Problems in studying loans

GARLAND CANNON
Texas A&M University

Borrowing raises at least four major problems: delimitation of corpus, nomenclature, "quality" of items, and antedating and etymology. Delimitation is complicated by the fact that loans, a rubric term for loanwords and various kinds of translation, are essentially unlimited. That is, the total loans from one language into another include huge numbers of biographical or geographical terms which belong in encyclopedias and atlases rather than in the lexicon.

In updating my dictionary of Arabic loans in English (1994a, 1997, 1998a), I rechecked the first complete English translation of the Koran, which Alexander Ross rendered in 1649 from the French version in that same year. The French version employs many transliterations and translations. Ross, who acknowledged that he did not know Arabic, introduced several of the items into English via French, as in his Arabic kitab ‘book’. When I checked the Arabic text, as elaborated in George Sale's notes to his 1734 English translation directly from the Arabic, it was clear that Muhammad believed in the Arab classes of benevolent or malevolent spirits with supernatural powers said to inhabit the earth. Presumably not knowing the Arabic names jamn, shaitan, efreet, Marid, and jinn, Ross used bland English words like spirit and devil. However, since these names appear in subsequent English versions of the Koran and in some recent unabridged dictionaries, though in few of the space-restricted college dictionaries, they remain in my Arabic corpus.

E. J. Brill's ongoing new edition of The encyclopaedia of Islam (1960- ) and English concordances to the Koran could add thousands of obscure religious words that belong in any English corpus of Islamic loans. But their exclusion from nonencyclopedic dictionaries indicates that they would skew a general corpus. Similarly, a corpus concerned with Persian rugs could gain 73 specialized loans from the index alone of Ford's The Oriental carpet (1981:348). To prevent such dilution, any item not recorded in standard dictionaries, which are the usual source for a general corpus, should be rejected. Oral forms can be ephemeral, have rare variants, and lack attested currency.

Modern readers demand more and more encyclopedic-type words in their dictionaries, particularly the unabridged ones. Needing a larger vocabulary to compete with Webster's third new international dictionary of the English language (W3, 1961), the first edition of the Random House dictionary of the English language (1966) uniquely included 30 items denoting Japanese swords and sword fittings. My reasoning was that the words' appearance in this modern, generally currency-based collection could outweigh the fact that almost all were found only in Gunsaulus' specialized monograph (1923). Though they were excluded from Random House's second edition in 1987, I accepted them in my dictionary of Japanese loans (1996:65-6), and thereby falsely elevated the category of Swords to the seventh highest among the 40 semantic categories comprising the 1,425 loans. Recent reconsideration of the technical usage of the 30 rare items and their recording solely in an encyclopedic-
type dictionary led me to delete them and thus drop the Swords category to its proper low ranking among the Japanese cultural aspects lexically represented in English.

The subtitle of the famous Century dictionary: An encyclopedic lexicon of the English language (1889-91) suggests the unique inclusions that somewhat disqualify it as a general source. The name of the Iranian president, Muhammad Khatami, is likely now being added to comparable new collections, especially in view of his call for a people-to-people dialogue with Americans that could end the U.S. effort to isolate Iran. By ultimately substantially reducing the flow of Farsi words into English, this effort demonstrates that politics can affect language contacts in once mutually beneficial situations (Cannon 1998c). Names like Khatami in the sense of 'current Iranian president' and Tehran as 'Iranian capitol' do not belong in a general corpus. Mao flu or names like Nehru and Mao 'a kind of coat' do throw light on neology by exhibiting semantic shifts motivated by English speakers' observation of the coat that these two Asians wore. But as these leaders did not call the coats by their own names, there was no lexical transfer; such words are not loans.

Proper nouns should not automatically be rejected (Mufwene 1988). An encyclopedic-type dictionary records some names that belong in a general corpus, because they provide information about dating, currency, word-formation, and inflections. The preface to Webster's new international dictionary of the English language (W1, 1909.vi) describes Merriam-Webster's innovative 'pearl' section, containing "various minor words, foreign words and phrases, abbreviations", mere variants, and "less-common and less looked-for terms" in narrow columns and small type on every page. An effort was made to place the purely geographical and biographical terms in a gazetteer and pronouncing biographical dictionary (pp. 2375-2545). Webster's second (W2, 1934) reduced the number of pearl items.

Then, especially to save space, W3 adopted the principle of the Oxford English dictionary (OED, 1933), by excluding encyclopedic-type terms and, unfortunately, the pearl section, which permits one to track such items. Of the 132 loans in my 1994 Arabic corpus not recorded in the OED and/or Webster's tenth (1993), the fact that only 23 were not originally pearls validates the OED's principle. W1 is the earliest record of 76 of these, of which 68 were pearls (Cannon 1998a:108). Nine lacked the currency to continue in W2 even as pearls, but were admitted to W3. Italian melongene 'eggplant' and Ar. Liyyanite 'an ancient Semitic language' were validated by inclusion in the second edition of the OED (OED2, 1989). Spanish almagra 'Philippine timber tree' and German zwetschgenwasser 'plum brandy' began as pearls in W1, but became regular entries in W2 and W3. Ar. Kitalpha 'a star in Equeules' experienced the reverse, because, after gaining regular status in W2, it was excluded from W3. Such evidence about "first" dates and currency led me to reduce my Arabic corpus by 191 encyclopedic items found only in the unabridged dictionaries that accept such terms (Cannon 1998a:119).

Some otherwise encyclopedic entries contain information about derivation
and pluralization. Though rejecting the Sassanian name *Yezdegird III*, I accepted its adjectival form because of the -ian suffix, which exhibits a common derivational pattern. The simultaneous 1954 transfer of both *Chicana* and *Chicano* preserves the Spanish gender-differentiation. Loans like *Chicanismo* and *Sandinista* (but not *Chicanism* and *Sandinist*) and *Fidelista* - *Fidelista* give insight into the loss or retention of borrowed bound forms. W3 eschews the noun form of purely geographical names, but records the usually unchanged adjectival form. Some of these functional shifts are acceptable when they reveal contrasts, as in *Tehran* ‘in the manner of Tehran ... etc.’, vs. the also recorded noun *Tehrani* ‘Iranian inhabitant’, which pluralizes with -s.

Common derivatives like *Chinese* and *Japanese* also merit inclusion to show that they take a zero plural. Inflections are seldom borrowed. Yet the many loans from English into High German since World War II have evidently completed the standardizing of the borrowed -s plural alongside the historical G. -Ø, -e, -er, and -(e)n, which had been initiated by Low German and French loans. Thus *der Account Executive* takes -s; and *die Ability*, -ies. And many of the large number of abbreviations from English have not received a German article, mainly for international precision and economy. That is, does G. *die USA* convey the same meaning as G. *USA*? Transfers like the article-less *Aids* have considerably expanded the number of German nouns exhibiting the continued leveling of grammatical gender (Pfeffer and Cannon forthcoming, Carstensen 1993-96, Wennrich 1976-80).

Nomenclature, the second problem, can be approached by updating Haugen's article (1950) and Deroy's book (1980), though the great variety of loans in the world's languages precludes a wholly empirical classification. As said, **loans** consist of loanwords and various types of translation. The term *loanword* is itself a loanword, naturalized from the German compound *Lehnwort* to fit English phonetic and graphemic patterns. Loanwords are usually introduced in one of three major types, based on the degree of differences between the etymon and features in the target language. First, the only change is in any needed pronunciation and transliteration. Thus in recent transfers like It. *Gran Turismo*, Japanese *yakuza* ‘a Japanese gangster’, Malay *ringgit* ‘Malaysian monetary unit’, and Sp. *numero uno* (see Cannon 1987:69-97), bilinguals have influenced the preserving of some native phonetic features. **Adaptation**, the second type, involves relatively minor formal change. It ranges from the loss only of the German initial capital letter, as in changing the plural forms of *Auge* ‘eye’ and *Griebe* ‘goose crackling’ to *augen* and *grieben*. Almost as straightforward are the adapting of G. *Aktivismus* ‘a philosophical theory’ to *activism* (chiefly clipping the terminal syllable and anglicizing the k to c), and the respelling of the k and ie in G. *Laparoskopie* as *laparoscopy* ‘visual examination of the peritoneum’ (see Pfeffer and Cannon 1994:119-22). **Alteration**, the third type, can extend into a gray area, where the change is nearly large enough to constitute a translation. For example, the bilingual's pronunciation of *-wort* with a /v/- is
outweighed by the majority, monolingual reader's word, with a respelling also of Lehn to give a good translation of the German etymon. Within these three types of change, loanwords may undergo initial, medial, and/or terminal clipping, either at the time of transfer or in later productivity, as in Taj from Urdu Taj Mahal (< Per., lit. ‘distinguished place’), and Mir ‘a fine Saraband rug’ from the Persian place-name Mirabad. Expansion, the reverse, is often motivated by the addition of a terminal noun to redundantly explain the foreign element, as in chenar tree, afghani rupee, and Shiraz lamb.

The three major types of translation also belong in a loan corpus because they entail contact with the lending language, when there is no possibility of independent coinage semantically uninfluenced by another language. A loan translation may be the trickiest, even when one can prove that the putative source sufficiently antedates its would-be rendering. While items like E. as if (< G. als ob) are fairly straightforward, Fr. la SUBLIME PORTE might have been a French speaker's spontaneous creation upon seeing the majestic gate. This phrase is usually analyzed as a loan translation of Turkish Babiali ‘the Sublime Gate’ (< Per. < Ar. Bab ‘gate’). If Fr. la PORTE was the transfer of only part of the Turkish source, it was a partial translation, rather than a later French clipping of the originally full translation. When English took Fr. PORTE as a loanword, the otherwise impressive French structural equivalence was lost. In recent decades English has been borrowing both the loan translation and its source at roughly the same time: Chinese (wallposter, dazibao), French (black comedy, comédie noire), German (guest worker, Gastarbeiter), Spanish (refried beans, frijoles refritos), etc. (cf. Sasquatch < Salish se’xxac ‘wild men’ c. 1929 vs. Bigfoot in 1962 - see Cannon 1994b:49). The third type, expanded translation, can pose problems. Was tower of silence created by simultaneously adding the prepositional phrase of silence terminally to the translated Per. dakhma ‘tower’? Or did an early English speaker coin the entire phrase, tower of silence, upon seeing the distant Malebar towers eerily disturbed only by the carrion birds?

Two old philological terms can be useful. Hybrid can diachronically characterize the German source of E. Autobahn, borrowed from the German compounding of the clipped Fr. automobile and G. Bahn. Graeco-Latin terms are often viewed as being outside the lexicon. These usually originally technical compounds or derivatives, coined from classical elements, may contain few or even no features of the neologizers' language, in systematically characterizing the object named. A dubious, still unaccepted synonym is International Scientific Vocabulary (ISV), which Philip Gove introduced in W3 for items whose "language of origin is not positively ascertainable but they are known to be current in at least one language other than English" (p. 7a).

Moeritherium (1902), etymologized in W3 and OED2 as New Latin, will illustrate. OED2's all-German citation guided Pfeffer and me (1994:xxvii) to the 1901 Verhandlungen of the Fifth International Zoological Congresses (v:528). The
archaeologist Sir William Andrews named this extinct proboscidean mammal, in his German paper presented there. Compounded from Gk. Moiros ‘the name of an Egyptian lake’ and thērion ‘wild beast’, Moeritherium employs no special features from German or English. Yet, like many other polysyllabic words created for a technical international audience, it has moved from restricted usage into the general language and now appears in college dictionaries. Inasmuch as Greek and Latin have provided a word-horde of elements for English, German, Italian, French, Arabic, etc., the artificial origin of such items with a more general currency makes them a special type of loan. Görlach (1996) criticized Pfeffer-Cannon's 1994 German corpus for accepting some Graeco-Latin terms. However, ours have considerable usage and usually contain German features. G. der Aktograph ‘a device that records experimental animals' movements’ has a gender-marking article and a capitalized first letter, requiring change for borrowing into English and other languages. Though English dropped the masculine der when transferring actograph, languages often retain the article, as the Romance languages have done with Ar. al. Portuguese albacora ‘albacore’ did so in taking Ar. al-bakūrah, as did English when taking the mediated Portuguese form and recent direct loans like Al Fatah. Many ultimately Arabic items in English like alchemy and algebra retain the al (Cannon 1994:119-34). An English speaker who buys rotenone ‘insecticide’ or Dekontaminol ‘graffiti-remover’ probably neither knows nor cares that these mixtures were created and specially named by a Japanese and a German speaker, respectively, from non-English elements. Word-formation categories like abbreviations include huge numbers of items coined for special purposes (Cannon 1989, 1998b:21-3).

The concept of special types raises the third problem, the "quality" of loans. Of course, no loan is abstractly inferior or superior to any other. The appropriate, major criticism of the dictionaries of German, Arabic, and Japanese loans was the rarity of many items, by comparison with high-frequency ones like biology, statistics, alcohol, Arab, and soy. Actually, I first relied on the old German tripartite scale of Gastwort, Fremdwort, and Lehnwort, emending it with the naturals, denizens, aliens, and casuals jocularly advanced by OED1, which recognized that three levels were insufficient. The need to include translated items in a loan corpus and resolve other inadequacies eventuated in my somewhat fluid four-degree scale, numbered 1-4, as partly determined by the four general levels found in a 368-item Malaysian corpus. With a primary base of productivity and currency, the assignment of a given degree to a word is chiefly determined by the word's admission to current editions of the eight constantly updated college dictionaries (listed in the References). The scale was first applied to the Malay loans (1992:137) and successively refined for the German and later collections. It is readily adaptable to measure the currency of nonloans as well.

Some Turkish loans will illustrate. After 24 encyclopedic terms like charik ‘rawhide sandal’ (found only in W2) are excluded, the extremes are easily
recognizable. New loans or older, seldom-used items normally requiring glosses and/or special punctuation like italics or quotation marks receive a [1], whereas fully naturalized loans receive a [4] (Cannon 1997:178, Pfeffer and Cannon 1994:xxxiii). The obsolete Mussulmanlik ‘Islam’ is a [1] because it appears only in a 1625 citation in OED2. By contrast, jackal is not only in OED2, but also in the unabridged dictionaries and all eight college ones. It has been functionally shifted to an adjective and a verb, has produced the compound jackal buzzard and the adjective jackalled, and has developed three additional senses since its 1603 transfer denoting a kind of wild dog. While some Turkish items with a [4] are household words, all lack the high currency and productivity of old loans like sultan (as in the name Sultan of Swat given to Babe Ruth), desk, and art, which occupy the top layer of [4]. But, as is normal for lexical transfers, even these three words experienced decades of usage in rising from the introductory degree of [1], unlike the remarkable, near-instantaneous [4] merited by a few recent loans into English like Ar. intifada and Russian perestroika. Even if a higher degree is reached, the currency can change. Various items in Burchfield’s Supplement to OED1 (1972-86) demonstrate that their widespread usage during the Renaissance has sagged to a [1] or [2] in modern times.

Most of the 291 collected Turkish loans (Cannon 1999) merit only a [2], such as araba ‘wheeled carriage’. Like Mussulmanlik, it has been unproductive since its 1845 introduction. However, wider usage gained its admission to the OED, W1/2/3, and other unabridged collections, but to only one college dictionary. The verb chouse receives a [3] primarily by virtue of its acceptance in all the unabridged dictionaries and four college ones, but without semantic extension since its borrowing in 1659 as a restricted item (slang and chiefly British) denoting a cheat or trick. Unlike araba, chouse later provided forms like choused and chousing and a functional shift to a noun.

If a historical record like OED2 reliably dates a loan, and if there is adequate time between its original transfer and its productive forms, assignment to this naturalization scale is relatively simple. Reliable dates are necessary to prove that there has been any translation, development of extended senses, functional shift, compounding, derivation, and/or clipping (Pfeffer and Cannon 1994:119-20). For example, English borrowed the German spelling Kaffeeklatsch in 1888. Was the E. coffee klat(s)ch five years later a partial translation of a reborrowing, which would entail a new exposure to the German source and thus require a tabulation of two loans rather than one? Or was it a routine alteration or adaptation of the 1888 transfer? Was the klat(s)ch of 1953 a modern clipping of the 1888 transfer or of a reborrowed etymon? Such perhaps trivial details can determine the precision of an etymology in unabridged dictionaries.

If reborrowing cannot be demonstrated by much later exposure to the original source in an unchanged sense and form, the new occurrence should probably be considered as a revival, which itself can be a century or so later. If an altered
meaning is still sufficiently related to the original one possibly to be a semantic extension, reborrowing is difficult to prove. Thus the old Urdu *shalwar-kameez* ‘women's loose-fitting trousers and long tunic’ was first used in English by colonial residents on the Indian subcontinent. After a long hiatus, the word reappeared in 1955 with roughly the same sense, except that this attire is worn by Western rather than Asian women. As for linguistic productivity, English received the verb *plunder* from G. *plündern* by 1632; was there adequate time for the English noun usage in 1643 to be a functional shift rather than a borrowing of the German noun *plünder*? G. *Allergin* was transferred as *allergen* in 1910; was the E. *allergy* of 1911 from G. *Allergie* or a derivative of E. *allergen*?

Description can be further complicated by antedating and etymology, the fourth problem. Dates are again critical, especially in a mediated transfer. The editors of the third edition of the OED estimate that they have antedated a fourth of the subsenses of words ("OED3: An overview" 1997). Concordances for successively translated works like the Koran and the Bible can mislead, as in several supposed occurrences of the Turkish *jackals* in the King James Version, but which actually did not appear until the modern replacement of the original *dragons* conveying the "jackals" meaning. OED2's earliest record of *inshallah* is 1857, with citations as recent as 1971. But as 1857 is greatly antedated by Gothard Arthus' *Dialogues in the English and Malayan languages* (London, 1614:2), Malay was technically the original source for this religious word in English, with Arabic remaining the chief source. The same is true for some other religious words mediated by non-Arabic. The earliest OED2 record may contextually identify the transmitting language, but an antedated context can correct that standard etymology (Cannon 1998a), as when both English and Persian borrowed Ar. *shūrā* 'religious consultation'. After Afghan Persian extended the meaning to include the Afghan parliament, both Arabic and English borrowed this Afghan sense. Each thereby has two senses from the same ultimate source, and the English senses require two dictionary entries so as not to suggest semantic extension of the original transfer from Arabic.

French, Spanish, and Portuguese transmitted numerous Arabic and Persian items to other languages. In historically language-mixed areas like the Middle East and Central Asia, it can be impossible to distinguish among Persian, Turkish, Arabic, and Hebrew as the major and especially the first known source of some loans. Zenker (1866-76) records various near-identical sets of loans. Most Islamic senses originated in Arabic, just as the Talmudic senses came from Hebrew; but the Semitic cognate words for the two may be quite similar. This huge subject permits only the generalization here that at least unabridged dictionaries should attempt to specify the earliest source, even if it is later eclipsed by another language. E. *choppstikes* was a 1615 translation of Jp. *hashi*—in 1894 Lafcadio Hearn wrote of "hashi (chopsticks)". But actually, this was submerged after 1699 by the dialectal Chinese form, which provided the *chochaku* for the Sino-Japanese reading of the Japanese character for
chopsticks, and then became the chief conveyer of chopsticks to English (Cannon 1996:104). An item may name a ubiquitous thing and thereby become a kind of lingua-franca term. Revealing studies can be made of the worldwide representations of orange (ult. < Dravidian), algebra (ult. < Arabic), etc.

Se a loan can have a direct, single/multiple, and/or mediated transmission. When a mediated form differs substantively from its etymon, it becomes a distant loan—e.g., the many hybrids that Ar. qahwah ‘coffee’ has motivated, such as the several phrases like café au lait that French created and gave to English. If a Spanish dictionary, say, does not specify the distant Arabic source in etymologizing café con leche, it may imply that café is natively Spanish. Similarly, the addition of the Swahili ki- prefix and other changes obscure the distant Ar. nur ‘light’ in kinara ‘Kwanza candle holder’, though not in Kiswahili. In the English borrowing of Ar. Iblis ‘devil, chief of the wicked jinn’, the Arabic dropping of the original /d-/ in Gk. diàbolos has shrouded that distant Greek etymon of Ar. Iblis. Diàbolos is not the source of E. Iblis; Arabic is.

Numbers of Japanese neologisms taken into English are etymologically troublesome for a different reason. As long as an item constructed wholly from English elements is not borrowed by English, the etymology presents no difficulty, as in the Japanese alteration of mass communication to masû-komyunikēshon, which, of course, was not needed in English. But after Japanese blended E. camera and tape recorder into Jp. kamukōdā, to name an original Sony product that was quickly exported to the world, camcorder came into English as an instantaneously naturalized loanword exemplifying English patterns, with an immediate currency of [4]. A lexicographer who does not know the Japanese source is unlikely to suspect that loans like camcorder, AutoCut VCR, glocal, multivalve engine, and Walkman are not native English (Cannon 1995). The non-Asian context and immediate loss of the Japanese phonetic features have led some dictionaries to misetymologize probable calques like high/low profile (< kō-shisei ‘high posture’ and tei-shisei).

In summary, a general model is needed for studies of loans. A refined model utilizing principles like those sketched here might help to resolve the four major problems and permit more incisive comparisons among languages and language families. My data from Chinese (1988) and some other languages show that they are receptive to lexical borrowing (see Décsy's 1973 scale for levels of receptivity). These data offer comparative information on semantic fields, degree of naturalization, word-formation, variant forms, percentages of form-classes, productivity, chronology and rate of inflow, and other aspects. French, Japanese, German, and Spanish, in that order, seem to be the leading lenders to English since 1961 (1998b:41). Such data from languages in contact show how the cultures represented have impacted and continue to influence each other. The widening linguistic interaction may be moving us all closer to a world citizenship.
References


The Koran; Commonly called the Alcoran of Mohammed; Translated into English immediately from the original Arabic, with explanatory notes. (1734). Transl. by George Sale. London: J. Wilcox.


_____ forthcoming. Abbreviations: Continued leveling of gender in Modern High German.


Verlagbuchhandlung, 1967).