Toward a typological profile of Lingua Franca: A view from the lexicon and word formation[1]

Natalie Operstein

ABSTRACT

Although the Mediterranean contact language Lingua Franca (LF) is generally classified as a pidgin, a closer examination of the specialist literature reveals some doubts regarding this categorization. This paper approaches the classification of LF from the viewpoint of its vocabulary structure and word formation processes. The basis for the study is the lexicon of some 2000 words recorded in the anonymous didactic dictionary of 1830 that constitutes the most detailed source of information about LF. The study finds that the LF word formation processes are a detailed continuation of those of its Romance lexifiers. This finding calls for a refinement of our understanding of LF, and with it of our "typology of contact outcomes" (Winford 1997: 3).

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

1.1. Lingua Franca lexicon

Lingua Franca (LF)[2] is a Romance-based contact vernacular that was used for interethnic communication in the Mediterranean area until the second half of the nineteenth century. It appears to have achieved structural stability in the context of the slave societies of the Maghreb between the sixteenth and nineteenth centuries (Cifoletti, 2004; Castellanos, 2007). The documentation of LF comes mostly in the form of brief textual samples given as mere curiosities, or to provide a local color, in pre-scholarly written sources whose genres range from travelers' accounts to former slaves' narratives of captivity, to jocular poems, and to plays set in such locations as North Africa, Venice or the Levant (see Arends, 1998). Against this background, two sources in particular tend to stand out: the monumental *Topographia e historia general de Argel*, published as Haedo (1612) but likely composed several decades earlier (Camamis, 1977), and the slender volume titled *Dictionnaire de la langue franque ou petit mauresque, suivi de quelques dialogues familiers et d'un vocabulaire des mots arabes les plus usuels; à l'usage des Français en Afrique*, published anonymously in 1830 (Anonymous, 1830; henceforth the *Dictionnaire*). The former describes the sociolinguistic setting of LF and provides textual samples totaling about one hundred distinct lexical items (Cifoletti, 1989: 163–164; Cornelissen, 1992). The latter is a learner's dictionary, published in Marseilles and intended for the use of the French in

North Africa. It consists of a 6-page grammar of LF (this section has unnumbered pages), an 82-page French-LF glossary, 6 pages of French-LF dialogues, and a 9-page French-Maghrebi Arabic glossary. The value of these sources is enhanced by the fact that both document the variety or varieties of LF used in Algiers, albeit at a distance of about two and a half centuries.

The profoundly Romance character of LF is revealed by the etymological composition of its lexicon. Cornelissen (1992: 220) estimates that from among the about one hundred distinct lexical items documented by Haedo (1612), 41% derive from Spanish, 17% from Italian, 39% are traceable to multiple Romance sources, and 3% derive from Turkish and Arabic. With respect to the about 2000 LF lexical items in Anonymous (1830), he estimates that 58% derive from Italo-Romance, 6% from Spanish, 4% from French, 27% may have multiple Romance sources, and 3% derive from Arabic; the remaining 2% are composed of Turkish, Portuguese and Catalan words (Cornelissen, 1992: 221). Operstein's investigation (2017a) of the Swadesh list vocabulary in the *Dictionnaire*'s LF reveals only one non-Romance lexical item – Lat. *cinis* 'ash' – in both the 100- and 200-item Swadesh lists.

The non-Romance portion of the *Dictionnaire*'s vocabulary of LF is conspicuous by the absence of verbs and of words of Berber origin (Cifoletti, 1980: 26; Aslanov 2014: 124). The most substantial non-Romance lexical component is Arabic; there is also a small number of Greek, Latin and Turkish words as well as one suspected Anglicism, *flinta* [3] (Cifoletti, 1980: 29–35; 1980: 62–70; 2004: 58–73). In terms of its size, the Arabic component of LF does not exceed those of Sicilian, Spanish or Portuguese, the Romance languages that historically have been in direct contact with Arabic (Aslanov 2014: 124); Aslanov's assessment resonates with Schuchardt's (1909 [1980]: 30) earlier observation that "[t]he Romance vocabulary of Lingua Franca appears to have been enriched by a number of Arabic words, but for comparative purposes the number is probably not greater than that of the Arabic loans in Spanish . . .".[4] The immediate source for some of the Arabic words is not Arabic but rather the Romance lexifiers of LF. This includes such items as *harem* and *magazino*, whose phonological shape points to their acquisition from Italian (< It. *harem*, *magazzino*); as well as *sultan* and *minaret*, likely acquired from French (< Fr. *sultan*, *minaret*) (Schuchardt, 1909 [1980]: 30; Cifoletti, 2004: 53; Aslanov 2014: 126–127). Arabic words are also somewhat marked in the LF phonology: for example, words of Arabic origin that end in a consonant do not acquire a final vowel in LF, whereas consonant-final words (non-verbs) of Romance origin often do (Cifoletti, 2004: 37–38). Some of the Turkish and Arabic ("oriental") words are also singled out in the *Dictionnaire*'s orthography of LF (Operstein, 2017b).

The main Romance lexical components in the *Dictionnaire* are Italian, Spanish and French, in that order. In terms of their diachronic layering, the Spanish component appears to be the earliest of the three, and the French component the most recent (Cornelissen, 1992; Cifoletti, 2004; Operstein, 2017a). The Italian component includes over 94% of words that are compatible in form with standard written Italian; Cornelissen (1992: 221) has identified only about 60 words, or under 6% of the total number of Italianisms, that differ enough from written Italian to be qualified as "informal", "archaic" or "dialectal". The Italo-Romance contributors to the *Dictionnaire*'s LF, other than Italian, include Ligurian, Venetian and Southern Italo-Romance. Other Romance lexical contributors to LF include Catalan, Occitan and Portuguese, with the Portuguese component being the least researched to date (Schuchardt, 1909 [1980]; Coates, 1971; Cifoletti, 1989, 2004; Cornelissen, 1992; Castellanos, 2007).

1.2. Classification of Lingua Franca

In the literature on contact languages, including textbooks, LF is usually categorized as a pidgin (see Foltys, 1984: 1–2; Bakker, 1994: 27; Smith, 1994: 355; Mufwene, 1997: 56; Thomason, 2001: 162; Holm, 2004: 15; Vellupilai, 2015: 151). Pidgins, in turn, are conceptualized as a distinct type in the typology of contact languages (e.g. Thomason, 1997, 2001; Bakker and Matras, 2013), though, as summarized by Parkvall and Bakker (2013: 19ff), attempts at defining this language type satisfactorily have generally been unsuccessful. In their own cross-linguistic typological study of pidgins, the first of its kind in its extent and scope, Parkvall and Bakker distill a small set of linguistic and social criteria which they consider "essential parts of pidginhood" (21) and use them to set pidgins apart from such other linguistic systems as L2 varieties, creoles, and natively spoken languages. The structural criterion in this set merely states that a pidgin "is highly reduced lexically and grammatically compared to its input languages" (22). The criterion of structural reduction forms part of the provisional definition of pidgins that Parkvall and Bakker adopt:

> A pidgin is a language which (a) functions as a **lingua franca**, and which (b) is lexically and structurally **extremely limited** in its communicative possibilities. (Parkvall and Bakker, 2013: 25; the font and emphasis are original)

Among the morphosyntactic characteristics that are "typically absent from pidgins", Parkvall and Bakker include the following:

- In the area of morphology: inflection, derivation,[5] reduplication, infixation, suprafixation, allomorphy, any synthetic structures;

---

[3] Aslanov (2014: 127) derives *flinta* from Fr. *platine* 'platinum'. This etymology seems less likely in view of the absence of the change from *pl- > fl-* in LF *plata* 'silver, metal', *platzà* 'square' and *platzar* 'to place' (< Sp. *plata*, Sp. *plaza*/Fr. *place*, Fr. *placer*).

[4] In quantitative terms, Sayahi (2014: 157–158) mentions the following figures for Iberian Romance: 1188 direct loans from Arabic in Spanish (according to the 2001 *Diccionario de la Real Academia Española*), around 1000 in Portuguese, and around 450 in modern Galician. For Spanish, Lapesa's earlier estimate (1981: 133 fn. 5bis) comprises about 850 direct loans, 780 derivatives and over 1500 toponyms.

[5] But see Crowley's (2008: 84) observation that derivational morphology is better retained than inflectional morphology in pidgins.