Loanword Typology:
Steps toward a systematic cross-linguistic study of lexical borrowability

MARTIN HASPELMATH
June 2003

1. Motivation and goals

One of the most important tasks of diachronic linguistics is to establish **general constraints on language change**. There are two main types of constraints on language change: **paths of change**, which limit the direction that changes can take, and **rates of change**, which give us an idea about the frequency or speed with which certain types of changes occur. Constraints on language change are of interest for at least two reasons:

(i) A **theoretical** reason: Understanding the nature of language change presupposes identifying constraints on language change. If there were no such constraints, if anything goes, then we would have a difficult time understanding how and why change occurs.

(ii) An **applied/practical** reason: Constraints on language change are a prerequisite for reconstructing unattested changes and unattested linguistic situations.

In the context of the research areas of the Max Planck Institute for Evolutionary Anthropology, an applied area of particular interest is the reconstruction of linguistic family trees. Linguistic family trees are increasingly found relevant by researchers in other disciplines such as archeology and molecular anthropology (e.g. Renfrew et al. 2000, McMahon & McMahon 2003). Whether genealogical classification is based primarily on shared innovations (as in the comparative method) or on shared retentions (as in lexicostatistics), constraints on language change are absolutely crucial for the reconstruction. For instance, historical phonologists routinely make use of information about likely phonological changes. Consider a hypothetical situation as in (1), where A, B, and C stand for three related languages.

(1)	A	B	C	reconstructed:
	p	*p*	*f*	*p
	h	*s*	*h*	*s
	i	*y*	*y*	*y
	s	*ʃ*	*ʃ*	?*s ?*ʃ

Since historical linguists know that $p > f$, $s > h$, and $y > i$ are quite likely unconditioned changes, they would reconstruct *p, *s, and *y, regardless of the number of daughter languages that preserve these sounds. Where nothing is known about directionality (as in the case of *s* and *ʃ*), we do not know which sound to reconstruct. In phonology, this knowledge seems to be largely impressionistic, even after almost two centuries of research on sound changes in the world's languages. No handbook of attested phonological changes exists. Likewise, we have very little systematic information about general tendencies of lexical semantic change, although again a lot of research has been devoted to this

topic (e.g. Blank 1997). The only area where the research of the last few decades has been summarized in a handbook-like publication is the area of grammatical semantics (grammaticalizatin, see Heine & Kuteva 2002). However, while we know quite a bit about paths of change for grammaticalization, we know next to nothing about the rate of change here.

In the area of lexical change, a lot of work has been done on the rates of change in the lexicostatistical research tradition. While Swadesh (1955) and Lees (1953) seem to have assumed that the replacement rate for lexical items is identical for all the lexical meanings under consideration, it became soon clear that this is not the case (e.g. Sankoff 1970). Moreover, it was often pointed out that the rate of replacement of words on the Swadesh lists is different in different historical situations (more recently e.g. by Blust 2000). Some proposals have been made for identifying the most stable lexical meanings (e.g. Dyen et al. 1967, Lohr 1999), but none of these has been based on a broad database from a representative sample of the world's languages.

One of the most important confounding factors for any type of approach to genealogical classification is **lexical borrowing.** Researchers in lexicostatistics have long recognized this, and at the beginning the hope was apparently that the 200 words on the Swadesh list (or at least the 100 words on the reduced list, Swadesh 1955) are so resistant to borrowing that this confounding factor can be neglected for these word meanings. However, the Swadesh lists do not seem to have been based on any kind of systematic research, but just on Swadesh's intuitive sense of which word meanings would be the most easily identified across languages and at the same time likely to be highly conservative. In any event, subsequent research on language contact showed that borrowing can be quite massive, and that the situation of well-studied Indo-European languages such as German, French or Russian (where loanwords are easy to identify and occur in rather circumscribed domains) may be quite atypical (cf. Thomason & Kaufman 1988).

Thus, it is of paramount importance for lexicon-based historical linguistics to get a clearer idea about the **differential borrowability** of different types of words. It should not be difficult to advance our knowledge beyond what current textbooks have to offer in this regards. Typically they invoke a vague notion of **"basic vocabulary"** (or "**core vocabulary"**), e.g.

> "From a purely linguistic perspective, the most important fact is that different spheres of the vocabulary are borrowed more easily, others significantly less easily. For instance, the most successful resistance to borrowing is offered by BASIC VOCABULARY, words referring to the most essential human activities, needs, etc., such as *eat, sleep; moon, rain; do, have, be,...*" (Hock & Joseph 1996:257)

> "Here it should be noted that the distinction between basic and nonbasic vocabulary is a rough practical distinction, not a well-supported theoretical notion. Several decades ago ... Morris Swadesh devised two lists of basic vocabulary items... His goal was to include only items that are unlikely to be borrowed... There was, and is, no theoretical foundation for this notion of universal-and-thus-hard-to-borrow basic vocabulary, and in fact all the items on Swadesh's list can and have been borrowed. Still, the lists are useful, because in most cases are at least less likely to be borrowed than more culture-specific vocabulary..." (Thomason 2001:71-72)

Hock & Joseph give a sketchy definition of "basic vocabulary" and a short list of examples, but neither is particularly helpful. And Thomason makes no attempt to go beyond Swadesh.

The idea of the planned **"Loanword Typology" project** is that it should be possible to get a clearer idea of lexical borrowability by examining the loanwords in a reasonably representative and reasonably large set of languages (say, 30-40 languages), and by making inductive generalizations over the data assembled in this way. The outcome of the project would be an edited volume (or even more than one volume) consisting of 30-40 language-particular chapters and a number of more general chapters that explain the methodological choices and discuss the results. Each language-particular chapter would be authored by a specialist of the language who knows enough about neighboring languages and historical-comparative linguistics of the family to identify the loanwords in the language. The project identifies a fixed list of word meanings (for instance, the IDS list, based ultimately on Buck 1949) that is checked for each language. Each chapter consists of a data part and a discussion part: the data is a list of those words on the project list that can be identified as loanwords, plus perhaps other loanwords whose meanings are more specialized and hence do not appear on the project list. In addition, the source of each loanword is identified to the extent that it is known. The discussion part attempts to generalize over the data and puts the loanword into the relevant context (structural, historical, cultural, etc.), trying to explain why these words and no others were borrowed. While the data part is relatively standardized, authors will have a lot of freedom in the discussion part.

The remainder of this paper will mention a few general issues that will be relevant for any project that studies lexical borrowabiliy in a comparative perspective. They are not intended to be complete or even particularly coherent. The main purpose is to serve as a first basis of the discussion in the initial stages of the collaborative project.

2. Kinds of loanwords

Let me start with some terminology. It is now customary to use the terms **recipient language** for the language that acquires a loanword and **donor language** for the language that is the source of the loanword. A **loanword** can be defined as a word that is transferred from a donor language to a recipient language, and it should not necessarily be equated with "borrowed word", because some linguists define **borrowing** in a narrow way that excludes the effects of shift-induced interference or substrate (e.g. Thomason & Kaufman 1988:37ff.). More general terms for contact-induced change are **transfer** and **copying** (Johanson 2002).

According to Ross (1991), two other kinds of contact situations need to be distinguished, in addition to typical borrowing and typical shift-induced interference. He notes that typical borrowing is created by <u>native speakers</u> who <u>consciously</u> **import** a word from another language, whereas typical shift-induced interference is created by <u>non-native speakers</u> who <u>unconsciously</u> **impose** features of their native language to the <u>recipient language.</u> But imposition may happen to native speakers as well, especially when their native language is not their dominant language. In such cases, according to Ross, native speakers may transfer syntactic

features from a dominant language to their native language, resulting in *metatypy* (see also Ross 1997, 2001). Moreover, in addition to importing words from a language spoken by a different group into their language, speakers may also import words from a language of their own group into the majority language, thus creating a new variety of the majority language that expresses the minority group's cultural identity. An example of this might be the Yiddish words used in English by Jewish Americans.

		agents of change:	
		native speakers	non-natives
motivation for transfer:	processing ease [imposition]	**metatypy**	**substrate** (= shift-induced interference)
	"culture" [importation]	**"exo-borrowing"** (= adoption)	**"endo-borrowing"** (= retention)

Table 1: *Ross's (1991) four types of contact-induced change*

One of the best-known taxonomies of borrowed items comes from Haugen (1950), who distinguishes between **loanwords** (form and meaning are copied completely), **loanblends** (words consisting of a copied part and a native part), and **loanshifts**, where only the meaning is copied. Loanshifts fall into two subtypes: **loan translations** (like Spanish *rasca-cielos* from *sky-scraper*) and **semantic borrowings**, where only the meaning is copied (like German *kontrollieren*, which originally only meant 'check', but is now also used in the sense 'have control over', as in English).

Myers-Scotton (2002:239) distinguishes between **cultural borrowings** and **core borrowings**, which according to her have very different origins. Cultural borrowings are words for new objects (e.g. *espresso*) or words for new (non-object) concepts (e.g. *zeitgeist*), and they usually appear abruptly when influential groups use them. Core borrowings, by contrast, are words that more or less duplicate already existing words (e.g. *OK* in German, which replaces *gut*, or *einverstanden*). Core borrowings "usually begin life in the recipient language when bilinguals introduce them as singly occurring codeswitching forms in the mixed constituents of their codeswitching".

3. Factors for differential borrowability of word meanings

3.1. Borrowability scales

The most important type of constraint on borrowing that has been discussed in the literature is the **borrowability scale** (also called *borrowing hierarchy*).[1] For instance, Matras (1998) proposes the scale in (2) for coordinators:

---

[1] Since the term *hierarchy*, when used outside language typology, is generally reserved for taxonomic hierarchies, I prefer the synonymous term *scale*.

(2) 'but' > 'or' > 'and'

Similarly, Field (2002:38) proposes the scale in (3):

(3) content item > function word > agglutinating affix > fusional affix

Such scales can be interpreted in three ways:
    (i) **Temporal**: A language borrows elements on the left before it borrows elements further to the right.
    (ii) **Implicational**: A language that contains borrowed elements on the right also contains borrowed elements further to the left.
    (iii) **Quantitative**: A language borrows more elements belonging to the types on the left than elements belonging to the types further to the right.
    (iv) **Probabilistic**: Elements belonging to the types on the left are more likely to be borrwed than elements further to the right.
    The temporal and implicational interpretations are generally difficult to distinguish, as are the quantitative and probabilistic interpretations. It should be noted, however, that absolute quantities are of little interest. Thus, Haugen (1950:224) notes that 75.5% of all American Norwegian loanwords in his corpus are nouns, but he does not say what percentage of all the words are nouns. If the entire corpus contains 75.5% nouns, then loanwords show no special behavior and there is no basis for saying that nouns are borrwed preferentially.

3.2. Morpheme type

It is widely acknowledged that lexical items are more likely to be borrowed than grammatical items, and that words are more likely to be borrwed than bound morphemes (e.g. Moravcsik 1978). Field (2002) adds the claim that agglutinative affixes are borrowed more easily than fusional affixes (see (3) above). Van Hout & Muysken (1994) cite data on content vs. functions word types from their Quechua corpus (which contains many Spanish borrowings):

	not borrowed	borrowed	total
function word	105 (80.8%)	25 (19.2%)	130 (100%)
content word	592 (63.7%)	338 (36.3%)	930 (100%)
total	697	363	1060

p = 0.0001174

3.3. Parts of speech

It is widely acknowledged that **nouns are borrwed more easily than other parts of speech** (e.g. Whitney 1881, Moravcsik 1978, Myers-Scotton 2002:240). Van Hout & Muysken (1994:42) give the following explanation:

> "A very important factor involves one of the primary motivations for lexical borrowing, that is, to extend the referential potential of a language. Since reference is established primarily through nouns, these are the elemnets borrowed most easily."

According to Myers-Scotton (2002:240), nouns are borrowed preferentially "because they receive, not assign, thematic roles", so "their insertion in another language is less disruptive of predicate-argument structure". Van Hout & Muysken (1994) again cite data from their Quechua corpus:

	Quechua native words	Quechua Spanish loans	total
Types			
Nouns	194	184 (49%)	378 (100%)
Verbs	214	81 (27%)	295 (100%)
Tokens			
Nouns	1,101	823 (42%)	1,924 (100%)
Verbs	1,929	241 (11%)	2,170 (100%)

These data show that in Quechua, a much higher percentage of noun types and tokens are loanwords than verb types and tokens.

The difficulty of borrowing verbs as verbs has sometimes been addressed in the literature. For instance, it has been said that verbs cannot be borrowed in French because of their elaborate inflection, so that it is difficult to incorporate other languages' verbs into French (Meillet 1921, cited in Thomason & Kaufman 1988:348; however, French does have loan verbs such as *shooter*). Moravcsik (1975, l978: 111-112) observed that if verbs are borrowed, they seem to be borrowed as if they were nouns: the borrowing language employs its own means of denominal verbalization to turn the borrowed forms into verbs "before" using them as such (see also Moravcsik 2003+).

Not much is known about adjective borrowing, but adjectives remind us of the fact that when we talk about borrowability of different parts of speech, we have to specify whether we mean donor part of speech or recipient part of speech (cf. Curnow 2001:415). As is well known, part-of-speech systems differ quite dramatically when it comes to adjectives, so here both the donor and the recipient systems should be taken into account. A concrete example is an observation made by Dik Bakker (p.c.): Quechua borrows considerably more (donor-language) adjectives from Spanish than Otomí does, which may well have to do with the fact that traditional Otomí lacks adjectives (see the corpus counts in Hekking & Muysken 1995 and Hekking & Bakker 1999).

3.4. The role of token frequency

One way in which the notions "basic vocabulary" and "core vocabulary" can be interpreted is as the words which are used most frequently. It would not be surprising if they were resistant to borrowing, because it is well known that high-frequency items are resistant to other types of language change such as analogy. Van Hout & Muysken (1994) find some evidence for this in their Quechua corpus. The following table gives the percentage of Spanish loans in 7 frequency classes

token frequency	frequency class	Quechua native types	Quechua Spanish loan types	total types	% borrowed from Spanish
1	1	185	170	355	47.9
2-3	2	116	99	215	46.0
4-7	3	86	46	132	34.8
8-15	4	72	29	101	28.7
16-31	5	27	11	38	28.9
32-63	6	17	7	24	29.2
> 63	7	14	1	15	6.7
total					

Van Hout & Muysken conclude that frequency in the recipient language may operate as an inhibiting factor for borrowing.

3.5. Lexical semantic field

One area where little systematic research has been done (and no universals have been proposed) is the lexical semantic fields that loanwords tend to come from. However, it seems clear that there are many regularities here. For instance, victorious invaders will typically borrow placenames, names for local plant and animal species, and languages of peoples ruled by a foreign invaders will typically adopt military terms (see, e.g., Vennemann's 2000 inference that the Germanic peoples must once have been dominated by a foreign ruling class, perhaps of northern African origin).

4. Factors for differential borrowing behavior among different languages

4.1. Intensity of language contact

The most obvious sociolinguistic factor favoring borrowing is widespread bilingualism. This is often called "intensity of contact". Thomason & Kaufman (1988) propose a five-point scale of intensity of contact:
   (1): casual contact,
   (2): slightly more intense contact,
   (3): more intense contact,
   (4): strong cultural pressure,
   (5): very strong cultural pressure).
They claim that beginning with stage 3, we also find nonbasic vocabulary among the loanwords.
   Brown (1999) finds that there is significantly more borrowing into Native American languages from Spanish than from English or French. The likely explanation, according to Brown, is that Native Americans have often been bilingual in Spanish because they were integrated into Spanish society much more and earlier than Native Americans in the British and French colonies (and later the U.S. and Canada).

An additional factor is probably the prestige of a language, although this of course correlates with widespread bilingualism (people are more likely to learn another language if it is prestigious). However, widespread bilingualism without great prestige does occur (e.g. Spanish and Guarani in Paraguay, where Guarani is spoken by many speakers of non-Guarani origin, but Paraguayan Spanish has very few Guarani loanwords), and languages may be widely regarded as prestigious but still few people speak it (e.g. French in 19th century Europe, or English in much of the world today).

4.2. Purism

It is sometimes claimed that different cultures have different attitudes toward borrowing than others, which would explain differential borrowing behavior (e.g. it is claimed that Icelander are purists and hence their language has very few loanwords, cf. *tölva* 'computer', *útvarp* 'radio', etc.). However, unless there is legislation or language academies with a high degree of social acceptance, it seems to be difficult to find evidence for the exact role of speaker attitudes, and we must be careful to avoid circular reasoning.

4.3. Structural incompatibility

Structural incompatibility has often been invoked as explaining reststance to borrowing, although in recent years it has come under attack (especially by Thomason & Kaufman 1988). For grammatical borrowing, it seems undeniable that it plays a role (e.g. it seems very unlikely that an isolating language like Vietnamese would borrow a case suffix), but it is not clear whether it might be relevant for lexical borrowing.

4.4. Genealogical relatedness

McMahnon (1994:204) implies that related languages are more likely to borrow from each other, especially if they are so closely related that mutual intelligibility is relatively easy to establish. This would explain why English borrowed "basic vocabulary" items such as *skin, sky, get, they, them, their* from Old Norse.

5. Establishing borrowability through language comparison: two examples

I this section I give two examples of the kinds of results that a systematic cross-linguistic study of loanwords can yield.

5.1. Items of acculturation in languages of the Americas

Brown (1999) examined words for 77 "items of acculturation" (things/concepts unfamiliar to Native Americans before the European invasion) in 292 Native American languages. This research resulted in tables such as Table 2, where lexical meanings are ranked by borrowability.

Table 2: Items of acculturation, ranked by index of borrowability (European loan percentage)

item	percentage	item	percentage
coffee	81	apricot	37
coriander	72	box	36
cat	70	school	35
garlic	69	scissors	35
orange	67	Wednesday	35
cheese	66	match	34
donkey	64	nail	33
lemon	60	candle	32
apple	59	spoon	31
cow	58	shovel	30
Saturday	56	bread	29
pig	55	flour	29
soldier	55	board	28
peach	54	butter	28
mule	54	wagon	28
goat	53	peas	28
sugar	53	cup	27
tea	53	mile	27
horse	52	pistol	27
rice	52	window	27
table	52	clock	25
soap	51	grapes	25
bottle	51	oats	23
watermelon	49	book	22
ribbon	48	needle	21
cabbage	47	paper	20
lettuce	47	hundred	19
hour	45	beets	18
sheep	44	thread	17
onion	41	chicken	16
key	41	town	15
barley	39	rich	13
turnip	39	money	11
button	38	fork	10
wheat	38		

Among many other things, he observed the following tendencies:

(1) "Words for natural kinds tend more strongly than those for introduced artifacts to be associated with high borrowability." (p. 56)

(2) Within the category of words for introduced living things, terms for animals tend to show greater borrowability scores than words for plants.

Brown's (1999:66) explanation for the first tendency is that Native Americans encountered living things more often during interaction with Europeans, while artifacts were more often encountered in other contexts.

Brown's study seems to be the only systematic comparative study of loanwords so far in the literature. It is more limited than the present project in that it considers only 77 lexical meanings, but this allows him to take into account a very high number of languages.

5.2. Words of the IDS list in 80 Austronesian languages

The Comparative Ausronesian dictionary (Tryon (ed.) 1995) contains lexical information on about 1300 lexical meanings for 80 Austronesian languages, including information on whether a word is known to be a loanword. The 1300 meanings come from the list of the Intercontinental Dictionary Series (IDS).[2] My team went through all four volumes of the dictionary and identified the percentage of languages showing loanwords for each lexical meaning.[3] Table 3 shows the average percentages for each lexical field.

Table 3: IDS lexical fields, ranked by average percentage of loanwords in the 80 Austronesian languages of Tryon 1995

Chapter 4: PARTS OF THE BODY; BODILY FUNCTIONS AND CONDITIONS	2.33
Chapter 15: SENSE PERCEPTION	2.65
Chapter 2: MANKIND: SEX, AGE, FAMILY RELATIONSHIP	3.02
Chapter 12: SPATIAL RELATIONS: PLACE, FORM, SIZE	3.85
Chapter 16: EMOTION (WITH SOME PHYSICAL EXPRESSIONS OF EMOTION); TEMPERAMENTAL, MORAL, AND AESTHETIC NOTIONS	4.59
Chapter 1: PHYSICAL WORLD IN ITS LARGER ASPECTS	5.36
Chapter 13: QUANTITY AND NUMBER	5.36
Chapter 10: MOTION; LOCOMOTION, TRANSPORTATION, NAVIGATION	5.95
Chapter 17: MIND, THOUGHT	9.70
Chapter 19: TERRITORIAL, SOCIAL, AND POLITICAL DIVISIONS; SOCIAL RELATIONS	12.04
Chapter 18: VOCAL UTTERANCE, SPEECH; MUSIC	12.47
Chapter 20: WARFARE AND HUNTING	13.16
Chapter 8: AGRICULTURE, VEGETATION	13.34
Chapter 11: POSSESSION, PROPERTY, AND COMMERCE	13.64
Chapter 9: MISCELLANEOUS PHYSICAL ACTS AND THOSE PERTAINING TO ARTS AND CRAFTS, WITH SOME IMPLEMENTS, MATERIALS, AND PRODUCTS	14.73
Chapter 14: TIME	17.28
Chapter 7: DWELLING, HOUSE, FURNITURE	20.26
Chapter 21: LAW	21.09
Chapter 3: ANIMALS	22.99
Chapter 5: FOOD AND DRINK; COOKING AND UTENSILS	23.16
Chapter 22: RELIGION AND BELIEFS	28.22
Chapter 6: CLOTHING; PERSONAL ADORNMENT AND CARE	31.83

Table 4 is an excerpt from the list of lexical items.

---

[2] The Intercontinental Dictionary Series is a long-term project founded by Mary Ritchie Key (University of California, Irvine) that aims to publish electronic lexical databases for a large number of languages from around the world. All these databases are based on the IDS WordList, which is itself an adaptation of the list used by Buck (1949).

[3] Thanks to Ulrike Gurt and Jenny Seeg for their invaluable help.

Table 4: Some randomly selected IDS word meanings, ranked by percentage of loanwords in the 80 Austronesian languages of Tryon 1995

01.212	earth=ground, soil; terre, sol; suelo	0
01.222	cliff, precipice; falaise; barranca	0
01.270	shore; rivage, littoral; ribera, playa	0
01.310	water; eau; agua	0
01.323	rough (of sea); trouble, agité (mer); agitado (mar), borrascoso	0
01.342	reef; récif; arrecife	0
01.352	tide; marée; marea	0
01.353	lowtide; marée basse; bajamar	0
01.430	wood; bois; madera	0
01.440	stone, rock; pierre, roc, rocher; piedra, roca	0
01.530	moon; lune; luna	0
01.550	lightning; éclair; relámpago	0
01.560	thunder; tonnerre; trueno	0
01.570	lightning (as striking); foudre; rayo	0
01.620	darkness; obscurité, ténèbres; oscuridad	0
01.630	shade, shadow; ombre; sombra	0
01.640	dew; rosée; rocío	0
01.720	wind; vent; viento	0
01.740	fog; brume, brouillard; neblina, niebla	0
01.750	rain (noun); pluie; lluvia	0
01.852	burn (vb intrans); brûler; arder	0
01.880	firewood; bois (pour le feu), bûches; leña	0
02.210	man (vs.woman); homme; hombre	0
02.220	woman; femme; mujer	0
...		
05.220	boil (vb); bouillir; hervir	3.89
07.560	post, pole; poteau, piquet; poste	3.89
09.110	do, make; faire; hacer	3.89
14.450	noon, midday; midi; mediodía	3.89
15.220	smell (vb trans); sentir; oler, olfatear	3.89
11.820	sell; vendre; vender	3.94
...		
15.440	sound, noise; son, bruit; ruido, sonido	7.5
18.120	sing; chanter; cantar	7.5
18.210	speak, talk; parler; hablar	7.5
17.440	suspect; soupçonner; sospechar	7.57
...		
14.760	summer; été; verano	11.62
02.520	aunt; tante; tía	11.66
01.280	cave; grotte; cueva	11.68
12.540	measure; mesurer; medir	11.68

20.460	surrender; céder, se rendre; rendirse	11.76
02.760	widow; veuve; viuda	12
13.440	three times; trois fois; tres veces	12
...		
19.110	country; pays; país	20
19.370	citizen, subject; citoyen, sujet; ciudadano, súbdito	20
20.310	armor (defensive); armure; armadura	20
07.470	shelf; étagère; repisa	20.28
08.830	citrus fruit; citrique; cítrico	20.89
22.220	preach; prêcher; predicar	21.21
...		
14.780	season; saison; estación	31.14
07.580	arch; arche; arco	31.57
09.422	tool; outil; herramienta	31.57
06.730	ring (for finger); bague, anneau; anillo, sortija	31.74
05.370	spoon; cuiller; cuchara	31.94
03.370	he-goat; bouc; chivo	32.07
...		
03.260	ram; bélier; carnero, morueco	55.10
06.920	brush; brosse; cepillo	55.31
08.691	pipe; pipe; pipa	55.55
20.170	soldier; soldat; soldado	56.92
05.270	kettle; bouilloire; caldera	57.37
08.240	shovel; pelle; pala	57.40
...		
14.530	clock, timepiece; horloge, montre; reloj	71.42
18.560	paper; papier; papel	72
03.560	goose; oie; ganso	72.5
03.410	horse (equine); cheval; caballo	73.13
07.240	key; clef; llave	73.13
05.630	sausage; saucisse; salchicha	73.91
...		
05.930	beer; bière; cerveza	91.30
03.780	camel; chameau; camello	93.33
05.760	grape; raisin; uva	94.11
03.460	ass, donkey; âne; asno, burro	94.59
03.470	mule; mulet; mulo	100
03.770	elephant; éléphant; elefante	100
05.880	cheese; fromage; queso	100
05.890	butter; beurre; manteca, mantequilla	100

6. Some choices in the planned project

6.1. Which languages?

Since the goal of the project is to discover universals of lexical borrowing, the fundamental requirement is that the languages should be as diverse as possible, not only genealogically and geographically, but also sociolinguistically. There should be national languages with large numbers of speakers and great prestige, and there should be tribal languages with few speakers and little prestige for outsiders. There should be languages with a long written tradition, and unwritten languages. Moreover, the contact situations should be diverse: There should be languages that have undergone extensive lexical enrichment from outside sources, but also languages that have largely gotten by with their own lexical resources. There should be cases where the loanwords came in exclvely through the spoken language, and cases where many loanwords were introduced through writing.

   Of course, the choice of languages will to a large extent be determined by the linguists who are willing to collaborate on this project. Procedures for identifying linguists who are willing to participate need to be established.

   The general requirement of genealogical diversity does not exclude the possibility of including pairs of closely related languages. Such pairs may actually be particularly instructive when the two closely related languages (or varieties of the same language) have been associated with very different sociolinguistic circumstances (David Gil, p.c.). Possible pairs of this kind are Hindi/Urdu, Yidish/German, Iranian Azerbaijani/Azerbaijani of the Republic of Azerbaijan, Kuala lumpur Malay/Jakarta Malay.

   How many languages should be represented? A priori, as many as possible, but practical considerations will of course limit the number of languages. It remains to be seen how much can be achieved. The number of languages should not in any event be less than 25.

6.2. Which lexical meanings?

If only a small number of lexical meanings were covered (say a 200-word list), it would be easier to get data for a large number of languages. However, it is one of the main goals of this project to find out which lexical meanings are resistant to borrowing, so by including only those that have been thought to be resistant, we cannot really test Swadesh's claim that his words are particularly resistant. It seems that a list such as the IDS WordList (consisting of 1310 word meanings) is better suited for this project.

   However, it is conceivable that shorter word lists will be used in addition to the full 1300-word lists. Many more linguists would be willing to contribute data for a 200-word list, so such lists could be used in addition to the larger word lists. In this way, the project would combine both lexical breadth (for a somewhat narrow range of languages) with linguistic breadth (for a somewhat small number of word meanings).

6.3. What is a loanword?

A loanword is defined as a word that at some point came into a language by transfer from another language.[4] Thus, not only *manga* is a loanword in English (from Japanese, first attested in the OED in 1951), but also *very* (from French, first attested in 1250) and *mill* (from Latin, first attested in 962, but probably borrowed several centuries earlier).

Evidently, identifying early loanwords such as English *mill* is not possible in languages that do not have a long written history. Thus, in order to make the data more comparable, it might be advisable to limit oneself to more recent loanwords (say, of the last 300-500 years) even in languages where we would be able to identify older loanwords.

Loanwords can be more or less integrated phonologically and morphologically. It is unclear whether the degree of integration should play a role in this project. Much of the current literature on loanwords is devoted to studying regularities of integration as a way of gaining insight into the nature of phonological and morphological structure; the focus of this project is clearly very different.

The main focus of the project is on loanwords, rather than loanshifts such as calques and semantic borrowings. It remains to be seen to what extent these can be taken into consideration.

References

Blank, Andreas. 1997. *Prinzipien des lexikalischen Bedeutungswandels am Beispiel der romanischen Sprachen.* Tübingen: Niemeyer.

Blust, Robert. 2000. "Why lexicostatics doesn't work: the 'universal constant' hypothesis and the Austronesian languages." In: Renfrew et al. 2000:311-331.

Brown, Cecil H. 1999. *Lexical acculturation in Native American languages.* New York: Oxford University Press.

Buck, Carl Darling. 1949. *A Dictionary of Selected Synonyms in the Principal Indo-European Languages.* Chicago: The University of Chicago Press.

Burrow, Th. 1946. "Loanwords in Sanskrit." *Transactions of the Philological Society*, 1946, 1-30

Cavalli-Sforza, L.L. & W.S.-Y. Wang. 1986. Spatial distance and lexical replacement. *Language* 62: 38-55.

Curnow, Timothy Jowan. 2001. "What language features can be 'borrowed'?" In: Aikhenvald, Alexandra Y. & Dixon, R.M.W. (ed.) 2001. *Areal diffusion and genetic inheritance.* Oxford: Oxford University Press, 412-36.

Dyen, Isidore & James, A.T. & Cole, J.W.L. 1967. "Language divergence and estimated word retention rate." *Language* 43: 150-171.

Field, Fredric W. 2002. *Linguistic borrowing in bilingual contexts.* Amsterdam: Benjamins.

---

[4] This definition presupposes continued existence of a language with no break in transmission. If creole languages are regarded as languages with a break in transmission and no straightforward genealogical ancestor, then the notion of loanword does not apply to creole languages.

Haugen, Einar. 1950. "The analysis of linguistic borrowing." *Language* 26: 210-31.

Hekking, Ewald & Bakker, Dik. 1999. "El Otomí y el español de Santiago Maxquititlán: Dos lenguas en contacto." *Foro Hispánico* 13: 45-74.

Hekking, Ewald & Muysken, Pieter. 1995. "Otomí y Quechua: una comparación de los elementos gramaticales prestados del español." In: Zimmermann, Klaus (ed.) *Lenguas en contacto en Hispanoamérica: Nuevos enfoques.* Frankfurt am Main: Vervuert, 101-118.

Hock, Hans Henrich & Joseph, Brian D. 1996. *Language history, language change, and language relationship.* Berlin: Mouton de Gruyter.

Hymes, Dell. 1960. "Lexicostatistics so far." *Current Anthropology* 1: 3-44.

Johanson, Lars. 2002. *Structural factors in Turkic language contacts.* London: Curzon.

Kapitan, M.E. 1994. Influence of various system features of Romance words on their survival. *Journal of Quantitative Linguistics* 1: 237-76.

Lees, Robert B. 1953. "The basis of glottochronology." *Language* 29: 113-27.

Lohr, Marisa. 1999. *Methods for the genetic classification of languages.* Ph.D. dissertation, University of Cambridge.

Matras, Yaron. 1998. "Utterance modifiers and universals of grammatical borrowing." *Linguistics* 36:281-331.

McMahon, April & McMahon, Robert. 2003. "Finding families: quantitative methods in language classification." *Transactions of the Philological Society* 101: 7-55.

Moravcsik, Edith A. 1975. "Verb borrowing." *Wiener Linguistische Gazette* 8: 3-30.

Moravcsik, Edith A. 1978. "Universals of language contact." In Joseph H. Greenberg et al. (ed.) *Universals of human language, volume 1, Method and theory*. 93-122. Stanford, CA; Stanford University Press.

Moravcsik, Edith A. 2003+. "Borrowed verbs." Ms., University of Wisconsin, Milwaukee.

Myers-Scotton, Carol. 2002. *Language contact: Bilingual encounters and grammatical outcomes.* Oxford: Oxford University Press.

Oswalt, R.L. 1975. "The relative stability of some syntactic and semantic categories." *Working Papers on Language Universals (Stanford)*, vol. 19: 1-19.

Oswalt, Robert L. 1971. "Towards the construction of a standard lexicostatistic list." *Anthropological Linguistics* 13: 421-34.

Poplack, Sandra & Sankoff, David. 1984. "Borrowing: the synchrony of integration." *Linguistics* 22:99-136.

Poplack, Shana & Sankoff, David & Miller, Christopher. 1988. "The social correlates and linguistic processes of lexical borrowing and assimilation." *Linguistics* 26:47-104.

Poplack, Shana & Sankoff, David & Miller, Christopher. 1988. "The social correlates and linguistic processes of lexical borrowing and assimilation." *Linguistics* 26:47-104.

Renfrew, Colin & McMahon, April & Trask, Larry (eds.) *Time depth in historical linguistics.* 2 vols. Cambridge: McDonald Institute for Archaelogical Research.

Ross, Malcolm. 1991. "Refining Guy's sociolinguistic types of language change." In: *Diachronica* 8.1: 119-29.

Ross, Malcolm. 1997. "Social networks and kinds of speech-community events." In: Blench, Roger & Spriggs, Matthew (eds.) *Archaeology and language I: Theoretical and methodological orientations.* London: Routledge, 209-61.

Ross, Malcolm. 2001. "Contact-induced change in Oceanic languages in north-west Melanesia." In: Aikhenvald, Alexandra Y. & Dixon, R.M.W. (eds.) 2001. *Areal diffusion and genetic inheritance.* Oxford: Oxford University Press, 134-66.

Sankoff, David. 1970. "On the rate of replacement of word-meaning relationships." *Language* 46: 564-69.

Swadesh, Morris. 1955. "Towards greater accuracy in lexicostatistic dating." *International Journal of American Linguistics* 21: 121-37.

Szemerenyi, Oswald. 1974. "The origins of the Greek lexicon: Ex oriente lux." *Journal of Hellenic Studies* 94: 144-157.

Thomason, Sarah G. 2001. *Language contact.* Washington, D.C.: Georgetown University Press.

Thomason, Sarah Grey. & Kaufman, Terrence. 1988. *Language contact, creolization, and genetic linguistics.* Berkeley: University of California Press.

Thomason, Sarah Grey. 1997+. "On the unpredictability of contact effects." Ms., University of Michigan.

Tryon, Darrell T. (ed.) 1995. *Comparative Austronesian dictionary: An introduction to Austronesian studies.* Parts 1-4. Berlin: Mouton de Gruyter.

van Hout, Roeland & Muysken. 1994. "Modeling lexical borrowability." *Language Variation and Change* 6: 39-62.

Vennemann, Theo. 2000. "Zur Entstehung des Germanischen." *Sprachwissenschaft* 25.3: 233-269.

Whitney, William Dwight. 1881. "On mixture in language." *Transactions of the American Philosophical Association* 12:1-26.