Lexical contact phenomena: Schuchardt (and beyond?)

Pieter Muysken, Rijksuniversiteit Leiden

In much of the recent language contact literature, the emphasis has been on syntax, particularly in the constraint-oriented tradition established by Poplack (1980). A second emphasis has been discourse, as evident in the work of scholars such as Gumperz (e.g. 1983) and Auer (e.g. 1995). Much less attention has been paid to the lexicon in recent years, in spite of the enormous amount of preparatory work in this area by the founders of modern language contact research, Haugen (1950; 1956) and Weinreich (1953). The reasons for this are particularly the general difficulties linguists encounter when studying the lexicon (cf. e.g. Johanson 1993). (It should be said that in recent years the role of phonology and phonetics in language contact has likewise been neglected; however, that area will need further study at some moment in the future). The purpose of this contribution is two-fold: first to draw attention to the formalisms developed by Hugo Schuchardt to deal with lexical language contact phenomena, and second to critically analyze those formalisms to see where we need to go from there. Since the Neogrammarians explicitly modelled their work on the hard sciences and Schuchardt is generally regarded as one of their main critics at the time, it is ironic that it is in his work that we find one of the first formalisms in linguistics (see Seuren 1998 for a discussion of the general reluctance to use formalisms in linguistics).

The formalisms are tucked away in one of Schuchardt’s longest and most complex writings (the reason probably that they have escaped notice so far), his Kreolische Studien IX, dedicated to Malayo-Portuguese (1890:177-180). In this work he deals with all sorts of lexical aspects of language contact. After a
lengthy historical and sociolinguistic introduction (pp. 1-23), there is a presentation of and commentary on songs (pp. 24-42), followed by prose texts with extensive notes (pp. 42-78). Subsequently there is a list of separate words (pp. 78-90), common expressions (pp. 90-97), narrative (pp. 97-100), a reproduced brief grammatical sketch (pp. 100-102), followed by isolated words found by Schuchardt in the texts (pp. 102-147). Then there is an extensive discussion of the material, which includes the formalisms discussed here.

The formalism developed by Schuchardt can be illustrated with the most simple case, that of “unzweideutige Entlehnung” (unambiguous borrowing), characterised as:

\[
\begin{array}{c}
\text{l} \\
\text{b} \\
= \\
\text{L}
\end{array} \quad \begin{array}{c}
\text{L} \\
\text{B}
\end{array} \quad \text{L.}
\]

This formula should be interpreted as follows: the sound shapes \((l, L < \text{Laut} \text{sound})\) of the lexical items in the base language (“Grundsprache”) and the influencing language (“einwirkende Sprache”) are different \([l](L)\), but their meanings are the same \([b = B]\), and the final result is the borrowing of the sound shape of the word from the influencing language \((L)\). In the original the base language is italic, the influencing language gothic; here I have replaced these by lower case and capital. A full list of the symbols used is given in the appendix.

Notice that formula (1) only covers those cases where we already have a word in the base language available, as in some cases of ‘intimate borrowing’ (Bloomfield 1930). It does not cover the situation where the word is entirely new. The latter situation could be described (in a formula not given by Schuchardt) as:

\[
\begin{array}{c}
\text{-} \\
\text{L} \\
= \\
\text{B}
\end{array} \quad \text{L}^B.
\]

The great merit of Schuchardt’s formalism lies in its treatment of the ‘sound’ and ‘meaning’ components of a word as two separate dimensions, prefiguring the Saussurean conception of the duality of signs.

Schuchardt goes on to sketch six other possibilities of lexical contact. In (3) the meanings are identical, and the sound shapes are similar; the resulting lexical item represents a fusion of the two original sound shapes:
3) \( L b = B \)
\( L \) or \( 1+L \).

This situation characterizes many dialect contact settings, to be sure. Schuchardt gives the example of Asian Portuguese *fula* 'flower', which can be traced back both to Hindi *phul* 'flower' (*phula* 'flourishing') and Portuguese *flor* 'flower' (compare also Brazilian Portuguese *fulor, fulô*).

The mirror image of (3) is (4), where sound shapes are identical but meanings are slightly different:

4) \( L b \sim B \)
\( [Lb (L^b)] \).

The example Schuchardt gives is of Portuguese *lá*, which resembles Malay emphatic *lab* in some of its uses, and is used frequently in Malay Portuguese.

When both sound shape and meaning are similar but not identical all kinds of forms can result, as in (6):

5) \( L b \sim B \)
\( L^b \) or \( l^b \) or \( l+L^b \) or \( l+L^b \).

Again, this is characteristic of a dialect contact situation.

A shift in meaning of an otherwise identical word form is given as:

6) \( L b \sim B \)
\( L^b (l^b) \).

A similar case is:

7) \( L b \sim B \)
\( L^b \) or \( l^b \) or \( l+L^b \) or \( l+L^b \).

In these two cases no examples are given by Schuchardt. However, it is easy to think of cases involving related languages.

Schuchardt then turns to a number of cases in which the meaning changes without there being necessarily a close sound correspondence between the
word forms. In the first two cases, the word has multiple set of meanings in the influencing language, but not in the base language, and these meanings are transferred. In the first instance, the sound forms are still somewhat similar:

\[
\begin{align*}
8) \quad & l \quad = \quad L \\
& b \quad = \quad B_1 \\
& \quad = \quad B_2 \\
& \quad \Rightarrow \quad I^B_2.
\end{align*}
\]

However, in the second case, there is only a partial meaning correspondence:

\[
\begin{align*}
9) \quad & l \quad \Rightarrow \quad L \\
& b \quad = \quad B_1 \\
& \quad = \quad B_2 \\
& \quad \Rightarrow \quad I^B_2.
\end{align*}
\]

This is a specific instance of the more general pattern in (10): semantic transfer without formal correspondence:

\[
\begin{align*}
10) \quad & l \quad \Rightarrow \quad L \\
& b \quad \sim \quad B \\
& \quad \Rightarrow \quad I^B.
\end{align*}
\]

Of the nine logical basic input possibilities, Schuchardt mentions six, disregarding (8) for the moment, since it is a mixed case. Two possibilities he does not mention are:

\[
\begin{align*}
11) \quad & l \quad \Rightarrow \quad L \\
& b \quad \Rightarrow \quad B \\
12) \quad & l \quad = \quad L \\
& b \quad = \quad B
\end{align*}
\]

These are not so interesting. Formula (11) may be taken as an alternative for (2) above to describe unambiguous borrowing. Formula (12) is trivial, reflecting full identity in the two languages.

Formula (10) does merit a more detailed discussion, however. It is quite interesting in indirectly reflecting the case of what was later called relexification (Muysken 1981). Relexification is assumed to involve the importation (to use Haugen's term) of word forms without adopting the original meanings and uses of these words. Massive relexification would lead, for instance, to a language with African, Fongbe, word meanings and structures but French lexical shapes. This has been claimed to be the situation in Haitian (Lefebvre 1998). Formula
(10) is relevant to this scenario because it reflects a situation that could be the outcome of relexification without reflecting the actual process. It would describe Haitian as a form of French that has massively borrowed Fongbe word meanings.

The formula directly describing relexification is (13), but it is not given by Schuchardt.

\[
\begin{align*}
\text{Lb} & \rightarrow \text{B} \\
\text{L} & \rightarrow \text{b} \\
\end{align*}
\]

In (13), the word shapes are distinct, the meanings similar, and a new sound shape is grafted onto an existing meaning.

Given Schuchardt's fascination with creoles, it is somewhat ironic that the contact pattern (also described as "calquing" in much literature) that was the subject of much research and debate a century after Schuchardt's own work does not receive a formalized treatment, although it easily could have been.

Schuchardt is cautious in claiming the possibility of relexification for Malayo-Portuguese, because it only involves meanings, not sound forms. He does not want to invoke it for cases where the meaning change may have taken place independently elsewhere in the Portuguese-speaking world or involves a general creole feature. Thus the TMA markers \textit{lo} 'unrealized' (< Port. \textit{logo} 'soon') and \textit{dja} 'perfective' (< Port. \textit{já} 'already') are not claimed by him to derive from Malay \textit{nanti} and \textit{sudah}, however plausible this might seem semantically, because they resemble TMA markers in other creoles.

In his treatment of semantic shift Schuchardt distinguishes lexical (pp. 182-193) from grammatical (pp. 193-247) features, and it is clear from the fifty-odd page discussion of grammatical features that he takes these extremely seriously. He makes a convincing argument in the closing section of \textit{Kreolische Studien} IX for the claim that Malayo-Portuguese is closely modeled upon Malay grammatical patterns, on the basis of formula (12).

The definition of relexification needs a more complicated treatment than Schuchardt's model implies, and this is where it reaches the end of its usefulness. Both sound and meaning turn out to be too complex to be caught in a simple formula. The reason is that words are entities in themselves, but also function within structured systems (Myers-Scotton and Jake 1995). The extra structure is both phonological and syntactic in nature.

To see what I mean take the case of \textit{dizi-} in Media Lengua (Muysken 1981; 1996), derived from Spanish \textit{decir}'say', but modelled upon Quechua structures. The sound shape of the Media Lengua verb is a compromise between the
original Spanish lexical shape /desi/, the infinitive of which entered Media Lengua in lexicalized form, and the phonotactic requirements of Quechua and Media Lengua itself. The latter include: (a) there are only three vowels /a, i, u/; (b) intervocalic /s/ is sometimes voiced. The latter process is not well understood. However, Media Lengua retained the /d/ of decir, even though Quechua does not have /d/ as a phoneme. Media Lengua has developed several cases of optional reduction of lexical items. Two cases in point are the alternation yuya-ni ~ ya-ni 'I think' (< Q yuya- 'think'), and the alternation dizi- ~ zi- 'say, want' (< Sp decir 'say'). When 'I think' is affixed to a statement, as a parenthetical, the result is ya-ni; otherwise, it is yuya-ni. Interestingly enough, the alternation is absent in Quechua (even though the verb yuya- occurs in Quechua), and with yuya- it is limited to the 1st person. The choice between zi- (73 occurrences in a one-hour corpus) and dizi- (50 occurrences) seems to be determined by considerations of emphasis and lento/allegro style.

To consider the syntax and semantics of dizi- we have to take the Media Lengua verb kiri- 'wish' into consideration as well. In Quechua two verbs exist which express wishing and wanting:

14) muna-ni 
   ni- 
  'wish', 'say' 
  (Q)

These verbs occur in constructions such as:

15) a. papa-da 
     potato-AC 
     muna-ni 
     wish-1 
     'I want potatoes'

b. papa-da 
     ni-ni 
     'I want potatoes'

c. miku-na 
     eat-NOM 
     muna-ni 
     wish-1 
     'I want to eat'

d. miku-sha 
     eat-1FU 
     ni-ni 
     'I say I'll eat', 'I want to eat' 
     say-1

In Quechua both verbs, ni- and muna-, can take NP complements and infinitival complements. In the latter case, ni- selects -sha on the infinitive verb, and muna- selects -na or another marker.
In Media Lengua we find that:

a) Quechua *muna-* has been relexified as Media Lengua *kiri-* ‘want’ (Spanish ‘querer’), and Quechua *ni-* has been relexified as Media Lengua *(di)zi-* ‘wish’, ‘say’ (Spanish ‘decir’);

b) with NP complements only Media Lengua *kiri-* occurs;

c) with infinitival complements *kiri-* often gets -*na* or another nominalizer, and *(di)zi-* often, but not always gets -*sha* complements.

We notice that the relexification process has been only partial here. Whereas in Quechua the verb *ni-* can take NP complements, the corresponding Media Lengua item *(di)zi-* cannot. More details, including more detailed quantitative data, are given in Muysken (1981).

I have introduced a single example from Media Lengua to illustrate the complexities involved in relexification, complexities only partially captured by Schuchardt’s formulas. However, as a first attempt to provide a comprehensive formal typology of lexical *Einwirkung* ‘influence’, the fairly insightful formalism employed remains unsurpassed, testimony to Schuchardt’s genius as the founder of creole and language contact studies.

**Appendix:** The Symbols Used in the Formulas

- \( l \): sound shape of the word in the base language
- \( b \): meaning of the word in the base language
- \( L \): sound shape of the word in the influencing language
- \( B \): meaning of the word in the influencing language
- \( \)\(=(\) = non-identity
- \( ~ \) = similarity (“Aehnlichkeit”)
- \( = \) = identity

**References**


Poplack, Shana 1980: Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL. *Linguistics* 18, 581-618.