Sino-Moroccan citrus: Borrowing as a natural linguistic experiment

Jeffrey Heath

appeared in: Lutz Edzard & Mohammed Nekroumi, eds. 1999. *Tradition and innovation: Norm and deviation in Arabic and Semitic linguistics*, pp. 168-176. Wiesbaden: Harrassowitz.[*]

1. Introduction

In the physical sciences, we study the essential nature of a substance by subjecting it to interventions of our devising. We heat it, freeze it, burn it, bombard it with ions, or pour sulphuric acid on it, then observe and quantify its reaction, and use this information to make inferences about its internal structure prior to the intervention.

In linguistics, true experimental manipulation is largely confined to acoustic phonetics, where we can administer synthesized signals and tightly control subject responses. Most other apparent "experiments" are really just devices for gathering ordinary language data with a particular focus, as when a sociolinguist steers subjects through a sequence of interview components to gather "style" data, or when a syntactician elicits sentences to see how a particular combination of elements is expressed.

Interventions that actually subject grammatical systems to external shocks are harder to imagine. We can, however, think of the historical evolution of languages as consisting of a series of "natural experiments" that we can monitor and assess. For example, if we think of sound change as being driven by grammar-external forces, we can observe how grammars respond to these changes, especially those which (threaten to) disrupt previous structural symmetries or neutralize categorial oppositions.

But language-contact phenomena have an especial attraction. Uriel Weinreich's influential book (1953) applied structuralist principles to this topic. His method was to first submit the phonological systems of two Ianguages (such as Swiss German and Romansh) to analysis, make predictions about how borrowings (in either direction) would be nativized (by phonemic under- and over-differentiation and similar processes), and then test these hypotheses against the field data. Although Weinreich was not responsible for the demographic movements that brought the two populations together, the research strategy can be summarized as the observation of a

"natural experiment". Here, the linguist is closer methodologically to an astronomer or seismologist than to a chemist.

To be sure, linguists have occasionally dreamed of intervening more forcefully. A few decades ago, at the height of creole fever in linguistics, two prominent linguists proposed, in earnest, an experiment whereby monolingual couples from six different nationalities (carefully chosen to get a mixture of word-order patterns and other typological variables) would be paid handsomely to create a Robinson-Crusoe type mini-community on a desert island for several years. The point was to see what kind of creole emerged among their intermingled offspring. Alas, the project was not funded.

Many specialists now believe that so-called creoles, rather than being dramatic new creations, were originally produced by intense, but structurally ordinary language-contact processes involving substrates and lexifying languages. In any event, since modern conditions are not conducive to creole formation and since the on-site documentation of the first stages of historical creoles is very limited, our best bet is to study "ordinary" language contact in cases where we know precisely what the component languages were, preferably in fairly modern times.

Arabic is an outstanding "laboratory" for such study, primarily for two reasons. First, the great Arab military expansion across North Africa, the Levant, Mesopotamia, and elsewhere brought Arabic into long-term contact with numerous substrate or adstrate languages (Berber, Coptic, Aramaic, Persian, Turkish, Romance, etc.), nearly all of which are extant or at least well-described. In the few other instances of comparably rapid and large- scale colonization (the ancient Inda-Europeans, then the Romans), we know much less about the substrates.

Secondly, a borrowing into Arabic involves not only the usual problems of first-order phonological nativization, it also feeds into (and may challenge) a complex system of derivational ablaut. A nominal borrowing usually comes in as a singular stem, but its plural (perhaps) and its diminutive (certainly) are formed by mapping this stem onto an ablaut template that may have to be stretched to accomodate a previously unfamiliar canonical shape or segmental sequence. This can set off a complex phonological and morphological chain reaction, as the following "juicy" example from Moroccan Arabic (hereafter: MA) will show.

2. The initial borrowing

The affricate *c* (IPA [tʃ]) was absent from Classical Arabic but has developed in several dialects (e.g. Iraqi) through a combination of a) adstratum or substratum lexical borrowings and b) Arabic-internal palatalizations of *k or *t. In Morocco, *c* is an established (minor) phoneme in the northern Muslim dialects (e.g. Tangiers, Tetuan), but is essentially absent from mainstream dialects and from the emerging national koiné (which is based on varieties spoken in the central region including Meknes and Fes). The occurrence of *c* in northern dialects is probably due to Latinate/Romance

influence, but whether to locate this influence at the earliest formative period (assuming a Late Latin-speaking substratum population in Tangiers and Volubilis at the time of the initial Arab conquest) or at the time of Andalusi migration / expulsion from Spain to Morocco (mainly 15th and 16th centuries) will have to be discussed elsewhere.

My concern here is rather with a single loanword into MA meaning 'orange (fruit)', the source being Spanish *la China* 'China'. I should note, to begin with, that some MA dialects have a distinct term for 'orange', either *lim* (common in Jewish dialects of the Casablanca-Rabat area and of some towns east and south of Marrakesh), *limun* (widespread in western Muslim dialects in and north of Rabat and Meknes, on much of the Atlantic coast south of Rabat, and in the pre-Saharan oases), and *znbuʕ* (Jewish dialects of Tafilalt in the southeast and of Skoura). Of course, *lim* and *limun* are also Romance (probably Spanish) borrowings. For all 'orange' terms, the unsuffixed form is collective, and adding Feminine Singular *-a* produces a count noun (individuative singular): *lim-a, limun-a, znbuʕ-a*.

Reflexes of *la China* occur in the remaining zones, for example in the northern dialects (Muslim), the Jebli dialects of the western and southern fringe of the Rif (Muslim), the Fes-Sefrou area (Muslim and Jewish), and Marrakesh (Muslim and Jewish). It is reasonable to assume that this Spanish term first took hold in the north and worked its way southward as the citrus industry grew in importance, particularly during the Protectorate period (1912-1956).

The form *la China* posed no particular phonological problem for northern dialects, which already had a *c* phoneme. Moreover, these northern dialects also have a phonetic voiced affricate [dʒ], which functions as the usual pronunciation of geminated /žž/, as in *žbəl* [ʒbəl] 'mountain', but affricated Definite *ž-žbəl* [d(:)ʒbəl]. The closely related Jebli dialects (certainly Chaouen, perhaps also Taounate and the village dialects north of Taza) have similar systems. The usual form for 'orange' in these northern dialects is *ləCCin(-a) (-a* being individuative, as usual), pronounced [lət:ʃin(a)] or [lit:ʃin(a)]. CaC$_x$C$_x$iC(-a) with geminated C$_2$C$_3$ is a typical canonical shape for fruit terms. This helps to account for the gemination of *cc* and for the inclusion of *l* from the Spanish feminine definite marker *la* in the MA stem. But the term then worked its way south into dialects that lacked phonemic affricates, though they had *t-s* sequences at morpheme boundaries : *ka-t-šuf* 'you see', *ma mat-š* 'he did not die'. In these dialects, which also tend strongly to syncopate *ə* (schwa) in nonfinal syllables, speakers had to decide whether to keep the affricate in 'orange' or replace it. One possibility was to keep the same basic pronunciation but reinterpret it, since a phonological representation /ltšin(-a)/ with *tš* instead of geminate *cc* would result in at most a minor pronunciation change (reduced duration of the stop component in the affricate articulation). However, a *tš* cluster is otherwise rare (nonexistent?) in C$_2$C$_3$ position of a CCCVC noun. On the other hand, C$_2$C$_3$ in such nouns is very often a geminate, a pattern that is particularly common for fruit terms, e.g. *bttix(-a)* 'melon' and *tffaḥ(-a)* 'apple'.

The actual forms attested are as follows, in a somewhat surface-y quasi-phonemic transcription, leaving the phonemic interpretation of "c" and "cc" open for the moment. The Definite prefix is shown in parentheses; we omit -*a* suffixes.

		form	dialects (J = Jewish, M = Muslim)
(1)	a.	*(l-)lččin, (l-)ləččin*	J: Taroudant (Sous Valley), rarely in Meknes
			M: north, Jebli, Rabat, Fes-Sefrou, Oujda
	b.	*(l-)lčin*	M: mainly southern and eastern: Marrakesh, Safi (Atlantic), Taza, Oujda, Erfoud (Tafilalt region in southeast)
		(l-)ləčin	M: in the north (as a simplification of ləččin): common in Chaouen but rare in Tangiers
	c.	*(l-)čin*	J: Oujda (northeast)
			M: common in Marrakesh
	d.	*(∅-)čin*	M: Safi (Atlantic)
		(∅-)ččin	J: Taroudant (Sous), as a variant
	e.	*(l-)ečin / (l-)ičin*	M: pre-Saharan oases (Gulmine, Tata, M'hamid)
(2)	a.	*(l-)lššin*	M: rural towns: Souk El Arba (northeast of Rabat), Lemhaya (between Fes and Meknes)
	b.	*(l-)lssin*	J: Essaouira (Atlantic), reflecting sibilant merger in J dialects
(3)	a.	*(l-)lttin*	J: dominant overall (Ouezzane to Marrakesh)
			M: Rissani (Tafilalt)
	b.	*(l-)lttim*	J: Tafilalt (southeast)

Dialects with variants (2) or (3) have resolved the problem of interpreting the status of affricates by transforming the original *čč to either *šš* (which becomes *ss* in sibilant-merging dialects) or *tt*. These replacements are rare in M dialects, but the *tt*-type is predominant among J dialects.

Dialects with forms of type (1), on the other hand, retain some version of the affricated [tʃ] pronunciation. In (la,b) the stem begins with *i*, as in (2-3). The affricate is heard as geminated in (1a), as ungeminated in (1b), though in the absence of other *c*-words in the (1b) dialects we have no yardsticks for distinguishing phonemic *c* from *cc* (or *ts)* and therefore cannot determine the phonological representation from a simple surface transcription. In (1e), most common in Marrakesh-M, the stem begins with the

affricate, the lateral (if present) being interpreted as the Definite prefix. Definite *l-* is normally assimilated fully to a following noun-stem-initial coronal consonant {*d, t, s, ž,* …), so the form *l-Cin* is somewhat anomalous morphophonernically. Variant (1d) resembles (le), but solves the morphophonemic problem by not allowing Definite *l-* to appear; arguably the Definite is phonemic /č-čin/ at least for Safi-M, but I heard no phonetic distinction between this and the unprefixed *čin*. Finally, in (1e) we find a pattern limited to the pre-Saharan oases where the stem begins with a vowel *e* (phonologically /ăy/) or *i*.

3. The borrowing as input to ablaut derivation

The simple forms for 'orange' described in §2 are already somewhat more complex cognitively than the classic cases of interlinguistic borrowing phonology described by Weinreich, since the existence of an abstract pattern CC_xC_xVC with a geminate in C2C3 position seems to have played a significant role in determining the phonological shape. But borrowings into Arabic vernaculars pose further problems for speakers insofar as previously nonexistent stem forms feed into the derivational ablaut system. (A formal model of MA ablaut is given in Heath 1987.) In the case of a noun stem, the relevant derivations are plural and diminutive. For 'orange', ablaut plural does not apply, since (like other fruit terms) it is basically a collective and takes suffixal forms (singular *-a,* plural *-at*) to mark grammatical number. However, these number-marked forms should be able to form diminutives, as do other fruit terms (e.g. *tffaḥ-a* 'apple', diminutive *tfifḥ-a / tfifiḥ-a*). For a CCCVC-a noun, the regular diminutive is CCiCC-a (chiefly in J, northern M, and some central urban M dialects), CCiCiC-a (many southern and rural M dialects), or CCăyCiC-a (Saharan and some pre-Saharan M dialects). In all cases input C_2 and C_3 are separated by an ablaut vowel.

Speakers with *tt* or *šš/ss* in 'orange' (2,3) have no difficulty producing diminutives, e.g. *ltitn-a* in most J dialects and *lšišin-a* in Lemhaya-M. There is also no problem for speakers with affricate-initial stems (1c,d), where the *t* counts as a C_1C_2 cluster, giving diminutive *čiyn-a* (Safi-M, Marrakesh-M) or *čiwn-a* (Marrakesh-M, one attestation). Pre-Saharan speakers with (1e) *ečin-a* have the same type of diminutive, hence *ečăy(y)n-a* (Guelmine-M) or *ečin-a* (Tata-M). Either the *eč* sequence is treated as equivalent to C_1C_2 (Tata?), or the initial *e* is disregarded and does not affect the rest of the output (Guelmine?).

However, other speakers with *čč* or (phonetic) *č* in $C_2(C_3)$ position (1a,b) are forced by the diminutive ablaut to show their hand. The forms in (4) are attested for these speakers. All dialects cited are M, except for Taroudant-J in (4c). Fractions like "2/4" give the number of informants who gave the form divided by the total number of informants who use (1a,b) nondiminutive forms in the community indicated.

        form   dialects

(1)   a.   *lčiwn-a*   M: rare in north (Chaouen 1/5, Tangiers 1/3
                lčiyyn-a   M: Oujda (1/1)

       b.   *lčičn-a*   M: dominant in northern and Jebli regions,
                                       Taza, and Sefrou; recorded once in Meknes
                                       (1/2)
           lčičin-a   M: Fes (2/4), recorded once in the Jebli
                                       village Taounate (1/2)

       c.   *lčitn-a*   M: Fes (2/4)

       d.   *ltitn-a*   M: Meknes (1/2), Tetuan (1/3)

       e.   *ltišin-a*   M: dominant in Marrakesh (4/4)
           ltisn-a   J: Taroudant (Sous Valley)

(4a) is explainable by taking *c* in the input as a unit phoneme, i.e. as C_2 rather than as a C_2C_3 gerninate cluster. This is because a CCVC-a noun stern has an insufficient number of lexical consonants to fill four output C positions, and therefore has a diminutive CCiwC-a or CCiy(y)C-a, with an inserted (nonlexical) semivowel filling output C_3 position. One might expect the distribution of (4a) to match that of (1b) nondirninutive forms, which have no phonetically obvious gemination of the affricate. In fact, (4a) has a very limited distribution, indicating that even speakers with (1b) nondiminutive sterns tend to interpret the affricate as a geminate (or other cluster).

      Diminutive type (4b) treats the affricate as gerninate *čč*. This is regular in the northern dialects and extends into the central cities. The variant *lčičn-a* is slightly awkward, since here speakers must pronounce the second affricate at the beginning of a consonant cluster (with following alveolar nasal). This seems to cause no problems in the northern and (closely related) Jebli dialects, where *c* is a well-established phoneme. It was also elicited from several speakers in the central cities, but here *c* is less well-established and there are indications of resistance to this form. Note that Fes-M speakers were divided between *lčičin-a* (4b), where the extra ablaut vowel in CCiCiC-a keeps the C_3 affricate prevocalic, and *lčitn-a* (4c), where the shorter diminutive shape CCiCC-a puts the second affricate before another C, which forces deaffrication of *c* to *t* (/lčičn-a / → *lčitn-a)*.

      The *lčitn-a* form solves the problem of local pronounceability of the C_3C_4 cluster, but only by sacrificing the normal surface $C_2 = C_3$ identity in diminutives produced by breaking up geminate C_2C_3 clusters. Of course, one could point out that *lčitn-a* is a surface expression of underlying /lčičn-a/, which does respect $C_2 = C_3$ identity. But in MA, there is usually resistance to allowing low-level CC-cluster rules to obliterate an identity relationship between one of these C's and a nonadjacent C. (I will demonstrate this in detail in another paper on diminutives of *byəḍ* 'white'.)

This conundrum seems to be behind diminutives of type (4d). Here it is necessary to recall that all the data in (4) are from speakers who have (1a,b) nondiminutives, so the diminutive *t…t* sequence in (4d) is the result of splitting up input *cc* (not *tt*). It is as though *ltitn-a* (4d) were produced by first converting *lččin-a* into diminutive /lčičn-a/, then making the second CC cluster pronounceable by converting *I*čn/ into *I*tn/ in /1čitn-a/, and then passing this *t* back to the (autosegmentally linked) C_2 to give the final output *ltitn-a,* which is easily pronounceable and also respects $C_2 = C_3$ identity. This heavenly situation is purchased at the expense of a divergence in consonantism between diminutive and nondiminutive forms. This too could be solved by restructuring the nondiminutive form as *lttin* (3a), but since the diminutive is merely a discourse option and can simply be avoided, it would be adventurous to argue that nondiminutive forms were reshaped to resolve an engineering problem with diminutives.

In (4e), the input *c* is analysed as a phonemic nongeminate cluster *tš,* resulting in a *…tiš…* diminutive. This type is most clearly attested in Marrakesh-M (Taroudant is not far away). Note that the relevant Marrakesh-M nondiminutive form is *lčin* (lb) with no obvious extra duration (i.e. gemination) of the affricate, which makes the *tš* analysis attractive.

Conclusion

We conclude that the borrowing from Spanish *la China* presented a series of analytical problems for MA speakers, beginning with the pronunciation and phonological representation of the initial borrowed form and continuing with its interaction with the ablaut apparatus. Even the initial borrowing seems to have been adapted to MA canonical-shape patterns, rather than by a simple segment-by-segment transfer process. After this initial adaptation, the term for 'orange' continued to pose problems in dialects with a geminated affricate pronunciation, particularly where the local diminutive pattern was CCiCC-a. The diminutive of 'orange' forced speakers to resolve (or minimize) tensions between local consonant-cluster rules, maintaining the identity relationship linking the two (secondarily separated) elements of a gerninate cluster, and retaining the consonantism of the non-ablauted input form. Close study of such cases, particularly in the context of a regional dialectological study that reveals the various ways local communities deal with such problems, is one of the great attractions of Arabic linguistics.

References

Heath, Jeffrey. 1987. *Ablaut and ambiguity: Phonology of a Moroccan Arabic dialect.* Albany NY: State University of New York Press.

Weinreich, Uriel. 1953. *Languages in contact.* New York: Linguistic Circle of New York.