A recently published article on the Chaozhou dialect of Min spoken in Singapore states that the word sak⁷ buŋ² “soap” (sap⁷ bun² in other Min dialects of China and Taiwan) is a borrowing of Malay sabun (Li Yongming 1991:58). This claim may very well be true as far as it goes. But in directing our attention to only one small corner of the globe, it overlooks the interesting position occupied by the Min and Malay terms within the much larger and very colorful global mosaic to which Min and Malay have contributed only a few of the pieces. A close look at the lexeme “soap” in languages around the world reveals that many languages have inherited or borrowed the same word, viz., many of the modern forms meaning “soap” are ultimately derived from the same etymon, Proto-Indo-European *solb-on “soap”. This remarkable transmission of “soap” into the world’s languages has now reached global proportions, so that it seems to be one of the world’s most widely borrowed words of material culture.

Cultural transmission is one of several mechanisms for introducing new words into a language. When a society adopts a new cultural object and an activity associated with it, the language of the borrowing society may follow one of two processes (or even both) to name the new object and activity. i.e., phonetic transliteration or semantic translation. In the first process the borrowing language uses its own phonetic resources to transliterate the name of the new object and activity as they are pronounced in the donor language. Japanese well illustrates this process. On the inside back cover of a Japanese-English dictionary for Japanese students of English (Nakajima 1985), there is a colorful page labelled “Games and Sports” with cartoon-like figures of boys and girls engaged in various activities depicting nine kinds of games and sports. Beneath the pictures the names of these games and sports have been transcribed in both Japanese (in kana and kanji) and English; the Japanese terms can be divided into the following two groups on the basis of their origin:

125
1. English loanwords:
   (1) saikuringu < "cycling"
   (2) suke-tobodo < "skateboard"
   (3) furizubī < "frisbee"
   (4) sāfuin < "surfing"

2. Japanese words
   (1) henka = "fight"
   (2) sakunatsuri = "fishing"
   (3) nawaitobi = "jump rope"
   (4) onigokko = "tag"
   (5) umatobi = "leapfrog" 1

The group of English loanwords comprises the transliterated names of two objects and two activities associated with objects which have been adopted from English-speaking, American culture (aside from the question of who invented the bicycle, skateboard, frisbee, surfboard, etc., America is probably responsible for the commercial development of these various instruments of sport). The group of Japanese words includes two activities performed with instruments, i.e., fishing line and rope, which probably have a very long history and an independent origin within Japanese culture.

Standard Chinese, on the other hand, has generally preferred to use its own lexical resources to translate semantically the names of new inventions borrowed from the West; cf. the following two examples:

huō "fire" + che "cart" → huōche "steam-powered locomotive, train"

chū "out" + zū "rent" + qīche "car" → chūzūqīche "car for renting out" = "taxi"

In contrast, Hong Kong Cantonese, which has been in intimate contact with English over the past 150 years, more resembles Japanese by favoring the phonetic transliteration of English loanwords; cf. tiksi "taxi", fillám "film", tsykulik "chocolate".

Lexical borrowing via phonetic transliteration is a common, pervasive phenomenon among languages in contact. I have been intrigued by the possibility that a loanword could be passed along from one language to another so that eventually it has girdled the globe. On the basis of both cognate and loan relationships between an etymon and its modern reflexes, which etymon has the greatest number of modern reflexes in the world's languages? If we consider all modern reflexes related through cognition and borrowing as being the "same" word, the question can be rephrased as follows: What is the world’s most widely-transmitted etymon? Which etymon displays the most widespread geographical distribution across the world’s languages? Finally, can a loanword’s transmission among borrowing

1 Literally "leap-horse". [Ed.]
languages exhibit global dimensions? I think the answer to the first three questions is Proto-Indo-European *soiβ-on “soap”, and the answer to the last question is “yes” with cognate and loan reflexes of *soiβ-on demonstrating how this is so.

This discussion begins with words for soap in the Min dialects of Chinese which are phonetically quite different from those in other neighboring Chinese dialects, and therefore set themselves apart as a unique group. Several publications on the Min dialects have attributed the origin of the Min forms to the borrowing of Malay sabun “soap”, e.g. Li Yongming 1991:58; Hanyu Fangyan Cihui (Lexicon of Chinese Dialects), Beijing University 1964:153; and Yuan 1983:299. There is no question that the Min and Malay forms are so phonetically similar that they must be related in some way. However, as will be described below, the etymological story of soap has attained global dimensions: there are many more links in the linguistic chain besides this one between Min and Malay.2

The Appendix (Global Distribution of Modern Reflexes for SOAP) indicates that the phonetic shape sabun which is phonetically similar to forms in the Min dialects appears in a number of genetically related and unrelated languages that encompass the globe. The modern forms for soap listed in the Appendix can be classified under four main phonetic shapes, viz., sabu1, sabu2, sabu, sop (and their variants). One etymon underlies these modern reflexes in the languages of the Indo-European, Uralic, Semitic, Niger-Congo, Dravidian, Altaic, Sino-Tibetan, Austroasiatic, Austronesian, Kada, Amerind, and Indo-Pacific families. According to Buck (1949:453), the etymon of “soap” is Germanic *saip(1)ōn; he claims the ancient Greeks and Romans did not possess soap but borrowed it and its name from Germanic tribes. Watkins (1985:56) defines the Germanic root as “dripping thing, resin” and states that it referred to “a reddish hair dye used by Germanic warriors to give [themselves] a frightening appearance.” From this Germanic root Watkins has reconstructed Proto-Indo-European *soiβ-on.

Similar forms for “soap” have passed into a number of languages distributed across Europe, Africa, the Middle East, South, East, and Southeast Asia, the Pacific including Australia and Polynesia, and North America. The precise details associated with the adoption of the word for soap by so many languages must be quite complex; but it is reasonable to believe that the mode of earliest lexical transmission was via trade contacts

---

2 In addition to “soap”, the lexemes “chocolate” and “coffee” may also belong to the pantheon of loanwords which have achieved the status of global distribution.
among Europeans, Arabs, Indians, and Asians. Some of the forms from languages distributed across Asia may be related to the arrival into this area of early European sailor-explorers, probably Portuguese, who carried soap with them and introduced it to the peoples with whom they came in contact. Portugal’s colonial presence in Asia began in the early 16th century, with the establishment of Goa on the west coast of India in 1510 and Macao on the southeast coast of China in 1557. Portugal’s first diplomat to Siam arrived in 1511, and a contingent of 120 Portuguese mercenarites were employed as bodyguards and military advisers by King Chai Racha of Ayudhya (ruled 1534-1447) (Wyatt 1984:88-9).

In attempting to identify the donor languages from which the loanwords were borrowed, we need to keep in mind the varying phonetic shapes of words in the possible donor and borrowing languages. Our assumption is that the phonetic shape of the loanword will be as similar as possible to that of its source. On the basis of their phonetic shapes, loanwords for “soap” in Asian languages can be divided into two types, sabu and sabun, one with an open second syllable and one with a second syllable closed by a dental nasal consonant. We can see that among the modern European languages, Portuguese sabão with its nasalized second syllable is phonetically closest to sabu(n). The Portuguese etymology has already been adduced for Siamese sabū: (Haas 1964:521), for Akha saⁿbyaⁿ, and Mpi saⁿbuⁿ, two Loloish languages spoken in Thailand (Bradley 1979:332-3), and for Lahu ᵏṭ̥̄p̥̄ (Matisoff 1988:1168). We can also compare Cambodian saⁿbū:, Lao saⁿ buⁿ, and Dai-Dehong saⁿuⁿ pēu³.

The words for “soap” in Arabic, Malay, and Min terminate in a dental nasal stop consonant. Malay sabun may have been borrowed from Arabic or some Indo-Iranian language: cf. Arabic sabuun, Nepali sābun, Hindi/Urdu saⁿbun, Persian sābun. Arabic in turn could have borrowed its term from Greek or a language belonging to the Romance or Indo-Iranian branches. Among the Altaic languages, we find Turkish sabun, Uighur sapun, Kazakh saben, Tartar saben, Kirghiz samen (with nasalization of the bilabial stop). Although Japan’s first European contacts in the Japanese islands began with Portuguese sailors who arrived in Japan in 1543 (and were followed soon after by St. Francis Xavier who landed in Kyushu in 1549 to embark upon his mission of Christianizing the local population), the phonetic shape of Japanese sabon is closer to the Malay, Arabic, and Turkish forms, which suggests one of these languages as its source.

3 For the moment, until I have learned more about Portuguese historical phonology and the historical development of nasalized vowels from earlier final nasal consonants in Proto-Romance, I am assuming that the pronunciation of the lexeme for “soap” in 16th century Portuguese was fairly similar to the modern pronunciation.