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

Martin Haspelmath

1. Introduction

This paper gives an overview of some of the general issues arising when one studies lexical borrowing across languages. It discusses the motivations and goals (Section 2), kinds of loanwords (Section 3), factors influencing borrowability (Section 4), and factors determining the borrowing behavior of different languages (Section 5). The context is a collaborative cross-linguistic project on loanwords and lexical borrowability coordinated by the Max Planck Institute for Evolutionary Anthropology (2007–2008).[1]

2. 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 (cf. Haspelmath 2004), 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.

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 and 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 so far.[2] 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. Wilkins 1996; 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 (grammaticalization, see Heine and 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 assumption is overly optimistic (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 lexico-

statistics 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 and 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 and 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 and 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 ongoing 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 planned outcome of the project will be an edited 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 will 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 which are translated into each language. The list consists of 1460 lexical meanings, most of which are taken from Mary Ritchie Key's *Intercontinental Dictionary Series*.[3] 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 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 borrowability in a comparative perspective.

3. 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 and 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 native speakers who consciously import a word from another language, whereas typical shift-induced interference is created by non-native speakers who uncon-

sciously impose features of their native language to the recipient language. 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.

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

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

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".

4. Factors for differential borrowability of word meanings

4.1. 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*).[4] For instance, Matras (1998) proposes the scale in (2) for coordinators:

(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 borrowed 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 borrowed preferentially.

One of the goals of the Loanword Typology project is to find implicational borrowability scales for (sets of) lexical meanings (see Section 4.5, and Table 6 in Section 6.2).

4.2. 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 borrowed 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 and Muysken (1994) cite supporting data on content vs. functions word types from their Quechua corpus (which contains many Spanish borrowings), shown in Table 2.

Table 2. Native and borrowed function vs. content words in Quechua

	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

4.3. Parts of speech

It is widely acknowledged that nouns are borrowed more easily than other parts of speech (e.g. Whitney 1881, Moravcsik 1978, Myers-Scotton 2002: 240). Van Hout and 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 elements 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 and Muysken (1994) again cite data from their Quechua corpus, shown in Table 3.

Table 3. Native and borrowed nouns vs. verbs in Quechua

	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 and Kaufman 1988: 348; however, French does have loan verbs such as *shooter*). Moravcsik (1975, 1978: 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, Wichmann and Wohlgemuth 2008).

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 and Muysken 1995 and Hekking and Bakker 1999).

4.4. 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 and Muysken (1994) find some evidence for this in their Quechua corpus. The following table gives the percentage of Spanish loans in 7 frequency classes.

Table 4. Word frequency and borrowing rate

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

4.5. Lexical semantic field

One area where little systematic research has been done 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). It is here that the Loanword Typology project can make an important contribution.

5. Factors for differential borrowing behavior among different languages

5.1. Intensity of language contact

The most obvious sociolinguistic factor favoring borrowing is widespread bilingualism. This is often called "intensity of contact". Thomason and 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 Guaraní in Paraguay, where Guaraní is spoken by many speakers of non-Guaraní origin, but Paraguayan Spanish has very few Guaraní 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).

5.2. 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.

5.3. Structural incompatibility

Structural incompatibility has often been invoked as explaining resistance to borrowing, although in recent years it has come under attack (especially by Thomason and 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 at present whether it might be relevant for lexical borrowing.

5.4. Genealogical relatedness

McMahon (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.

6. 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.

6.1. 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 5, where lexical meanings are ranked by borrowability. The number following each lexical meaning is the percentage of languages in which it is a loanword from a European language. Thus, "coffee" is a loanword in 81% of the languages for which Brown has information (not always all the 292 languages), whereas the remaining 19% have native words for "coffee" (compounds, derivatives, or simple words whose meaning was extended or shifted).

Among many other things, Brown observed the following tendencies:

(1) "Words for natural kinds tend more strongly than those for introduced artifacts to be associated with high borrowability." (Brown 1999: 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 Loanword Typology project in that it considers only 77 lexical meanings, but this allows him to take into account a very high number of languages.

Table 5. Items of acculturation, ranked by index of borrowability (European loan percentage)

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

6.2. Words of the IDS list in 80 Austronesian languages

The *Comparative Austronesian Dictionary* (Tryon 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 (see note 3). We went through all four volumes of the dictionary and identified the percentage of languages showing loanwords for each lexical meaning.[5] Table 6 shows the average percentages for each lexical field.

Table 6. 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 7 is an excerpt from the list of lexical items. Each lexical item is preceded by its IDS number and is followed by the percentage of Austronesian languages in the dictionary in which it is a loanword.

Table 7. Some randomly selected IDS word meanings, ranked by percentage of loanwords in the 80 Austronesian languages of Tryon 1995

01.212	earth=ground, soil	0	12.540	measure	11.68
01.222	cliff, precipice	0	20.460	surrender	11.76
01.270	shore	0	02.760	widow	12
01.310	water	0	13.440	three times	12
01.323	rough (of sea)	0	.		
01.342	reef	0	19.110	country	20
01.352	tide	0	19.370	citizen, subject	20
01.353	lowtide	0	20.310	armor (defensive)	20
01.430	wood	0	07.470	shelf	20.28
01.440	stone, rock	0	08.830	citrus fruit	20.89
01.530	moon	0	22.220	preach	21.21
01.550	lightning	0	…		
01.560	thunder	0	14.780	season	31.14
01.570	lightning (as striking)	0	07.580	arch	31.57
01.620	darkness	0	09.422	tool	31.57
01.630	shade, shadow	0	06.730	ring (for finger)	31.74
01.640	dew	0	05.370	spoon	31.94
01.720	wind	0	03.370	he-goat	32.07
01.740	fog	0	…		
01.750	rain (noun)	0	03.260	ram	55.10
01.852	burn (vb intrans)	0	06.920	brush	55.31
01.880	firewood	0	08.691	pipe	55.55
02.210	man (vs.woman)	0	20.170	soldier	56.92
02.220	woman	0	05.270	kettle	57.37
…			08.240	shovel	57.40
05.220	boil (vb)	3.89	…		
07.560	post, pole	3.89	14.530	clock, timepiece	71.42
09.110	do, make	3.89	18.560	paper	72
14.450	noon, midday	3.89	03.560	goose	72.5
15.220	smell (vb trans)	3.89	03.410	horse (equine)	73.13
11.820	sell	3.94	07.240	key	73.13
…			05.630	sausage	73.91
15.440	sound, noise	7.5	…		
18.120	sing	7.5	05.930	beer	91.30
18.210	speak, talk	7.5	03.780	camel	93.33
17.440	suspect	7.57	05.760	grape	94.11
…			03.460	ass, donkey	94.59
14.760	summer	11.62	03.470	mule	100
02.520	aunt	11.66	03.770	elephant	100
01.280	cave	11.68	05.880	cheese	100
			05.890	butter	100

7. Some choices of the Loanword Typology project

7.1. Which languages?

Since the goal of the Loanword Typology 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 exclusively through the spoken language, and cases where many loanwords were introduced through writing. (Of course, in practice the choice of languages is to a large extent also determined by the linguists who are willing to collaborate on this project.)

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, Yiddish/German, Iranian Azerbaijani/Azerbaijani of the Republic of Azerbaijan, Riau Indonesian /Standard Indonesian.

7.2. 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. Thus, a list such as the IDS word list (consisting of 1310 word meanings) is better suited for the Loanword Typology project. The IDS list does not contain any lexical meanings relating to the modern world ("radio", "truck", "hospital", "election", etc.), and it contains a disproportionate number of lexical meanings that are relevant primarily to

the European (and similar) ecoregion. To make the list more balanced in these regards, 150 meanings have been added, so that the list now consists of 1460 lexical meanings.

7.3. What is a loanword?

A loanword is defined as a word that at some point came into a language by transfer from another language.[6] 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. The Loanword Typology project is taking a more sophisticated approach: For each loanword, authors are asked to record the approximate time at which the word came into the language. This allows us to filter out older loanwords when we want to compare languages with a well-known history with languages about whose history little is known, and in which only young loanwords are likely to be recognizable as such.

Notes

1. See http://www.eva.mpg.de/lingua/files/lwt.html.
2. However, Juliette Blevins (Max Planck Institute for Evolutionary Anthropology) has begun an effort in this direction ("Handbook of Phonological Change").
3. The Intercontinental Dictionary Series (IDS) 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 Word List (consisting of 1310 items), which is itself an adaptation of the list used by Buck (1949). The IDS is currently being developed further by Bernard Comrie (Max Planck Institute for Evolutionary Anthropology) (see http://lingweb.eva.mpg.de/ids).
4. Since the term "hierarchy", when used outside language typology, is generally reserved for taxonomic hierarchies, I prefer the synonymous term "scale".
5. Thanks to my assistants Ulrike Gurt and Jenny Seeg for their invaluable help.

6. 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.

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 *Time depth in historical linguistics*, Colin Renfrew, April McMahon and Larry Trask (eds.), 311–331. Cambridge: McDonald Institute for Archaelogical Research.

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, Luca L. and S.-Y. Wang William
1986 Spatial distance and lexical replacement. *Language* 62: 38–55.

Curnow, Timothy Jowan
2001 What language features can be 'borrowed'? In *Areal diffusion and genetic inheritance*, Alexandra Y. Aikhenvald, and R.M.W. Dixon (eds.), 412–436. Oxford: Oxford University Press.

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

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

Haspelmath, Martin
2004 On directionality in language change with particular reference to grammaticalization. In *Up and down the cline: The nature of grammaticalization*, Olga Fischer, Muriel Norde and Harry Perridon (eds.), 17–44. (Typological Studies in Language 59.) Amsterdam: Benjamins.

Haugen, Einar
1950 The analysis of linguistic borrowing. *Language* 26: 210–231.
Heine, Bernd and Tania Kuteva
2002 *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
Hekking, Ewald and Dik Bakker
1999 El Otomí y el español de Santiago Maxquititlán: Dos lenguas en contacto. *Foro Hispánico* 13: 45–74.
Hekking, Ewald and Pieter Muysken
1995 Otomí y Quechua: una comparación de los elementos gramaticales prestados del español. In *Lenguas en contacto en Hispanoamérica: Nuevos enfoques*, Klaus Zimmermann (ed.), 101–118. Frankfurt a. M.: Vervuert.
Hock, Hans Henrich and Brian D. Joseph
1996 *Language history, language change, and language relationship.* Berlin/New York: 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: 237–276.
Lees, Robert B.
1953 The basis of glottochronology. *Language* 29: 113–127.
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
1994 *Understanding language change.* Cambridge: Cambridge University Press.
McMahon, April and Robert McMahon
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.
1978 Universals of language contact. In *Universals of human language, volume 1, Method and theory*, Joseph H. Greenberg et al. (eds.), 93–122. Stanford, CA: Stanford University Press.
2003 Borrowed verbs. Ms., University of Wisconsin, Milwaukee.

Myers-Scotton, Carol
2002 *Language contact: Bilingual encounters and grammatical outcomes*. Oxford: Oxford University Press.
Oswalt, Robert Louis
1975 The relative stability of some syntactic and semantic categories. *Working Papers on Language Universals* (Stanford) 19: 1–19.
1971 Towards the construction of a standard lexicostatistic list. *Anthropological Linguistics* 13: 421–434.
Poplack, Sandra and David Sankoff
1984 Borrowing: the synchrony of integration. *Linguistics* 22: 99–136.
Poplack, Shana, David Sankoff and Christopher Miller
1988 The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics* 26: 47–104.
Renfrew, Colin, April McMahon and Larry Trask (eds.)
2000 *Time depth in historical linguistics.* 2 vols. Cambridge: McDonald Institute for Archaelogical Research.
Ross, Malcolm
1991 Refining Guy's sociolinguistic types of language change. *Diachronica* 8(1): 119–129.
1997 Social networks and kinds of speech-community events. In *Archaeology and language I: Theoretical and methodological orientations*, Roger Blench, and Matthew Spriggs (eds.), 209–261. London: Routledge.
2001 Contact-induced change in Oceanic languages in north-west Melanesia. In *Areal diffusion and genetic inheritance*, Alexandra Y. Aikhenvald, and R.M.W. Dixon (eds.), 134–166. Oxford: Oxford University Press.
Sankoff, David
1970 On the rate of replacement of word-meaning relationships. *Language* 46: 564–569.
Swadesh, Morris
1955 Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics* 21: 121–137.
Szemerenyi, Oswald
1974 The origins of the Greek lexicon: Ex oriente lux. *Journal of Hellenic Studies* 94: 144–157.
Thomason, Sarah Grey
2001 *Language contact*. Washington, D.C.: Georgetown University Press.
Thomason, Sarah Grey and Terrence Kaufman
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/New York: Mouton de Gruyter.
van Hout, Roeland and Pieter 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.
Wichmann, Søren and Jan Wohlgemuth
2008 Loan verbs in a typological perspective. This volume.
Wilkins, David
1996 Natural tendencies of semantic change and the search for cognates. In *The comparative method reviewed*, Malcolm Ross and Mark Durrie (eds.), 264–304. New York: Oxford University Press.