Modeling lexical borrowability

ROELAND VAN HOUT
Tilburg University

PIETER MUYSKEN
University of Amsterdam

ABSTRACT

In this article, we develop analytical techniques to determine borrowability—that is, the ease with which a lexical item or a category of lexical items can be borrowed. The analysis is based on two assumptions: (1) the distribution of items in both the host and donor language should be taken into account to explain why certain items are, and others are not, borrowed; (2) the borrowability of a lexical category may result from a set of (underlying) operative factors or constraints. Our analysis is applied to Spanish borrowings in Bolivian Quechua on the basis of a set of bilingual texts.

Borrowing of lexical items is a subject that has given rise to many offhand remarks, but to little systematic study. The reasons for this are twofold. First, lexical borrowing has not been perceived as particularly interesting from a structural perspective, but rather as a cultural phenomenon. Second, the study of lexical borrowing has suffered from the fact that the lexicon is difficult to study using the standard tools of structural linguistic analysis. Here we focus on one particular aspect of the study of borrowing—borrowability. To what extent can a given item or class of items be borrowed into another language? Are there differences between different lexical items as to the ease with which they are borrowed? If so, what causes these differences? In what follows we combine two different traditions in dealing with the issue of borrowability: (1) a tradition relating borrowability to grammatical patterns, (2) corpus-based sociolinguistic research treating borrowability from a quantitative and statistical perspective.

The present study looks at the borrowing of Spanish elements in Bolivian Quechua on the basis of a bilingual corpus. We are not looking at the history or fate of individual words (e.g., the very early borrowing of Spanish *parlar* for 'to speak' and its subsequent disappearance from the donor language, where now *hablar* is the common form), but at general factors or constraints. Our approach fits into the general constraints framework developed by Sankoff and Labov (1979). Before we approach the problem of borrowability in more detail, let us briefly survey a number of questions involved in the general study of lexical borrowing. These questions include problems of delimitation, which we cannot begin to deal with in any satisfactory way.

The first problem involves multiple-word borrowing. Although generally lexical items are single words, sometimes they correspond to phrases, as in the case of idioms and fixed expressions. These may again be borrowed as wholes, thus resembling multiword switches. In this article, we do not deal with this problem. We took the existence of Spanish borrowed fixed expressions (such as *así que* 'so that', *a ver* 'let's see', and *ya está* 'it's there') into account in the analysis of the data, but we did not analyze the component parts of the expressions. This means that we did not take fixed expressions into further consideration.

The second problem involves the degree of adaptation. Words may be borrowed and then undergo various degrees of adaptation. We do not yet know what determines this process of integration, but it may be relevant to borrowability because the latter may change over time. There may be long-term integration of items, and there may be the development of channels for integration, as suggested by Heath (1989). We should state right away that we do not deal with phonological aspects of the adaptation of borrowings, however important they may be from all points of view. In fact, we collapsed different forms of a borrowed word, including diminutives, in our data base (e.g., *burro*, *burru*, *burriquitu*, *borriquito*, *burrito* 'donkey').

Third, often lexical borrowing will go together with phenomena such as syntactic convergence or influence, and it is hard to separate the effects of the two in individual cases. A case in point is the Quechua numeral *uj* 'one'. This is a numeral, but it can also be used as the indefinite article, possibly under the influence of Spanish *un* 'a', which is related to *uno* 'one'.

The fourth issue concerns the types of borrowing. The most complex typology of borrowing is due to Haugen (1950), who introduced a number of concepts such as *loan-blend*, *loan-shift*, and so forth. From an anthropological perspective, a different basic distinction in lexical borrowing is made by Albó (1968), who distinguished between *substitution* and *addition* of vocabulary. There is substitution if the borrowed item is used for a concept that already exists in the culture (and, of course, is expressed by a lexical item in the host language), and addition if it is a new concept. This relates directly to the *embedding* problem of Weinreich, Herzog, and Labov (1968): the structural and social embedding of the borrowed elements. What is the relation of a borrowing with existing semantic fields or structured meaning domains?

Finally, a fundamental problem for any typology of lexical borrowings is how to distinguish between words that are taken from another language in discourse only accidentally, in which case we speak of *lexical interference* or *nonce borrowings* (cf. Sankoff, Poplack, & Vanniarajan, 1990), and words that become fully integrated into the receptor language. We look at nonce borrowing later, in that we consider the issue for which word categories borrowing is a productive process. In other respects, we see no reason to make a fundamental distinction between nonce borrowings and ordinary borrowings.

HIERARCHIES OF BORROWABILITY: AN OVERVIEW

The words of a language are loose elements, but at the same time they are part of a system: the lexicon itself is partly structured, and the context in which words occur in the sentence may also impose structural constraints on their occurrence. These constraints may manifest themselves in the fact that some categories appear to be borrowed more easily than others or at least are borrowed more frequently than others. This fact was observed by the Sankritist William Dwight Whitney (1881), who arrived at the following hierarchy:

(1) nouns — other parts of speech — suffixes — inflection — sounds

This hierarchy was elaborated on by Haugen (1950), using data from Norwegian immigrants in the United States, to include:

(2) nouns — verbs — adjectives — adverbs — prepositions — interjections — . . .

Nouns are borrowed more easily than verbs according to this perspective, verbs more easily than adjectives, and so on. Independently from Haugen, Singh (1981) came to a comparable hierarchy on the basis of English borrowings in Hindi:

(3) nouns — adjectives — verbs — prepositions

On the basis of data from Spanish borrowings in Quechua, Muysken (1981) tentatively concluded that there may be something like the following hierarchy:

(4) nouns — adjectives — verbs — prepositions — coordinating conjunctions — quantifiers — determiners — free pronouns — clitic pronouns — subordinating conjunctions

The data Muysken used included, among other things, absolute numbers of Spanish words in a given corpus of recorded speech (types, not tokens). These numbers are given in Table 1.

Several main issues come to the fore immediately in relation to meaning of the concept of hierarchy. First, what is precisely claimed in setting up such hierarchies as in (1)–(4)? Can they be translated into the kind of *implicational universals* of borrowing developed by Moravcsik (1978) (e.g., when a language has borrowed verbs, it also has borrowed nouns)? Is there a *quantitative* claim (e.g., in the set of borrowed elements, there are more nouns than verbs)? Is there a *temporal* claim (e.g., a language first borrows nouns, and only then verbs)? These interpretations are compatible, but clearly separate.

TABLE 1. *Spanish borrowings in*
Ecuadorian Quechua:
Number of types, absolute figures

Nouns	221
Verbs	70
Adjectives	33
Sentence adverbs	15
Quantifiers	7
Conjunctions	6
Prepositions	5
Interjections	5
Negation	2
Manner adverbs	1
Greetings	1

Source: Muysken (1981).

We limit ourselves to quantitative analysis, given the fact that we are doing neither a typological nor a historical study.

Next, how are hierarchies of type (1)–(4), insofar as they hold true, to be explained? What factors or constraints can be appealed to in order to explain the hierarchies found? In this article, we explore a number of such factors.

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. More generally, the class of words most closely involved with the culture of a language are the *content words*, such as adjectives, nouns, verbs. They may be borrowed more easily than *function words* (e.g., articles, pronouns, conjunctions) because the former have a clear link to cultural content and the latter do not. In some cases, borrowing extends beyond cultural content words, however, and there may well be other constraints on borrowing (e.g., distinguishing among different kinds of content words).

Another explanatory factor to consider is the *frequency* of lexical items, perhaps both in the donor language and in the recipient language (cf. Poplack, Sankoff, & Miller, 1988).

The next factor, in fact a cluster of factors, is *structural* in nature. To what extent do syntagmatic and paradigmatic constraints on lexical items, again both in the donor language and in the recipient language, influence their borrowability? Here we focus on the role of inflection, the extent to which forms are part of a paradigm, the role words play in the clause, and so on.

Finally, there is the factor of *equivalence* to consider. Weinreich (1953:61) noted that resistance to borrowing is always a function, not so much of properties of recipient and source languages by themselves, but of the difference in structures of the recipient and source languages.

METHODOLOGICAL APPROACH:
PRELIMINARY CONSIDERATIONS

It is not possible to establish hierarchies of borrowability simply by count-
ing elements in a corpus or dictionary. There are four important questions
relevant to studying differences in borrowability between different types of
words empirically:

1. Which categorial distinctions do we make?
2. Do we count types or tokens?
3. To what do we compare a given number of borrowed items?
4. What type of corpus do we use?

In the present study, we look at the borrowing of Spanish elements belong-
ing to different word classes in Bolivian Quechua, basing our work on a bilin-
gual corpus consisting of about 40 fairly short folkloric texts recorded in the
Department of Potosí by the Jesuit linguist Federico Aguiló (1980). They are
literally transcribed, and one of the advantages of the corpus is that Aguiló
had the stories translated into Spanish by bilingual peasants of the same
region. Thus, we have a source corpus of Spanish available that corresponds
in content to the Quechua corpus and also to the type of Spanish that will
be the source for at least the more recent borrowings into Quechua (which,
according to Albó, 1968, will be a substantial majority, given that massive
bilingualism dates from the post-1952 era). To give but one example of the
type of matching possible, consider the sentences taken from the Quechua
corpus (5a) and the Spanish corpus (5b).

(5) a. *pagaway uj   qolqe durowan   willasqaykitaj*
       pay-me   one silver hard-with what-you-told
    b. *se había comprometido pagarle con plata dura*
       He had promised to pay him with hard silver.

Even though the translation is not a literal one, it is clear that a subset of the
items in the Spanish translation also occurs, as borrowings, in the Quechua
text.

Almost all studies of borrowing so far have focused on the main syntac-
tic categories (e.g., nouns, verbs, etc.). There has been some work that has
taken semantic fields into account. In our research, we limit ourselves to the
morphosyntactic categories. In Table 2 we illustrate the categories used in our
analysis. As becomes clear, however, the morphosyntactic categories play a
role at the level of description, not explanation. Indeed, the thrust of our
argument is to study why certain word classes behave the way they do in the
borrowing process.

The main issue in dealing with the hierarchies discussed in the previous sec-
tion is how to analyze quantitative borrowing data of the type in Table 1, the

TABLE 2. *The categories employed to classify the elements in the corpus*

		Spanish	Quechua
a	adjective	*triste* 'sad'	*hatun* 'big'
adv	adverb	*ahora* 'now'	*ahina* 'thus'
aux	auxiliary	*esta-* 'be'	
comp	complementizer	*que* 'that'	
conj	conjunction	*y* 'and'	
cop	copula	*se-* 'be'	*ka-* 'be'
dem	demonstrative	*esa* 'that (f)'	*kay* 'this'
det	determiner	*los* 'the [m.pl.]'	
excl	exclamative, interjection, ideophone	*carajo* 'damn'	*huay* 'hey'
		bueno 'well'	*ari* 'yes'
			ch'in 'scratch'
n	noun	*mujer* 'woman'	*chaupi* 'half'
name		*Manuel*	*Sik'imira*
neg	negation	*ni* 'nor'	*mana* 'not'
num	numeral	*cuatro* 'four'	*huq* 'one'
p	preposition	*sin* 'without'	
	postposition		*hina* 'like'
p + det	preposition + determiner combination	*al* 'at the'	
p + pron	preposition + pronoun combination	*contigo* 'with you'	
poss	possessive pronoun	*tuyo* 'yours'	
pron	pronoun	*ella* 'she'	*qan* 'you'
pron-cl	clitic pronoun	*me* 'me'	
q	quantifier	*todo* 'all'	*wakin* 'some'
v	verb	*come-* 'eat'	*puri-* 'walk'

list of Spanish elements borrowed into Ecuadorian Quechua. Obviously they cannot be directly used to establish a hierarchy of the type in (1)–(4), since there may be differing amounts of elements of each category available for borrowing. Spanish has many more nouns than verbs, and the fact that three times as many nouns than verbs were borrowed could be interpreted, if we take the percentage of elements of a category borrowed, as meaning that verbs are easier to borrow than nouns.

The data in Table 1 give a distorted picture for yet another reason: *types* are counted, not *tokens*. This makes a big difference, because one word may be used many times. In the Sango corpus studied by Taber (1979), 508 French loans accounted for 51% of the types, but for only 7% of the tokens: they are used relatively infrequently. A token analysis of the elements in Table 1 shows that Spanish nouns are much more frequent than this table suggests, and elements such as prepositions, adverbs, and quantifiers are much less frequent. Should one look at the number of different elements borrowed (the types), or should one take into account the number of tokens (occurrences)? In terms of the vocabulary of the recipient language and the speakers using the vocabulary, it seems more relevant to determine how many lexical items are borrowed. In speaking the recipient language and actually producing borrowed elements, the frequency of occurrence of borrowed lexical items (num-

ber of tokens) can be a critical factor. It implies that it is relevant to compare borrowed items both on the number of types and tokens. It also implies that word-class distribution can be measured both on the level of number of items (types) and number of tokens.

There are several possible approaches to the issue of statistically comparing borrowings to other categories in the lexicon. If we have a language sample containing borrowed elements or more particularly borrowed lexical items, the next step is to compare this set of elements to something else. A purely set-internal, word-class distribution (i.e., where only a set of borrowed items is taken into account), as adopted in Table 1, is not revealing. A set-external comparison is called for, involving either the recipient or the donor language.

A first approach, then, is set-external comparison with word-class distributions of the recipient language (both word-class distributions involved belong to the same language, i.e., the recipient language). Taking such an approach, there are two options.

As is done in Poplack et al. (1988), who investigated English-origin vocabulary in Ottawa French, the borrowed elements can be calculated in terms of percentages of the recipient language vocabulary (tokens; Fr = French origin, En = English origin).

(6) Nouns(Fr) : Verbs(Fr) : Adjectives(Fr) : ... 100%
    Nouns(En) : Verbs(En) : Adjectives(En) : ... 100%

In this way, the percentage of nouns in the French-origin vocabulary in Ottawa French is contrasted with the percentage of nouns in the English-origin vocabulary, and so on. The structural possibilities and the frequency patterns of the recipient language are taken as the point of comparison. For this procedure to be most revealing, there needs to be a basic correspondence between the recipient language categories and the categories of borrowed elements.

A similar approach is to look within each category at the proportion of borrowed elements (in the corpus studied).

(7) Nouns(Fr) Verbs(Fr) Adjectives(Fr) ...
    :   :   :
    Nouns(En) Verbs(En) Adjectives(En) ...
    100%   100%   100%

To take once again the example of English and French, here we would simply take the category of nouns in the recipient language corpus and see what percentage is of French origin and what is of English origin. This approach has the advantage that the lack of correspondence of recipient and source language categories is no problem from a computational point of view. If either language lacks a category, we simply have a 100% distribution.

TABLE 3. *The distribution of different clause types in 341 Yagua clauses (running text)*

Only a verb	238
V + subject NP + object NP	23
V + three participant NPs or PPs	1
V + one NP (subject or object)	79

Source: Payne (1986:442).

However, when alien categories are borrowed (e.g., Spanish conjunctions and prepositions in many Amerindian languages), a comparison with the distribution in terms of the guest language is meaningless. This type of borrowing is a manifestation of the fact that, in some cases, lexical borrowing is linked to or correlated with structural borrowing. And then a comparison with the distributional properties of the source language is more to the point.

A more fruitful approach, and in fact the one we opt for, is therefore set-external comparison with a source-language corpus (the word-class distributions involved belong to different languages: the borrowed items in the recipient language and the items of the source language). Here the distribution in the source language is taken as a point of comparison. Then the question is: What is the chance an individual source element will end up in the recipient language?

There are no particular properties a bilingual corpus must have, apart from the ones usually suggested. However, it should be clear that different languages will show different distributional patterns. In many Amerindian languages, there is extensive cross-referencing morphology on the verb and hence little need to express the arguments of the clause with lexical means or pronouns. A typical example is Yagua, a Peruvian Amazonian language which Payne (1986) described as follows: "it is reasonable to suggest that the basic clause type of Yagua is an inflected verb and nothing more, that the use of full noun phrases or free pronouns is somehow a marked usage" (441).

A frequency count of a number of Yagua texts gives the distribution in Table 3, which amounts to 341 verbs and 128 nouns or pronouns for the text sample analyzed. In addition, the number of pronouns will be modest, when compared to a language like English. Similar observations can be made with respect to auxiliaries and modals in languages such as Yagua. Whereas in English tense categories are frequent in most types of texts, in Amerindian languages (such as Yagua) the notions of tense, mood, and aspect are marked on the verb with affixes. Although there are still some elements that may be classified as auxiliaries, their frequencies in texts will be much lower than in English.

Most studies on borrowing give some indication of the word-class distribution of the lexical elements borrowed. What is the use of comparing the

elements being borrowed in terms of word classes, and how are the figures evaluated? What figures are in fact necessary to say anything meaningful about the role of word-class categories?

It will be clear that using a bilingual corpus has great advantages for the particular approach taken. For one thing, it makes it possible to obtain significant results, even though the size of the corpus is relatively modest.

ANALYSIS

The analysis of the borrowings took place in several steps. We first had to establish the relevant types and tokens in the data base before we could analyze the data quantitatively.

Data bases

The lexical data bases used in the analyses are standardized, in that endings (including flection and derivational affixes) have been taken off and spelling variants have been regularized. Because Quechua morphology is very rich, many different forms of a word were counted as one lexical type. This also holds for the Spanish borrowings in that data base, since they were often marked with all relevant Quechua suffixes as well. We started with two separate data bases.

1. Quechua: containing both Quechua word types and Spanish word types and their respective number of tokens; the data are derived from the corpus of Quechua texts.
2. Spanish: containing Spanish word types and their number of tokens; the data are derived from the corpus of Spanish texts.

These two data bases were combined into one general data base.

3. Combined: containing all word types of the Quechua and the Spanish data bases and their respective number of tokens; word types not occurring in one of the original data bases simply have an occurrence of zero for the language corpus in question.

As can be seen in Table 4, the Quechua data base contains 908 word types and 6,870 tokens; 28 types could not be classified as either Quechua or Spanish.[1] Of the remaining 880 types, 40.8% are Spanish-based. The relative share in tokens is much lower: 23.1% are Spanish-based. Still the amount of Spanish elements is fairly high, particularly when we compare it to other corpora such as the Ottawa corpus (Poplack et al., 1988) or the Brussels corpus (Treffers, 1991).

The number of hapaxes (word types that occur only once) is also given in Table 4. Their frequency divided by size of the sample (the number of tokens)

TABLE 4. *Number of tokens, types, and hapaxes in the Quechua data base related to their origin*

Language	Types	Tokens	Hapaxes
Undeterminable	28	80	22
Quechua	517	5,224	185
Spanish	363	1,566	170
Total	908	6,870	377

TABLE 5. *Number of tokens, types, and hapaxes in the Spanish data base related to their origin*

Language	Types	Tokens	Hapaxes
Undeterminable	4	28	0
Quechua	37	231	14
Spanish	992	10,617	408
Total	1,033	10,876	422

gives an impression of the productivity of a specific word category (cf. Baayen, 1989). One problem is that our sample is not particularly large, especially not if the number of Spanish words is considered.[2] Nevertheless, one can conclude that the number of Spanish hapaxes in relation to the number of Spanish word tokens is fairly high. This result indicates that the number of possible Spanish words (the types) in Quechua is not exhausted at all by our sample. If a larger sample were drawn (in fact, it is more realistic to say that if a larger corpus would have been available), new Spanish word types would be found.[3] The outcomes in our sample suggest that Spanish-based words in Quechua is a rather extended category. The contribution of Spanish words to Quechua is not restricted to a small and specific set of borrowed Spanish word types.

As Table 5 shows in comparison with Table 4, the Spanish data base contains more word tokens than the Quechua data base, due to the fact that in Spanish many concepts are expressed by separate words, including determiners and clitic pronouns, that are affixal in Quechua.

Table 5 shows too that several Quechua word types are present in the Spanish data base. That is not a surprise, given the fact that the Spanish texts were translations of the original Quechua texts by bilingual speakers from the same region in Bolivia. The word-class distribution of the Quechua word types is as follows: 7 exclamations, 25 nouns, 5 names, 1 verb. The most frequent Quechua word in the Spanish data base is the noun *cóndor* (number of tokens = 73), a word which was already borrowed from Quechua (*kuntur*)

long ago. We pay no further attention to this Quechua subset in the Spanish data base because this subset is relatively small and not very interesting from a more structural point of view.

With respect to lexical richness, normally one may expect that the Spanish and Quechua data bases are about equally rich in the lexical means used, despite the difference in the number of different word types. Measures of lexical richness relate in some way the number of types to the number of tokens. One of the better measures for lexical richness (cf. Broeder, Extra, & van Hout, 1993; van Hout & Vermeer, 1989) is the number of types divided by the square root of the number of tokens (especially when the size of the sample is relatively small); a particularly bad measure would be the type/token ratio.[4] Applying the square root measure, the Quechua corpus gives an outcome that is even higher than the one obtained for the Spanish data base. The Quechua corpus gives an outcome of 10.95 ($908/\sqrt{6870}$); the Spanish corpus gives an outcome of 9.91 ($1033/\sqrt{10867}$). However, this difference can most probably be ascribed to the high number of Spanish borrowings in the Quechua data base, since the extensive borrowing has led to the availability of Quechua and Spanish words for expressing similar concepts. The relatively high correspondence in lexical richness of the two data bases illustrates, as could have been expected, that the two language corpora involved are quite equivalent.

A random distribution?

When we consider the set of borrowed items in a given recipient language sample, there is the obvious possibility that the distribution may be random with respect to word class. The words borrowed into a recipient language might be the outcome of a real random process by which every element has an equal chance of being borrowed. In terms of the recipient language, one might say that each element has an equal chance of being Spanish. Despite the fact that this hypothesis seems to be trivial and uninteresting from a linguistic point of view, the rejection of this hypothesis must be approached from different angles. First, this hypothesis can be related to the number of types or the number of tokens occurring in Quechua. Second, the randomness of borrowing can be related to frequency distributions of the number of types and tokens typical of Spanish.

In order to test the random model for the recipient language, the distribution figures of Quechua and Spanish word types and tokens in the Quechua data base are given in Table 6 for all word classes distinguished. Table 6 also includes the figures for the Spanish lexical items in the Spanish data base. If borrowability is really random with respect to the linguistic properties of the recipient language, the distribution of Spanish words in the Quechua data base should be independent from the word class a word belongs to. Given the figures in Table 6, it is not difficult to reject the randomness hypothesis. We can restrict ourselves to the data on the category of nouns and verbs. The

TABLE 8. *Number of types and tokens of Spanish words in the Quechua and Spanish data bases by word class*

	Types		Tokens	
	Quechua Data Base	Spanish Data Base	Quechua Data Base	Spanish Data Base
Nouns	184	371	823	1,179
Verbs	81	270	241	1,985

Quechua and Spanish data base is given in Table 8. Again, the category of nouns and verbs can be compared. Now the distribution over word classes of the Spanish words in the Quechua data base has to be compared with the distribution in the Spanish data base. The results are obvious. Both the number of verb types and verb tokens are too low in the Quechua data base. The picture would become even more evident if the Spanish words in the categories of aux and cop were incorporated into the category of verbs, since aux occurs in the Spanish data base only (with a high token frequency; $n = 809$) and cop has a high token frequency only in the Spanish data base ($n = 226$; the token frequency of Spanish-based copulas in the Quechua data base is 4).

We have to conclude that the random model has insufficient explanatory power. It does not explain differences in the amount of borrowing between word classes in a satisfactory way. Given this result, we now discuss the way we studied the four possible sets of factors influencing borrowability: lexical content, frequency, structural coherence factors, and equivalence.

Lexical content

We tested the hypothesis that lexical items that potentially carry cultural meaning are primarily borrowed by distinguishing between clear content word types (verbs, nouns, names, exclamatives, adjectives, and adverbs—many of which are to some extent content-bound) and the remaining word types (including clear function words like pronouns, determiners, and quantifiers). We can test this hypothesis on the Spanish word types in the data base. For each Spanish word type we know whether it can be classified as a borrowing in our data base. The distinction between + and −content word types can be made by using the morphosyntactic word-class information available in the data base. Cross-tabulating ±borrowing and ±content word gives the outcomes presented in Table 9.

We notice a strong effect in Table 9. The percentage of borrowings in the category of +content word types is 36.3 (338/930); the percentage in the category of −content word types is 19.2 (25/130). The distinction in ±content words is an interesting candidate to investigate further, especially in relation to other candidate factors. The effect of the content factor is strong, but it

TABLE 9. *Cross-tabulation of ±borrowing and the ±content word distinction of the Spanish word types*

Spanish Word Types	−Borrowing	+Borrowing	Total
−Content word	105	25	130
+Content word	592	338	930
Total	697	363	1,060

is also a matter of degree. Even within the category of −content words, many borrowings were found.

Frequency

A frequency model claims that the chance or probability of a specific word being borrowed is determined by the frequency of the word in question in the donor language. An even stronger interpretation of the frequency model can be formulated by adding the stipulation that not only is borrowing determined by the frequency factor, but so is the number of occurrences of a borrowed word in the recipient language (the number of tokens).

The most natural interpretation of the frequency factor seems to be to take frequency as a kind of pushing factor from the source language. To test the impact of the frequency factor we have to investigate whether the borrowed Spanish words have a relatively high frequency of occurrence in the Spanish data base. The more obvious interpretation of the frequency factor may not prevent us from facing a recipient language version of the frequency factor. In theory, two opposite patterns are possible:

Word frequency in the recipient language operates as a pulling factor; frequent words readily accept borrowed competitors.

Word frequency in the recipient language operates as an inhibiting factor; frequent words resist source language competitors.

The data base available is not set up in terms of a replacement model; a replacement model would require spelling out which borrowed Spanish words compete with and replace which Quechua words. We have to look for indirect evidence if we want to investigate the recipient language interpretation of the frequency factor. All in all, it can be concluded already that an extended interpretation of the frequency model allows a range of various possible processes going on in borrowing. And if frequency plays a role on both sides of the borrowing process, we may be faced with the following "paradox":

Frequency operates as a promoting factor: frequent words in the source language are better candidates for being borrowed.

Frequency operates as an inhibiting or blocking factor: frequent words in the recipient language resist borrowing competitors from the source language.

TABLE 10. *Token frequency of word types in the Quechua data base in relation to the relative share of Spanish word types*

Token Frequency	Frequency Class	Quechua Types	Spanish Types	Total Types	% 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		517	363	880	41.2

The opposite operation of these two forces might imply that especially lex-emes with an average frequency in both the recipient and source language are involved in processes of borrowing.

Table 6 indicates a frequency pattern in the Quechua data base with respect to the token frequency of Quechua versus Spanish word types. The Spanish words have a much lower token–type ratio. For instance, the Quechua verbs have a token–type ratio (= mean frequency) of 9.01 (1929/214), whereas the Spanish verbs have a token–type ratio of 2.98 (241/81). Is this pattern consistent, in the sense that the higher the frequency of a word the lower the probability that a word is Spanish?

In order to get a good impression of the frequency pattern, the word types in the Quechua data base were split up into seven frequency classes. Next, the number of Quechua word types and the number of Spanish word types were counted. The outcomes are presented in Table 10, which shows a step-wise decrease of the share of Spanish word types as token frequency increases. The more frequent a word in the Quechua data base, the less the chance that it is Spanish. This suggests indirectly that frequency in the recipient language may operate as an inhibiting factor. Direct evidence would involve an analysis based on a replacement approach, implying detailed analysis of the Quechua and Spanish lexicon. We do not pursue this here.

What is the role of word frequencies in relation to the source language? For any Spanish word type, we know its token frequency in the Spanish data base. The relationship between frequency and borrowing can be investigated by computing the correlation between the token frequency in the Spanish data base and whether a Spanish word type is being borrowed. Because word frequencies have a rather skewed distribution (with a small number of word types having a very high frequency), it is better to use a transformed frequency score as well — for instance, the log of the frequency. Because a log value of zero does not exist, a value of .5 is added to all frequencies: log frequency = log (frequency + .5). The correlations between ±borrowing and the frequency and the log frequency can be found in Table 11.

TABLE 11. *Correlations between word frequencies of Spanish
words in the Spanish data base and ±borrowing*

	Word Frequency	Log Word Frequency
Yes/no borrowing ($n = 1,060$)	.047*	−.043*
±Borrowing + content words ($n = 930$)	.022	−.064
±Borrowing − content words ($n = 130$)	.320*	.380*

*$p < .01$, one-tailed.

Table 11 gives rather low correlations for all Spanish word types together. The correlation coefficient is significant because of the high number of observations, but the relationship is weak. The log frequency even contradicts the correlation found for the raw frequency; the correlation coefficient is negative. It indicates that the frequency factor for the whole data set can be said to be rather minimal, if not absent altogether. This result is not surprising if one takes into account that highly frequent words are often function words, which are generally resistant to borrowing. This effect can be investigated by distinguishing between clear +content words (a, adv, n, v) and −content words, a distinction we used before. The correlations are given in Table 11.

This table shows that the distinction between ±content words results in a clear difference. The frequency factor operates as a pushing or favoring factor within the class of −content word types. There is no notable effect within the class of +content word types. We must draw the conclusion that frequency effects possibly play a role in the chance of a word of being borrowed, but also that this frequency effect should be interpreted with care because its impact may depend on other factors.

Structural coherence factors

We explore the importance of structural factors from a number of perspectives: peripherality in the clause, case assignment, paradigmatic coherence, and inflection.

It is clear from a number of recent studies on language contact that words that play a peripheral role in sentence grammar, particularly the grammar of the recipient language (e.g., interjections, some types of adverbs, discourse markers, and even sentence coordination markers), are borrowed relatively easily. Note that this is the same class of elements that participates in emblematic switching, the type of phenomenon halfway between intersentential and intrasentential code switching (cf. Poplack, 1980). What this suggests is that switching and borrowing may to some extent be subject to the same type of constraints; both are difficult when the coherence of the clause is disturbed.

A related way to approach the same question is to see to what extent categories are directly implied in the organization of the sentence: for example,

a verb is more crucial to that organization than a noun, and perhaps there-
fore it may be harder to borrow verbs than nouns. Thus, for a noun/verb
asymmetry in borrowability, the principal explanation could lie in the differ-
ent role that these categories play in the organization of the sentence. Nouns
denote elements referred to, and verbs link the elements referred to to each
other. In other words, nouns are inert as far as the syntactic makeup of the
clause is concerned (−structure building). A limited subclass, action nomi-
nalizations, shares the property of allowing a complement with verbs. None-
theless, this complement is never obligatory. Verbs are active in the syntax
and form the nucleus of the clause. More generally, complementizers, aux-
iliaries, copulas, and verbs play a role in structuring the clause, and pre-/post-
positions, determiners, and demonstratives help structure the argument
constituents in the clause (+structure building). We take this feature to be
comparable for Quechua and Spanish.

The central role of the verb is also reflected in its assigning different cases,
which may be specific to that verb and idiosyncratic, to different elements
in the sentence. This also inhibits their being taken from one system to
another. Prepositions share this property with verbs, which may inhibit their
being borrowed. On the other hand, prepositions are rarely inflected them-
selves. Nonetheless, they may not be frequently borrowed. Additional fac-
tors hindering their borrowability include the fact that often their meaning
is grammaticalized (and hence language-specific), that they are sometimes
paradigmatically organized (systematically subdividing a semantic field such
as space in a language-specific way), or that they themselves are part of the
subcategorization of a verb or adjective (*angry with*, *afraid of*, *wait for*,
attend on). This line of thinking would predict that elements such as transi-
tive verbs and prepositions would be harder to borrow than, for example,
nouns.

In addition to these factors deriving from syntagmatic coherence (i.e.,
peripherality, structure-building, case-marking), there is also paradigmatic
coherence. Paradigmatic coherence is the tightness of organization of a given
subcategory. The pronoun system is tightly organized, and it is difficult to
imagine English borrowing a new pronoun to create a second person dual in
addition to second person singular and plural. For this reason, determiners,
pronouns, demonstratives, and other paradigmatically organized words may
be harder to borrow. Although, in principle, the paradigmatic cohesion of
elements of the recipient language is the crucial one, a number of categories
relevant here are simply absent in Quechua, so we took cohesion in either or
both languages as the criterion. Notice also that paradigmatic organization
in the donor language may also stand in the way of borrowing, because par-
adigmatically organized elements often have rather abstract, grammaticalized
meanings which are not accessible independently of the subsystem to which
they belong.[5] Thus, Spanish *este* 'this' and *ese* 'that' may be hard to borrow
due to two factors: first, their Quechua equivalents *kay* 'this' and *chay* 'that'
also form a tight subsystem (to which it would be difficult to add new mem-

TABLE 12. *Coding of the structural coherence factors*[a]

Class	PERI	STRUCT	TRANS	PARAD	INF-D	INF-H
a	−	−	−	−	+	−
adv	+	−	−	−	−	−
aux	−	+	−	−	+	−
comp	+	+	−	−	−	−
conj	+	−	−	−	−	−
cop	−	+	−	−	+	+
dem	−	+	−	+	+	−
det	−	+	−	+	+	−
excl	+	−	−	−	−	−
n	−	−	−	−	−	−
name	−	−	−	−	−	−
neg	+	−	−	−	−	−
num	−	−	−	+	−	−
p	+	+	+	−	−	−
p + det	−	+	+	+	+	−
poss	−	−	−	+	+	−
pron	−	−	−	+	+	−
pron-cl	−	−	−	+	+	−
q	−	−	−	+	−	−
v	−	+	−/+	−	+	+
wh	−	−	−	+	−	−

[a]See Table 2 for description of word-class categories.

bers); second, the meaning of *este* is defined in opposition to that of *ese* (and of *aquel* 'yonder'), and thus the element is not quite independently transportable into another system. However, we were not able to separate these two factors.

Often, the different elements in the clause are marked on the verb, which may be morphologically quite complex for this reason. Borrowing will imply morphological integration as well, and this often is a hindering factor. A separate dimension, then, will be inflection: agreement (subject/object/verb and adjective/noun agreement) and case affixes (cf. Moravcsik, 1978; a critical survey of the literature on borrowability of inflection is given in Thomason & Kaufman, 1988). We predicted that uninflected elements would be easier to borrow than inflected ones. In addition to the inflection of the donor language, we assumed that the inflection of the host language plays a role. It may well be easier to incorporate elements into the lexicon that do not have to become integrated morphologically as well.

The structural factors to be used in the analysis are summarized in Table 12. The factors are: peripherality (PERI), structure-building (STRUCT), transitivity (TRANS), paradigmatic (PARAD), inflection donor (INF-D), and inflection host (INF-H). Moreover, Table 12 shows how the factors are coded for the different morphosyntactic classes. To test the transitivity factor, all verbs

TABLE 13. *Cross-tabulation of ±borrowing and the structural factors discussed*

Spanish Word Types	−Borrowing	+Borrowing	χ^2	Contingency Coefficient
−Peripherality	610	322	0.61	.016
+Peripherality	86	41		
−Structure-building	448	269	10.34*	.099
+Structure-building	248	94		
−Transitive	582	314	1.52	.038
+Transitive	114	49		
−Paradigmatic	621	354	22.49*	.146
+Paradigmatic	75	9		
−Inflection donor	333	237	29.21*	.164
+Inflection donor	363	126		
−Inflection host	486	281	6.87*	.081
+Inflection host	210	82		

*$p < .05$.

were coded for the property of being transitive or not. We decided to code nouns as not inflected for the language pair involved and to code adjectives as inflected only in the donor language. In Quechua, nouns and adjectives can be affixed with case markers or postpositions, and, in addition, they can carry person markers. There is also a plural marker, particularly used when nouns are high on the animacy hierarchy. There is no adjective–noun agreement. Indeed, many researchers take nouns and adjectives to form one single category in Quechua (cf. Lefebvre & Muysken, 1988). In Spanish, nouns can be marked for number, and they are either feminine or masculine— something that is visible in their -*a* or -*o* ending. The reason we have taken nouns to be uninflected is that they need no inflection (contrary to verbs) and are interpreted in the language contact situation involved as uninflected (again, contrary to verbs). Spanish nouns are either borrowed as singulars (*plata* 'money') or as invariant plurals (*aritis* 'earrings', *xabas* 'lima beans') (cf. Albó, 1968). The end vowels marking gender are treated as part of the stem. Marking nouns as inflected would have obfuscated the large difference with verbs, which are always inflected.

The results for the separate analyses on these factors are given in Table 13. The six factors were cross-tabulated with ±borrowing, and a chi-square was calculated to determine the strength of the dependency.[6] The figures, the chi-square values, and the contingency coefficients in Table 13 make clear that particularly paradigmaticity and inflection in the donor language have a fairly strong effect, taken by themselves. Structure-building and inflection in the recipient language also have an effect, but apparently less strong, given the resulting contingency coefficients. All significant effects have the expected direction: +structure-building, +paradigmatic, +inflection donor, and +inflection host turn out to be inhibiting factors for borrowability.

TABLE 14. *Cross-tabulation of ±borrowing and
the ±equivalence distinction[a]*

Spanish Word Types	−Borrowing	+Borrowing	Total
−Equivalence	49	18	67
+Equivalence	647	345	992
Total	696	363	1,059

[a]$\chi^2 = 1.73$, not significant.

Equivalence

The final factor we want to consider is equivalence between word classes. There is little precise information available about the universality of word classes (a classic study is Steele et al., 1981). Two points on which there is a moderate amount of consensus are: (1) there are verbs and nouns in all languages, and (2) there is no exact match between the minor categories in different languages. Well-documented problem areas in categorial equivalence include: full pronouns (West Germanic) versus clitics (Romance) (Kayne, 1975), auxiliaries (Modern English) versus main verbs (Old English, Dutch) (Lightfoot, 1979), predicate adjectives (Indo-European) versus stative verbs (Kwa, Caribbean Creoles) (Winford, 1993), clauses (Indo-European) versus nominalizations (e.g., Turkish/Quechua) (Lefebvre & Muysken, 1988), and cases (e.g., Turkish, Finnish) versus adpositions (e.g., English) (van Riemsdijk, 1978). All oppositions in this list correspond to notional equivalents with different categorial realizations. The list can be extended once we get into determiner systems. We ignore for the moment the issue of equivalence of morphosyntactic and morpholexical categories, such as case, conjugation class, gender, although these are clearly relevant as well.

It is clear that a number of Spanish categories are either part of a larger class (auxiliaries) or expressed by affixes (complementizers and conjunctions, determiners, prepositions, clitic and possessive pronouns) in Quechua. Thus, there is no equivalence for these categories. The results for borrowing are given in Table 14.

Equivalence may seem to have some minor effect on borrowing; the chi-square value (1.87, $df = 1$) and the contingency coefficient (.041), however, are not significant. This may perhaps be interpreted as implying that the lexical classes of a language are somewhat separate from the grammar of that same language.

The constraints model

Now that we have looked at the individual, separate contribution of factors of lexical content, frequency, structural dimensions of cohesion, and equiv-

TABLE 15. *Results of logistic regression on the constraints*

	χ^2	df	Significance
−2 log likelihood	1,255.287	1,054	.0000
Model χ^2	90.061	4	.0000
Goodness of fit	1,054.644	1,054	.4886

Variable	B	S.E.	Wald	df	Significance	R	Exp(B)
Paradigmatic	−2.8271	.5967	22.4508	1	.0000	−.1233	.0992
Inflection donor	−.8484	.1452	34.1315	1	.0000	−.1545	.4281
Peripherality	−.8217	.2291	13.5944	1	.0002	.0928	2.2743
Log frequency	.1126	.0577	3.8148	1	.0508	.0367	1.1192
Constant	−1.0535	.2147	24.0782	1	.0000		

alence, we may wonder how these factors combine to determine the borrowability of a given source-language item. We have studied this through the statistical technique of logistic regression, a technique that matches the approach in variable rule analysis (cf. Hosmer & Lemeshow, 1989; Rietveld & van Hout, 1993). In this analysis, the dependent variable is the property of being borrowed or not, which is a binomial variable (a 0 indicating −borrowing and a 1 indicating +borrowing). The question to be answered is: to what extent is a set of independent or criterion variables useful in predicting the chance of a lexical item in the source language being borrowed? The independent or criterion variables used are: ±content, log frequency (using the natural logarithm), ±transitivity, ±peripherality, ±structure-building, ±paradigmatic, ±inflection donor, ±inflection host, ±equivalence. These two-valued variables are coded as 0 (−) and 1 (+).

In the regression analysis, the criterion variables (our factors) were evaluated step by step as to their possible contribution to the explanation of the borrowing variable. Beginning with the variable with the strongest impact, the analysis then successively evaluates the remaining factors. The final outcomes of the logistic regression analysis are given in Table 15.[7]

The outcomes in the top half of Table 15 give information on a general level of the successfulness of the analysis. The significance of the model chi-square means that, given the variables entered in the analysis, the degree of explanatory force is acceptable—although the degree of explanation could have been much better, given the residual amount of chi-square under the heading of −2 log likelihood. The goodness of fit is acceptable as well, given its nonsignificance, which means that there is a reasonable match between the predicted probability values of being borrowed and the borrowings actually found.

The second part of Table 15 gives detailed information on the variables entered in the analysis. First, we can observe that only four variables were entered, which implies that the remaining variables do not improve the out-

TABLE 16. *Ordering of the relevant constraints and the morphosyntatic word classes from easy to difficult to borrow*

Paradigmatic	Inflection Donor	Peripherality	Word Class[a]
−	−	−	n name
−	−	+	adv comp conj excl neg p
−	+	−	a aux cop v
+	−	−	num q wh
+	+	−	dem det p + det poss pron pron-cl

[a]See Table 2 for description of word-class categories.

comes in a significant way. The values under B inform us about the strength of the factor. The significance of the individual variables is also evaluated. It turns out that the frequency factor (log frequency) has a probability value of just above .05. Should this variable be removed from the model? Given the fact that the frequency factor is measured very roughly because of the small size of the corpus, we can be satisfied with the result obtained.

The B values show that paradigmaticity is the strongest structural factor in our model. The second strongest structural factor is inflection in the donor language. Frequency also has a (somewhat weaker) effect, whereas peripherality has a clear effect, but opposite to what we predicted. Not included are the factors of content, structure-building, transitivity, inflection host, and equivalence.

The role of the factors can be made more understandable by ordering the morphosyntactic classes from more easily borrowed to more difficult. The order is based on the variables of paradigmaticity, inflection donor language, and peripherality, as well as the B values in the regression analysis. Table 16 contains the results. As this table shows, the results are fairly plausible. Most easily borrowed are nominal heads without structure-building properties. Functional heads occur lower in this hierarchy. The most difficult categories to borrow consist of functional elements that are nominal in nature and form tightly organized subsystems.

In addition, the frequency factor is operative in all categories distinguished. Frequency favors borrowing quite generally. The one puzzling result of our analysis is that peripherality has the opposite effect from the one predicted for the differentiation of the class {n, name} and the class {adv, comp, conj, excl, neg, p}. Apparently nominal nuclei of constituents can be quite easily borrowed in spite of not being at all peripheral.

CONCLUSION

We can conclude that the constraints model, operating on the basis of a comparison between a donor language and a recipient language corpus, seems to

be a promising way of studying the process of lexical borrowing. The results may be interpreted in such a way as to set up a new hierarchy of borrowability, which would simply result from classifying the individual categories in terms of their value for the factors inhibiting or furthering the borrowing process.

It should be stressed that the results we obtained for Spanish borrowings in Bolivian Quechua are not meant to be independent of this particular set of languages. In other language pairs, quite different factors may turn out to be operant, depending on sociolinguistic factors and different contrasting typological properties. The same holds for the particular factors chosen and the way they are applied to classify the borrowings. The ones we chose relate, in part, to particular properties of Spanish and Quechua and to our first impressions of which factors may have been steering the borrowing process.

NOTES

1.   A significant part of the types not classified for language consisted of onomatopoeic exclamatives. Their occurrence is not unexpected, given the type of language corpus.

2.   The language corpus used by Poplack et al. (1988), for instance, is much larger. We do not want to discuss here which criteria could be set for a large or a sufficiently large corpus. For global research on word frequencies the corpus should be fairly large. Even a corpus of one million words can be insufficient for reliable estimations of infrequent words (cf. Baayen, 1989).

3.   The number of Spanish hapaxes (170) is about the same as the number of Quechua hapaxes (185) in the Quechua corpus (see Table 4). This indicates that, in enlarging the sample (the corpus), new word types of Spanish origin (Spanish word types not occurring in the available corpus) would show up at about the same rate as new word types of Quechua origin.

4.   The problem with the type/token ratio is that the number of types is not related in a linear way to the number of tokens (cf. Rietveld & van Hout, 1993:171–174). Longer texts (with more tokens) will produce smaller type/token ratios (for the same conclusion in relation to language acquisition data, see van Hout & Vermeer, 1989). This effect can also be noticed in our bilingual data base. The type/token ratio for the Spanish data (1033/10867 = .095) is much lower than the type/token ratio for the Quechua data (908/6870 = .132).

This aspect of the type/token ratio as a measure for lexical richness has nothing to do with the validity of calculating mean frequencies of word types (which is in fact the token/type ratio). We use the mean frequency in this article as an indicator of differences in the frequency pattern of Quechua and Spanish words.

5.   When the paradigmatic organization in two typologically related languages is similar, paradigmatic coherence is not an inhibiting factor, as illustrated by the borrowing of pronouns in Old English from Scandinavian.

6.   The total of 1,059 words relates to the words that could be classified without any problem for word class (see category ? in Table 6 for the Spanish corpus). In comparison with the figures in Tables 9–11, one Spanish word type has been left out.

7.   The statistical package used is SPSS. The procedure applied is logistic regression, with forward stepwise entry. Applying a backward stepwise approach results in the same model as the one given in Table 15.

REFERENCES

Aguiló, S. I., F. (1980). ¿Los cuentos: Tradiciones o vivencias? Cochabamba: Editorial Amigos del Libro.

Albó, X. (1968). Social constraints on Cochambamba Quechua. Doctoral dissertation, Cornell University.

Baayen, R. H. (1989). *A corpus-based approach to morphological productivity. Statistical analysis and interpretation.* Doctoral dissertation, University of Amsterdam.

Broeder, P., Extra, G., & van Hout, R. (1993). Richness and variety in the developing lexicon. In C. Perdue (ed.), *Adult language acquisition: Cross-linguistic perspectives: Volume I. Field methods.* Cambridge: Cambridge University Press. 145–163.

Haugen, E. (1950). The analysis of linguistic borrowing. *Language* 26:210–231.

Heath, J. (1989). *From code-switching to borrowing: Foreign and diglossic mixing in Moroccan Arabic.* London: Kegan Paul.

Hosmer, D., & Lemeshow, S. (1989). *Applied logistic regression.* New York: Wiley.

Kayne, R. (1975). *French syntax.* Cambridge, MA: MIT Press.

Lefebvre, C., & Muysken, P. (1988). *Mixed categories: Nominalizations in Quechua.* Dordrecht: Kluwer Academic.

Lightfoot, D. (1979). *Principles of diachronic syntax.* Cambridge: Cambridge University Press.

Moravcsik, E. (1978). Language contact. In J. H. Greenberg (ed.), *Universals of human language.* Stanford: Stanford University Press. 93–122.

Muysken, P. (1981). Quechua en Spaans in het Andesgebied. *TTT* 1:124–138.

Payne, D. L. (1986). Basic constituent order in Yagua causes: Implications for word order universals. In D. C. Derkyshire & G. K. Pullum (eds.), *Handbook of Amazonian languages: Volume 1.* Berlin: Mouton de Gruyter. 440–465.

Poplack, S. (1980). Sometimes I'll start a sentence in English Y TERMINO EN ESPAÑOL. *Linguistics* 18:581–618.

Poplack, S., Sankoff, D., & Miller, C. (1988). The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics* 26:47–104.

Rietveld, T., & van Hout, R. (1993). *Statistical techniques for the study of language and language behaviour.* Berlin: Mouton de Gruyter.

Sankoff, D., & Labov, W. (1979). On the uses of variable rules. *Language in Society* 8:189–222.

Sankoff, D., Poplack, S., & Vanniarajan, S. (1990). The case of the nonce loan in Tamil. *Language Variation and Change* 2:71–101.

Singh, R. (1981). Aspects of language borrowing: English loans in Hindi. In P. H. Nelde (ed.), *Sprachkontakt und Sprachkonflikt.* Wiesbaden: Steiner. 113–116.

Steele, S., with Akmajian, A., Demers, R., Jelinek, E., Kitagama, Ch., Oehrle, R., & Wasow, T. (1981). *An encyclopedia of* AUX. *A study of cross-linguistic equivalence.* Cambridge: Cambridge University Press.

Taber, C. R. (1979). French loanwords in Sango: The motivation of lexical borrowing. In I. Hancock (ed.), *Readings in creole studies.* Gent: Storia Scientia. 121–126.

Thomason, S., & Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics.* Berkeley: University of California Press.

Treffers, J. (1991). *French–Dutch language mixture in Brussels.* Doctoral dissertation, University of Amsterdam.

van Hout, R., & Vermeer, A. (1989). Spontane taaldata en het meten van lexicale rijkdom in tweede-taalverwerving. *Toegepaste Taalwetenschap in Artikelen* 30:67–80.

van Riemsdijk, H. (1978). *A case study in markedness.* Dordrecht: Foris.

Weinreich, U. (1953). *Languages in contact: Findings and problems.* The Hague: Mouton.

Weinreich, U., Herzog, M., & Labov, W. (1968). Empirical foundations for a theory of language change. In W. P. Lehmann & Y. Malkiel (eds.), *Directions for historical linguistics.* Austin: University of Texas Press. 95–188.

Whitney, W. D. (1881). On mixture in language. *Transactions of the American Philosophical Association* 12:1–26.

Winford, D. (1993). *Predication in Caribbean English creoles.* Amsterdam: Benjamins.