Borrowing: the synchrony of integration*

SHANA POPLACK and DAVID SANKOFF

Abstract

The notion of loanword assimilation is operationalized in a number of different ways, focusing on both linguistic and social aspects. The indices of integration thus constructed are applied to a set of lexical data elicited from Puerto Rican children and adults from East Harlem, New York. The results of this survey are analyzed statistically using the method of principal components. We interpret the output in terms of the social and linguistic trajectory of words during the borrowing and integration process. Of particular importance are the relatively close relationship between increase in usage frequencies and the processes of phonological integration, the transient nature of inconsistencies in gender assignment, and the fates of competing lexical items for a single referent.

The lexical stock of languages may contain a considerable proportion of words borrowed from one or more other languages. The historical record, together with methods of historical and comparative linguistics, can help us infer which words were borrowed, from what language, and approximately when. On the synchronic level, however, making such inferences can be more difficult, particularly because there is no unequivocal way of deciding when a lexical item from one language, used during discourse in another language, whether by a single speaker, or repeatedly in a community, should be considered a loanword. It may constitute all or part of a code-switch, which is a phenomenon quite distinct from borrowing. It may be a manifestation of incomplete acquisition of one of a bilingual's two languages. It might be a momentary lapse of the type often classified as 'interference'. Or we might want to characterize it as still another kind of result of language contact.

It has been claimed that 'from synchronic examination [i.e. without comparative or etymological evidence] no loans are discoverable or describable' (Fries and Pike 1949; see also Haugen 1950a; Weinreich

Linguistics 22 (1984), 99–135 0024–3949/84/0022–0099 $2.00
© Mouton Publishers

1953), presumably because they are so perfectly assimilated to recipient-language patterns. We dispute this all-or-nothing viewpoint of the borrowing process, focusing specifically on the mechanisms by which an item is gradually converted from a foreign element to a nativized one. It is during this transition that it is difficult to recognize and distinguish loanwords. Though there is a large literature on the topic of borrowing, there remain many unanswered questions about the linguistic, sociological, acquisitional, and other aspects of borrowing, as well as the subtle methodological dilemmas involved in detecting, defining, identifying, and characterizing loanwords.

In this paper we first review the pertinent literature, focusing on three aspects of the borrowing process: the linguistic mechanisms involved, the social dynamics, and the definitional and analytic problems in studying lexical transfer. We will abstract from this work a number of concepts and hypotheses which have been developed to describe and account for the importation of forms from one language to another. We will proceed to operationalize these concepts on the basis of a corpus of potential loanwords collected from a group of bilingual Puerto Rican children and their parents. We then analyze these data statistically to evaluate the dimensionality of the borrowing process — the degree to which processes of phonological, syntactic, lexical, and sociological integration of a foreign element occur in unison or proceed independently. We pay special attention to the intergenerational transmission of borrowed material, and in particular to the roles of older and younger, monolingual and bilingual speakers in its propagation and eventual phonological shape.

Traditional as well as more recent studies of borrowing stressed that adaptations of foreign items of any linguistic level to the pattern of the recipient language, such as incorporation of verbal and nominal suffixes, assignment of gender, etc., were indications that the forms had been integrated into that language (e.g. Haugen 1950a: 396, 440; also implicit in Bloomfield 1933: 450, 453; see also Hyman 1970; Lovins 1974). Fries and Pike incorporate phonological, grammatical, and social criteria in their assumption (1949: 39) that a loan sequence of phonemes can be considered completely assimilated when (a) it parallels the sequences occurring in native materials, or is analogous to them; when (b) its occurrence in relation to grammatical boundaries is the same as sequences in native words; and when (c) the words containing it are in common use by monolinguals; or a loan sequence may be considered completely assimilated when it serves as a pattern for the development of new sequences in the native language.

This state of complete assimilation, however, obviously does not come about instantaneously. And during the period, months, years, or genera-

tions, over which it is attained, little is known about which of these criteria is satisfied first, or last, or about the intermediate stages in this attainment.

The linguistic integration of loanwords is but one aspect of their assimilation into a language. The sociological process of acceptation is another. Studies of languages in contact have focused on two basically sociological or sociolinguistic distinctions as being relevant to the incorporation of borrowed material into the linguistic repertoire of the community. One is the differential role, as carriers of innovations, of monolinguals versus bilinguals of varying degrees of competence; the other involves the changing shape of loanwords across successive generations of speakers.

A common view is that bilinguals tend to use loanwords before monolinguals, who learn them from the former. Bilingual speakers are also thought to assimilate new sounds sooner than their monolingual counterparts (Fries and Pike 1949: 39). Haugen distinguishes between monolinguals and those who became bilingual as adults on the one hand and childhood bilinguals on the other. Observing that borrowed items tend to retain an uncertain linguistic status for some time after their first adoption (1956: 55), he attributes part of this vacillation to the awareness on the part of the bilingual of the origin of the borrowed word, and presumably to his indecision as to whether to produce it according to recipient or donor language rules. In addition, both monolinguals and nonfluent ('adult') bilinguals make phonic adaptations (or 'distortions') of loanwords, while fluent ('childhood') bilinguals reproduce the patterns of the donor language (Haugen 1950a, 1956). These individual differences in ability and desire are presumably responsible for the alternative forms often shown by loans reported by Haugen. He posits three stages of loanword integration (e.g. 1969: 394) ranging from 'pre-bilingual', when forms are reproduced according to recipient language patterns with great irregularity in the results, to 'adult bilingualism', at which point the loans are produced more systematically, through 'childhood bilingualism', during which sound types (and presumably other patterns) from the donor language are introduced into the recipient language. Indeed, childhood bilinguals or younger speakers (who seem to be thought of as indistinguishable), according to this schema, are the speakers responsible for introducing NEW patterns into the recipient language.

Whichever speakers first make use of loanwords, and whatever the mechanism of spread within the lexicon through time and throughout the community, an important diagnostic for the incorporation of a form into the native lexicon is the increased frequency of its usage. Even the degree to which the loanword is linguistically integrated has been attributed to

the frequency of its use within the community (Kreidler 1979: 143). Holden (1976: 131) has suggested that 'most loanwords which show even a minimal degree of acceptance by the target language immediately assume a phonetic shape which is identical to that of the native vocabulary'.

Independent of these hypotheses correlating frequency of use with morphophonemic adaptations, frequency of use in itself has been considered a criterion of integration. Hasselmo suggests, for Swedish-English bilinguals, that 'some of the English discourse units introduced into Swedish discourse are used with such regularity that it may be necessary to regard them as in some sense *integrated* with an American Swedish mode of speaking' (1970: 179). More specifically 'if certain instances of interference are repeated often enough in discourse in a certain language to be regarded as habitualized, the forms and/or patterns involved can be referred to as (socially) integrated with the language of the community' (1970: 179; see also Mackey 1970). But as Mackey has observed (1970: 204), there is no necessary indication in the (single) occurrence of a given element in discourse of whether it represents a case of interference or a case of integration.

1.0. Analytical distinctions

At first glance it might seem to be an easy matter to detect material borrowed from one language in the discourse of the other, merely by comparison with the standard variety of the recipient language, or with communities not in contact with the donor language. This type of external comparison is, however, quite inadequate. The co-occurrence of forms from two languages may be due to a number of processes other than borrowing, the most important of which are code-switching and incomplete second-language acquisition by native speakers of another. Code-switching is simply the alternate use of the two languages in discourse, and even in a sentence, without any necessary influence of one language on those stretches of discourse realized in the other. Partial acquisition of a second language may lead to the use of first-language items in intended second-language discourse, but on an idiosyncratic basis. They may be considered borrowings on the individual-speaker level, but not on the level of the community speech variety. As Haugen (1969: 371) has pointed out, innovations made by language learners do not spread to native speakers of the language they learn; it is the innovations they make in their own language which spread.

Linguists have also tried to define borrowing disjunctively — by virtue

of being neither code-switching nor momentary or idiosyncratic uses of first-language terms to fill gaps in intended second-language competence. The latter behavior, along with others, has generally been subsumed under the label popularized by Weinreich (1953), 'interference', and occupies some nebulous area in between the other two extremes. In Haugen's (1956) schema these phenomena are located along a continuum of code distinctiveness, with switching representing maximal distinctness, integration representing maximal levelling of distinctions, and interference referring to an overlapping of two codes, contrary to contemporary norms. In determining whether adaptation had occurred or not (i.e. whether integration or switching was involved), Haugen suggested that the phonological and morphological shape of the borrowed form were the determining factors. However, it is rather the bilingual ability of the speaker which determines the pronunciation of the second language, so that this criterion will misidentify code-switches as loanwords and vice versa. Shaffer's (1978) claim that integration is more accurately measured by syntactic considerations also fails to unambiguously distinguish loanwords from (one-word) code-switches.

Indeed, as Hasselmo observes, although the intention of the speaker may be a binary choice between switching and integration, the stretches of speech actually produced are often ambiguous. Since code-switching is not identifiable on the basis of linguistic — phonological, morphological, or syntactic — features alone, the occurrence of a borrowed item that shows a high degree of social integration (i.e. acceptance and use by community members) could be interpreted as an instance of a loanword, while one that shows a low degree of social integration would be an instance of code-switching (1970: 180). In a similar vein, Mackey (1970: 211) suggests that the more an item is integrated (used), the less likely that its appearance is a case of interference.

2.0. An operational framework

We may abstract from the previous considerations four basic types of criteria for the characterization of loanwords:

1. Frequency of use. By this measure, as used by Fries and Pike, Hasselmo, Mackey, and Murphy, the more frequently a specific donor-language item is used in recipient-language discourse and by more people, the more reasonable it is to consider it as having become a bona-fide term of the recipient language.

2. Native-language synonym displacement, as measured by the translatability test used by Hasselmo and Mackey's availability test, and implicit

in Weinreich's discussion of lexical integration. If a borrowed term can be shown to displace in usage an indigenous term for the same concept, it can be considered to have taken over the latter's role in the lexicon.

3. Morphophonemic and/or syntactic integration. If a borrowed term takes on a phonological shape typical to the recipient language, acquires the morphological affixes appropriate to that language, and functions in sentences as a native word of some syntactic category, then it can be considered a well-established borrowing. This approach is embodied in the work of Fries and Pike, Bloomfield, Weinreich, Haugen, and others.

4. Acceptability. If native speakers judge a donor-language word to be an appropriate designation whether or not they are aware of its etymological origins, this is indicative that it may occupy a place in the recipient lexicon.

Not all of these criteria, however, will be satisfied in all cases which we may want to consider loanwords, and each of them may be satisfied by words which are not. For example, a word from one language may be used frequently in discourse which is predominantly of another language, but only because it occurs often in code-switches (e.g. the determiner *the* occurs frequently in switched NPs). A borrowed word may be phonologically, morphologically, and syntactically integrated into the recipient language but only because the speaker has little productive competence in the donor language or simply because of interlingual coincidence between donor and recipient codes. Acceptability is notoriously misleading, especially in contexts where the recipient language is socially inferior to the donor. Even in cases where neither language is stigmatized, Hasselmo documents for Swedish-English bilingualism cases where items were identified as being of English origin, yet showed low translatability, but high acceptability (1969: 71), results which are difficult to interpret in terms of integration into the linguistic repertoire, short of arbitrarily assigning supremacy to one of the criteria involved. Similar difficulties were encountered in a later replication of this study among Chicano bilinguals by Murphy (1974), leading him to suggest that not only these tasks but also the very attempt to get bilinguals to establish language boundaries are inappropriate (1974: 63–64). Synonym displacement may be a solid criterion, but only when a single borrowed word is displacing a single well-identified native word, a situation which is not necessarily the rule and which is very difficult to demonstrate, since the precise referent of the word may be impossible to reconstruct at the time of the analysis.

Most of these criteria are based on anecdotal evidence, albeit on the part of observant and highly insightful scholars, and remain empirically unsubstantiated. For example, the proposed correlation between frequency of use and degree of linguistic integration, or between degree of

acceptability and linguistic integration, quantitative hypotheses both, have never been quantitatively tested. The role of bilingual versus monolingual, or older versus younger speakers, in introducing and propagating loanwords has never been empirically investigated or established.[1] Parenthetical comments on the vacillations which mark the incorporation of transferred words into the vocabulary cannot capture the orderly heterogeneity which has been documented for all aspects of speech behavior at the community level. Conversely, the empirical studies which do exist, largely on acceptability and availability, do not take into account the more strictly linguistic concomitants of loanword assimilation.

Nevertheless, the four types of criteria discussed above are useful in that they are abstracted from the key processes which make up the phenomenon of lexical borrowing. It is in fact reasonable to assume that as a borrowed word is more and more used, it tends to become phonologically and morphologically integrated, to displace competing recipient language forms,[2] and at least eventually, to be accepted by its native speakers.

3.0. Data and methods

3.1. *The contact situation*

The data on which this paper is based were collected from 14 children and eight adult residents of a stable bilingual (Spanish-English) community in East Harlem, New York, one of the oldest continuous Puerto Rican communities in the United States. Though English has coexisted with Puerto Rican Spanish to some degree since the American occupation of the island in 1898, the most intense contact has occurred since the end of World War II with a massive population influx to the United States, concentrated almost wholly in the geographically circumscribed area in which this community is located. Despite widespread unfavorable attitudes of the non-Hispanic population at large toward Puerto Rican Spanish language and culture, and generally strong pressure to assimilate to the ways of the mainstream, there is as yet no conclusive evidence in this community of language shift, even among third-generation bilinguals, partially due to a pattern of circulatory movement between Puerto Rico and the United States and to a more recent influx of Spanish speakers from other parts of the Caribbean, both of which have the effect of replenishing the stock of Spanish-dominant or monolingual speakers and hence revitalizing the language. Puerto Ricans in New York may be

said, then, to be undergoing intense, though short-term 'cultural pressure' from source-language speakers (Thomason 1981: 14), a situation which generally results in copious borrowing of (at least) donor-language lexical items if not also foreign structure.[3] Though several aspects of the grammatical structure of local Spanish have in fact been shown to be free of influence from English (Poplack 1981), incorporations from that language, in the form of both code-switching and borrowing, are numerous enough to be remarked upon and stereotyped by Puerto Ricans and non-Hispanics alike.

3.2. *The informants*

The parents and children who participated in this study are members of a larger group which was the object of a long-term interdisciplinary study investigating language maintenance and language shift in a stable bilingual community, and the effects on the Spanish language of prolonged contact with English.[4] All but one of the adults are Spanish-dominant, while more than half of their children may be characterized as balanced bilinguals, in that they report and are observed using Spanish as frequently as English in a wide variety of settings and domains. An additional four children prefer English in all situations.[5] The sample was constructed to include parents and their children in order to investigate intergenerational transmission of bilingual skills in the context of stable bilingualism.

3.3. *The data*

It would seem most natural, in studying the usage properties of loanwords, to work with corpora of spontaneous bilingual discourse. As we shall see, however, this was not feasible.

It is obvious that in the speech of bilinguals, items may be borrowed from another language momentarily and never be heard again, or they may be used with great regularity. The position of a borrowed element within a host language system, as Mackey rightly points out (1970: 201), is a matter of degree: it may be completely integrated, partially integrated, or not integrated at all. In searching for a method to measure degree of integration, we first examined over 65 hours of speech recorded in a variety of naturalistic and more formal speech situations. We quickly found that even loanwords we had thought very common occurred quite infrequently across all speakers. This is because the occurrence of any lexical item depends largely on what the speaker is talking about, and

even in a large corpus of free undirected speech, there is never any guarantee that any two speakers will be talking about the same thing. Indeed, in the nearly 200 hours of tape-recorded speech of the children and adults who form the sample on which this study is based, less than half of the English words in Table 1 occurred in a Spanish context at all, only ten of them were uttered by more than one speaker, and only two by more than two speakers: *tape* and *building*.

An additional difficulty is that to measure frequency or degree it is necessary to know not only the number of times an item occurred but also all the times it did not occur when it might have. With free speech data this would involve identifying all semantic equivalents of every borrowed word in a corpus far larger than this one, and ascertaining the proportion of those uttered in English and those in Spanish, a formidable undertaking well beyond the scope of the present work.

Accordingly, we opted to elicit designations of concepts, or referents, while holding the stimulus constant. An elicitation instrument was constructed containing a randomized series of photographs of 45 everyday items which we expected to be easily identifiable and able to be designated by a concrete noun. The majority, though not all, of these photographs referred to items we had heard designated in English in otherwise Spanish contexts, either in our corpus of taped data or in our fieldwork experience. The actual frequency of mention, however, depended on the word.

This instrument, part of a battery of language-related tasks, was administered entirely in Spanish. Each informant was shown each picture and asked to name the object in Spanish: *Qué es esto?* 'What's this?' Respondents were encouraged to answer in a complete sentence (e.g. *es un rufo* 'It's a roof') to enable us to examine gender indications as well. After responding, informants were prompted as to whether they knew any additional words for the concept, in order to correct for memory limitations. These were noted as second and third choices.

The data elicited this way are similar to those obtained from the availability tests used by Hasselmo and Mackey, with the differences that the particular stimulus here was held constant instead of just the general semantic domain, thus obviating the problem of semantic equivalence described above, and that they yield much more information on both social and linguistic integration of English items.[6] Despite their relatively artificial nature, it should be noted that by the time the data were collected, close rapport had been established between interviewer and informants, thanks to the familiarity acquired in several years of prior fieldwork in the same neighborhood.

Though responses such as these can provide only indirect evidence for

spontaneous speech phenomena, they can nevertheless be highly indicative of the borrowing processes which interest us. First, we can see which, if any, English word is most frequently elicited for a concept. A word's frequency should be roughly related to how likely it is to be used for that concept in discourse, although this relationship may not be simple and could be mediated by other factors, such as acceptability. Since it may be difficult to elicit any very specific lexical item for certain concepts, it may be more appropriate to take into account the frequency with which ANY English word is elicited.

A different way of assessing the importance of English versus Spanish words for a concept is to see which comes to mind first, i.e. which is offered first, if several responses are forthcoming. The above measures not only pertain to the relative frequency of English versus Spanish terms for a concept, but they also indirectly indicate to what extent a term borrowed from English has displaced a Spanish synonym. A more direct way of measuring this is to count the number of speakers offering only English responses for a concept versus those who give both an English and a Spanish response.

Turning to the question of linguistic integration, careful examination of the elicited words can indicate degrees of phonological and morphological integration and, if there is a gender carrier, some indication of syntactic integration. It is known that pronunciation of a word in response to a formal elicitation instrument tends to depart from the vernacular, so some effort should be made not only to average the degree of phonological integration over speakers, but also to take account of the most integrated form of the word. With respect to gender, we would expect a frequently used word to show consistent gender assignment across speakers. However, it could be that the very choice of whether or not to assign a gender to an originally English word is a reflection of its status as an integrated loanword, as has been suggested by Barkin (1980), and this too can be measured.

3.4. *Indices of integration*

Our data base, then, consisted of between one and three designations of the 45 concepts for each of the 22 speakers in the sample. Following on the considerations at the end of section 2, we defined the following three types of indices:

1. Social measures of integration: for each test item these indicate what proportion of the sample of speakers gives responses manifesting the incorporation of English elements into their Spanish lexicon.

2. Linguistic measures of integration: these indicate the average or maximal extent to which forms of English provenance were rendered according to Spanish phonological, morphological, or syntactic (as manifested by gender assignment) patterns.

3. Test controls: these indices are not intended as measures of loanword acceptance, integration, or assimilation into Spanish. They measure other parameters of the set of responses which, for technical reasons, could interfere with or exaggerate the social and linguistic measures in the previous two categories.

We will define the indices in the last set first, although they are intrinsically the least interesting, because some of them enter as components into the definitions of the indices in the other two sets.

Test controls.

Total (tokens):[7] the total number of words, counting all repetitions by different speakers and each token in multiple responses separately.

Total (no response).

Total (failures): including unidentifiable responses (e.g. nonexistent *fluche* for 'switch') and failure to elicit desired semantic field (e.g. *New Jersey* for 'building').

Total (types): the total number of different forms elicited; repetitions of the same form by several speakers were always counted as one type, even if they involved somewhat different phonology (e.g. *suera*, *suéter* 'sweater') or different gender.

Total (multiresponse speakers): the number of speakers who offered at least two different designations of a particular referent.[8]

Social measures. The indices in this group are operationalizations of some of the ways discussed earlier of measuring the frequency of use and/or acceptability of borrowed designations for a specific concept, and for assessing the degree of synonym displacement.

Proportion (English tokens): the proportion of the total number of tokens identified as of English origin offered for a given concept.

Proportion (English types): the proportion of the types which are identified as of English origin.

Proportion (English first choice): the proportion, over all speakers offering at least one response, of those giving an English response first.

These are different ways of evaluating the importance of English words generally for the concept, in contrast to Spanish words.

Proportion (most common English type): the proportion of all the responses which consist of the most frequent English type.

Proportion (most common English type)/Proportion (English tokens): the proportion that the most common English type represents of all ENGLISH responses.

Proportion (English first choice)/Proportion (English tokens): the proportion of English tokens which are first choices.

These measures all indicate the extent to which there is a SINGLE English word which is the preferred designation of the concept.

Proportion (English-only responses): the proportion of speakers giving one or more responses which are all of English origin.

Proportion (bilingual responses): the proportion of speakers offering at least one English and one Spanish response.

Proportion (non-Spanish-only): the sum of the preceding two indices.

Contrasts between proportion of English-only, bilingual, and non-Spanish responses are directly indicative of the degree of synonym displacement and are also sensitive to overall rates of English and Spanish usage.

Linguistic measures. The third group of indices measures the phonological, morphological, and syntactic integration of English-origin words into recipient language patterns.

Mean (token integration): the average value, over all English tokens, of the phonological integration code for the token:
0 = unintegrated (i.e. rendered wholly via donor-language phonology, e.g. [ays Kriym] 'ice cream'); 1 = partially integrated (i.e. rendered with both donor and recipient language phonology, e.g. [aih Kriym]); 2 = completely integrated into recipient language patterns, including manifestations of recipient-language variability, e.g. [aih krĩŋ]).

Mean (type integration): the average value, over all English types, of the average phonological integration for the tokens of each type. The mean integration is first calculated for each type separately, and then the average is taken of these type values.

Maximum (integration): the maximum integration of any English token.

These indices represent different ways of summarizing the phonological integration of the responses.

Gender consistency: for the most frequent English type, the (absolute) difference between the number of tokens coded as feminine and those coded as masculine, divided by the total number of tokens coded for gender.

Use-weighted consistency: the absolute difference between the numbers

of feminine and masculine tokens divided by the total number of tokens, whether or not assigned gender, for the most common English type (0 = gender assignment is totally inconsistent, 1 = fully consistent).

Field consistency: over ALL English types coded for gender, the difference between the numbers of feminine and masculine tokens, divided by their sum.

Gender consistency is a measure of whether the most frequent loanword is assigned the same gender consistently. Use-weighted consistency measures both whether it is assigned gender regularly and whether the gender is consistent. Field consistency measures to what extent gender is consistent across all the loanwords elicited for the concept, i.e. whether gender is assigned by analogy with the gender of a recipient-language designation of the concept.

4.0. Results

4.1. *Preliminary tabulations*

Our method for choosing test items ensured only the possibility of eliciting some borrowed forms; but we had no way of knowing in advance whether they would be significant in number, or even whether the indices we defined would differentiate the items in any interesting way. In fact, in calculating the indices for the responses to each of the 45 stimuli, we discovered not only that most of the stimuli elicited a good proportion of English-origin forms, but that these manifested a great deal of variability in their assimilation into the recipient-language lexicon. This indicates the diversity inherent in the integration process.

Of the 45, the average number of blanks and anomalous responses was less than 1.5 each per stimulus. Seven of the concepts were designated in Spanish only by all respondents and so were excluded from further calculations ('boy', 'bicycle', 'plate', 'spoon', 'window', 'girl').[9] The remaining 38 items were designated in English and Spanish to varying degrees. The results may be seen in Table 1.

Table 1 shows that most of these stimuli (63%) were designated in English anywhere from half to all of the time (Proportion [English tokens]). A number of concepts may be seen from these calculations not to have any commonly used Spanish designation in this community at all (*tape*: 100% English designations, *hamburger*: 96%, *zipper*: 95%, *jacket*: 83%, *hot dog*: 82%, *truck*: 80%, *basement*: 76%, and *sweater*: 75%).

For most concepts, very few speakers offered BOTH Spanish and English

Table 1. *Indices of integration for 38 referents*

	Proportion (English tokens)	Proportion (English types)	Proportion (bilingual responses)	Proportion (English-only responses)	Proportion (non-Spanish responses)	Mean (type integration)	Mean (token integration)	Maximum (token integration)
'switch'	0.57	0.33	0.05	0.55	0.61	1.10	1.10	2.00
'frog'	0.33	0.39	0.15	0.15	0.31	0	0	0
'ice cream'	0.67	0.66	0.18	0.54	0.72	0.55	1.00	2.00
'turtle'	0.21	0.66	0	0.21	0.21	1.30	1.00	2.00
'pig'	0.08	0.19	0.04	0.04	0.09	0	0	0
'coat'	0.69	0.50	0.31	0.54	0.86	0.88	1.30	2.00
'puppy'	0.16	0.50	0.14	0.04	0.19	0.50	0.50	2.00
'hot dog'	0.82	0.75	0.13	0.77	0.90	0.63	1.50	2.00
'shade'	0.73	0.59	0.11	0.70	0.82	0.17	0.42	2.00
'sandwich'	0.67	0.50	0.42	0.52	0.95	0.58	1.10	2.00
'sofa'	0.12	0.39	0.19	0	0.19	2.00	2.00	2.00
'jeans'	0.41	0.62	0.44	0.19	0.64	0.80	0.30	2.00
'basement'	0.76	0.33	0.19	0.69	0.89	1.80	1.80	2.00
'tape'	1.00	1.00	0	1.00	1.00	1.10	1.10	2.00
'watch'	0.04	0.50	0.05	0	0.05	0	0	0
'suit'	0.67	0.33	0.14	0.59	0.75	1.00	1.10	2.00
'sweater'	0.75	0.25	0.23	0.66	0.90	1.90	1.90	2.00
'building'	0.62	0.33	0.38	0.47	0.85	1.40	1.40	2.00
'baby'	0.59	0.16	0.18	0.59	0.77	0.70	0.66	2.00
'zipper'	0.95	0.50	0	0.95	0.95	1.30	1.30	2.00
'cake'	0.29	0.33	0.19	0.14	0.34	1.10	1.10	2.00
'tape recorder'	0.57	0.66	0.04	0.54	0.59	0	0	0
'butterfly'	0.28	0.50	0	0.28	0.28	0.20	0.16	1.00
'truck'	0.80	0.66	0.14	0.75	0.89	1.60	1.50	2.00
'lipstick'	0.66	0.28	0.09	0.66	0.76	1.10	0.50	2.00
'hamburger'	0.96	0.75	0.04	0.95	1.00	0.67	0.75	2.00
'mattress'	0.66	0.59	0.28	0.57	0.85	1.30	1.80	2.00
'toilet'	0.74	0.19	0.23	0.71	0.95	1.60	1.60	2.00
'dime'	0.54	0.19	0.05	0.55	0.61	1.60	1.60	2.00
'bat'	0.09	0.50	0	0.09	0.09	0	0	0
'garbage can'	0.11	0.33	0.04	0.09	0.14	1.00	1.30	2.00
'swimming pool'	0.59	0.50	0.09	0.54	0.63	0.40	0.39	2.00
'roof'	0.70	0.50	0.19	0.66	0.85	0.90	1.70	2.00
'jacket'	0.83	0.59	0.09	0.77	0.86	1.50	1.20	2.00
'penny'	0.66	0.33	0.19	0.59	0.79	1.40	1.40	2.00
'nickel'	0.40	0.28	0.10	0.36	0.47	1.90	1.80	2.00
'quarter'	0.30	0.50	0.04	0.23	0.28	1.00	1.70	2.00
'bobby pin'	0.15	0.33	0.09	0.09	0.18	0.50	0.50	2.00

Proportion (+frequent English type) / Proportion (English tokens)	Proportion (+frequent English type)	Proportion (English 1st choice)	Proportion (English 1st choice) / Proportion (English tokens)	Gender consistency	Use-weighted consistency	Field consistency	Total (tokens)	Total (multi-response speakers)	Total types	Total (no response)	Total (failures)
1.00	0.57	0.55	0.90	1.00	0.82	1.00	19	1	3	2	2
0.75	0.25	0.26	0.62	1.00	0.50	1.00	24	5	5	2	1
0.95	0.64	0.72	0.76	1.00	0.10	1.00	31	9	3	0	0
0.75	0.15	0.21	1.00	0	0	0.33	19	0	3	3	0
1.00	0.08	0.04	0.50	1.00	0.50	1.00	24	3	5	1	0
0.73	0.51	0.72	0.69	1.00	0.88	1.00	33	11	4	0	0
1.00	0.16	0.09	0.50	1.00	0.75	1.00	24	3	2	1	0
0.79	0.65	0.90	0.79	1.00	0.53	0.66	29	6	4	0	1
0.85	0.63	0.76	0.92	0.33	0.17	0.66	19	2	5	5	0
0.95	0.64	0.61	0.61	1.00	0.80	1.00	31	10	4	1	2
0.75	0.09	0.04	0.25	1.00	1.00	1.00	32	12	5	1	1
0.69	0.29	0.50	0.76	1.00	0.44	1.00	31	10	8	2	0
1.00	0.76	0.69	0.69	1.00	0.80	1.00	26	7	3	2	1
0.76	0.76	1.00	0.85	0.90	0.60	0.87	21	3	2	0	4
1.00	0.04	0	0	1.00	1.00	1.00	20	1	2	3	0
0.88	0.59	0.73	0.82	1.00	1.00	1.00	25	6	6	2	3
1.00	0.75	0.66	0.63	0.37	0.32	0.36	29	9	4	1	1
1.00	0.62	0.69	0.77	1.00	0.72	1.00	29	11	3	1	4
1.00	0.59	0.72	0.88	1.00	0.72	1.00	30	6	6	0	0
1.00	0.95	0.95	0.91	0.88	0.74	0.89	24	4	2	0	2
1.00	0.29	0.36	1.00	1.00	1.00	1.00	24	5	3	2	1
0.91	0.52	0.59	1.00	1.00	0.82	1.00	21	3	3	2	2
1.00	0.28	0.28	1.00	0.50	0.33	0.50	21	0	2	1	0
0.90	0.73	0.84	0.80	1.00	0.84	1.00	26	8	3	1	4
0.87	0.58	0.66	0.87	1.00	0.64	1.00	24	3	7	0	1
0.68	0.66	0.94	0.65	0.76	0.65	0.84	30	9	4	1	1
0.88	0.59	0.85	1.00	0.86	0.81	0.87	27	9	5	1	4
1.00	0.74	0.80	0.84	1.00	0.90	1.00	27	5	5	1	0
1.00	0.54	0.55	0.90	1.00	0.91	1.00	20	7	5	1	8
1.00	0.09	0.09	1.00	1.00	1.00	1.00	22	1	2	0	1
0.66	0.07	0.09	0.66	1.00	1.00	1.00	26	7	6	2	1
1.00	0.59	0.59	0.79	0.83	0.67	0.83	25	4	2	0	2
0.94	0.66	0.68	0.68	1.00	0.89	1.00	27	8	4	1	4
0.67	0.56	0.77	0.67	1.00	0.82	1.00	30	8	5	0	0
1.00	0.66	0.75	0.93	0.75	0.75	0.75	24	4	3	1	1
0.88	0.36	0.44	0.88	1.00	1.00	1.00	22	3	7	3	1
0.85	0.26	0.29	0.85	1.00	1.00	1.00	23	2	4	1	1
0.50	0.07	0.13	0.75	1.00	0.50	1.00	25	2	6	0	0

designations. In fact, more than half of the speakers offered English designations only for 60% of the concepts.

Almost all of the designations offered in English were partially or completely integrated into Spanish phonological and/or morphological patterns (Mean [token integration]). Designations for only five concepts remained totally unintegrated into Spanish by all speakers: *frog* (6 occurrences), *pig* (2), *Timex* (1), *tape recorder* (11) and (baseball) *bat* (2).

For the large majority of concepts (84%), a single English designation accounts for between 75% and 100% of all English words offered (see Proportion [most frequent English type]/Proportion[English tokens]). Indeed, in well over half the cases (63%), the particular favored English designation represented half or better of all tokens, both English and Spanish, a good indicator of loanword status (Proportion [most frequent English type]). This despite the fact that a minimum of two types was offered per stimulus[10] and some received as many as eight, e.g. *jeans* (8), *nickel* (7), *lipstick* (6), *garbage can* (6), *baby* (6) (Total [types]). This reflects either the present state of flux in lexical assignment to these concepts or the inherent lack of one-to-one word–meaning relationships in these semantic fields.

Almost 75% of the most common English designations assigned a Spanish gender showed 100% agreement in gender assignment (Gender consistency). Some of the figures indicating less than perfect consistency obscure the fact that different lexical variants of a word may be assigned different genders, but within each variant there is total agreement. This is the case of *el hambérguer* and *la hamberga* 'hamburger'; *el suéter* and *la suera* 'sweater'. The single case of 0% gender agreement is due to sparse data: there is one token each of *el turtle* and *la térol.*

Of course these tabulations of index values do not bear on the broader lexicon of the respondents, given the way the items were selected. They do, however, indicate that diverse segments of the lexicon affected by borrowing have indeed been sampled, ranging from items which are only very occasionally designated by English forms, to those which are categorically expressed in English. Indeed, the figures in Table 1 provide a sensitive portrayal of the differential progress in the integration of foreign material into the lexicon for a given segment of the vocabulary at a given point in time. There may be clear (50:50) competition between the choice of English and Spanish designations for a concept (as for *switch*, *baby*, *dime*: Proportion [English tokens]), with no variation in the English designation (in this case *suiche*, *bebi*, *daim*: Proportion [most frequent English type]/Proportion [English tokens]). Other concepts may always or almost always be designated in English (*tape*, *zipper*, *hamburger*) but not always by the same English form, etc. This kind of information should

enable us to predict, on the basis of synchronic data, not only which of two competing languages will win out in the designation of a particular term, but also, in certain cases, which particular word. We are thus quite confident that the comparison of these responses will not exclude any key stage in the borrowing process, such as the early or final stages.

Though the actual value of each index may well be a function of its arbitrary definition and the particular choice of test item, nevertheless, and this is central to our study, the covariation among the indices can be measured in a way which factors out much of this arbitrariness. These measurements can then reveal genuine relationships among the different aspects of the borrowing process. Elucidation of this covariation will be the task undertaken in the ensuing sections.

4.2. *Relations among indices*

Some of the indices defined above are obviously closely related. As Proportion (English tokens) increases, so, very likely, will Proportion (English-only responses). Whenever Mean (token integration) is greater than 1.0, Maximum (integration) will necessarily be 2.0. To assess these and other relationships systematically, we calculate correlations among all the indices as shown in Table 2.

Four clearly demarcated clusters of indices emerge from these calculations, as indicated by the solid line.

1. The English-use cluster (measuring the importance of English designations for a concept): Proportion (English tokens), Proportion (English first choice), Proportion (English-only responses), Proportion (non-Spanish only), and Proportion (most common English type).

2. The phonological-integration cluster: Mean (token integration), Mean (type integration), Maximum (integration).

3. The gender-consistency cluster: Gender consistency, Use-weighted consistency, Field consistency.

4. The multiple-response cluster: Total (tokens), Total (multiresponse speakers), Proportion (bilingual responses).

All of these groupings are predictable to a greater or lesser extent, from the definition of their member indices, and the correlations within each cluster are all over 0.65, and for the English-use cluster, over 0.90. No other correlations across groups are even nearly as high, though there are some around 0.50.[11]

Other relatively high correlations appear to be more than definitional artifacts. For example, all the indices in the English-use cluster show statistically highly significant correlations with all the indices in the

Table 2. *Correlations among indices*

	Proportion (English-only responses)	Proportion (English 1st choice)	Proportion (English tokens)	Proportion (+frequent English type)	Proportion (non-Spanish responses)	Maximum (token integration)	Mean (type integration)	Mean (token integration)	Proportional (bilingual responses)
Proportion (English-only responses)	1.00								
Proportion (English (1st choice)	0.96	1.00							
Proportion (English tokens)	0.98	0.98	1.00						
Proportion (+frequent English type)	0.94	0.95	0.96	1.00					
Proportion (non-Spanish responses)	0.93	0.97	0.97	0.95	1.00				
Maximum (token integration)	0.44	0.49	0.47	0.46	0.52	1.00			
Mean (type integration)	0.32	0.32	0.33	0.37	0.38	0.66	1.00		
Mean (token integration)	0.35	0.38	0.38	0.42	0.43	0.69	0.87	1.00	
Proportion (bilingual responses)	0.02	0.23	0.19	0.22	0.39	0.32	0.23	0.27	1.00
Total (tokens)	0.16	0.28	0.28	0.22	0.40	0.35	0.23	0.33	0.68
Total (multiresponse speakers)	0.23	0.36	0.36	0.32	0.48	0.39	0.41	0.50	0.72
Use-weighted consistency	−0.09	−0.09	−0.10	−0.07	−0.07	−0.03	0.24	0.29	0.03
Field consistency	−0.15	−0.10	−0.11	−0.11	−0.07	−0.11	−0.06	−0.05	0.18
Gender consistency	−0.09	−0.02	−0.04	−0.04	0.00	−0.10	−0.03	0.04	0.20
Total (types)	−0.10	−0.02	−0.07	−0.14	0.03	0.22	0.19	0.09	0.30
Proportion (English (1st choice)	0.35	0.40	0.30	0.34	0.27	0.22	0.09	0.06	−0.13
Proportion (English tokens)									
Proportion (English types)	0.31	0.31	0.32	0.16	0.23	0.02	−0.24	−0.12	−0.14
Proportion (+frequent English type)	0.08	0.07	0.05	0.28	0.08	−0.17	0.01	0.02	0.03
Proportion (English tokens)									
Total (no response)	−0.25	−0.24	−0.25	−0.23	−0.23	−0.14	−0.05	−0.10	−0.03
Total (failures)	0.34	0.34	0.33	0.38	0.32	0.17	0.31	0.37	0.02

	Total (tokens)	Total (multiresponse speakers)	Use-weighted consistency	Field consistency	Gender consistency	Total (types)	Proportion (English 1st choice) / Proportion (English tokens)	Proportion (English types)	Proportion (+ frequent English type) / Proportion (English tokens)	Total (no response)	Total (failures)
Total (tokens)	1.00										
Total (multiresponse speakers)	0.86	1.00									
Use-weighted consistency	0.05	0.15	1.00								
Field consistency	0.20	0.23	0.66	1.00							
Gender consistency	0.30	0.28	0.67	0.92	1.00						
Total (types)	0.29	0.27	0.04	0.24	0.19	1.00					
Proportion (English 1st choice) / Proportion (English tokens)	−0.30	−0.27	−0.17	−0.24	−0.25	−0.00	1.00				
Proportion (English types)	0.01	0.01	−0.29	−0.22	−0.19	−0.37	0.07	1.00			
Proportion (+ frequent English type) / Proportion (English tokens)	−0.25	−0.17	0.15	−0.01	0.02	−0.47	0.11	−0.35	1.00		
Total (no response)	−0.47	−0.25	−0.10	−0.18	−0.38	0.14	−0.05	−0.06	0.02	1.00	
Total (failures)	−0.16	0.25	0.32	0.17	0.19	−0.09	0.22	0.03	0.21	−0.11	1.00

integration cluster, strongly indicating a relationship between the distinct phenomena measured by the two different clusters.

4.3. *Principal components analysis*

The correlations in Table 2 summarize the pairwise relationships among the 20 indices. Though we were able in the previous section to remark on some particularly high and particularly low correlations in this table, and to identify four groups of highly correlated indices, it is difficult if not impossible to appreciate the global patterns and relationships implicit in the data by trying to simultaneously compare all 190 separate correlation values. Thus, to obtain an overall perspective of these relationships, and to discover how the lexical representations of the various concepts are disposed in the space spanned by the indices, we undertook a principal components analysis of the correlations in Table 2.

Principal components analysis projects a multidimensional data configuration down to a subspace of low dimensionality, yet one which conserves as much of the variance in the data as possible. In this low-dimensional space, concepts, or referents, which have similar scores on all or most of the indices in Table 1 will be located close to each other, while those which have very different scores will be situated in regions remote from each other. This enables us to visualize grouping patterns not apparent in the original data format of Table 1, and to identify the one or two axes, or dimensions, or 'principal components', along which most of this variance occurs. We can also compare the projections of the 20 indices with respect to the principal components. Indices which are highly correlated will be projected close together on the space spanned by the principal components, while those which are negatively correlated will tend to appear far apart.

The coordinates of the 20 indices projected on the first four principal components account for 33%, 18%, 11%, and 10%, respectively, of the variance in the data (see Appendix). Subsequent components each account for 5% or less of the variance. These results can be more clearly described in graphical terms. In Figure 1, the projection of the indices on the two-dimensional space defined by the first two principal components shows a tight clustering of the indices within each of the four groups identified by examining the correlations in the previous section.

In addition, the somewhat elevated correlations in Table 2 between the English-use cluster at the upper right-hand quadrant of Figure 1 and the phonological integration cluster just below it show up in terms of the relative proximity of these two groups of indices. Indeed, the first

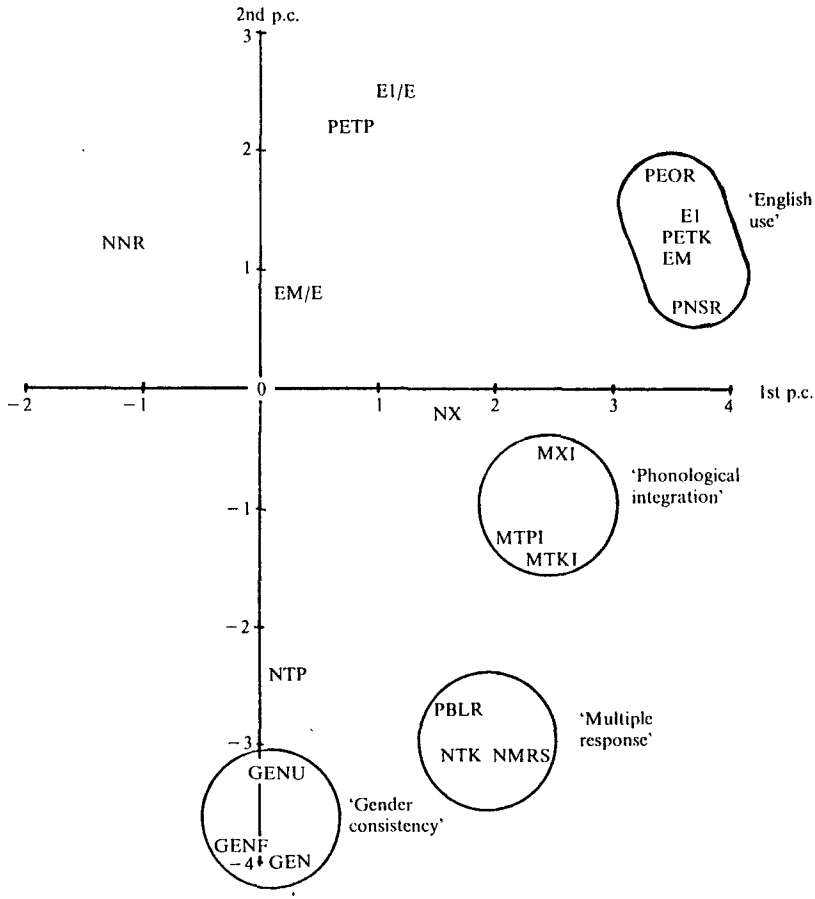

Figure 1. *Projection of indices on first two principal components*

Legend for Figures 1, 2, and 5:

PETK: proportion (English tokens)
PETP: proportion (English types)
PBLR: proportion (bilingual responses)
PEOR: proportion (English-only responses)
PNSR: proportion (non-Spanish responses)
MTPI: mean (type integration)
MTKI: mean (token integration)
MXI: maximum (token integration)
EM/E: $\frac{\text{proportion (most frequent English type)}}{\text{proportion (English tokens)}}$
EM: proportion (most frequent English type)
EI: proportion (English 1st choice)
EI/E: $\frac{\text{proportion (English 1st choice)}}{\text{proportion (English tokens)}}$
GEN: gender consistency
GENU: use-weighted consistency
GENF: field consistency
NTK: total (tokens)
NMRS: total (multiresponse speakers)
NTP: total (types)
NNR: total (no response)
NX: total (failures)

principal component falls in between these two clusters, with the English-use indices having somewhat higher scores on it than the phonological-integration cluster. Thus this component seems to be made up largely of a combination of the indices that most closely reflect the two key aspects of the incorporation of foreign material into a recipient language: frequency of use of a form and degree to which it is made to resemble recipient language forms. We thus interpret this component as the major axis of assimilation of English loanwords into Spanish.

Another important feature of Figure 1 is the proximity of the gender-consistency cluster, the multiple-response cluster and the Total (types) index at the bottom of the figure. In fact, the second principal component, which we call the diverse-response axis, seems to be determined largely by this set of indices. The proximity of these clusters, however, could not readily be predicted from the figures in Table 2, though the moderately elevated correlations involved are not inconsistent with the configuration. Though the association of Total (types) and the multiple-response cluster is directly interpretable in terms of the definitions of the indices (the more multiple responses, the more likely a diversity of types), the relationship of both to gender consistency is not. As we shall show below, however, this association is an indirect artifact of index definitions and the pattern of responses elicited by the questionnaire.[12]

Figure 2 is a plot of the second and third principal components. The major features are the integrity of the four main clusters, the continued position of the integration cluster between the English-use and multiple-response clusters, and the clear diassociation of the gender-consistency cluster from the multiple-response cluster and Total (types). Thus we see that the somewhat puzzling grouping of the latter two clusters is nowhere near as strong as the correlation between frequency of English use and degree of phonological assimilation.[13]

The fourth component, which we need not display graphically, sharply distinguishes the integration cluster from most of the other indices, including the ones in the English-use cluster.

In summary, from studying the relationship of the 20 indices to the principal components, as displayed in Figures 1 and 2, the following picture emerges. The first component, the major axis of loanword assimilation, is related to various indices both of frequency of English use and degree of phonological and morphological integration. In other words, of all the different indices we defined, those in the English-use and phonological clusters seem to be measuring phenomena which are closely related, and which proceed concurrently. We consider these to comprise the major pattern of the process of loanword assimilation. The second component, the diverse-response axis, which accounts for much less of the

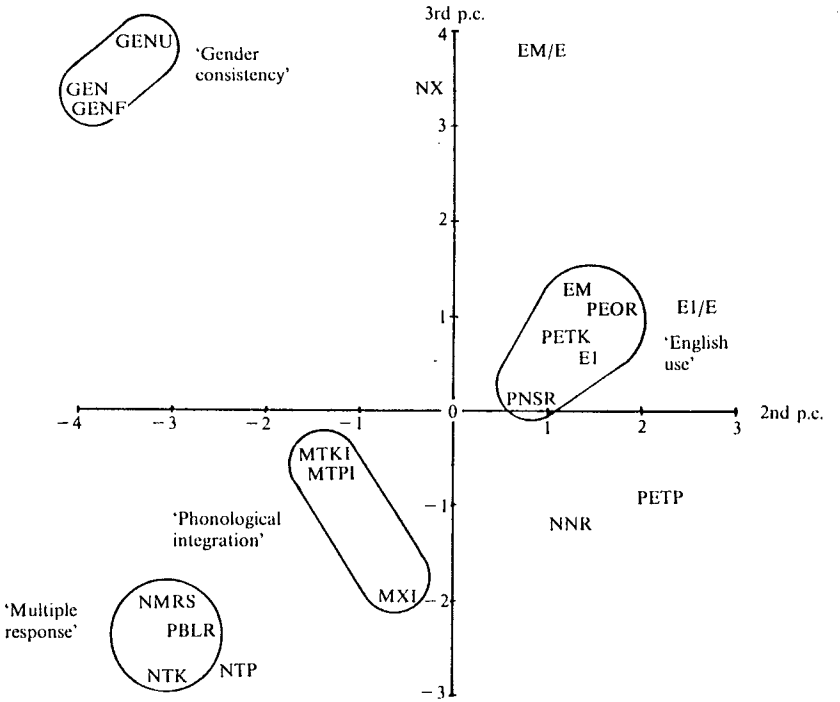

Figure 2. *Projection of indices on second and third principal components*

variance, reflects (as yet unexplained) association between gender consistency on one hand, and Total (types) and multiple responses on the other.[14]

4.4. *Principal components analysis of referents*

Having explored the significance of the first principal components in terms of their relationships to the 20 indices, we turn to another set of results produced during the same analysis. This is a study of where the test items (or concepts, or referents) are located in the space spanned by these components.

Since our choice of test items was not an attempt to sample the lexicon or parts of the lexicon in any systematic way, but was made for the practical reasons described in section 3.0, our aim here is not primarily to contrast different cultural domains or general types of nouns or of referents with respect to their susceptibility to borrowing, although the methodology we introduce here would be well suited to such an endeavor.

Rather we use the results from the 38 concepts studied to arrive at a more detailed understanding of how the assimilation and integration processes measured by the 20 indices actually operate on a given referent. This can then be abstracted to the broader lexicon.

Figure 3 plots the coordinates of the referents on the first two components. To illustrate the strong relationship of English use and phonological integration to each other and to the major axis of loanword assimilation, the referents whose designations were most (linguistically) integrated according to Mean (token integration) are enclosed in ellipses, and the figure is divided according to regions of high, medium, and low values of English use, according to Proportion (English tokens). With few exceptions, referents expressed by both highly phonologically integrated

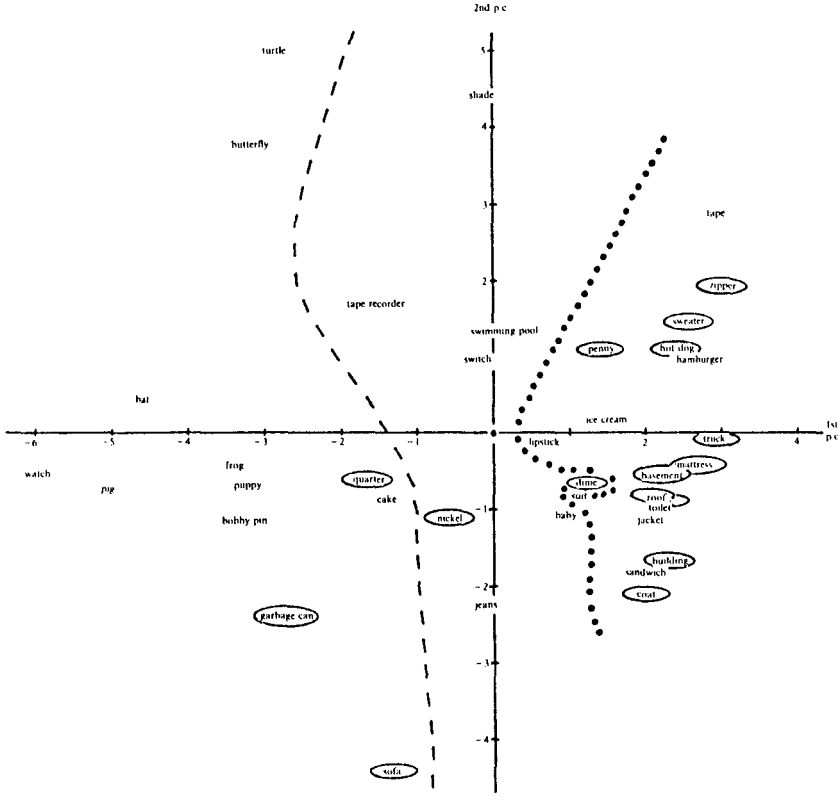

Figure 3. *Scores of 38 referents on first two principal components.* Proportion (English tokens) *is less than 1/3 to left of broken line and more than 3/5 to right of dotted line. Ellipses indicate average token integration is at least 1.25*

AND frequently used English words are at the right of the figure, while those only occasionally expressed by English words, and which undergo minimal phonological adaptation, are found to the left. This finding provides strong confirmation of the claims in the literature that borrowed words which are frequently used are made to conform with recipient language linguistic patterns.

That English use and phonological integration are nonetheless not perfectly correlated is illustrated by the four items in the lower left-hand quadrant of the figure (*quarter*, *nickel*, *sofa*, *garbage can*). Their designations show high integration indices, but the use of English forms has not displaced Spanish ones. Note also the location of *hamburger* and *tape*, for which virtually no Spanish forms are used, but where the unintegrated English form is often nonetheless given in response to the stimulus.

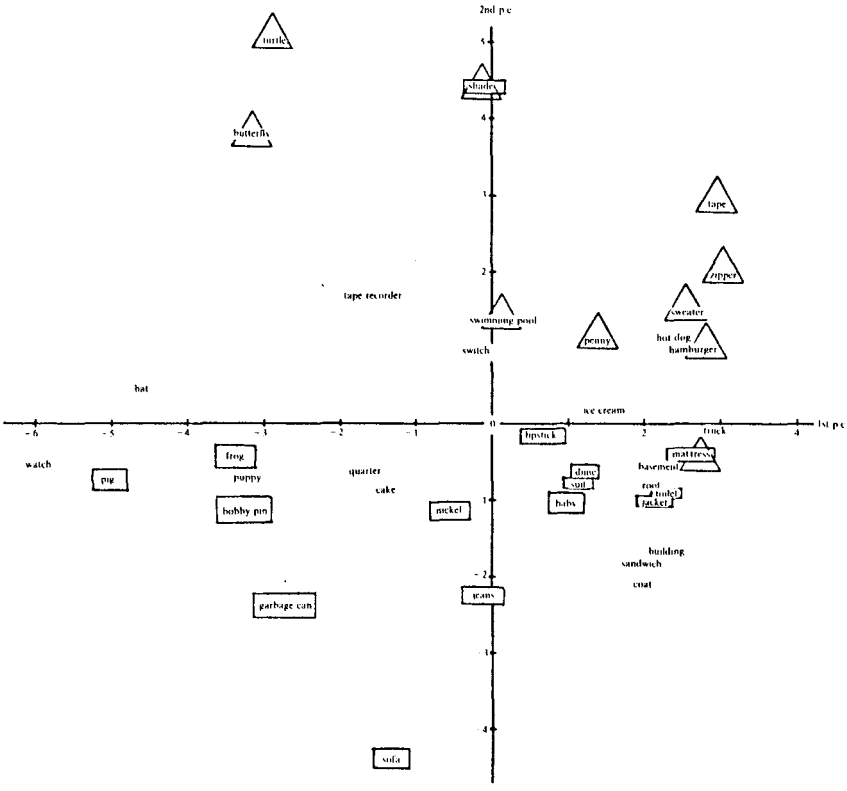

Figure 4. *Relationship between gender consistency and number of different types. Rectangles indicate at least five types elicited for concept. Triangles indicate at least one masculine and one feminine token elicited for most common English word*

Figure 4 is a copy of Figure 3 except that here we investigate the relationship of gender consistency and Total (types). It is at this point that we can best explicate the artifactual nature of this relationship. Those fields in the bottom half of the figure tend to be those which elicited the most diverse responses from the informants, e.g. the concept 'mattress' elicited the designations *matre*, *bed*, *pillow*, *colcha*, *caucho* (<*couch*) and *cóber* (<*cover*); 'building' elicited *bildin*, *apartamento*, *edificio*, *New Jersey*, *casa*. Since there were many different types, there could not be as many tokens for the most common English type as with the fields higher up in the figure. The smaller sample of tokens means that the possibility is reduced of detecting any gender conflict which may exist — a word which is feminine 90% of the time and masculine 10% of the time in discourse is not likely to show up as a gender conflict in the test if there are only three or four tokens of that word. Thus the apparent concentration of inconsistency in gender assignment to concepts designated by only a few different types must be interpreted as an indirect result of limited sample size.

However, one genuine pattern involving gender consistency can be seen by comparing Figures 3 and 4. Gender conflict (as measured by Gender consistency: the gender assigned to the most frequent English type offered for the concept) seems to afflict words midway along the assimilation axis (e.g. *turtle*, *butterfly*, *shade*, *swimming pool*) and/or types which have not been phonologically integrated into recipient language patterns (e.g. *tape*, *hamburger*). That *watch*, *pig*, and *bat* on the left show no conflict may be due to the low number of English tokens. But that so many of the frequently used and phonologically integrated words show no conflict cannot be explained away. Indeed, even the minimal apparent conflict that does exist among types with elevated phonological integration scores is largely due to divergent phonological forms which determine gender (*el suéter/la suera*, *el ziper/la zipa* 'zipper'),[15] and so gender consistency is actually higher than appears for these words. This result indicates that conflict in gender assignment is a transitory stage on the route to assimilation of certain loanwords, and tends to disappear as frequency of use and phonological integration increase. Indeed, this is precisely what was found in a study of gender assignment to borrowed nouns in natural (not elicited) speech among many of the same informants: gender inconsistency was almost nil (Poplack et al. 1982).

Summarizing the analysis of referents, the difference in the pattern of responses to the various test items is largely along a single dimension, from referents which are rarely represented by English words, and which in turn are relatively unintegrated into Spanish phonologically, to referents which are almost always represented by (one or more) phonologically

integrated forms of English origin. Of the variability among the test items not accounted for by this dimension, a large proportion is accounted for along a second dimension. The referents at one end of this latter axis are those which elicit a variety of different forms, which tend to be mostly Spanish (and hence do not constitute sufficient data for any single English form to manifest gender inconsistency), and at the other end are referents which are designated by few forms, so that one English form occurs in quantities sufficient to detect even a small degree of gender inconsistency. The association of response diversity and gender inconsistency is not, therefore, a linguistically significant fact but a consequence of the test format. On the other hand, it is noteworthy that gender conflict is actually restricted to words of English origin which either are used infrequently and/or retain donor language phonology.

4.5. *Intergenerational aspects of loanword assimilation*

The assimilation of loanwords is, of course, a diachronic process, best studied if possible at several points in time. When comparable historical data are not available, however, as is obviously the case with the quantitative materials we are using, we can have recourse to apparent time, i.e. intergenerational differences. In the case of borrowing, more, say, than in phonological or morphological change, this approach is complicated by acquisitional effects. Children may show different patterns of loanword use from their elders due to undeveloped vocabulary before arriving at the adult patterns. On the other hand, variation between the generations may be explicable in terms of differential handling of donor language material due to greater familiarity with it, as suggested in the literature.

To address these questions we chose 12 of our 20 indices and evaluated them separately for the 14 children and eight adults in our sample, resulting in 24 distinct measures.[16]

Table 3 compares children's and adult's mean scores on these 12 indices. That the indices in the English-use cluster are all higher for the children than the adults is clear evidence that English usage is advancing among the younger generation, at least insofar as the referents under investigation are concerned.[17] The equivocal results for the phonological-integration indices — children score lower on Mean (token integration) but higher on Maximum (token integration) than the adults — suggest that there is no diminishing of the Spanish phonological assimilatory mechanism in the children's speech, as might be expected if children regularly rendered words of English origin with donor language pho-

Table 3. *Children's and adults' average scores on twelve indices of loanword integration*

Indices	Children	adults
Total (tokens)	16.5	9.2
Total (no response)	0.6	0.6
Total (multiresponse speakers)	4.1	1.8
Proportion (English tokens)	62.6	51.2
Proportion (English-only responses)	56.4	44.4
Proportion (non-Spanish responses)	71.1	59.8
Proportion (English first choice)	64.2	52.6
Proportion (English first choice) / Proportion (English tokens)	80.3	73.7
Proportion (bilingual responses)	14.3	15.2
Mean (token integration)	107.2	115.6
Maximum (token integration)	179.4	150.0
Field consistency	90.1	78.6

nology. This finding thus disproves the contention of Haugen and others that childhood bilinguals (which is the case in varying degrees of the children in our sample) tend to reproduce borrowed material in a form which more closely approximates that of the source language (English) than speakers who acquired one of their two languages in adulthood, at least in the context of stable bilingualism. Rather, these results indicate that a term is transmitted across generations in the form under which it has become accepted into the speech community. Indeed, if we were to hear someone produce *roof* rather than *rufo* in an otherwise Spanish context, we would have good reason to consider this an instance of code-switching rather than borrowing.

Finally, note that the larger values among the children for Total (tokens) and Total (multiresponse speakers) are due to the fact that these are not normalized by the number of speakers: there are almost twice as many children as adults in the sample.

4.6. *Principal components analysis of children and adults*

When the 24 measures are entered into a principal components analysis, the results are as in Figure 5.

Three main tendencies emerge from these results. First, despite the reduction in the set of indices, and the distinction made between children and adults, the major patterns of the original analysis remain stable, confirming the reliability of the results. The English-use cluster, the integration cluster, and the multiple-response cluster are situated with

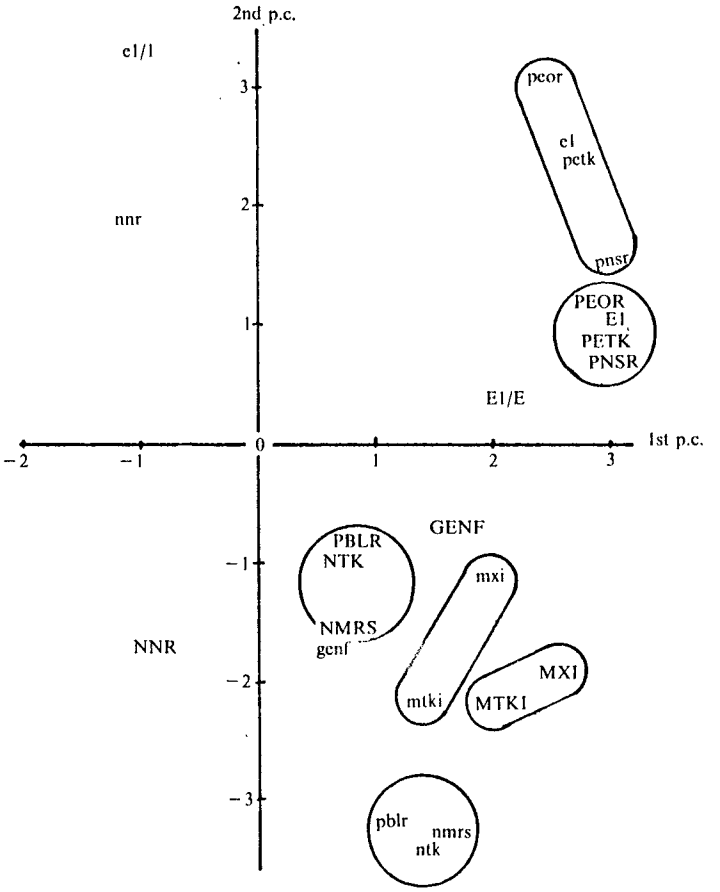

Figure 5. *Principal components analysis of children's (lower case) and adults' (upper case) indices: first two principal components*

respect to each other in much the same configuration as in the original analysis in Figure 2.[18]

Second, we note a striking congruence between the child and adult indices, indicated on Figure 5 with lower-case and capital letters respectively. The English-use cluster for adults is very close to that of the children, and similarly for the integration clusters and the Field consistency indices. The multiple-response clusters are somewhat farther apart, but in the same general area. Though Proportion (English first choice)/Proportion (English tokens) and Total (no response) indices show wide (vertical) separation between children and adults, they are in much

the same relationship with the rest of the indices. This congruence is proof that patterns of loanword assimilation are community-wide. They are not transitory tendencies on the part of first-generation immigrants, nor are they artifacts of the acquisition process, but rather a set of processes operating in a regular way in the New York Puerto Rican variety of Spanish.

Third, insofar as there are some intergenerational differences, these are almost all confined to the second principal component. On the major axis of loanword assimilation (the first component), the adults' indices and the children's indices are generally very close. But on the second component, the difference between adults and children is that the children's indices have much more extreme values, those above the horizontal axis being much farther above, and those below, generally much farther below.

In other words, with respect to the second component, which contrasts multiple-response indices against those associated with simple responses, the children's indices vary more widely. This indicates a greater variability in their responses regarding the number of different terms supplied for each concept. And this in turn can be related largely to the fact that more children were tested than adults. In Table 3, the average number of tokens and number of speakers who offered more than one response are almost double for children what they are for adults. It is obvious that these indices must thus have a greater range of variation for the former than for the latter. This difference between children and adults, then, reflects the construction and application of the questionnaire rather than any real intergenerational linguistic difference.

5.0. Discussion

We have defined a number of indices measuring what we assumed to be key components of the mechanism of integration of foreign material into a recipient language. These measured aspects such as (1) the degree to which English designations (more than one type) are displacing Spanish designations for a concept; (2) the degree to which a SINGLE English word is displacing other designations of the concept (synonym displacement); (3) various aspects of phonological and morphological integration of the English-origin form; (4) consistency in its assignment to a gender; and (5) the number of people who use a given designation.

These indices were applied to a data base consisting of English and Spanish designations for 45 referents proffered by 22 Puerto Rican bilinguals. The coordinates of our 20 indices of loanword integration projected onto a low-dimensional space enabled us to confirm that the

major pattern of assimilation of English loanwords into Spanish is made up of a combination of frequency of use and phonological and morphological integration indices, which showed up in the analysis as measuring phenomena which are closely related and which tend to proceed concurrently. Projection of the coordinates of the 38 referents retained in the analysis onto the space spanned by the principal components enabled us to confirm that it is precisely those concepts for which mostly English designations are offered which show the greatest degree of linguistic integration into the recipient language (Figure 3). We also saw that the apparent but puzzling relationship between gender inconsistency and synonym displacement (as expressed by designation of a referent via a reduced number of types) is in fact best understood as resulting from infrequent use and/or lack of linguistic integration of a concept (Figure 4). Figures 3 and 4 in conjunction also showed that concepts may be distinguished according to the number of English types by which they are designated. Obviously, if an English-origin word has displaced all other designations for a concept it is an unmistakeable candidate for loanword status. Thus, it is the concepts showing most English use, most phonological and morphological integration, and fewest different types for which the designations can be considered true loanwords.[19] Other concepts show less advanced stages of loanword incorporation.

This state of affairs can be interpreted in terms of inferred trajectories over time of the concepts in the space defined by the principal components. English designations may be used originally in relatively linguistically unintegrated form and at low frequency, as alternatives to the usual Spanish term. Of course, imposition of the phonology of the recipient language on borrowed forms often implies some degree of integration even at this early stage. In addition, some concepts may not have readily accessible designations in the recipient language, so that borrowed terms are used at the outset when these concepts came to be expressed.[20]

For most concepts, however, low-frequency, little-integrated terms gain in currency and lose their English phonology as the borrowing process proceeds. At early stages, analogical and/or phonological conditioning of gender is equivocal and some inconsistency may result. As integration proceeds, however, this inconsistency is completely resolved in one of two ways: either a single gender becomes the norm, or two phonologically distinct reflexes of the same etymon persist, each with appropriate gender (e.g. *el suéter/la suera*).

In the course of the linguistic and social integration of the words designating a referent, two divergent paths appear. The first involves the displacement of other Spanish and English designations by a single English term. This is the easily interpretable case, clearly identifiable as a

borrowing. The other still involves the exclusion of Spanish designations, but in favor of a multiplicity of English types. There are different ways of interpreting this outcome. It may well be that the borrowing process can pass through a phase where the Spanish designation is displaced by a number of competing English terms before any single one of the latter can predominate. Alternatively, synonymy or partial synonymy may be a viable long-term outcome, but the English influence extends to all concepts in the surrounding semantic domain. Finally, it could be in the nature of the data elicitation procedure that despite the use of photographs of concrete items, some referents are inherently more unambiguously connotated than others, and the form–function or word–meaning relationship tends to be one-to-one, either universally or in one or both of Spanish and English. The other items would tend more to elicit words for referentially closely related concepts. Again, this possibility implies the extension of English to all concepts in the semantic domain surrounding the target referent. Neither of these latter explanations necessarily precludes the 'true loanword' status of the terms involved.

Further research would be necessary to distinguish between these explanations — temporary flux as a stage in borrowing, persistent synonymy, or nonspecificity in elicitation.

A final important finding to emerge from this analysis is the relative homogeneity of the community with respect to loanword acceptance and utilization. This stems from the result that the older and younger speakers are not highly differentiated with respect to loanword usage. The fact that the children's indices of loanword integration are so close to those of the adults is especially striking, given that almost all of the adults in this study are Spanish-dominant, while most of the children are English-dominant or bilingual. Nonetheless, the younger speakers are not agents of importation of foreign phonological and morphological patterns into the recipient language in the context of a stable bilingual community,[21] although it is clearly within their competence in the English language to do so. Rather it appears that once a term is accepted into the speech community, and adapted into a particular phonological form, it is that form which is transmitted across generations in much the same way as monolingual neologisms. This is important evidence that the process of borrowing is carried out in a regular way on the community level and is not a series of random accidents.

The use of principal components analysis in this work has proved an extremely useful tool in discovering and illustrating regularities in the data. At the same time, we have seen that it is a technique which must be used with great care, and all apparent patterns should be carefully verified with the correlation matrix and interpreted in terms of the original data.

For example, the process of morphological/syntactic integration (Gender consistency) showed up in a distorted and difficult-to-interpret way, which could only be understood in terms of a critique of the original data. A projection of the data on a space of too low dimensionality resulted in a spurious grouping of some of the indices. Finally, a systematic but artifactual difference between children and adults appeared on the second principal component. This was merely an effect of sample size. These dangers, however, are more than compensated for if the analysis is used carefully, since not only can true patterns be recovered, but operational problems with the methodology at any stage of the study can be detected and resolved.

The traditional structuralist position, that loanwords are indistinguishable from the rest of the lexicon, clearly does not pertain to the transitional phenomena discussed here. But in a stable bilingual community, the problem is just the opposite: how to distinguish loanwords from an L_2 from the rest of the L_2-origin material also coexisting in the recipient language. Etymological considerations are of no help, since in any event they are known; speaker intuitions may be contradictory, and even empirical observation of usage at a given point in time is inadequate, given the superficial similarities between borrowing, code-switching, incomplete L_2 acquisition, interference, etc. These problems have led us to the fixed-referent elicitation protocol described here, which, though encumbered with the artificiality of the test situation, still enables us to probe rather directly the linguistic and social mechanisms affecting the integration of specific loanwords. We believe we have shown that when examined on the community level, rather than on an individual basis, candidates for loanword status do indeed manifest characteristics not shared by other L_2-origin words.

Received 18 May 1984

Department of Linguistics
University of Ottawa
78 Laurier St. E.
Ottawa, Ontario
K1N 6N5 Canada

Centre de Recherche de
Mathématiques Appliquées
University of Montreal
Montreal, Quebec
H3C 3J7, Canada

Appendix

Projection of 20 indices of loanword integration on the first four principal components

Indices	Components			
	1	2	3	4
Proportion (English tokens)	0.36	0.13	0.06	−0.14
Proportion (English types)	0.07	0.22	−0.09	−0.37
Proportion (bilingual responses)	0.16	−0.27	−0.23	−0.05
Proportion (English-only responses)	0.34	0.18	0.11	−0.10
Proportion (non-Spanish responses)	0.37	0.07	0.01	−0.11
Mean (type integration)	0.22	−0.13	−0.06	0.47
Mean (token integration)	0.25	−0.14	−0.04	0.41
Maximum (token integration)	0.26	−0.05	−0.20	0.26
Proportion (most frequent English type) / Proportion (English tokens)	0.02	0.08	0.38	0.15
Proportion (most frequent English type)	0.36	0.13	0.13	−0.05
Proportion (English first choice)	0.36	0.13	0.06	−0.12
Proportion (English first choice) / Proportion (English tokens)	0.11	0.25	0.11	0.15
Gender consistency	0.02	−0.40	0.33	−0.22
Use-weighted consistency	0.01	−0.32	0.39	0.13
Field consistency	−0.02	−0.39	0.31	−0.17
Total (tokens)	0.18	−0.31	−0.29	−0.25
Total (multiresponse speakers)	0.23	−0.31	−0.20	−0.09
Total (types)	0.02	−0.24	−0.27	0.12
Total (no response)	−0.12	0.12	−0.12	0.29
Total (failures)	0.16	−0.02	0.34	0.17

Notes

* This research has benefited from grants from the National Science and Engineering Research Council (Canada) to Sankoff, and from the National Institute of Education (USA), the latter as part of a larger research project on language contact carried out while Poplack was at the Center for Puerto Rican Studies, City University of New York. Alicia Pousada administered the questionnaire as part of the overall fieldwork for the project. We are grateful to François Grosjean, Don Hindle, Raymond Mougeon, Edouard Béniak, Daniel Valois, and researchers at the Center for Puerto Rican Studies, who read and commented on earlier versions of this paper.

1. See, however, two recent studies by Roy (1979) and Mougeon and associates (1983) of the use of English conjunctions in varieties of Canadian French.
2. Provided, of course, that (1) there exists an indigenous term to be displaced after its content becomes fully covered by the loanword, and (2) the native and borrowed term do not coexist with specialization in content. As Weinreich ((953) remarked, there is often insufficient information in loanword studies to decide whether the indigenous term has become specialized or discarded. The fixed-referent loanword elicitation procedure used here (section 3.2) allows us to control for these possibilities.

3. In Thomason's (1981) schema, short-term intense contact situations result in language death before recipient-language speakers are able to incorporate many donor-language features into it. As mentioned above, however, there are no signs here of community-wide language shift or death.
4. For a detailed description of the study and its findings, see e.g. Language Policy Task Force 1980, Poplack 1981, 1983.
5. Assessments of language preference were made through a combination of long-term ethnographic observation, self-report, and linguistic analysis.
6. The fact that the task was loaded in favor of Spanish responses to the stimuli may be taken as further indication of the degree of integration of English-origin terms into the Spanish repertoire of the informants.
7. We use the term 'token' to refer to every occurrence of a lexical item. Identical tokens belong to a single 'type'.
8. Failures and no response were not included in any calculations other than the ones specifically designated to account for them.
9. These were distractors; with the exception of 'window', we had no reason to expect any of them to be rendered in English.
10. Undoubtedly due to the elicitation procedure, which prompted the respondent for additional designations.
11. Certain of these latter are also somewhat predictable from their definitions. For example, Total (no response) and Total (tokens) are relatively highly but negatively correlated (-0.47). Similarly, Proportion (most common English type)/Proportion (English tokens) is negatively correlated (-0.47) with Total (types), since the more tokens that are concentrated on one type, the fewer tokens there are left to represent other types. Proportion (bilingual responses) correlates with Proportion (non-Spanish-only) responses at the 0.39 level.
12. We note as well in Figure 1 the distance of Total (no response) in the upper left quadrant from Total (tokens), in the lower right quadrant, as was predicted by their negative correlations.

    The apparent grouping of Proportion (English types), Proportion (English first choice)/Proportion (English tokens), and Proportion (most frequent English type)/Proportion (English tokens) in the upper central part of the figure is somewhat surprising and is not the result of any similarity among the indices, since they are only marginally or even negatively correlated. Their proximity is a consequence of each one's position being determined by a large number of small or negative correlations with the indices in two or more clusters: Proportion (English types) is positioned by virtue of its negative correlation with the integration and gender-consistency clusters. Proportion (English first response)/Proportion (English Tokens) is positioned high in the first quadrant by virtue of its negative correlation with all elements of the gender-consistency and multiple-response clusters and its moderate correlation with the English-use cluster. Proportion (most common English type)/Proportion (English tokens) is independent (near zero correlation) of most of the other indices and is positioned as a function of its high negative correlations with the multiple-response cluster and Proportion (English types). Such spurious groupings are a shortcoming of principal components analysis, but may be detected by reference to the correlations and by examining further components.
13. The spurious grouping discussed in note 12 is also dispersed when the third dimension is added.
14. The third and fourth components, each accounting for only about 10% of the variance, distinguish between clusters or indices which were closer together on the first two

components than is justified by their correlations. The projection down to two-dimensional space necessarily crowds together a few items which are not highly correlated, in order for the majority of the distances to correspond well to the correlation matrix. Allowing a third or fourth dimension alleviates some of these unwarranted proximities. A preliminary survey of the next few principal components reveals no interpretable patterns.

15. Spanish phonological rules for gender assignment require masculine gender for most words ending in *-r* and feminine for most words ending in *-a*. Due to variable syllable- and word-final *r* deletion in New York City English, borrowings ending in this consonant may be rendered in Spanish with final [er] or [a].
16. The computational tools we are using have limited capacity and so we eliminated indices which had proved redundant or not systematically patterned in the previous analysis, or could not be applied to the smaller data sets.
17. This can be at least partially understood in terms of the independent finding that these children use English in conjunction with Spanish, in more contexts and interactions than their elders in the community, who reported more situations requiring exclusive use of Spanish (Poplack 1983).
18. The Field gender consistency indices are somewhat more central than in the original, but the overall pattern remains.
19. From the analysis presented in Figures 3 and 4, the closest candidates for loanword status are *penny*: [pɛni], *hot dog*: [franfura], *sweater*: ['suɛtɛr/suera], *zipper*: ['sipɛr/sipa], *truck*: [tro], *basement*: ['beihmaŋ], *building*: ['bildiŋ], *coat*: [ko], and *roof*: [rufo].
20. Examples from this study include '(cassette) tape', 'hot dog', 'hamburger'.
21. Mougeon (1982 and personal communication) reports a case of French-English shift where English-dominant speakers are the rule rather than the exception. Their 'non-native' innovations in French appear to be spreading to French-dominant speakers. This is clearly a different situation from the situation of stable bilingualism examined here.

References

Barkin, F. (1980). The role of loanword assimilation in gender assignment. *Bilingual Review*.

Bloomfield, L. (1933). *Language History*, H. Hoijer (ed.). New York: Holt, Rinehart and Winston.

Fries, C., and Pike, K. (1949). Coexistent phonemic systems. *Language* 25, 29–50.

Hasselmo, N. (1969). On diversity in American Swedish. *Svenska Landsmäl och Svenskt Folkliv*, 53–72.

—(1970). Code-switching and modes of speaking. In G. Gilbert (ed.), *Texas Studies in Bilingualism*. Berlin: de Gruyter.

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

—(1950b). Problems of bilingualism. *Lingua* 2, 271–290.

—(1956). *Bilingualism in the Americas: A Bibliography and Research Guide*. American Dialect Society Monograph No. 26. University: University of Alabama Press.

—(1969). *The Norwegian Language in America*. Philadelphia: University of Pennsylvania Press.

Holden, K. (1976). Assimilation rates of borrowing. *Language* 52(1), 131–147.

Hyman, L. (1970). The role of borrowing in the justification of phonological grammars. *Studies in African Linguistics* 1(1), 1–48.

Kreidler, C. (1979). A study of the influence of English on the Spanish of Puerto Ricans in Jersey City. Unpublished Ph.D. dissertation, University of Michigan.

Language Policy Task Force (1980). Social dimensions of language use in East Harlem. Working Paper #7. New York: Center for Puerto Rican Studies.

Lovins, J. (1974). Why loan phonology is natural phonology. In A. Bruck et al. (eds.), *Papers from the Parasession on Natural Phonology*, 240–250. Chicago: Chicago Linguistic Society.

Mackey, W. (1970). Interference, integration and the synchronic fallacy. In J. Alatis (ed.), *Georgetown University Roundtable on Languages and Linguistics* 23, 195–227. Washington, D.C.: Georgetown University Press.

Mougeon, R. (1982). The problem of ambiguous change in a contact language. Paper presented at NWAVE-11. University of Pennsylvania.

—, Valois, D., and Béniak, E. (1983). For a quantitative study of linguistic borrowing. Paper presented at the Canadian Linguistic Association, University of British Columbia.

Murphy, R. P. (1974). Interference, integration and the verbal repertoire. *Linguistics* 128, 59–67.

Poplack, S. (1980). 'Sometimes I'll start a sentence in Spanish y termino en español': toward a typology of code-switching. *Linguistics* 18: 581–618.

—(1981). Bilingualism and the vernacular. In B. Hartford, A. Valdman, and C. Foster (eds.), *Issues in International Bilingual Education: The Role of the Vernacular*, 1–24, New York: Plenum.

—(1983). Intergenerational variation in language use and structure in a bilingual context. In C. Rivera (ed.), *An Ethnographic/Sociolinguistic Approach to Language Proficiency Assessment*, 42–70. Avon, England: Multilingual Matters.

—, Pousada, A., and Sankoff, D. (1982). Competing influences on gender assignment: variable process, stable outcome. *Lingua* 57, 1–28.

Roy, M.-M. (1979). Les conjonctions anglaises 'but' et 'so' dans le français de Moncton. Unpublished M.A. thesis, Université du Québec à Montréal.

Sankoff, D., and Poplack, S. (1981). A formal grammar for code-switching. *Papers in Linguistics* 14(1), 3–46.

Shaffer, D. (1978). The place of code-switching in linguistic contacts. In M. Paradis (ed.), *Aspects of Bilingualism*. Columbia, S.C.: Hornbeam.

Thomason, S. (1981). Are there linguistic prerequisites for contact-induced language change? Paper presented at the Tenth Annual University of Wisconsin-Milwaukee Linguistic Symposium.

Weinreich, U. (1953). *Languages in Contact*. The Hague: Mouton.

Whitney, W. D. (1881). On mixture in language. *Transactions of the American Philological Association* 12, 5–26.