Including Caribbean Languages in the World Loanword Database

Carlos Rodriguez-Iglesias
University of Puerto Rico, Río Piedras

In 2003, linguist, novelist and well-known evangelist of the Papiamento language, Frank Martinus Arion, wrote an essay titled "Creole Identity through Chinese Wall: Affinities between Papiamento and Chinese" where, among other topics, he explored some of the linguistic similarities between Chinese and creole languages. Towards the end of the article, Martinus touched on the topic of lexical expansion:

> I will in any case suggest the use of the Chinese way in the amplification of the lexicon and further standardization of Papiamento. It will make us much stronger against those agents of decreolization who believe that normalization of Papiamento means turning it into Spanish, Dutch or English. [1]

The example offered in his essay is the Mandarin Chinese word for "elevator," which was coined by joining the words for "electricity" (电 *diàn*) + "ladder" (梯 *tī*), resulting in the compound 电梯 (diàntī). According to Martinus, the analogous process in Papiamento would have yielded the term *trapi electrico*. Papiamentu speakers seem to have disregarded this recommendation and continue to use the word "lift" instead, which they perceive to be a Dutch loanword, though it is the preferred term in British English as well. Curiously, in Cantonese – a Chinese "dialect" different from Mandarin – the preferred term is actually 較 (pronounced "lip[1]"), which is clearly a case of lexical borrowing, presumably from English.

It goes without saying that all languages borrow lexical items from other languages, especially if they arrive alongside some sort of cultural importation: cf. English *kosher* from Yiddish, or Mandarin 咖啡 *kafei* "coffee" from Portuguese. Quechua borrowed and adapted *arrusa* "rice" and *riluju* "clock" from Spanish, to refer to items that did not exist in the Americas prior to European colonization. Similarly, in the 20th and 21st centuries, Anglicisms have easily found their way into many European and non-European languages, especially in connection with new technologies, despite efforts by the language academies of France and Spain to curtail this

trend by offering "native" alternatives for words such as *email*. One does not need to be a purist, however, to recognize that all languages have at their disposal internal mechanisms of word formation that make the adoption of foreign terms unnecessary. As Haspelmath points out: "...there is always the question why a borrowing had to take place at all, because all languages have the means to create novel expressions out of their own resources. Instead of borrowing a word, they could simply make up a new word. And of course there are many other cases where it is not at all clear why a language borrowed a word from another language, because a fully equivalent word existed beforehand."[2] At first glance, "cultural" loans such as the ones referenced above may seem almost mandatory, but in fact many Native American language communities did not follow the example of the Quechua and relied instead on their own language-internal processes to develop "native" terms to refer to rice and clocks.

The "Chinese way" of word formation that Martinus alludes to in his essay is certainly not unique to the Chinese language either. Hancock calls it "incoining," which he defines as "the creation of new terms by combining morphemes already available in the language, so that the meaning of the combination is different from either or any of its components."[3] Incoining is only one of the 12 processes of expansion (or "progression" as he prefers to call it) that languages may resort to when adding items to their lexicon, according to Hancock. Some of these processes involve morphological changes; others consist of semantic shifts (including extension, convergence, and divergence); and last but not least, languages may use an external model that is either calqued or adopted. This latter method of "adoption" results in what is commonly known as a loanword, or lexical borrowing. All languages employ most if not all of these processes of lexical expansion to some degree, but particular languages or language groups may rely on specific processes more extensively than others. For example, the paucity of affixes that has often been pointed out as an almost defining quality of pidgins and creoles might be counterbalanced in these languages by a more extensive use of alternative mechanisms such as reduplication. Reduplication in this context refers to repetition of morphemes resulting in a distinct lexical item with a slightly different meaning. For example, in Haitian *gran* means "big" whereas *gran gran* is "huge." Often there is not only an intensification but also a shift in meaning, as in the Negerhollands word for "morning" *vroevroe* (from Dutch *vroeg* "early").[4] Another strategy of semantic and/or syntactic differentiation (absent in the European lexifiers) is the use of tones, a relatively common feature in the phonology of many African and Asian languages.

Assuming that lexical expansion or progression does not take place in a completely random fashion, the following questions ensue: When and why do language communities resort to borrowing lexical items, as opposed to the creation of new words using their own language-internal processes? Do some languages "borrow" more readily than others? And for that matter, are Caribbean languages, particularly those commonly called "creoles," different from non-creole languages in this respect? Aren't ***all*** the words in the lexicon of "Creole languages" borrowed from other languages?

The search for answers to the first two of these questions motivated Martin Haspelmath and Uri Tadmor to create the World Loanword Database (http://wold.clld.org), a publication by the Max Planck Institute for Evolutionary Anthropology, following in the footsteps of the well-known World Atlas of Language Structures (WALS). WOLD provides vocabularies (mini-dictionaries of about 1000-2000 entries) for 41 languages from around the world, with comprehensive information about the loanword status of each entry. It allows users to examine the borrowing behavior of each language from a typological and comparative perspective, by indicating the source words and donor languages for each entry on the lists. The list of 1460 entries on which the vocabularies are based is called the Loanword Typology Meaning List, and it is in turn based on the list of the Intercontinental Dictionary Series. Each vocabulary was contributed by an expert on the language and its history. Among their general findings, the editors found that around half of the sample languages were classified "high borrowers" (between 25–50% of their lexicon can be traced to other languages), including Japanese and English. Only four languages are "low borrowers" with less than 10%. (Interestingly, Mandarin Chinese has the lowest borrowing rate in the sample pool: 1.2 %).

Any undertaking of this magnitude understandingly implies that not all languages can be included in the sample, and Saramaccan is the only Caribbean vernacular represented in the WOLD.[5] There is no information on Papiamento or Haitian, much less other Caribbean vernaculars such as Jamaican whose status as "languages" or "dialects" is a hotly debated issue among linguists. It is one of the central arguments of this essay to support and encourage the addition of Caribbean vernaculars to the WOLD database; alternatively, an analogous project may be developed independently, so that creolists, lexicographers and others interested in the linguistic situation of the Caribbean, may access this type of data for their own analyses.

Embarking on such a project, of course, requires tackling some challenging issues regarding the lexicon of "creole" languages. First, for most of these vernaculars, it is not easy to determine which words are

"officially" part of the lexicon. Even in the case of Papiamento, one of the most standardized "creoles" in the Caribbean and one of only two in the region that enjoys official government recognition, there are no monolingual dictionaries in existence. Some bilingual dictionaries are available, but they seem to vary significantly in their quality, and they certainly lack any "official" status or recognition as a normative standard. Lexicographers are likely to encounter the additional challenge that, in islands such as Aruba and Curaçao where a large portion of the population is multi-lingual, there is bound to be a considerable amount of code-switching taking place, either deliberate or at an unconscious level. Papiamento speakers often find themselves interspersing their speech with Spanish or Dutch vocabulary, sometimes even using a word they believe to exist in Papiamento but which actually does not figure in the lexicon, at least according to other speakers. Moreover, there may be attitudinal and social factors at play, whereby word choice is based on the speaker's perception of which of the input languages is more "prestigious." Eckkrammer, writing about the challenges facing the standardization effort in the case of Papiamento, speaks about these difficulties: "In conclusion the obstacles for the standardization committee are enormous considering the elevated number of more or less established synonyms. In written and oral discourse it is highly difficult to decide whether an expression is an unnecessary Hispanism or Anglicism or an established term in a certain area."[6]

To illustrate using a couple of examples from her article: the entry for "strawberry" in the English-Papiamento dictionary compiled and published in 1992 by Betty Ratzlaff[7] lists three options: 1. *artbei* (from Dutch); 2. *fresa* (from Spanish); and 3. *stroberi* (from English). Similarly, three possible translations are provided for the concept of "eyebrow": 1. *seha*; 2. *wenkbrou*/*wenbroul*; and 3. *kachu di wowo*. The first is clearly an adoption from Spanish, followed by two variants of the Dutch derivate. The third would surely be Martinus' preferred option, a compound of two words, presumably of African origin. Are we to conclude that the first two items are loanwords, and the third a "native" Papiamento word? Would that assessment change if we ascertained that the words *kachu* ("horn") and *wowo* ("eye") are borrowings from African languages? Or perhaps, although *kachu* and *wowo* may be lexical borrowings as well, their pairing in *kachu di wowo* is a semantic composition unique to Papiamento, thus reinforcing the conclusion that it represents the "native" option. Finally, when do *seha* and *wenkbrou* cease to be regarded as lexical borrowings and become Papiamento words in their own right?

The easiest of these questions to answer is the last one, and the response is: never, at least according to the WOLD editors. They define "loanwords"

(lexical borrowings) as follows: "A loanword is a word that was copied from another language, either by adoption or by retention, at some point in the history of the language. Even if a loanword is fully integrated, it is still a loanword, and a loanword never ceases to be a loanword."[8] Under this definition, and having seen that English and Japanese are classified in the WOLD as "high borrowers" due to the fact that 25% to 50% of their lexicon is comprised of loanwords, it seems reasonable to expect that Papiamento, Haitian and other Caribbean languages would turn out to be high borrowers as well, if not *very high* borrowers. According to Eckkrammer, the lexical distribution of modern Papiamentu is the following: almost two thirds of vocabulary has an Iberian (either Spanish and/or Portuguese) origin, 28% of the lexical borrowings derive from the Dutch language, a small number of words evolved from English and French, and minor influences are observed from African and indigenous languages, especially in names for plants, animals, food and music.[9] This would suggest that there is not a word in Papiamentu that cannot be traced to other languages, whether European, African or indigenous. Are we to conclude then that "creole" languages such as Papiamento exhibit the highest percentages of lexical borrowing among the languages of the world?

In addition to the lack of "native" vocabulary, there is also the widespread claim that "creoles" have little or no morphology, and thus, lacking these internal mechanisms of word formation, must resort to the adoption of foreign items in order to augment their lexicon. Such a claim, however, is simply not accurate, as it can be easily demonstrated that Papiamento and other Caribbean vernaculars do possess productive morphology. Continuing to use Papiamento as our primary example, the Aruba Department of Education's *Manual di Gramatica*[10] distinguishes three types of nouns in the language: simple nouns, compound nouns, and derived nouns. Compound nouns are those that consist of more than one "word": e.g., *awa di wowo* (tear, literally "water of eye"). Derived nouns are formed by adding prefixes or suffixes to another morpheme, typically called the root or base. These derived lexical entries could potentially be the result of internal word-formation processes, or they could have been adopted in their already derived form. The terms *acumulashon* ("accumulation") and *acusashon* ("accusation"), for example, are almost identical in spelling and meaning to their Spanish counterparts; "no evidence for productive morphology can be distilled."[11] Nevertheless, there are other cases where the productivity of bound morphemes can be demonstrated, as in the following examples:

perde-do	"loser"
uza-do	"user"
huur-do	"tenant"

The evidence that this is in fact a productive morpheme lies precisely in the fact that it can be attached to different stems regardless of the linguistic origin of the etymon. While in the first two examples, it is conceivable that the words were borrowed from Spanish or Portuguese in their already derived forms, the third example demonstrates that the *-do* suffix can be applied to stems of Dutch origin as well. A second productive suffix is *mento*, also used to create nominal forms but in this case referring to the act rather than the agent; the three lexical roots presented in the earlier example can be combined with *mento* to produce the nouns *perdemento* ("loss"), *uzamento* ("usage") and *huurmento* ("lease"), forms that are not attested in the input languages.

The *Manual* also makes reference to derived adjectives and adverbs but, unlike the case of the nouns described above, it cannot be conclusively shown that said affixes are being used productively in Papiamento since it is possible that those lexical items were adopted directly from the lexifier already in derived form. There is one notable exception, *lastimamente*, an adverb derived "natively" from the noun *lastima* (in Spanish and Portuguese, adverbs ending in *-mente* are derived from adjectives, not from nouns, and accordingly the Spanish/Portuguese equivalent here would be "lamentablemente"/"lamentavelmente"). This isolated case has led the Aruba Education Department to conclude: "*Posiblemente Papiamento a tuma hopi adverbio cu mente directamente for di Spaño, sin tene cuenta cu derivacion. Pero e proceso di derivacion ta existi internamente den nos idioma tambe, por ehempel den e caso di lastimamente.*"[12] ("It is possible that Papiamento took many adverbs ending in *mente* directly from Spanish, without derivation taking place. But the derivation process does exist in our language as well, for example, in the case of *lastimamente*.")

Even after acknowledging that Papiamento does in fact employ morphological processes responsible for the formation of "native" words such as *huurmento*, the percentage of loanwords in the language is likely to be extraordinarily high, compared to other languages in the WOLD. Having concluded that this is not due to a lack of internal productive mechanisms, an alternative explanation may be that we simply have better data regarding the history of Papiamento than we do for "older" languages. After all, we can never exclude that a word has been borrowed at some stage in the history of a language. Haspelmath writes:

Even for words that have been reconstructed for a very ancient proto-language, such as English mother (from Proto-Indo-European *mātēr) or ten (from *deḱm), we cannot be sure that they were not borrowed from another language at some earlier stage. Thus, we can identify loanwords, but we cannot identify "non-loanwords" in an absolute sense. A "non-loanword" is simply a word for which we have no knowledge that it was borrowed. [13]

In other words, the status of "native" words is always relative to what we know about the history of a language. It is likely that many non-creoles would display a much higher percentage of loanwords if we knew more about their history, especially their history of contact with other languages. Certainly, this is exemplified by English and Japanese (identified by the WOLD as high borrowers) for which there are abundant sources that document the importation of lexical items from French and Chinese, respectively. The only Romance language represented in the database, Rumanian, is also among the high borrowers at 48%.

Ultimately, adding Caribbean vernaculars to the WOLD would bring to the forefront difficult and potentially divisive discussions about the relationship between "creoles" and their "lexifiers." Are the French-derived items in Haitian, for example, to be seen as adoptions, or only the African ones? Jeff Good, the author of the Saramaccan vocabulary in the WOLD, explains how he managed this issue:

One crucial theoretically-driven assumption has been made about the Saramaccan lexicon for the purposes of this project. This is that it represents a continuation of the lexicon of English which branched off from that of standard English varieties at some point before the formation of Saramaccan itself. The theoretical position most closely associated with this assumption is one what can be called the "superstratist" perspective which treats creoles as varieties of their superstrate lexifiers. This basic position is also adopted by the other creole language database that is part of this project, Seselwa. [14]

Good concluded that the percentage of loanwords in Saramaccan is actually 37%, with the major donor languages being: Portuguese (13%), Sranan (12%), Dutch (4%), Loango Banu (3%), Carib (2%) and Gbe (2%). This situates Saramaccan as a high borrower, alongside English, Japanese, and Rumanian. The data on Saramaccan could of course be analyzed differently under a different set of theoretical assumptions.

This last point highlights one of the most valuable traits of the WOLD: it is available online so that researchers may freely access the data themselves and reach their own conclusions. This brings us back to the central argument of this essay, which is to lobby for the inclusion of more Caribbean languages into the Loanword Typology project, possibly

beginning with Papiamento and Haitian (the ones who have achieved official status in their respective countries) and eventually extending to other vernaculars. Though a formal analysis of the data has yet to be carried out, I expect that we will find that Afro-Caribbean vernaculars are not so exceptional after all in terms of their lexical borrowing behavior. As we have seen, some of the world's "major" languages (such as English) can be estimated to have borrowed about half of their lexicon from other languages; the actual percentage may be even higher, but we simply do not have enough information regarding the origin of certain words. Moreover, most if not all of the issues that have been discussed herein can be said to apply equally to the linguistic situation in other speech communities: pluri-lingualism, code-switching, co-existence of prestige vs. non-prestige variants, and so forth. Perhaps the lack of normative reference materials could be considered an obstacle, but even then, it represents an opportunity to look into the nature of languages before they achieve the level of formalization and standardization that we commonly associate with Western European languages. Incorporating more Caribbean vernaculars into the database should therefore be of interest to both creolists and theoretical-comparative-historical linguists at large, not only because it increases the overall size of the comparison pool, but because patterns observed in "creoles" today may be indicative of patterns that actually took place in "non-creoles" at an earlier time in history.

---

1. Frank Martinus Arion, "Creole Identity through Chinese Wall: Affinities between Papiamento and Chinese," in *A Pepper-Pot of Cultures: Aspects of Creolization in the Caribbean*, eds. Gordon Collier and Ulrich Fleischmann (Amsterdam: Editions Rodopi B.V., 2003), 165.

2. Martin Haspelmath, "Lexical borrowing: Concepts and Issues," in *Loanwords in the World's Languages. A Comparative Handbook*, eds. Martin Haspelmath and Uri Tadmor (Berlin, Boston: De Gruyter Mouton, 2009), 35.

3. Ian F. Hancock, "Lexical Expansion in Creole Languages" in *Theoretical Orientations in Creole Studies*, eds. A. Valdman and Andrew Highfield (New York: Academic Press, 1980), 69. Note that his definition of incoining encompasses processes described elsewhere as affixation and/or compounding.

4. John Holm. *An Introduction to Pidgins and Creoles*. (Cambridge University Press, 2000). See section 4.3.5 on reduplication, from where these examples are drawn.

5. There is however, another "creole" language included in the sample: Seselwa, spoken in the Seychelles.

6. Eva M. Eckkrammer, "The Standardisation of Papiamentu: New Trends, Problems and Perspectives," *Bulletin suisse de linguistique appliquée*, 69/1, 1999, 71. The examples provided in the following paragraph are also cited by Eckkrammer in her article.

7. Betty Ratzlaff, *English/Papiamentu – Papiamentu/English Dictionary*, (Bonaire: TWR Dictionary Foundation, 1992). Cited in Eckkrammer, "The Standardisation of Papiamentu," 69-71.

8. http://wold.clld.org/terms#loanword

9. Eckkrammer, "The Standardisation of Papiamentu," 61.

10. *Manual di Gramatica di Papiamento: Morfologia*. Edited by Elvia Solognier-Croes. Aruba: Departamento di Enseñansa, 2010.

11. Sylvia Kouwenberg and Pieter Muysken. "Papiamento." In *Pidgin and Creoles: An Introduction*, edited by Jacques Arends, Pieter Muysken and Norval Smith (Amsterdam: John Benjamins Publishing Company, 1995), 208. Some of the examples presented in this paragraph are also taken from this source.

12. *Manual di Gramatica*, 33.

13. Haspelmath, *Lexical borrowing*, 38.

14. Good, Jeff. "Loanwords in Saramaccan, an English-based Atlantic creole of Suriname." In *Loanwords in the World's Languages. A Comparative Handbook*, edited by Martin Haspelmath and Uri Tadmor (Berlin: De Gruyter Mouton,2009).www.acsu.buffalo.edu/~jcgood/jcgoodLoanwordsInSaramaccan.pdf .

References

Eckkrammer, Eva M. "The Standardisation of Papiamentu: New Trends, Problems and Perspectives." *Bulletin suisse de linguistique appliquée*, 69/1 (1999): 59-74.

Good, Jeff. "Loanwords in Saramaccan, an English-based Atlantic creole of Suriname." In *Loanwords in the World's Languages. A Comparative Handbook*, edited by Martin Haspelmath and Uri Tadmor, 918-943. Berlin: DeGruyter Mouton, 2009. http://www.acsu.buffalo.edu/ ~jcgood/jcgood-LoanwordsInSaramaccan.pdf.

Hancock, Ian F. "Lexical Expansion in Creole Languages." In *Theoretical Orientations in Creole Studies*, edited by Andrew Valdman and Arnold Highfield, 63-88. New York: Academic Press, 1980.

Haspelmath, Martin. "Lexical borrowing: Concepts and Issues." In *Loanwords in the World's Languages. A Comparative Handbook*, edited by Martin Haspelmath and Uri Tadmor, 35-54. Berlin: De Gruyter Mouton, 2009.

Holm, John. *An Introduction to Pidgins and Creoles*. Cambridge: Cambridge University Press, 2000.

Kouwenberg, Sylvia and Pieter Muysken. "Papiamento." In *Pidgin and Creoles: An Introduction*, edited by Jacques Arends, Pieter Muysken and Norval Smith. Amsterdam: John Benjamins Publishing, 1995.

Martinus Arion, Frank. "Creole Identity through Chinese Wall: Affinities between Papiamento and Chinese." In *A Pepper-Pot of Cultures: Aspects of Creolization in the Caribbean*, edited by Gordon Collier and Ulrich Fleischmann, 151-165. Amsterdam: Editions Rodopi B.V., 2003.

Ratzlaff, Becky. *English/Papiamentu – Papiamentu/English Dictionary*. Bonaire: TWR Dictionary Foundation, 1992.

Solognier-Croes, Elvia, ed. *Manual de Gramatica di Papiamento: Morfologia*. Aruba: Departamento di Enseñansa, 2010.

World Loanword Database (http:// http://wold.clld.org)