have little to do with the form in which an Arabic loanword was originally pronounced in Arabic, but instead with the way the recipient Indonesians dealt with it.

I do not attempt to give a comprehensive overview of the way all Arabic phonemes are pronounced, transformed, or transliterated into Indonesian, because this has already been done on many occasions. Nevertheless, I would like to take a closer look at the two emphatic phonemes *ḍād* (ض) and *ẓā'* (ظ), in particular because certain authors (Versteegh 2003, followed by Campbell 2007) have attached special significance to these phonemes as indicators of an earlier layer of loanwords.

The ḍād (ض)

In Klinkert's Malay-Dutch dictionary (1930) written in Arabic characters, all entries starting with *ḍād* are written or transcribed with *dl*: *dloeboe* (hyena), *dloha*, *dlarab*, *dlaroerat*, *dlaif*, *dlalalat*, *dlammah*, *dlamir* (Arabic: *ḍabu'*, *ḍuḥā*, *ḍarab*, *ḍarūrah*, *ḍa'if*, *ḍalālah*, *ḍammah*, *ḍamīr*). But when looking the other way around for the same words in Klinkert's Dutch-Malay dictionary (1926), we get different results: *doeboe'*, *daroerat*, *da'if*, *dalalat*, with no occurrence of the *dl*. One might conclude that *dl* functions as a way of transcribing the *ḍād*, rather than indicating the pronunciation. The actual pronunciation would probably be closer to the version in the Dutch-Malay dictionary, with *d* instead of *dl*. It is in this light that the two versions Nahdatul Ulama and Nahdlatul Ulama should be seen. The pronunciation is generally without the *dl* but with (plosive) *d* or (interdental) *dh*. The *dl* indicates in any case that the original Arabic phoneme must have been a *ḍād*.⁵

⁵ Stevens and Schmidgall-Tellings (2004) mention forms like *dlaif*, *dlamir* and *dluha* (but also of *laif*, *daif*, and *loha*).

Van der Tuuk (1877, 1880, 1884) does not give any clue as to the pronunciation of the *ḍād*, because he does not provide any transcriptions of his Arabic script.

The zā' (ظ)

In Klinkert's Malay-Dutch dictionary (1930), words starting with *zā'* are all written with an emphatic *ḷ*: *lālim*, *lāhir*, *lil*, *lālāmat*, *loeloemat*, *lohor*. (Arabic: *zālim*, *zāhir*, *zill*, *zulāmāh*, *zulmah*, *zuhr*). In Klinkert's Dutch-Malay dictionary (1926), we get more or less the same results: *lalim*, *lahil*, *lil*, *loelamat*, *loehoer*. I conclude from this that *zā'* is indeed pronounced *l* in all these words.

Van der Tuuk gives only two examples of *zā'* in Latin script. He transcribes the *zā'* as a bold printed *ḍl*: *mengḍlanken* ('to have a thought' from *zann*) and *ḍlahr* ('back' from *zahr*). All other words are given in Arabic script only. The question is whether Van der Tuuk transcribed the *zā'* in this way because those who provided him with these words were pronouncing them as such (as a velarized *ḍl*), or whether it was because the Indonesian recipient side (through Van der Tuuk in this case) transcribed these Arabic sounds as such because they did not have any Latin characters to indicate this phonetic feature. I conclude that similar to Klinkert's way of transcribing *ḍād* as *ḍl*, Van der Tuuk's *ḍl* should be seen as a way of transcribing the *zā'*, rather than reflecting its pronunciation in Malay. The same applies to Wilkinson (1903:436), who introduces yet another transcription for the *zā'*, namely *tlâ*.

In those (relatively few) loanwords in which the *zā'* is not pronounced /z/, it is generally pronounced /l/, as noted by Klinkert.

/F/ used to be generally pronounced /p/, because many Indonesians had difficulty pronouncing /f/, because /f/ was in most cases not part of the phoneme inventory of their mother tongues. Words containing an /f/ in Austronesian languages are generally loanwords. For that reason it was not unnatural for the /f/ in Arabic words like *farḍ* to be pronounced /p/ as in *perlu*. Many Indonesians are not aware that *perlu* is a loanword originating from *farḍ* (having the Indonesian equivalent of *fardu*). Klinkert (1930) notes that '*fardl*' (note his spelling) becomes *perloe* in colloquial Indonesian.

With respect to Javanese, Rochayah Machali (2006a, 2006b:55-6) has noted that realizing *ḍād* / *zā'* as /l/ 'becomes compulsory when they occur in final position, as such sounds would not normally occur finally in Javanese', for example *lapal* < *lafz* 'spoken word', *lila* 'sincere' < *riḍā* 'approval', *aral* < '*ard* 'hindrance'. Machali thereby underscores the importance of the role of the recipient language (as does Uhlenbeck 1949:71-82).

The South Arabian connection

In his article 'The Arabic component of the Indonesian lexicon', Kees Versteegh (2003) has put forward a fascinating thesis arguing that the /l/ reflex of the *ḍād* in Arabic loanwords in Indonesian is evidence of a very old *pre-Persianized* layer of borrowings. He classifies the earliest category of loanwords as having come through *the South Arabian connection*. Examples of the /l/ reflex for *ḍād* are: *rela* (from Arabic *riḍā*), *loha* (*ḍuhā*), *melarat* (*maḍarrāh*), *lalalat* (*ḍalālah*), *ilafi* (*'iḍāfi*), *aral* (*'arḍ*) and *perlu* (*fard*).

Jones's list (1978) contains 87 loanwords having a *ḍād* in the Arabic original. Only the above-mentioned seven of these have the /l/ reflex. The others have /d/. Does this mean that the seven words with the /l/ reflex for *ḍād* belong to a ('pre-Persianized') layer which is older than the other 80 words? Below, an attempt is made to answer this difficult question.

Versteegh (2003:223) notes that in Arabic this /l/ reflex for *ḍād*

was probably strongest in the southern part of the Arabian languages, [...] where it may have remained intact for some time in some Arabic dialects spoken in the South as well. One modern Arabic dialect is reported to have preserved *l* as reflex of the Old Arabic phoneme *ḍ*, namely the dialect of Daḥnā. The disappearance of the lateralized realization of *ḍ* and the merger of the two phonemes must have taken place in the earliest period of the Islamic expansion, but in a few cases early loanwords in some of the languages with which Arabic came in touch still show this realization in the form of a reflex *l*.

This applies, for instance, to Spanish, Yoruba, Hausa, and Malagasy (Versteegh 2001) and also to loanwords in Indonesian/Malay and in Javanese, Acehese and Minangkabau [...]. In all of these regions speakers of Arabic from South Arabia, specifically from Oman and Hadramaut, were probably involved' (see also Versteegh 1999, 2006 and Al-Saqqaf 2006:86-7).

Assuming that most Arabs in Indonesia are of Hadrami origin, the linguistic puzzle would fit exactly, for the *ḍād* to be pronounced /l/ in various Arabic loanwords in Indonesian. But after taking a closer look at the dialect geography of South Arabia today, we find this /l/ reflex for *ḍād* only in a few areas close to Hadramaut (such as Abyan and Habban), but not in Hadramaut itself (Vanhove 2009). On the other hand it cannot be ruled out that this was different in the past and that the /l/ reflex for *ḍād* was much more widespread at the time. The additional implication, however, is that the loanwords came to what is now Indonesia at a very early stage, whereas the Islamization of the Archipelago started only centuries after the older /l/ reflex for *ḍād* had already largely disappeared. And most Hadrami Arabs arrived in what is now Indonesia only in the eighteenth and nineteenth centuries.

Versteegh (2003:223) concludes that the ‘reflex of *l* for Arabic *ḍ* is an important argument for the existence of an early layer before Persian influence’, but that ‘unfortunately’ there is also a group of words exhibiting *l* for *zāʾ*, such as *hafal*, *lafal*, *lahir*, *lalim*, *lohor* and *nalar* (Arabic *ḥafz*, *lafz*, *zāhir*, *zālim*, *zuhr* and *nazar*). For the time being, he notes, he does not have an explanation for these seeming counter-examples.

Torsten Tschacher (2009) has found a solution by noting that *dād* and *zāʾ* have identical /1/ reflexes in Tamil,⁶ and that it is probable that Tamil is the source of those Arabic loanwords that exhibit /1/ for /z/:

Most interesting are the reflexes of Arabic /ḍ/ and [/z/], as these often allow one to distinguish loans [in Tamil] borrowed through Indian Ocean networks from those borrowed from northern India. [Arabic loans in Tamil from northern India do not have an /1/ reflex.⁷] The most widespread reflex of both phonemes, common in earlier Islamic Tamil literature, but also widespread in spoken Tamil among Muslims in both India and Sri Lanka, is a lateral, either dental /1/ or retroflex /l/, e.g. *parulu/paru* ‘duty’ (< *farḍ*). A lateral reflex of /ḍ/ is found also in several Southeast Asian and West African languages [...], and in the Tamil context obviously reflects borrowing through Indian Ocean networks. More surprising is that /ḍ/ and [/z/] have identical reflexes, as most languages exhibiting a lateral reflex of /ḍ/ treat the two phonemes differently. Yet, it is possible that there were originally different reflexes of /ḍ/ and [/z/], one represented by /1/ or /l/ and one by the retroflex approximant /l/ [ɭ] [...]. Whatever the case, it is probable that Tamil is the source of those Arabic loanwords in Malay that exhibit /1/ for [/z/].

Only in a relatively small number of Arabic loanwords in Indonesian does /1/ prevail over /d/ (for *dād*) and /z/ (for *zāʾ*), and these should be considered anomalies, or exceptions, not having been coined in that particular shape in the Indonesian archipelago itself.⁸

Of the roughly 30 Arabic loanwords that in their original Arabic form contained a *zāʾ* only one third have the /1/ reflex, and these, according to Tschacher, have probably come through Tamil. Most others have the /z/ reflex.

Only seven of the almost 90 Arabic loanwords have the /1/ reflex of *dād*, whereas most others have /d/. In the Tamil context, this /1/, according to Tschacher, clearly reflects borrowing through Indian Ocean networks. This corresponds with Versteegh’s theory that Arabic speakers from South Arabia were probably involved here.

⁶ There are various Arabic dialects in which *dād* and *zāʾ* have merged into an identical emphatic interdental *dād* (Versteegh 2006; De Jong 2000:60, 249, 332), but this phenomenon is not reflected in Arabic loanwords in Indonesian.

⁷ I owe this observation to Torsten Tschacher.

⁸ I owe this observation to Torsten Tschacher.

Another theory could be that the pronunciation of the *dād* as /l/ in some Arabic loanwords in Indonesian has to do with the fact that the *dād* is not part of the Indonesian phoneme inventory, as a result of which Indonesians replaced it by a phoneme in their own language that is phonetically close to it. But this does not explain why in a large majority of cases /d/ was preferred over /l/.

Why should the /l/ for *dād* have disappeared in the Arab world itself, whereas it has been retained in certain Arabic loanwords in Malay and Indonesian? One possible theory is that there have been other loans with /l/ for *dād* and *zā'* in the past, but that they have been 'updated'. It may be that those loanwords that were specifically recognized as Arabic loanwords were later given a new appearance, closer to classical Arabic. With respect to those loanwords that were not recognized as such (for instance *perlu*), no need was felt to reshape or 'update' them, and they therefore retain their original shape or something closer to it.

In as far as a South Arabian connection can be established at all, it is difficult to ascertain the period or periods in which the loanwords must have been transmitted. The Arabic loanwords having /l/ for *dād* seem to have been transmitted much earlier than the typically colloquial loans from the Arabian Peninsula and Yemen, mentioned above. Whatever the case, a more general and clearly identifiable *Hadrami* connection has so far not been convincingly established.⁹

Pseudo-classical endings in Arabic loanwords

In his article *The Arabic component of the Indonesian lexicon*, Versteegh (2003:222) uses the following argument concerning what at first sight appear to be residues of classical grammatical Arabic case endings in Arabic loanwords:

A special case is that of words in -u/-i such as *napsu* 'lust, passion' (= Arabic *nafs* 'mind, soul'), *salju/salji* 'snow' (= Arabic *ṭalj*), *waktu* 'time; when' (= Arabic *waqt* 'time'), *wahi/wahyu* 'revelation' (= Arabic *wahy*), *abdi/abdu* 'servant' (= Arabic 'abd), *rejeku/rejeki/rezeki* 'sustenance' (= Arabic *rizq*). Some of these may have been borrowed recently by learned people, who knew Arabic and tried to emulate the Arabic case endings. This may apply for instance to *salju*, and it applies almost certainly to *wahyu*. But the possibility should not be excluded that some of these words represent an older layer of loanwords in which the ending -u is the reflex of the Arabic personal suffix of the 3rd person masc. sing. -hu > -o, -u. In the case of *napsu*, for instance, the Indonesian meaning might stem from *nafsu-hu* with the

⁹ Given the strong links in modern history between the 'Arab community' in Indonesia and Hadramaut, it is not surprising that, conversely, Hadrami Arabic has incorporated several Indonesian loanwords, a few of which even turn out to be Dutch loans (Al-Saqqaf 2006:90-1).

presumed reading of '[it is] his mind, intention'. In the case of *perlu* this suggestion might provide an explanation for the grammaticalization of *fardu-hu* '[it is] his duty' > 'he must'. The learned loan *fardu*, which did not undergo this development but simply means 'moral obligation', probably has its ending as the result of recent re-arabization.

Campbell (2007:343), basing himself on Versteegh, makes similar remarks:

A handful of loanwords end in *-u* and/or *-i*, e.g. *napsu/nafsu* 'natural appetite or desire' < *nafsu*, *perlu*, *wahyu*, *salju/salji*. While some scholars have suggested that the *-u* ending is evidence of a South Indian intermediary source, Versteegh (2003) has more plausible explanations, e.g. naive attempts to emulate case endings or a reflex of the 3rd person masculine suffix *-hu*.

Note that the concept of 'naivety' was not introduced by Versteegh, but by Campbell.

Gonda (1973:59) concludes that 'many' Arabic loanwords 'often' betray a secondary Indian origin, but he provides very few examples.

Many [Arabic loan]words [...] often betray their secondary Indian origin by phonetic peculiarities; words like Mal.Jav. etc. *pěrlu* 'obligatory', which derives from the Ar. *fard* 'ordering anything to be observed, obligation', have been extended by *-u* in the Dravidian South of India, where descendants of Arabian merchants and their native converts speak Tamil and related languages [...].¹⁰

Tschacher (2009) notes a similar phenomenon for Arabic loanwords in Tamil:

The common final *-u* (rarely *-i*) [in Tamil] is not a reflex of Arabic endings, but a paragogic vowel added to avoid phonotactically restricted final consonants [...].

This does not necessarily imply, however, that the Arabic loanwords in Indonesian with final *-u*, and occasionally *-i*, must have come via Tamil.

Tadmor (2009) nevertheless suggests that the final *-u* in Arabic loanwords such as those mentioned above

seems to originate neither in Arabic itself nor in the Malay-Indonesian integration pattern. An alternative explanation would be that these forms were learned

¹⁰ From the page of Gonda's study (1973:59) quoted above, Bellwood (2007:139) concludes that 'linguistic evidence suggests that the Arabic and Persian loanwords in Austronesian languages came for the most part directly from India'. Campbell (1966:33), in turn, bases himself on Bellwood and notes that 'Arabic and Persian loanwords are said to have a mainly Indian source'. Gonda himself, however, cites no more than three examples in Malay (*pěrlu*, *ahli sunat* and *abdi*), which is too meagre to serve as 'linguistic evidence' to confirm this thesis.

from native speakers of a Dravidian language such as Tamil or Telugu, where *-u* is regularly appended to consonant-final loanwords. The linguistic evidence therefore supports the theory of the introduction of Islam to the Malay-Indonesian archipelago from southern India.

It seems to be rather far-fetched to base the history of the introduction of Islam to the Malay-Indonesian archipelago on the occurrence of the handful of Arabic loanwords of the *-u* type, whose origin is not even certain. If Tamils had really played such an important role, more Tamil loanwords would be expected. Tadmor's thesis, therefore, remains far from convincing.

Scholars have thus suggested a variety of theories about the origins of the final *-u* in Arabic loans in Indonesian and Malay. In my view, it is more likely that the clue to these final vowels is to be found neither in the original Arabic words, nor in the Tamil language, but rather in the structure of the recipient Malay/Indonesian language. This factor has not been sufficiently taken into account. By doing so, the explanation becomes much simpler and clearer. The endings of words like *salju/salji* and *waktu* can then simply be seen as forms that have been adapted so as to fit into the Malay/Indonesian phonology and syllable structure. As noted by James Sneddon (2003:164-5) in his study *The Indonesian language; Its history and role in modern society*, in 'early borrowings, final clusters were removed [...] by the addition of a vowel, as in *lampu* (lamp) and *pompa* (pump). [...] Early borrowings sometimes added a final vowel, such as *buku* (book).'¹¹

The same phenomenon is visible in the Arabic loanwords mentioned above. Other examples are: *abdu/abdi* (Arabic: 'abd), *amri* ('amr), *ardi* ('ard), *binti* (bint), *hamdu* (hamd), *ibnu/ibni* ('ibn), *ilmu* ('ilm), *kalbu* (qalb), *kasdu* (qaṣd), *mahzi* (maḥd), *nisfu* (niṣf), *parji/farji* (farj), *sabtu* (sabt), *syahru* (shahr), *syamsu* (shams) and (Jakarta) *wagtu* (waqt). Next to *nafsu/nafsi*, we find *nafas* and *napas* (Arabic: nafas) which conform to phonotactic constraints of the Malay/Indonesian language as well. It is remarkable that most of the Arabic loanwords of the *-u/-i* type mentioned here have an identical counterpart in Javanese and Sundanese. It is impossible, however, to tell with any certainty which language borrowed which word from which language in what period and in what sequence.

In Indonesian (as well as in Javanese and Sundanese) all these loans follow the word pattern CvCCv or vCCv (with the exception of *rejeku/rejeki*, because

¹¹ Tadmor (2009) categorizes words like *lampu*, *pompa* and *buku* as having been borrowed via Portuguese Creole, and explains the final *-u* in these words as a 'common phonological strategy in Portuguese Creole but not in Malay-Indonesian'. It is difficult, however, to conclude with any certainty whether or not these loanwords came through Dutch or through Portuguese. According to Jones (2007a) *buku* and *lampu* are Dutch loanwords, whereas *pompa* is ascribed to Dutch or Portuguese.

syllables should in principle not end with *j* in Indonesian¹²). Most of these loanwords (which all have the CvCC pattern in Arabic, including *rejeku* / *rejeku* which derives from Arabic *rizq*) come close to the classical Arabic forms and do not contain any remarkable colloquialisms.¹³ Had these loanwords really been transmitted through spoken Tamil, more deviant forms would be expected. At the same time, however, the possibility cannot be excluded that earlier existing 'deviant' traces were erased later through a process of re-arabization.

The origin of -ah and -at endings: a Persian connection?

Campbell (1996a) has published an important and often quoted study on 'The distribution of *-at* and *-ah* endings in Malay loanwords from Arabic' in *Bijdragen tot de Taal-, Land- en Volkenkunde*. The main aim of his study is to demonstrate that many Arabic loanwords in Malay and Indonesian did not actually come directly from Arabic, but rather through intermediaries who spoke Persianized Indian languages. In Campbell's own words (2007:344):

In a historiographical analysis of the evidence for the coming of Islam to the region, Drewes (1968) cites sources claiming Muslim influence from Bengal, Gujerat and South India. From a linguistic perspective, we need to exercise caution in looking for evidence of North and South Indian languages in the forms and meanings of Arabic loanwords. The most reliable conclusion that can be drawn is that numerous Arabic loanwords have come to the Malay world via languages that had been influenced by Persian, or via Persian itself; this conclusion is made on the basis of the distribution of the *-at* and *-ah* variants of *tā' marbūṭah*, which corresponds to the distribution in Persian or Persianized Indian languages (Perry 1991:158). Campbell argues that the distribution of *tā' marbūṭah* endings proves the existence of an early Persianized loan stock, but warns of the difficulty of untangling the Indian provenance of such words.' (Campbell 1996a:37.)

Campbell (1996a:40-1) ends his article with the following statement:

The interested reader is now invited to consider the historical implications of the fact that a Persianized Arabic origin for many Arabic loanwords can be shown statistically, and to test the general explanation offered here for the distribution of *-ah* and *-at* endings in B[ahasa] M[elayu] and B[ahasa] I[ndonesia].

¹² *Rezeki* apparently follows the Indonesian structure of *rejeku* here. Stevens and Schmidgall-Tellings (2004) mention the alternative forms *riz(e)ki* and *rizqi*.

¹³ *Azmu* (Pegasus: Arabic [al-faras al-] 'a'zam) follows the same word pattern in Indonesian (vCCv), but has a different pattern in Arabic (CvCCvC). *Perlu* deviates from all other -u/-i words in that the word stress is not on the first, but on the last syllable, and the first vowel is not /a/ but *schwa*.

What I want to do here is not to consider the historical implications of the outcome of Campbell's study, however interesting and important these may be, but rather to have another look at his statistical analysis, which is presented as 'the most reliable conclusion' (Campbell 2007:344).

Central to Campbell's research is the compilation of a checklist of *-ah* and *-at* words (for instance, Indonesian *masalah* from Arabic *mas'alah* and *istirahat* from Arabic *'istirāḥah*). He did so by comparing the respective words in two dictionaries and in two studies on Arabic loanwords in Malay and Indonesian: the *Kamus besar bahasa Indonesia* (Moeliono 1988), the *Kamus lengkap* (Hairul and Khan 1982), and the studies of Beg (1979) and Kasimin (1987).

Campbell's checklist was 'intended to be a broadly representative database, rather than an exhaustive one'. The database contains 'a core of 176 items that were attested by all four authorities' (Campbell 1996a:27). Campbell (1996a:40) concludes that

it can be shown statistically that a considerable core of *-at* words must have been borrowed not directly from Arabic but from a Persianized source. Historical evidence of a Perso-Indian rather than Arab origin for early Islam in the Malay world helps to independently confirm this. It is logical to assume that many *-ah* words must have been borrowed from the same source, but this cannot be conclusively shown through the comparative technique.

In other words, the high percentage of *at-* words in Indonesian/Malay coinciding with the same *-at* words occurring in Persian, as mentioned in *Persian vocabulary* by Ann K.S. Lambton (1961), used as Campbell's main source for Persian, leads Campbell to his conclusion. Campbell (1996:39) provides the following table to support his argument:

	Malay <i>-at</i>	Malay <i>-ah</i>	Malay contradictory
Persian <i>-e</i>	9	46	13
Persian <i>-at</i>	41	8	26

One might have thought that Campbell would make use of the extensive, much more comprehensive, list of loanwords provided by Jones (1978). Campbell (1996:26) gives, however, the following explanation for not using it:

It may be asked why Jones (1978) was not used as a source, given its comprehensiveness. In fact, its comprehensiveness, systematic lexicographic approach and range of sources argued against its use; the value of Beg and Kasimin lies in the differences between them, and what these reveal about the instability and uncertainty of Arabic loanwords in Malay.

Campbell preferred a less comprehensive list composed of words containing an element of instability and uncertainty, or irregularity. Accepting only the similarities among the 'instability and uncertainty' cannot be expected, however, to produce any certainties or regularities. How Campbell thought he could subsequently carry out a reliable statistical analysis on such a weak basis is not clear to me.

That made me curious to find out what the outcome would be if I compared Jones's comprehensive list with the *Comprehensive Persian-English dictionary* by Francis Joseph Steingass (2007), which is much more extensive than the *Persian vocabulary* (Lambton 1961) used by Campbell. This comparison results in a picture quite different from Campbell's:

	Indonesian -at	Indonesian -ah
Persian -e	5	46
Persian -at	151	107

The similarity between Campbell's statistics mentioned above, and my statistics (based on Jones and Steingass) is that the number of words having -at in both Persian and Indonesian is very high indeed (39% in Campbell, when leaving the contradictory items out of consideration, and even higher in Jones/Steingass: 49%). The number of words ending in Persian with -e and having -at in Indonesian is very low (Campbell: 9% and Jones/Steingass: 2%). This would theoretically support Campbell's thesis.

The striking difference is that when comparing the lists of Jones and Steingass the number of words having -at in Persian but -ah in Indonesian is also quite high in Jones (49%, as compared to 8% in Campbell). This would appear to indicate much more direct borrowing from Arabic.

As Campbell's study was mainly based on modern works (all dating from 1979 and the 1980s), I thought it might be useful to compare his findings with some older dictionaries, dating from a period in which modern Bahasa Indonesia was not yet officially established. Such an approach would also exclude various modern influences, such as 'language building' of Bahasa Indonesia and the introduction of all kinds of loanwords by Indonesian language commissions and the later influence of pesantren.

With this background in mind, I used the older Malay-Dutch dictionaries of Van der Tuuk and Von de Wall (1877, 1880, 1884), as well as those of Klinkert (1926, 1930).

These dictionaries provide us with yet another, and completely different, picture of the proportion of -ah and -at loans.

Van der Tuuk and Von de Wall note that:

979 words end with *tā' marbūṭah*. Of these 979 words

354 are indicated as ending with *-ah*;

598 are written only in Arabic script as a result of which the pronunciation of the ending cannot be established with certainty;

23 are indicated as ending with *-at*; and

4 have both *-ah* as well as *-at* ending.

If we leave the 598 words in only Arabic script out of consideration (even though they are probably considered by the authors to end with *-ah*), some

93 % of the remaining 381 words end with *-ah*,

6 % end with *-at* and

1 % end with *-ah / -at*.

Klinkert, on the other hand, notes a total of:

385 words ending with *tā' marbūṭah*.

Of these 385 words

318 are indicated as ending with *-at*, and

67 with *-ah*.

This means that

83% of the words with *tā' marbūṭah* end with *-at* and

17% with *-ah*.

The picture that emerges from my research is that the two dictionaries provide opposite outcomes as to the transcription or pronunciation of the *tā' marbūṭah*. Can we conclude anything from this? I presume that Klinkert as well as Van der Tuuk and Von de Wall all did a painstaking job, trying to be as accurate as possible. Although the two sets of results contradict one another to a very great extent, I do not think it is a matter of one of the dictionaries being mostly 'correct' and the other being mostly 'wrong'. Let me add, however, that a person with a strong background in the Arabic language (like Van der Tuuk) would tend to favour Arabic forms when compiling a dictionary, and therefore would prefer *-ah* over *-at*, when having to choose. Klinkert did not have an Arabist background.

The main conclusion which presents itself to me is that 'the issue of *-at* or *-ah*' had not yet crystallized in a written form in the nineteenth century at the time of Van der Tuuk/Von de Wall and Klinkert, as it exists today. The present-day situation in Indonesian, therefore, is not necessarily the result of centuries of linguistic contacts with language sources from the outside world, but may rather be the result of choices made more recently, particularly in the twentieth century, including by the relevant language committees.

When I combine the outcome of my findings in the dictionaries of Van der Tuuk/Von de Wall and Klinkert with the lists provided as an appendix

in Campbell's article, only 12 of the 176 words definitively share the same *-ah* or *-at* endings. This, however, is much too small a sample to provide a basis for reliable conclusions.

Irrespective of all the numbers and statistics mentioned above, my conclusion about Campbell's research is that his statistical analysis is based on grounds which are not reliable enough. The sources he used cannot really be used as reliable linguistic historical sources, even apart from the fact that Campbell deliberately chose a corpus of words that contained an element of instability and irregularity.

In addition, the *Kamus besar bahasa Indonesia* (Pusat Bahasa 2008), for instance, is much more a reflection of what its editors want the Indonesian language to be, than a reflection of the linguistic reality of current Indonesian, let alone that of the historical past. It is much more a prescriptive dictionary, even when looking into the past, than a descriptive one (Van Dam 2009a, 2009b). Using the *Kamus besar bahasa Indonesia*, therefore, introduces yet another element of uncertainty, weakening the basis of present linguistic historical research.

More generally, the question may be posed as to how reliable dictionaries can be as documents of the history of a language. It cannot always be determined which words entered the language in which period in which form and in which regions, to what extent certain varieties were influenced by particular informants, and so on.

Furthermore, statistical analysis itself is not sound enough to provide any proof. The examples given are not really always comparable, and may date from different periods.

Another essential factor is that there is not any substantial reference group of Arabic loanwords with an undeniably demonstrated Persian origin (Jones 1978:v-vi, xxxvi-xxxvii, 2007:xxvi; see also Azad 2008). This implies that there is not enough ground here on which to base a sound linguistic theory. Finally, the number of words seems too small, compared to the whole vocabulary, to form a basis for reliable statistical analysis (Van Dam 2009ab). The question of how representative the examples are, therefore, cannot be answered satisfactorily.

To conclude: I have not yet seen any convincing linguistic or statistical evidence to support Campbell's thesis on the influence of the spoken Persianized Indian languages, which is not to deny that it might be true from a historical perspective (see also Campbell in Jones 2007a:xxiii-xxiv; Steenbrink 2006; Van Dam 2008).

I want to make some final remarks on the *-ah* ending.

First it should be noted that the /h/ which is very clearly pronounced in Indonesian as *-ah* has little or nothing to do with the Arabic pronunciation of the *tā' marbūṭah*, because it is generally not pronounced in Arabic at the end of a word (except in construct state), although it is present in many dialects

of the Arabian Peninsula (Jastrow 1980:121).¹⁴

In general, *-ah* should rather be considered a reading sign. Therefore, the pronunciation of the /h/ should be considered a peculiarity of Indonesian in the capacity of recipient language. Maybe it is a result of the efforts of Islamic scholars or others in Indonesia who imagined that the *tā' marbūṭah* should be pronounced /h/. But most Arabs would not pronounce the /h/ in a word like *Makkah*, *al-Madinah*, or *mas'alah*, or whatever other example, as Indonesians do. The Indonesian pronunciation must be a result of the way the words ending with a *tā' marbūṭah* are written, not of how they are generally pronounced in Arabic.

Secondly, there are some 'exceptions' in Indonesian in words in which the *tā' marbūṭah* is only written and pronounced /a/. Words like *dakwa*, *gereba*, *hiba*, *jauza*, *kahwa*, *marhala*, *maskhara*, *menara*, and *mesara* (allowance) could actually be considered to be 'normal', because they conform to the original pronunciation in Arabic, in which the *tā' marbūṭah* is not pronounced *ah*, whereas the majority of so-called 'normal' words (ending with a *tā' marbūṭah* pronounced *-ah*) should theoretically be considered to be exceptions.

Finally, there are some Indonesian words that end with an *-ah* that is not based on a *tā' marbūṭah* but on a *hamzah*. The Arabic word *zinā'* (adultery), ending with a *hamzah* is, for instance, written in Indonesian both as *zinah* and *zina* (or *jinah* and *jina*). *Zinah* (and *jinah*) may have appeared by way of a mistake, although it could also be interpreted as a kind of 'hypercorrection'. And it may also point in the direction of colloquial Arabic, because in various dialects the 'classical' final *-ā'* has merged with *-ah*.¹⁵ Other examples are *kibriyah* (arrogance, from Arabic *kibriyā'*) and *wabah* (epidemic, from Arabic *wabā'*).

One might also consider an alternative method to detect some of the influence of Persian or Persianized languages in loanwords in Indonesian that are linguistically of Arabic origin: by paying more attention to words or word forms which, although Arabic according to dictionaries, are not commonly used by Arabs, but rather point in the direction of Arabic loanwords in Persian or Persianized languages.

Gonda (1972:59) has noted that 'in such expressions as the Malay *ahli sunnat* "an observer of the traditional law" the *-i* is Persian'. This *-i* appears to be a clear reflection of the Persian *ezāfe* and occurs in a series of combinations with *ahli*, such as *ahli bait*, *ahli kitab*, *ahli kubur*, *ahli nikah*, *ahli nujum*, and *ahli suluk*. In all these combinations the definite article before the second word has

¹⁴ In some dialects of the Arabian Peninsula, for instance in Dōsirī, the *tā' marbūṭah* is occasionally pronounced *-at* (even when not in construct state), like in '*ghawat*' (coffee) and '*ilḡārat*' (predatory incursion) (Jastrow 1980:121). There is no relationship, however, with the Arabic *-at* loanwords ending in *-at* in Indonesian.

¹⁵ De Jong 2000:255-7, 341, 509, 538, provides various examples.

been dropped from the original Arabic equivalent, as is the case in Persian.¹⁶ Several Indonesian expressions have been borrowed directly from Arabic, as can be seen in those cases where the definite article has been preserved, for instance, *ahlulbait* (Arabic: *'ahl al-bayt*), *ahlulkitab* (*'ahl al-kitāb*), *ahlulkubur* (*'ahl al-qubūr*), *ahlulnikah* (*'ahl al-nikāḥ*), *ahlunnujum* (*'ahl al-nujūm*) and *ahlulsuluk* (*'ahl al-sulūk*). It is not clear, however, whether the latter forms, which are closer to Arabic, were incorporated into Malay/Indonesian at a later stage than their Persian counterparts. The number of examples with *ahli* is rather restricted, and I did not find similar examples reflecting a clear Persian origin.

Gonda (1973:59) maintains that the *-i* in *abdi* 'slave' (Arabic: *'abd*) is the same Persian *-i* as that in Malay *ahli sunnat*. But the Persian *ezafe* cannot occur independently at the end of a word without a subsequent qualifier, and Gonda's thesis concerning the *-i* in *abdi* therefore must be incorrect.

Jones (1978:vi) has warned that 'one can decide that a given Arabic loan probably came to Indonesian via Persian' only *occasionally*, and he gives the example *haḍm*, which in Persian came to be pronounced *hazm*, and was later transformed into Indonesian as *hajam* (indigestion), because /z/ in Indonesian often changes into /j/. At first sight this looks like a promising method, but when checking Jones's list, the change from *ḍād* to /j/ occurs only in two of the 87 words that originally had a *ḍād* in Arabic: in *hajam* (which is now virtually obsolete) and in *jamin* (to guarantee, from Arabic *ḍāmin* 'guarantor'). Two other examples are *wazeh* (plain proof, from Arabic *wāḍiḥ*) and *mahzi* (pure, genuine, from Arabic *maḥḍ*), which may also have come through Persian (*vazeh* and *mahz*), be it that the change from *ḍād* to /j/ has not taken place here. In most other cases *ḍād* was transformed into /d/, with some exceptions where it resulted into /l/, as mentioned above.

If the Persian connection had really been that important, one would expect at least some further examples of the *ḍād* -/j/ shift. And if *-at* words containing the *ḍād*, like *darurat* and *mudarat*, had really come through Persian or Persianized languages, one would also expect an outcome like **zarurat* / *jarurat* or **mazarat* / *majarat*, whereas the original Arabic characteristics apparently prevailed by maintaining /d/.

¹⁶ When exploring the alternative theoretical possibility of these combinations being direct loans from Arabic, it could be argued that a loanword like *ahli nujum* might have entered Malay/Indonesian as **ahl innujum* (including the definite article *al-*, which in this view would have merged with an Arabic 'sun letter' like the *nūn* of *nujūm*). Since double consonants do not fit into Indonesian phonology, the length of the 'double' *n* would have been reduced, resulting in *ahli nujum*. But such a theory is undermined by other combinations like *ahli bait*, *ahli kitab* and *ahli kubur*, in which the second Arabic word starts with a 'moon letter', as a result of which the definite article (*al-*) would be expected to remain visible. Examples like *ahlulnikah* and *ahlulsuluk* clearly indicate that these words have been transferred through written Arabic, because the *l* of the definite article has been maintained as *l* in the case of 'sun letters'.

It is unrealistic to expect that it would really be possible to calculate the percentage, or even to give a rough estimate, of Arabic loanwords that have a Persian connection, because most of the Arabic loanwords – if having a Persian link at all – have a purely Arabic form. Jones (2007a:xxvi) therefore concludes that the ‘problem of deciding which loan-words [...] came directly from Arabic, and which came from Persian, is not really soluble’. And it was not without reason that Jones (1978:v-vi) decided to treat both Arabic and Persian in one list without attempting ‘to decide which of the two languages was the *direct* donor of a particular word to Malay/Indonesian’.

References

- Abdul Chaer
2009a *Kamus dialek Jakarta*; Edisi revisi. Depok: Masup Jakarta. [First edition 1976.]
2009b *Kamus ungkapan dan peribahasa Betawi*. Depok: Masup Jakarta.
- Adelaar, K.A.
1985 *Proto-Malayic; The reconstruction of its phonology and parts of its lexicon and morphology*. PhD thesis, Leiden University.
- Al-Saqqaf, Abdullah Hassan
2006 ‘The linguistics of loanwords in Ḥaḍramī Arabic’, *International Journal of Bilingual Education and Bilingualism* 9-1:75-93.
2009 ‘Wādī Ḥaḍramawt Arabic’, in: Kees Versteegh (ed.), *Encyclopedia of Arabic language and linguistics*. Vol. 4, pp. 687-99. Leiden: Brill.
- Azad, Mohammad Kh.H.
2008 *The background of Iran-Indonesia relations*. Embassy of the Islamic Republic of Iran, Jakarta, 2 November. [Contains a list of ‘Persian vocabulary in Melayu’.]
- Beg, Muhammad Abdul Jabbar
1979 *Arabic loan-words in Malay; A comparative study*. Kuala Lumpur: University of Malaysia Press.
- Behnstedt, Peter
2006 *Die nordjemenitischen Dialekte; Teil 2: Glossar; Fā’-Yā’*. Wiesbaden: Reichert.
- Bellwood, Peter
2007 *Prehistory of the Indo-Malaysian archipelago*. Revised edition. Canberra: Australian National University. [First edition 1985.]
- Berg, L.W.C. van den
1886 *Le Hadhramout et les colonies arabes dans l’archipel indien*. Batavia: Imprimerie du Gouvernement.
- Campbell, Stuart
1996a ‘The distribution of -at and -ah endings in Malay loanwords from Arabic’, *Bijdragen tot de Taal-, Land- en Volkenkunde* 152:23-44.
1996b ‘The Arabic element in Indonesian; What do students need to know about it?’. <http://your.usc.edu.au/wacana/4/campbell.html> (accessed 4-7-2010).

- 2007 'Indonesian/Malay', in: Kees Versteegh (ed.), *Encyclopedia of Arabic language and linguistics*. Vol. 2, pp. 340-5. Leiden: Brill.
- Dam, Nikolaos van
2007a 'Arabic language in contemporary Indonesian (1-2)', *The Jakarta Post*, 5-6 October.
- 2008 Review of Russell Jones (general editor), C.D. Grijns and Jan W. de Vries (eds), *Loan-words in Indonesian and Malay*. Leiden: KITLV Press 2007, in: *Bijdragen tot de Taal-, Land- en Volkenkunde* 164:305-8.
- 2009a Review of *Kamus besar bahasa Indonesia Pusat Bahasa*; Edisi keempat. Jakarta, *The Jakarta Post Sunday Edition*, 15 February.
- 2009b 'Bahasa! "Kamus besar yang agak terlalu sempurna"', *Majalah Tempo*, 30 March-5 April.
- 2009c 'Bahasa! "Kata serapan Arab; Apakah ejaan mengikuti pelafalan?"', *Majalah Tempo*, 15-21 June.
- 2010 'Bahasa Arab di Indonesia kontemporer', in: *Membaca Takdir; Pemikiran dan jejak STA*, pp. 209-15. Jakarta: Dian Rakyat.
- Drewes, G.W.J.
1968 'New light on the coming of Islam to Indonesia?', *Bijdragen tot de Taal-, Land- en Volkenkunde* 124:433-59.
- Gonda, J.
1973 *Sanskrit in Indonesia*. Second edition. New Delhi: International Academy of Indian Culture. [First edition 1962.]
- Grijns, C.D.
1991 *Jakarta Malay; A multidimensional approach to spatial variation*. Vol. 1. Leiden: KITLV Press. [Verhandelingen 149.]
- Hairul, Awang Sudjai and Yusoff Khan
1982 *Kamus lengkap*. Selangor: Pustaka Zaman.
- Hogervorst, Tom Gunnar
2008 'Basa Jawa Surabayaan; Describing Surabaya's linguistic ecology'. MA thesis, University Leiden.
- Jastrow, Otto,
1980 'Die Dialekte der arabischen Halbinsel', in: Wolfdietrich Fischer und Otto Jastrow, *Handbuch der arabischen Dialekte*, pp. 103-39. Wiesbaden: Harrassowitz.
- Johnstone, Thomas M.
1965 'The sound change j>y in the Arabic dialects of peninsular Arabia', *Bulletin of the School of Oriental and African Studies* 28-2:233-41.
- Jones, Russell
1978 *Arabic loanwords in Indonesian; A check-list of words of Arabic and Persian origin in bahasa Indonesia and traditional Malay, in the reformed spelling*. London: School of Oriental and African Studies.
- 2007a (general ed.) *Loan-words in Indonesian and Malay*. Compiled by the Indonesian Etymological Project. Leiden: KITLV Press. [Editors: C.D. Grijns, J.W de Vries.]
- 2007b (ed.) Website: *Loan-Words in Indonesian and Malay*. Compiled by the Indonesian etymological project. <http://sealang.net/indonesia/lwim>, <http://sealang.net/indonesia/lwim.old> (accessed 4-7-2010).

- Jong, Rudolf de
2000 *A grammar of the Bedouin dialects of the Northern Sinai Littoral; Bridging the linguistic gap between the eastern and western Arab world.* Leiden: Brill.
- Kasimin, Amran
1987 *Perbendaharaan kata Arab dalam bahasa Melayu.* Bangi: Penerbit Universiti Kebangsaan Malaysia. [Siri Islam dan Pembangunan Malaysia.]
- Klinkert, H.C.
1926 *Nieuw Nederlandsch-Maleisch woordenboek.* Derde druk herzien en bijgevoerd door C. Spat. Leiden: Brill. [First edition 1885.]
1930 *Nieuw Maleisch-Nederlandsch woordenboek met Arabisch karakter, naar de beste en laatste bronnen bewerkt.* Vierde druk. Leiden: Brill. [First edition 1885.]
- Lambton, Ann K.S.
1961 *Persian vocabulary.* Cambridge: Cambridge University Press.
1963 *Persian grammar.* Cambridge: Cambridge University Press.
- Machali, Rochayah
2006a 'Javanese', in: Kees Versteegh (ed.), *Encyclopedia of Arabic language and linguistics.* Vol. 2, pp. 472-7. Leiden: Brill.
2006b 'Cases of linguistic assimilation and etymological doublets in Javanese words of Arabic origin', *Pacific Linguistics* E6. <http://pacling.anu.edu.au/series/SEALS-PDFs/SEALS-16.pdf> (accessed 4-7-2010). [Edited by Paul Sidwell and Uri Tadmor.]
- Mobini-Kesheh, Natalie
1999 *The Hadrami awakening; Community and identity in the Netherlands East Indies, 1900-1942.* Ithaca, NY: Southeast Asia Program, Cornell University. [Studies on Southeast Asia 28.]
- Moeliono, Anton M. et al.
1988 *Kamus besar bahasa Indonesia.* Jakarta: Balai Pustaka.
- Perry, John R.
1991 *Form and meaning in Persian vocabulary; The Arabic feminine ending.* Costa Mesa, Calif.: Mazda. [Persian Studies Series 12.]
- Pusat Bahasa
2008 *Kamus besar bahasa Indonesia Pusat Bahasa.* Edisi ke-empat. Jakarta: Departemen Pendidikan Nasional, Gramedia Pustaka Utama. [Originally published as *Kamus besar bahasa Indonesia*, Jakarta: Balai Pustaka, 1988.]
- Saidi, Ridwan
2007 *Glossari Betawi.* Jakarta: Betawi Ngeriung.
- Sneddon, James
2003 *The Indonesian language; Its history and role in modern society.* Sydney: University of New South Wales Press.
- Steenbrink, Karel
2006 'Indonesia', in: Kees Versteegh (ed.), *Encyclopedia of Arabic language and linguistics.* Vol. 2, pp. 330-3. Leiden: Brill.
- Steingass, F.
2007 *A comprehensive Persian-English dictionary, including the Arabic words and phrases to be met with in Persian literature: Being Johnson and Richardson's Persian, Arabic and English dictionary revised, enlarged and entirely reconstructed.* London: Routledge and Paul. <http://dsal.uchicago.edu/dictionaries/steingass/> (accessed 4-7-2010).

- Stevens, Alan M. and E. Ed. Schmidgall-Tellings
2004 *Kamus lengkap, Indonesia-Inggris; A comprehensive Indonesian-English dictionary*. Athens, OH: Ohio University Press.
- Tadmor, Uri
2009 'Loanwords in Indonesian', in: Martin Haspelmath and Uri Tadmor, *Loanwords in the world's languages; A comparative handbook*, pp. 686-716. Berlin/New York: De Gruyter.
- Tjia, Johnny
2007 *A grammar of Mualang; An Ibanic language of Western Kalimantan, Indonesia*. [Utrecht]: LOT. [PhD thesis, University Leiden.]
- Tschacher, Torsten
2009 'Tamil', in: Kees Versteegh (ed.), *Encyclopedia of Arabic language and linguistics*. Vol. 4, pp. 433-6. Leiden: Brill.
- Tuuk, H.N. van der
1877-84 *Maleisch-Nederlandsch woordenboek op last van het Goevernement van Nederlandsch-Indie, samengesteld door wijlen H. von de Wall, en, met weglating van al het overtollige, uitgegeven door H.N. van der Tuuk*. Batavia: Landsdrukkerij. Three vols.
- Uhlenbeck, E.M.
1949 *De structuur van het Javaanse morpheem*. Bandoeng: Nix. [Koninklijk Bataviaasch Genootschap van Kunsten en Wetenschappen, Verhandelingen 78.]
- Vanhove, Martine
2009 'Yemen', in: Kees Versteegh (ed.), *Encyclopedia of Arabic language and linguistics*. Vol. 4, pp. 750-8. Leiden: Brill.
- Versteegh, Kees
1997 *The Arabic language*. Edinburgh: Edinburgh University Press.
1999 'Loanwords from Arabic and the merger of *d/d*'. *Israel Oriental Studies* 19:273-86.
2001 'Arabic in Madagascar', *Bulletin of the School of Oriental and African Studies* 64-2:177-87.
2003 'The Arabic component of the Indonesian lexicon', Lilie Suratminto and Munawar Holil (eds), *Rintisan dalam kajian leksikologi dan leksikografi; Laporan Sanggar Kerja Internasional tentang Leksikologi, Depok, 16-17 Desember 2002*, pp. 216-29. Depok: Fakultas Ilmu Pengetahuan Budaya Universitas Indonesia, [Kuala Lumpur]: Dewan Bahasa dan Pustaka. [Publikasi 2.]
2006 'Dād', in: Kees Versteegh (ed.), *Encyclopedia of Arabic language and linguistics*. Vol. 1, pp. 544-5. Leiden: Brill.
- Vikør, Lars S.
1988 *Perfecting spelling; Spelling discussions and reforms in Indonesia and Malaysia, 1900-1972*. Dordrecht: Foris. [KITLV, Verhandelingen 133.]
- Wilkinson, R.J.
1903 *A Malay-English dictionary*. Singapore: Kelly and Walsh.
<http://nla.gov.au/nla.gen-an3363573> (accessed 4-7-2010).
- Woidich, Manfred and Liesbeth Zack
2009 'The g/ġ-question in Egyptian Arabic revisited', in: Enam Al-Wer and Rudolf de Jong (eds), *Arabic dialectology; In honour of Clive Holes on the occasion of his sixtieth birthday*, pp. 41-60. Leiden/Boston: Brill. [Studies in Semitic Languages and Linguistics 53.]