



Language contact, borrowing and codeswitching

Stefano Manfredi, Marie-Claude Simeone-Senelle, Mauro Tosco^o

► To cite this version:

Stefano Manfredi, Marie-Claude Simeone-Senelle, Mauro Tosco^o. Language contact, borrowing and codeswitching. Amina Mettouchi; Martine Vanhove; Dominique Caubet. Corpus-based Studies of lesser-described Languages: the CorpAfroAs Corpus of spoken AfroAsiatic languages., 67, John Benjamins, pp.283-308, 2015, Studies in Corpus Linguistics, 9789027203762. hal-01249783

HAL Id: hal-01249783

<https://hal.science/hal-01249783>

Submitted on 2 Aug 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

John Benjamins Publishing Company



This is a contribution from *Corpus-based Studies of Lesser-described Languages*.

The CorpAfroAs corpus of spoken AfroAsiatic languages.

Edited by Amina Mettouchi, Martine Vanhove and Dominique Caubet.

© 2015. John Benjamins Publishing Company

This electronic file may not be altered in any way.

The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible to members (students and staff) only of the author's/s' institute, it is not permitted to post this PDF on the open internet.

For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).

Please contact rights@benjamins.nl or consult our website: www.benjamins.com

Tables of Contents, abstracts and guidelines are available at www.benjamins.com

Language contact, borrowing and codeswitching

Stefano Manfredi, Marie-Claude Simeone-Senelle and
Mauro Tosco

SeDyL (UMR 8202), Inalco, CNRS / LLACAN (UMR 8135), Inalco, CNRS,
PRES Sorbonne Paris-Cité / Università di Torino

Within the larger rubric of language contact we will analyze in this chapter the two phenomena of lexical borrowing and codeswitching as represented in the languages of the CorpAfroAs database. After establishing a theoretical background concerning the difficult distinction between borrowing and codeswitching (§ 1), the study analyzes the semantic, phonological and morphological integration of lexical borrowings in different languages of the corpus (§ 2). The core of the paper (§ 3) focuses on the relation between morphosyntactic and prosodic constraints of codeswitching in CorpAfroAs. Finally, the study argues (§ 4) that, even though syntactic constituency admittedly tells us a great deal about the types of boundaries where speakers are likely to codeswitch, prosodic segmentation plays a pivotal role in the definition of codeswitching. Furthermore, we will show that variation in intonation contours provides a good litmus test for telling the two phenomena of borrowing and codeswitching apart.

1. Theoretical framework

One of the very first issues faced in CorpAfroAs was the presence of “foreign” elements in the languages of the corpus. In the context of our work, the problem mainly centered around the morphosyntactic status of such material — and, therefore, its glossing: should such “aliens” be treated according to the rules and glosses of the recipient language? Or rather according to those of the donor language? Or should they not be glossed at all, marking their “foreign” status through the very absence of glossing? The first solution would imply that the foreign element is, all things considered, not that foreign at all (because a glossing in the target language is possible); the second choice would mean, if taken to its logical conclusion, that two separate morphosyntactic systems co-exist within one and the same language.

This in turn would entail the momentous implication that these systems qualify as mixed languages (along the lines of such well-known and disputed cases as Michif or Media Lengua; cf. Bakker and Matras 2003 for a general discussion). As to the third solution, it looks not only like an easy escape hatch: it could also be taken to imply that foreign material is unanalyzed in the target language — and therefore retains its grammatical independence. An even more basic question then comes to the fore: just how “foreign” must material be to be considered foreign? And who is to assess its foreignness?

Generally speaking, linguists view borrowing (hereafter BORR) and codeswitching (hereafter csw)¹ as forming a continuum, with code-switching providing the means by which new words can be introduced into the recipient language (Heath 1989; Romaine 1989; Myers-Scotton 1992). We consider lexical borrowing as a synonym of loanword (cf. Tadmor 2009) and an integral part of the language in which it occurs, without delving into its etymology and its ancientness. For instance, in English both *fellow* (ultimately from Old Norse *fēlagi* and attested before the 12th century) and *kamikaze* (from Japanese and first attested in English in 1945) are part of the same wide category of lexical borrowings. The distinction between older and newer borrowings, as well as the degree of their adaptation in the recipient language, are therefore largely immaterial, and even when the alien status of such foreign material is morphophonologically apparent and well known to the speaker. Lexical borrowings are therefore simply glossed according to the rules of the recipient language, with the optional addition of the label BORR in the rx tier (see § 2). Obviously, the impact of foreign material may be considerable: for example, 40% of the Kabyle lexicon is thought to be of Arabic origin, and made up of both ancient and recent borrowings. The same is true of Hausa, again with Arabic as the donor language, and possibly other languages of the corpus.

As to csw, it has been variously defined as ‘the juxtaposition within the same speech exchange of passages of speech belonging to two different grammatical systems or subsystems’ (Gumperz 1982: 59), or as ‘the alternative use by bilinguals of two or more languages in the same conversation’ (Milroy & Muysken 1995: 7). Matras, for his part, simply defines csw as ‘the alternation of languages within a conversation’ (Matras 2009: 101). While the alternation of material from different codes is certainly part of any definition of csw, the weakness of Matras’s definition rests on a poorly-defined concept of ‘conversation.’ It could for example encompass the case of a conversation among speakers of closely related languages or

1. We use the abbreviation csw for codeswitching, rather than the more usual CS, because csw is the label for codeswitched material in the rx line of CorpAfroAs. Moreover, CS is actually used as a gloss for the Construct Case in different Semitic and Berber varieties of the corpus.

varieties of the same language where each speaker uses her/his code and a certain degree of reciprocal accommodation is made: if such an accommodation consists, for example, in avoiding words and constructions which are foreign to each other's variant and does not involve the use of the other's variety within one's own utterances, by our definition there is no csw.

Any definition of csw must further take into account its relationship with BORR. The notion of a continuum between BORR and csw is stressed by Matras, who proposes the following multi-dimensional scale (Figure 1):

In Matras's view, there is therefore no theoretical boundary between BORR and csw. There are certainly many borderline cases between csw and BORR, as the following sections will exemplify with data from the corpus. Still, in contrast to Matras, we maintain that a distinction is necessary and possible (on both theoretical and heuristic grounds, as detailed below). The definition of csw as 'linguistic or discourse practice in which elements and items from two or more linguistic systems, or *codes* — be they different languages or varieties of a language — are used in the same language act or interaction' (Mejdell 2005: 414) is inadequate to distinguish it from BORR: the crux of the matter lies, of course, in the use of the English preposition *from*, which can imply either a synchronic or a diachronic transfer. For the same reason, the even broader definition of csw as 'the juxtaposition of

Bilinguality bilingual speaker ↔ monolingual speaker
Composition elaborate utterance/phrase ↔ single lexical item
Functionality special conversational effect, stylistic choice ↔ default expression
Unique referent (specificity) lexical ↔ para-lexical
Operationality core vocabulary ↔ grammatical operations
Regularity single occurrence ↔ regular occurrence
Structural integration not integrated ↔ integrated
codeswitching ↔ borrowing

Figure 1. A bidirectional Codeswitching-Borrowing Continuum (Matras 2009: 111)

elements from two (or more) languages or dialects' (McCormick 1994:581) is also insufficient.

For our part, we propose to define csw broadly as follows:

CSW is the presence of lexical or sentential material belonging to different linguistic systems, provided that its different origin is still transparent in the speaker's output in one or more grammatical domains.

The definition implies that the distinctive feature of csw is the simultaneous presence of two (or more) *codes* (Matras's 'alternation of languages') but it makes no references to a necessary boundary as to the stretch of linguistic material (the sentence, the utterance, or the conversation) encompassed by csw; in order to distinguish csw from BORR, reference is made here to the native or foreign status of the material in the speaker's output, and not to its origin. The speaker's output, in turn, can be assessed on the basis of the morphophonological integration of the foreign material, which is generally considered pivotal in distinguishing csw from BORR: BORR is morphophonologically integrated in the recipient language; csw is *typically* not. The borrowing process can affect the recipient language, causing it to change its phonomorphological rules (i.e., the canonical shape of words). Therefore, we do not make a distinction between integrated and non- (or partially) integrated borrowing. Some degree of integration is a necessary component of BORR, but only a possibility in csw, where the phonology of the switched elements can be influenced by the speaker's native language (see § 3.1.2., § 3.2.).² It must be noted that our definition does not make reference to discourse and sociolinguistic conditions, which are nevertheless crucial in the rise of csw as a social phenomenon.

Switched elements are part and parcel of another language, with which the speaker must be at least partially conversant: while lexical borrowings can be used by monolinguals, csw is always the production of (at least partial) bi- and multilinguals. Of course, bi- and multilingualism is a necessary but not sufficient condition for csw. In our corpus, Beja represents a case of widespread societal bilingualism with Sudanese Arabic and very little csw (s. also § 3.1.1.).

We further assume that, unlike BORR, csw is provisional and determined mainly by pragmatic factors: it is at least in principle the result of a choice. It has social and psychological values, and these values are at least partially shared by

2. It should be also remarked that a former BORR might become the object of csw when contact with the donor language persists or is renewed. For example, in 'Afar of Djibouti, the very toponym Djibouti (whose etymon is contested between Somali, 'Afar and Arabic) is phonologically integrated as /gebo:ti/. When it is pronounced /dʒibu:ti/ it could instead be considered a csw, even if very probably a very common one. The same applies of course to many toponyms, and not only in 'Afar.

the community of speakers: this is the reason why a (at least partially, imperfectly) bilingual community is necessary in order to have csw. The “community” itself may be minimal — consisting just of the participants in a dialogue, provided they share the same language or variety. This is also why csw is, by our definition, a less common phenomenon than BORR.

BORR, of course, does not exclude individual and societal bilingualism — it simply does not require it. Even a minimal degree of exposure to a foreign culture and its language — hardly qualifying as bilingualism — may trigger a significant amount of cultural borrowing. The flood of American loans in Western European languages in the post-WW2 period — although hardly stemming from a “minimal degree of cultural exposure” — is a case in point, as it certainly was not triggered by massive bilingualism.

This is also evident in the case of the only pidgin language in our corpus, i.e., Juba Arabic (and this notwithstanding that for a part of the community of speakers Juba Arabic is a creole (Tosco and Manfredi 2013)): the speakers of a pidgin are by definition (first, or native) speakers of “something else.” In our case, the Juba Arabic corpus was collected from speakers who are trilingual in Juba Arabic, Sudanese Arabic, and their ethnic, and presumably native, Nilotic language. The case of Arabic “dialects” in the corpus is similar but not identical, because Arabic diglossia is a matter of functional register. Still, we agree with Mejdell (2005) in considering Modern Standard Arabic as a separate language and it therefore follows that diglossia in Arabic dialects is an instance of csw.

Table 1. Borrowing and codeswitching in CorpAfroAs

Language	(Mainly) borrowing from	(Mainly) switching towards
(Djibouti) 'Afar	Arabic, French	French
Beja	Arabic	Arabic
Gawwada	Amharic	Amharic
Hausa	Arabic, English	English, Arabic
Hebrew	English, Arabic	–
Juba Arabic	English, Bari	English, Arabic
Kabyle	Arabic	French, Arabic
Moroccan Arabic	French, Spanish, English	French, Modern Standard Arabic, Spanish
Tripoli Arabic	Italian, Turkish, English	Modern Standard Arabic
Ts'amakko	Amharic, Hamar	Hamar
Wolaytta	Amharic	–
Zaar	Hausa, Arabic	Hausa, English

Table 1 shows the main donor of (at least recent) borrowings and the main target language in CSW for each language of the CorpAfroAs corpus. In this context, early lexical borrowing in ancient layers of the languages is largely disregarded.

The following sections will show how different cases of CSW and BORR were handled in the languages of the corpus.

2. Lexical borrowing

In CorpAfroAs as a whole there are very few lexical items glossed as BORR and in some of the corpora (e.g. Gawwada, Tsamakko, Wolaytta, Tamasheq and Tripoli Arabic), none has been singled out as such. This is not because these languages never integrated ‘foreign’ lexical elements, but rather because we decided that, being a diachronic phenomenon, lexical borrowing did not represent a retrieval priority and we thus chose to *optionally* mark it by means of the label BORR. Furthermore, the identification of a borrowed item always depends on both structural and socio-historical assessments. With regard to this, it should be remarked that all the languages of CorpAfroAs belong to the same genetic phylum, and that some of them have been in contact with each other for centuries. Given this overall situation, it may be very difficult to evaluate the degree of morphophonological integration of a given lexical item from one Afro-Asiatic language to another, and therefore to identify it as BORR.

In the Kabyle corpus, it is noteworthy that the verb *ruh* ‘leave’, which is borrowed from the Arabic imperfective paradigm **i-rūh*, is largely used by monolingual speakers and has the same paradigm as any Berber verb. That being so, it can be considered as an instance of lexical borrowing. However, it is never marked as BORR because it belongs to an ancient layer of the recipient language. In contrast, the Arabic vernacular phrase *la bās* ‘all right’, which has been integrated in Kabyle as a verb *labas* ‘be in good health’ that can be normally inflected with personal affixes (see KAB_AM_NARR_01_994), is overtly marked as BORR due to its relatively recent integration into the recipient language. This means that, independently from the degree of morphophonological integration, the criteria for labelling borrowings as such mainly depended on the linguist’s evaluation of the different socio-historical dynamics related to their integration.

The vast majority of borrowed lexical items in CorpAfroAs are nouns. In point of fact, nouns are more easily borrowed than any other word class (Haspelmath 2008: 50). With regard to this, Matras (2009: 150) has observed that “*the high borrowability of nouns is a product of their referential functions since nouns cover the most differentiated domains for labelling concepts, objects, and roles*”. These nominal lexical fillers are also called ‘non-core borrowings’ (Myers-Scotton 2002: 239)

This non-core borrowing presumably finds its etymology in the Arabic verb **xajāt*.⁴ The word is completely integrated into the Beja consonant system in which both the voiced velar fricative /x/ and the voiceless dental emphatic /t/ are absent. As a further matter the borrowed item is incorporated with a long vowel [i:], which is absent in the original Arabic verb.

The integration of borrowings passes also through their assimilation into the derivation system of the recipient language. In Hebrew, as well as in other Semitic languages, a borrowed item can be subjected to the intra-categorical derivation for expressing different meanings from a same root. For example, the borrowed root *filosof* shapes a feminine noun related to the abstract notion of ‘philosophy’ if followed by the gender marker *-ja* as in example (2).

- (2) tx giliti sefilosofja zelo kazenoša //
- | | | | | | | | |
|-----|--------------|----------|-------------|------------|-----------------|-------------|---------|
| mot | giliti | | sefilosofja | | kaze | nosa | // |
| mb | gili | -ti | fel= | filosof | -ja | kaze | nosa // |
| ge | discover\PFV | -SBJ.1SG | COMP= | philosophy | -F.SG like_this | terrible | // |
| rx | v | -PNG | CONJ= | N.BORR | -G | PRO.ADJ ADJ | // |
- ft I discovered that philosophy is not such a terrible thing.
- HEB IM CONV 3 SP1 317

However, the same borrowed root can also express the concrete meaning of ‘philosopher’ if marked by the plural masculine suffix *-im* as in example (3).

- (3) tx jefam exad seafilosofim seomeṣ se /
 mot jefam exad
 mb sel= sam exad
 ge be there one
 rx EXS ADV.DEICT NUM
 mot seafilosofim seomeṣ se /
 mb sel= ha= filosof -im se= omeṣ se /
 ge COMP= DEF= philosophy -PL.M NMLZ= say\ACT.PTCP NMLZ /
 rx CONJ= DET= N.BORR -PNG REL= v1\TAM REL /
 ft There is also one of the philosophers that says that
 HEB IM CONV 1 SP1 229

CorpAfroAs also displays a minor number of ‘core borrowings’, i.e. ‘*words that duplicate elements that the recipient language already has in its word store*’ (Myers-Scotton 2002: 240). Contrary to non-core borrowings, these lexical items are systematically marked as BORR in CorpAfroAs. For instance, Moroccan Arabic has

4. Beja people are originally camel herders and dressmaking is not a part of their traditional activities. For this reason, the Beja language does not possess a verb for expressing the meaning of 'sew'.

borrowed the word *semana* ‘week’ from Spanish, which is used nowadays alongside Arabic **(u)sbu:r* also as a result of the influence of Modern Standard Arabic.

- (4) tx *fi se:ma:na i:jəh fi se:ma:na laʔ //*
 mot *fi si:ma:na i:jəh fi si:ma:na laʔ //*
 mb *fi si:ma:n -a i:jəh fi si:ma:n -a laʔ //*
 ge INDF2 week -F yes INDF2 week -F no //
 rx DET N.BORR.SPA -PNG ADV DET N.BORR.SPA -PNG ADV //
 ft Every other week.
 ARY_AV_NARR_1_214

Irrespective of the semantic reference expressed by *si:ma:na* in Moroccan Arabic, the examples above show that the original term is also remodelled on the phonology and morphosyntax of the recipient language. More particularly, the Spanish vowel /e/, absent in Moroccan Arabic, is replaced by /i/ (even if it is phonetically realized as [e]). Furthermore, the syllabic structure of the borrowed noun varies from the original Spanish (CvCvCv > Cv:Cv:Cv). The borrowed noun follows the agreement rules of Moroccan Arabic: the noun is integrated as feminine because of the presence of a final *-a*, and it is determined by the indefinite article *fi*. In example (5), the same borrowed item is modified by the definite article *əl-*, which is assimilated with the first alveolar consonant (*s-*), as in Arabic.

- (5) tx *ba:f məlli ka:jəmfi di:ksse:mana /*
 mot *ba:f məlli ka:i:msfi*
 mb *ba:f məlli ka:= j- mfi:*
 ge in_order_to when REAL= 3- go\IPFV
 rx CONJ ADV TAM= PNG V
 mot *di:kssi:ma:na*
 mb *di:k= əl= si:ma:n -a /*
 ge DEM.DIST= DEF= week -F /
 rx DET.M/F= DET= N.BORR.SPA -PNG /
 ft In order to go that week.
 ARY_AV_NARR_3_240

As a final remark, it is noteworthy that CorpAfroAs also displays some instances of verb borrowing. The relatively low number of borrowed verbs in the corpus finds good parallels crosslinguistically: in the World Loanword Database,⁵ over 31% of all nouns are loanwords, while less than half this figure (14%) are verbs, and only two out of 41 languages in the sample have a smaller proportion of loan nouns than loan verbs (Tadmor 2009:61). Structural constraints certainly play a role in

5. World Loanword Database, <<http://wold.livingsources.org/>> (22 August 2013).

disfavoring the borrowing of verbs crosslinguistically (isolating languages, on the contrary, can apparently borrow verbs quite easily). The difficulty of integrating verbs into the complex inflectional/derivational system of Afro-Asiatic languages would thus explain both the small number of borrowed verbs in our corpus and the fact that the few borrowed verbs are usually integrated as bare forms to which personal indexes of the recipient language are added. The high degree of morpho-syntactic integration of borrowed verbs does not allow us to differentiate them from any other verb of the recipient language. In 'Afar, this is the case of the verb *jitrihe:nih* in example (6).

- (6) *tx badak addal jitrihe:nih haban ti: //*
mot badak addal jitrihe:nih
mb bada -k adda -l j- itrihe -eni -h
ge sea -at depth -in 3- drop\PFV -PL -COOR
rx N.M -POSTP N.F -POSTP PNG- V1.TAM -PNG -POSTP
mot haban ti: //
mb haba -n ti //
ge let\IPFV -3PL thing //
rx V2.TAM -PNG N //
ft Something that is let and dropped offshore.
 AAR_MCSS_NARR_1_

The 'Afar verb root *itrihe* 'drop' is etymologically related to the Arabic derived verb form **aṭraḥ* 'put'. In the previous example, it is noteworthy that /t/ is the reflex of the Arabic emphatic dental plosive **/ṭ/*, which is absent in 'Afar. Morphologically speaking, the borrowed item belongs to a verbal sub-class in which person and gender markers are prefixed (and number indexes are suffixed). As with any 'Afar verb, aspect is marked by vowel apophony: the open vowel *a* marks the Imperfective while close vowels signal the Perfective. Here, *itrihe* is used with the meaning *drop*, but its use is limited to the semantic sphere related to fishing techniques ('Afar already possesses a verb to express the notion of 'throw, drop').⁶ Apart from some 'Afar/Arabic bilingual fishermen the word is not recognized as a foreign lexeme.

Similarly, Beja displays several instances of verb borrowing from Arabic (see also example 1). In example (7), the verb *a:mal* 'do' is borrowed from Arabic **ʕamal*. This verb is phonologically well integrated, as shown by the elision of the initial pharyngeal consonant ʕ that has given rise to a long open vowel *a:* in Beja. In morphosyntactic terms, the verb also takes the same TAM and agreement markers as any Beja verb.

6. In 'Afar, as it was previously noticed, the technical vocabulary concerning the sea-life (names of boats, material, etc.) is borrowed from English or French. On the contrary, the vocabulary related to fishing techniques and fish names is borrowed from Arabic.

- (7) *tx* *o:dhe:j da:jib a:malin ako: /*
 mot *o:dhe:j* *da:jib* *a:mali:na*
 mb *o:=* *dhe:j* *da:ji* *=b* *a:mal* *-i:na*
 ge DEF.SG.M.ACC= people good =INDEF.M.ACC do -AOR.3PL
 rx DET= N.M ADJ =DET V2.BORR.ARA -TAM.PNG
 mot *ako:* /
 mb *ako:* /
 ge *be\CVB.SMLT /*
 rx *PTCL* /
 ft Since the people have good relationships with him.
 BEJ_MV_NARR_2_farmer_308

3. Codeswitching

In this section we aim at presenting a syntactic and prosodic overview of csw in CorpAfroAs. Broadly speaking, the annotation system allows a linear syntactic analysis of csw, at the same time as prosodic segmentation enables investigating the prosody-syntax interface of switched clauses. As to the syntactic grammaticality of csw, several studies (Poplack 1980; Sankoff & Poplack 1981; Muysken 1995) have shown that there are preferential syntactic sites for language switch. Given that csw is, among other things, a discourse phenomenon, we hold the view that its investigation needs to take prosody into account in order to explain the occurrence of switched utterances in natural discourse. We will thus follow the analytic approach of Shenk (2006) and Mettouchi (2008), who integrated the well-known (although disputed) opposition between intersentential and intrasentential csw with a prosodic investigation based on the distinction between monolingual and bilingual Intonation Units (§ 4.1). It has also been demonstrated that csw serves as a contextualization cue that tends to build up contrasts in discursive contexts (Auer 1998). In this regard, Gumperz (1982:98) has already noted that ‘*codeswitching signals contextual information equivalent to what in monolingual settings is conveyed through prosody or other syntactic and lexical processes*’. Zentella (1997:96), for her part, affirms that ‘*what monolinguals accomplish by repeating louder and/or slower, or with a change of wording, bilinguals can accomplish by switching languages*’. Bearing this in mind, we will also analyze the intonation contours of intrasentential csw and compare them with those related to the presence of lexical borrowings (§ 3.2).

3.1 Prosodic segmentation and codeswitching

3.1.1 Intersentential codeswitching

Intersentential csw involves a switch at a clause or sentence boundary, where each clause or sentence is either in one or the other language(s) (Romaine 1989). Intersentential csw represents by far the minority of csw types in CorpAfroAs. This is probably because the production of an entire sentence in a ‘foreign’ language requires a relatively high degree of bilingual proficiency, and in CorpAfroAs this holds true for both narrative and conversational texts. It is not a coincidence that intersentential csw mainly concerns Moroccan Arabic / French bilingual speakers who have been much more exposed to the use of the former colonial language than other linguistic communities represented in CorpAfroAs.

In syntactic terms, we can define any instance of intersentential csw as a syntactically coherent clause that lacks syntactic obligatoriness with regard to the preceding sentence. We agree with Watson, Breen and Gibson (2006: 1047) that syntactic obligatoriness is a better predictor than semantic closeness for the definition of prosodic boundaries. It follows that intersentential csw is prosodically isolated and systematically yields monolingual Intonation Units. Example (9) shows a clear instance of Moroccan Arabic / French intersentential csw in which the embedded clause contains the French idiomatic sentence ‘qui se ressemble, s’assemble’ [ki sə Rəsābl sassābl] (i.e., ‘birds of a feather flock together’). The embedded clause is separated from the previous Moroccan Arabic Wh-question by a major prosodic boundary and covers two French monolingual Intonation Units.⁷

- (9) tx mʃa mən ʃəbbəhtək // < ki sə Rəsābl > / < sassābl > //
- mot mʃa mən ʃəbbəhtək //
- mb mʃa mən ʃəbbəh -t =ək //
- ge with who compare\PFV -1SG =OBJ.2SG.M //
- rx PREP PRO V.TAM -PNG =PRO.PNG //
- ft With whom did I compare you?
- mot < qui se ressemble > / < s’assemble > //
- mb < qui se ressemble > / < s’assemble > //
- ge < qui se ressemble > / < s’assemble > //
- rx < CSW.FRA CSW.FRA CSW.FRA > / < CSW.FRA > //
- ft < birds of a feather > < flock together > .
- ARY_DC_NARR_04_59–61

7. Differently from borrowed items, csw is not glossed in the \ge tier. On the contrary, csw is highlighted by means of chevrons < ... >. With regard to intrasentential csw (§ 3.1.2.), it should be remarked that chevrons do not signal the embedded language as against the matrix language, but they rather indicate the presence of ‘foreign’ elements with regard to the sample corpus. See the online Manual for further information about the annotation of csw in CorpAfroAs.

Example 9 illustrates that intersentential csw may embed a complete verbal phrase. However, the majority of syntactically coherent switched clauses in CorpAfroAs include just a single self-standing constituent. Example (10) shows an instance of Beja / Arabic intersentential csw taken from a narrative text in which the narrator is reproducing a conversation between two speakers. The switch towards Arabic is the reaction of one of the speakers to a preceding Beja exhortative form. More specifically, the embedded clause shows an emphatic reiteration of the Arabic self-standing adverb *abadan* 'never' and is prosodically produced as a monolingual Intonation Unit with a major prosodic boundary. However, the narrative register used by the speaker might have favoured this prosodic segmentation of the switched material.

- (10) tx o:n wʔo:r nija:wa // < *abadan abadan* > //
- mot o:n wʔo:r nija:waj //
- mb o:n w= ʔo:r ni- jaw -aj //
- ge PROX.SG.M.ACC DEF.SG.M= child 1PL- give\IPFV -so //
- rx DEM DET= N PNG- v1.IRG -PTCL //
- ft "So, let's give it to that boy!"
- mot < *abadan abadan* > //
- mb < *abadan abadan* > //
- ge < *abadan abadan* > //
- rx < CSW.ARA CSW.ARA > //
- ft < *abadan abadan* > //
- ft < "it's out of the question" >.
- BEJ_MV_NARR_10_rabbit_59–61

Other examples of syntactically self-standing constituents occurring as intersentential csw are related to the use of Arabic formulaic expressions by Hausa speakers. In example (11), the embedded clause corresponds to the Arabic religious formula *alḥamdu lillāhi* 'God be praised' that, being integrated into the Hausa segmental and tonal system, is realized as [àlhamdùlilla:hì]. Also in this case, intersentential csw covers a monolingual Intonation Unit with a major prosodic boundary.⁸

- (11) tx sun dāuki wani za:màni: / sun dɔ:ʔo: wà ʔà:juwansù //
- < àlhamdùlilla:hì > //
- mot sun dāuki wani za:màni: /

8. In this connection, it should be stressed that the same Arabic religious formulas are considered as genuine borrowings in other languages (i.e. Kabyle, Beja). The different treatment of these complex items mainly depends on the different degree of bilingualism characterizing the Muslim communities represented in CorpAfroAs.

<i>mb</i>	<i>sun</i>	<i>dàuka: -i</i>	<i>wani</i>	<i>za:màni:</i>	/
<i>ge</i>	3PL.PFV.FOC	take	-ACC1	some.M	present_time /
<i>rx</i>	PNG.TAM	v2	-ACC1	DET.INDF	N /
<i>ft</i>	They took the period when they lived				
<i>mot</i>	<i>sun</i>	<i>dauɔ: wà</i>	<i>ɾà:juwansù</i>	//	
<i>mb</i>	<i>sun</i>	<i>dauɔ wà</i>	<i>ɾà:juw:</i>	<i>-n</i>	<i>-sù //</i>
<i>ge</i>	3PL.PFV.	tie.DIR	DAT	existence	-GEN L.GEN//
<i>rx</i>	PNG.TAM	v6	PTCL.SYNT	N	-SYNT -PNG //
<i>ft</i>	and they created their own destiny				
<i>mot</i>	< alḥamdu liḷlahi > //				
<i>mb</i>	< alḥamdu liḷlahi > //				
<i>ge</i>	< alḥamdu liḷlahi > //				
<i>rx</i>	< CSW.ARA > //				
<i>ft</i>	< God be praised >.				
HAU BC CONV 4 SP2 328–330					

3.1.2 *Intrasentential codeswitching*

Intrasentential csw involves a switch within the clause or sentence boundary and also includes switching within word boundaries (Romaine 1989). Intrasentential csw indubitably represents the most common type of csw within CorpAfroAs. The syntactic standpoint we use for the analysis of intrasentential csw is that of the Matrix Language Frame proposed by Myers-Scotton (1993, 2001), in which the matrix language (ML) is defined as the language from which the greatest number of morphemes is drawn, while the embedded language(s) (EL) refers to the other language(s) used in the conversation. The underlying assumption is that ML sets the frame for the different types of morphosyntactic constituents occurring in intrasentential csw.⁹ As far as CorpAfroAs is concerned, intrasentential csw is mainly enacted through the insertion of single, high-frequency lexical items from the embedded into the matrix language (a phenomenon also referred to as tag-switching; cf. Caron 2002).

In contrast to intersentential csw, intrasentential csw embeds obligatory constituents that would be syntactically incoherent if placed outside the sentence in which the language switch occurs. Therefore, if we accept the idea that syntactic obligatoriness is an operational predicator for the occurrence of prosodic boundaries (§ 3.1.1.), intrasentential csw is more likely to give rise to bilingual Intonation Units. Moreover, in line with Mettouchi's (2008:185) observation for

9. According to Myers-Scotton (2001), there are three types of constituents in sentences showing intersentential csw: (1) ML+EL constituents that show morphemes from the two or more participating languages. (2) ML islands that are composed entirely of ML morphemes. (3) EL islands that are instead composed entirely of EL morphemes.

Berber / French csw, we note that in CorpAfroAs too intrasentential language switch tends to occur at the boundaries of bilingual Intonation Units. Example (12) shows two instances of intrasentential csw with Moroccan Arabic as ML and French as EL. In this case, the switches take place within a complex sentence covering four Intonation Units. Both the EL constituents *les thèmes* [le tɛ:m] and *les morceaux* [le moʁso:] occur at the end of two minor Moroccan Arabic / French bilingual Intonation Units.

- (12) tx nna:s ʕəʒbu:hum < le tɛ:m > / ʕəʒbu:hum < le moʁso: > /
 u ʕəʒba:thum əlhəðɾa / ʕəʒba:thum əlmusi:qa //
- mot nna:s ʕəʒbu:hum < les themes > /
 mb əl= na:s ʕəʒb -u =hum < les themes > /
 ge DEF= people please\PFV -3PL =OBJ.3PL < les themes > /
 rx DET= N.PL V.TAM -PNG =PRO.PNG < CSW.FRA CSW.FRA > /
 ft People appreciated < the themes > ,
- mot ʕəʒbu:hum < les morceaux > /
 mb ʕəʒb -u =hum < les morceaux > /
 ge please\PFV -3PL =OBJ.3PL < les morceaux > /
 rx V.TAM -PNG =PRO.PNG < les morceaux > /
 ft they liked < the tracks > ,
- mot u ʕəʒba:thum əlhəðɾa /
 mb w ʕəʒb -at =hum əl= həðɾ -a /
 ge and please\PFV -3SG.F =OBJ.3PL DEF= discourse -F /
 rx CONJ V.TAM -PNG =PRO.PNG DET= N.F -PNG /
 ft and they liked what we had to say ,
- mot ʕəʒba:thum əlmusi:qa //
- mb ʕəʒb -at =hum əl= musi:q -a //
- ge please\PFV -3SG.F =OBJ.3PL DEF= music -F //
- rx V.TAM -PNG =PRO.PNG DET= N.F -PNG //
- ft they liked the music.
- ARY_DC_NARR_2_11–14

The same correspondence between intrasentential csw and prosodic boundaries can be observed in example (13), where Hausa is the ML and English the EL. In this case, the EL constituent *later*, which is phonologically integrated as [letti], occurs at the boundary of a bilingual Hausa / English Intonation Unit. However, this occurrence might also have been favoured by the syntactic overlap between Hausa and English, in which adverbs of time usually appear sentence finally.

- (13) tx àkwai wata ɾa:na: / 1277 mukà jiwo: <letti> //
- mot àkwai wata ɾa:na: / 1277
- mb àkwai wata ɾa:na: / 1277

ge COP3 some.F day / 1277
rx PTCL.SYNT DET.IND N / 1277
ft One day, 1277
mot *mukà* *jiwo:* < late > //
mb *mukà* *jiwo:* < late > //
ge 1PL.PFV.FOC do.DIR < late > //
rx PNG.TAM V6 < CSW.ENG > //
ft we were < late >.
 HAU_BC_NARR_3_SP2_2-4

It should be remarked that the tendency for intrasentential language switch to correspond to the boundaries of bilingual Intonation Units is not always necessarily realized because of the intervention of discourse and/or pragmatic factors. Actually, intrasentential csw can also occur within monolingual Intonation Units. This is typically the case of embedded discourse markers, that typically occur as prosodically independent words. Example (14) shows an instance of Juba Arabic / English intrasentential csw due to the insertion of the English discourse marker *so*, which alone constitutes a minor English monolingual Intonation Unit.

- (14) *tx* < so > / *min* *henák* //
mot < so > / *min* *henák* //
mb < so > / *min* *henák* //
ge < so > / from there //
rx < CSW.ENG > / PREP ADV.LOC //
ft < So > after that.
 PGA_SM_NARR_2_SP1_505-506

Like discourse markers, hesitation too can cause intrasentential csw to correlate with monolingual Intonation Units. Example (15) shows another instance of intrasentential csw with Juba Arabic as ML and English as EL. In this case, hesitation is signalled by vowel extra-lengthening on the preposition *le* [le:::] occurring at the boundary of a Juba Arabic monolingual minor Intonation Unit. The embedded clause which follows is made up of the English nominal phrase *displaced schools* [disple:si sku:l] and covers an Intonation Unit ending with a major prosodic boundary.

- (15) *tx* *medrése de tában tábe le:::* / < *disple:si sku:l* > //
mot *medrésa de tában tábe le* /
mb *medrésa de tában tábe le* /
ge school PROX.SG obviously belong to /
rx N PRO.DEM ADV V PREP /
ft This school belongs to

mot < *displaced schools* > //
mb < *displaced schools* > //
ge < *displaced schools* > //
rx < CSW.ENG CSW.ENG > //
ft < the displaced schools >.
 PGA_SM_NARR_1_39–40

Example (16) displays a more complex switch, in which English represents the ML and Juba Arabic the EL. The first Intonation Unit encloses both ML and EL constituents, the latter conforming to the general tendency of intrasentential csw to occur at prosodic boundaries. Here, hesitation is signalled by the Juba Arabic discourse marker *jáni* ‘that is to say’ which makes up a separate Intonation Unit. As for the predicative adjective *green* [gri:l] which follows, it covers an English monolingual Intonation Unit with a major prosodic boundary.

- (16) *tx* *u* < *iz olmos* > / *jáni* / < *gri:l* > //
- mot* *úo* < *is almost* > / *jáni* /
mb *úo* < *is almost* > / *jáni* /
ge 3SG < *is almost* > / *that_is_to_say* /
rx SBJ.PRO.IDP < CSW.ENG CSW.ENG > / DM /
ft It < *is almost* >, I mean,
mot < *green* > //
mb < *green* > //
ge < *green* > //
rx < CSW.ENG > //
ft < *green* >.
 PGA_SM_NARR_2_SP1_119–120

In contrast with example (16), in which prosodic segmentation is induced by the presence of the discourse marker *jáni* ‘that is to say’, if intrasentential csw covers more than one Intonation Unit, monolingual Intonation Units with a major prosodic boundary tend to start in the same language as the one in which the preceding minor bilingual unit ends.¹⁰ Such prosodic constraints are evident in example (17) in which French is the ML and Moroccan Arabic the EL. The first minor bilingual Intonation Unit ends in French. This is followed by a French monolingual Intonation Unit ending with a major prosodic boundary, signalled in turn by a pause and an intake of breath.

10. Mettouchi (2008:187) observed instead that ‘the tendency is for (bilingual and monolingual) intonation units to start consistently in the same language as the beginning of the preceding one, with occasional switches that are pragmatically motivated’.

- (17) tx wa:lɑ:kin < ila fe boku dʒo:z e il fo ǎ > / < ǎ minimō dǎ rǎkonesǎ:s > //
 BI_278 lɑ:kʃi lli dɑ:r //
- | | | | | | | |
|-----|-----------------------------|----------------|-----------------|---------|---------------|-----------------|
| mot | wa:lɑ:kin | < il | a | fait | beaucoup de | choses |
| mb | wa:lɑ:kin | < il | a | fait | beaucoup de | choses |
| ge | but | < il | a | fait | beaucoup de | choses |
| rx | CONJ | < CSW.FRA | CSW.FRA | CSW.FRA | CSW.FRA | CSW.FRA CSW.FRA |
| ft | But < he has done a lot | | | | | |
| mot | et | il | faut | un | > / < un | minimum |
| mb | et | il | faut | un | > / < un | minimum |
| ge | et | il | faut | un | > / < un | minimum |
| rx | CSW.FRA | CSW.FRA | CSW.FRA | CSW.FRA | > / < CSW.FRA | CSW.FRA |
| ft | and you need > < a minimum | | | | | |
| mot | de | reconnaissance | > // BI_278 l= | dɑ:k | ǎl= | |
| mb | de | reconnaissance | > // BI_278 l= | dɑ:k | ǎl= | |
| ge | de | reconnaissance | > // BI_278 to= | DIST.M | DEF= | |
| rx | CSW.FRA | CSW.FRA | > // BI_278 | PREP= | DEM | DET= |
| ft | of recognition > BI_278 | | | | | |
| mot | fɪ | lli | dɑ:r | // | | |
| mb | fɪ | lli | dɑ:r | // | | |
| ge | thing that do\PFV.2.SG.M // | | | | | |
| rx | N.M | REL | V.TAM | // | | |
| ft | he has done. | | | | | |
- ARY DC NARR 5 57–60

Lastly, we should also consider the sporadic presence of intra-word csw, in which the switch does not only occur within clause boundaries, but also within word boundaries. In view of this, the occurrence of intra-word csw does not systematically coincide with prosodic boundaries, since it can appear everywhere within a bilingual Intonation Unit. However, the switched component is generally signalled by an emphatic pitch rising (§ 3.2.). As example (18) shows, in CorpAfroAs intra-word csw is mainly realized through the insertion of lexical verbs from the EL (i.e., the English verb *work* [wo:k]), to which TAM markers and/or personal indexes from the ML are affixed (i.e., the Juba Arabic irrealis preverbal marker *bi=*).¹¹ This represents an instance of language switch rather than an integrated verb borrowing because ‘to work’ is regularly expressed by the verb *istákal* in Juba Arabic (see for example PGA_SM_CONV_1_SP2_446).

11. It should be remarked that, because of retrieving reasons, all the phonological words interested by intra-word csw are enclosed between chevrons (including affixes and clitics), but only the switched component is marked by the label csw in the /rx tier.

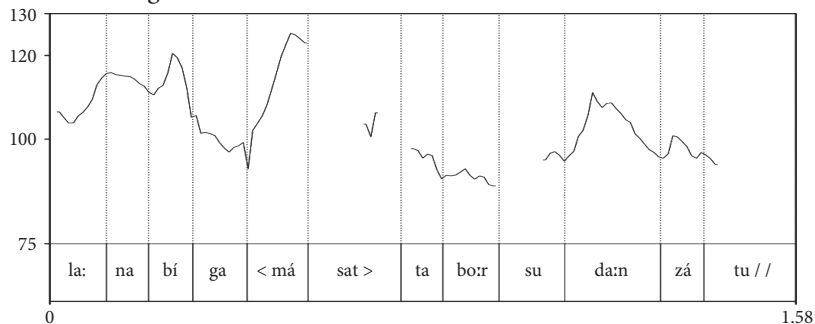
- (18) *tx éta kan ma < biwo:k > de éta ma:: / 238*
ámin henák dámman be henák //
- | | | | | | |
|------------|-------------------------|--------------|-------------------------------|--------------------------|-------------|
| <i>mot</i> | <i>íta</i> | <i>kan</i> | <i>ma</i> | <i>< biwork</i> | <i>></i> |
| <i>mb</i> | <i>íta</i> | <i>kan</i> | <i>ma=</i> | <i>< bi= work</i> | <i>></i> |
| <i>ge</i> | 2.SG | if | NEG= | <i>< IRR= work</i> | <i>></i> |
| <i>rx</i> | SUBJ.PRO.IDP | CONJ.POT | PTCL.NEG= | <i>< TAM= CSW.ENG</i> | <i>></i> |
| <i>ft</i> | If you don't < work > , | | | | |
| <i>mot</i> | <i>íta</i> | <i>ma</i> | <i>/ 238 ámin henák dáman</i> | | |
| <i>mb</i> | <i>íta</i> | <i>ma=</i> | <i>/ 238 ámin henák dáman</i> | | |
| <i>ge</i> | 2.SG | NEG= | <i>/ 238 safe there up_to</i> | | |
| <i>rx</i> | SUBJ.PRO.IDP | PTCL.NEG= | <i>/ 238</i> | ADJ | ADV.LOC ADV |
| <i>ft</i> | you won't be | | safe at all there (in Juba). | | |
| <i>mot</i> | <i>be</i> | <i>henák</i> | <i>//</i> | | |
| <i>mb</i> | <i>be</i> | <i>henák</i> | <i>//</i> | | |
| <i>ge</i> | by | there | <i>//</i> | | |
| <i>rx</i> | PREP | ADV.LOC | <i>//</i> | | |
- PGA_SM_CONV_1_SP2_193–195

3.2 Intonation and codeswitching

We have seen in the preceding paragraphs that there is a sizeable correlation between types of csw and prosodic boundaries. We want to show now that variation in intonation contours can represent a further structural discriminant for the identification of intrasentential csw. Even though many authors (Gumperz 1982; Zentella 1997; Karrebæk 2003) have already noted a connection between intonation and csw, there is a lack of research on this matter. In this regard, Olson and Ortega-Llebaria (2010) showed that the perceptual relevance of csw is more significant in the absence of other contextualization cues, such as narrow focus intonation. However, our data from CorpAfroAs point out that intrasentential tag-switching systematically correlates with some form of intonation emphasis.

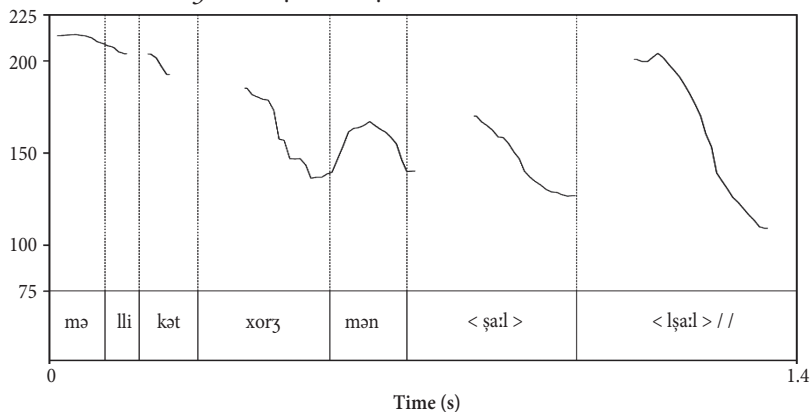
Example (20) shows an instance of Juba Arabic / English intrasentential csw occurring within an Intonation Unit enclosed within a major prosodic boundary. With regard to the intonation curve, it is noteworthy that the declination of F0 related to the declarative status of the utterance¹² is suspended by an emphatic pitch rise occurring on the first stressed syllable of the switched item 'master.' This word is well integrated into the Juba Arabic syllable structure as [masat]. This emphatic high pitch corresponds to the highest point of the intonation curve and reaches 125.1 Hz.

12. See the chapter 'The intonation of Topic and Focus in Zaar, Tamasheq, Juba Arabic and Tripoli Arabic' for a description of the intonation curve of declarative clauses in Juba Arabic.

(19) *tx la:na bíga < másat > ta bo:r suda:n zátu //*

mot le ána bíga < master > ta bo:r suda:n zátu //
mb le ána bíga < master > ta bo:r suda:n zátu //
ge to 1SG become < master > POSS Port Sudan FOC1 //
rx PREP SUBJ.PRO.IDP V < CSW.ENG > PTCL N.PR N.PR PTCL //
ft Until I became < the master > of Port Sudan
 PGA_S_NARR_1_447

Example (20) displays an instance of Moroccan Arabic / French intrasentential cs w occurring in a bilingual Intonation Unit with a major prosodic boundary. In this case, as before, the two EL constituents are phonologically integrated into the consonant system of the ML and they correlate with an emphatic high pitch, reaching 70 Hz max and 204.3 Hz respectively, on the last switched word.

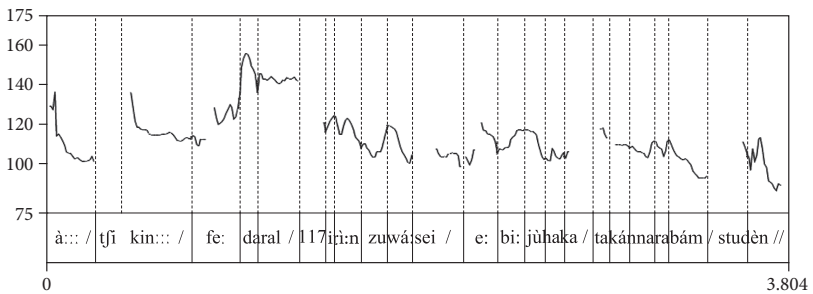
(20) *tx mǎlli katxrǝʒ mǎn < ʃa:l > < lʃa:l > //*

mot mǎlli katxrǝʒ mǎn < salle >
mb mǎlli ka= t- xrǝʒ mǎn < salle >
ge when REAL= 2- go_out\IPFV from < salle >
rx CONJ TAM= PNG- V PREP < CSW.FRA >
ft When you go from < hall >

mot < lsalle > //
mb < l= salle > //
ge < to= salle > //
rx < PREP= CSW.FRA > //
ft < to hall >.
 ARY_DC_CONV1_SP1_20

A similar change of fundamental frequency can be also observed in Hausa / English intrasentential csw. Example (21) contains three instances of English tag-switching occurring in a complex sentence covering seven Intonation Units. Each switched item (i.e. *federal*, *ABU* and *students*) corresponds to a new intonation unit, with an emphatic high pitch on the second syllable of English 'federal', and, in each case, with a suspension of F0 downdrift.

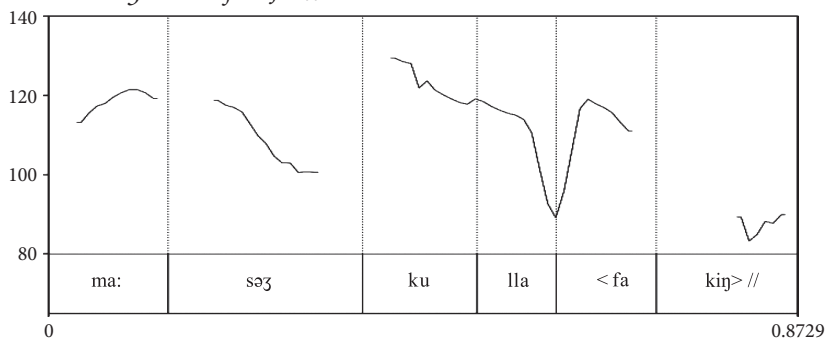
- (21) *tx* à::: / *tfikin*::: / < *fe:daral* > / 117 *iɾin zuwà: sei* / < *e:bi:jù* > *hakà* / *takàn rarràba:mà* / < *stu:dèn* > //



mot à::: / *tfikin*::: / < *federal* > /
mb à::: / *tfikin* -::: / < *federal* > /
ge FILL / inside -LENGTH / < *federal* > /
rx FILL / PREP -LENGTH / < CSW.ENG > /
ft (She would take it) aah, in < *federal* > (college)
mot 117 *iɾi:n zuwà: sei* / < *ABU* > *hakà* /
mb 117 *iɾi: -n zuwà: sei* / < *ABU* > *hakà* /
ge 117 type -GEN going to / < *ABU* > like_this /
rx 117 N -SYNT N.V0 PREP / < CSW.ENG > ADV /
ft 117 and she can go up to < *ABU* > (university) for example
mot *takàn rarràba: mà* / < *students* > //
mb *takàn rarràba: mà* / < *students* > //
ge 3F.SG.HAB distribute DAT / < *students* > //
rx PNG.TAM V1.PL PTCL.SYNT / < CSW.ENG > //
ft she would distribute (it) to the < *students* >.
 HAU_BC_CONV_1_SP2_42-49

The preceding analysis of the csw intonation supports the idea that ‘*the most common function of tag-switching is pragmatic: highlighting an event, setting off a personal reaction to what has been said*’ (Caron 2002: 22)¹³ and that csw functions as a contextualization cue among others cues such as narrow focus intonation (Auer 1998; Olson and Ortega-Llebaria 2010). In this regard, example 22 shows an instance of Tripoli Arabic / English tag-switching csw co-occurring with BORR and narrow focus. The highest pitch falls on the focused quantifier *kull* and reaches 137 Hz. The high pitch on the first syllable of the switched element ‘fucking’ barely reaches 83 Hz, and the lexical borrowing *ma:sə3* (from English *message*) does not display any intonation prominence. Therefore, contrastive focus appears to be more prominent prosodically than csw.

(22) tx *ma:sə3 kulla < fakiŋ > //*



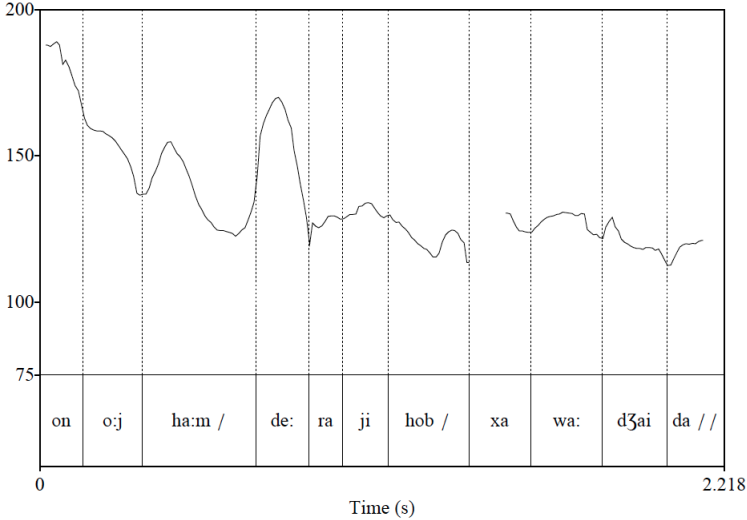
mot *ma:sə3 kulla < fucking > //*
 mb *ma:sə3 kull =h < fucking > //*
 ge message all =OBJ.3SG.M < fucking > //
 rx N.BORR.ENG QNT =PRO.PNG < CSW.ENG > //
 ft The message was all < insult >
 HAU_BC_CONV_1_SP2_42–49

The majority of the previous examples show that, contrary to what is generally assumed, morphophonological integration cannot be used to pinpoint the distinction between BORR and csw, since switched items may well be integrated into the phonology of the speaker’s native language. With this background, the change of f0 linked to the occurrence of the switched items is of particular interest when compared to the ordinary intonation contours related to BORR, as already seen in example (22). Example (23) further shows the Arabic nominal borrowing *xawa:dʒa* ‘westerner’ in Beja. That the borrowed item is not integrated into the phonology of the recipient language is shown by the retention of the Arabic

13. See also Ziamari (2010) for an analysis of the pragmatic implication of Moroccan Arabic / French.

voiceless velar fricative /x/ and by the absence of variation in the declining pattern of the first two Intonation Units.

(23) *tx* *ono:jham / de:r aji hob / xawa:d3ajda //*



mot *o:n* *o:jha:m* /
mb *o:n* *o:=* *jha:m* /
ge PROX.SG.M.ACC DEF.SG.M.ACC= leopard /
rx DEM DET= N.M /
ft The leopard
mot *de:ra* *aniho:b* /
mb *de:r* *-a* *a-ni* *=ho:b* /
ge kill\INT -IMP.SG.M say\PFV.1SG =when /
rx der.v1 -TAM.PNG v1.IRG =CONJ /
ft When I said: 'Kill it!'
mot *ixawa:d3ajda* // BI_425
mb *i=* *xawa:d3a* *-i* *=da* // BI_425
ge DEF.M= foreigner -GEN.SG =DIR // BI_425
rx DEM= N.M.BORR.ARA -CASE =POSTP // BI_425
ft to the foreigner // BI_425
 BEJ_MV_NARR_6_foreigner_12-14

4. Conclusions

In this chapter, we have argued for a neat separation between BORR and CSW: while the former is part and parcel of the recipient language, and is glossed and analyzed

as such, CSW is an instance of one or more alien elements entering a recipient language without — at least in principle — being integrated into the latter. In the CorpAfroAs corpus as a whole few borrowings have been indicated because, on the very basis of our definitions, all of them are integrated as a part of the lexicon of the recipient language — often, obviously, with semantic changes. On the morphosyntactic level borrowings in CorpAfroAs frequently — but not always — keep the word category of the donor language: borrowed nouns are integrated as nouns and verbs as verbs in the recipient language. The degree of integration in the system of the recipient language is very variable, and it does not seem to depend on either the donor or the recipient language. BORR is always predictable because in any given context it is the only tool for speakers to express what they want to say. Between CSW and BORR runs a very thin line, especially when adjustment to the recipient language is only phonetic.

Different types of CSW have different prosodic segmentation patterns: intersentential CSW is systematically related to monolingual Intonation Units, while intrasentential CSW tends to occur at the end of bilingual Intonation Units. In a minority of cases, intrasentential CSW can also occur in monolingual Intonation Units due to the intervention of discourse factors. Tag-switching is regularly highlighted by prosodic prominence (emphatic high pitch). This is a major result of our investigation and can be taken as a new constraint for distinguishing CSW from BORR. Of course, all these conclusions will have to be confirmed through further quantitative analysis.

References

- Auer, Peter. 1998. The pragmatics of codeswitching: A sequential approach. In *One Speaker, Two Languages: Cross-disciplinary Perspectives on Code-switching*, Lesley Milroy & Peter Muysken (eds), 115–135. Cambridge: CUP.
- Bakker Peter & Matras, Yaron (eds). 2003. *The Mixed Language Debate: Theoretical and Empirical Advances*. Berlin: Mouton de Gruyter.
- Caron, Bernard. 2002. Zaar, Hausa, English codemixing. In *Lexical and Structural Diffusion. Interplay of Internal and External Factors of Language Development in the West African Sahel*, Robert Nicolai & Petr Zima (eds), 19–25. Nice & Prague: Publications de la Faculté de Lettres, Arts et Sciences Humaines de Nice et de la Faculté des Etudes Humaines, Université Charles de Prague.
- Caron, Bernard. 2012. 'Hausa Corpus'. Corpus recorded, transcribed and annotated by Bernard Caron. In Amina Mettouchi & Christian Chanard (eds). *The CorpAfroAs Corpus of Spoken AfroAsiatic Languages*. DOI: <http://dx.doi.org/10.1075/scl.68.website>. Accessed on 01/07/2013. (= HAU_BC)
- Caubet, Dominique. 2012. 'Moroccan Arabic Corpus'. Corpus recorded, transcribed and annotated by Dominique Caubet. In Amina Mettouchi & Christian Chanard (eds). *The*

- CorpAfroAs Corpus of Spoken AfroAsiatic Languages*. DOI: <http://dx.doi.org/10.1075/scl.68.website>. Accessed on 01/07/2013. (= ARY_DC)
- Gumperz, John J. 1982. *Discourse Strategies*. Oxford: OUP. DOI: 10.1017/CBO9780511611834
- Haspelmath, Martin. 2008. Loanword typology: Steps toward a systematic cross-linguistic study of lexical borrowability. In *Aspects of Language Contact: New Theoretical, Methodological and Empirical Findings with Special Focus on Romancisation*, Thomas Stolz, Dik Bakker & Rosa Salas Palomo (eds), 43–62. Berlin: Mouton de Gruyter.
- Haspelmath, Martin. 2009. Lexical borrowing: Concepts and issue. In *Loanwords in the World's Languages. A Comparative Handbook*, Martin Haspelmath & Uri Tadmor (eds), 35–54. Berlin: Mouton de Gruyter.
- Heath, Jeffrey. 1989. *From Codeswitching to Borrowing. A Case Study in Moroccan Arabic*. London: Kegan.
- Heath, Jeffrey. 1994. Borrowing. In *The Encyclopaedia of Languages and Linguistics*, Vol. 1, Ron E. Asher (ed), 383–394. Oxford: Pergamon Press.
- Karrebæk, Martha Sif. 2003. Iconicity and structure in codeswitching. *International Journal of Bilingualism* 7: 407–411. DOI: 10.1177/13670069030070040401
- Malibert, Il-Il. 2012. 'Hebrew Corpus'. Corpus recorded, transcribed and annotated by Il-Il Malibert. In Amina Mettouchi & Christian Chanard (eds). *The CorpAfroAs Corpus of Spoken AfroAsiatic Languages*. DOI: <http://dx.doi.org/10.1075/scl.68.website>. Accessed on 01/07/2013. (= HEB_IM)
- Manfredi, Stefano. 2012. 'Juba Arabic Corpus'. Corpus recorded, transcribed and annotated by Stefano Manfredi. In Amina Mettouchi & Christian Chanard (eds). *The CorpAfroAs Corpus of Spoken AfroAsiatic Languages*. DOI: <http://dx.doi.org/10.1075/scl.68.website>. Accessed on 01/07/2013. (= PGA_SM)
- Manfredi, Stefano & Pereira, Christophe. 2013. Arabic Youth Languages in Africa: A Comparative Overview. Paper presented at the African Urban and Youth Languages Conference, University of Cape Town, July 5–6.
- Matras, Yaron. 2009. *Language Contact*. Cambridge: CUP. DOI: 10.1017/CBO9780511809873
- McCormick, Kay M. 1994. Code-switching and mixing. In *The Encyclopaedia of Languages and Linguistics*, Vol. 2, Ron E. Asher (ed), 581–587. Oxford: Pergamon Press.
- Mejdel, Gunvor. 2005. Code-Switching. In *Encyclopaedia of Arabic Language and Linguistics*, Vol. 1, Kees Versteegh, Mushira Eid, Alaa Elgibali, Manfred Woidich & Andrzej Zaborski (eds), 414–421. Leiden: Brill.
- Mettouchi, Amina. 2008. Kabyle/French codeswitching: A case study. In *Berber in Contact: Linguistic and Sociolinguistics Perspectives*, Mena Lafkioui & Vermondo Brugnatelli (eds), 187–198. Cologne: Rüdiger Köppe.
- Mettouchi, Amina. 2012. 'Kabyle Corpus'. Corpus recorded, transcribed and annotated by Amina Mettouchi. In Amina Mettouchi & Christian Chanard (eds). *The CorpAfroAs Corpus of Spoken AfroAsiatic Languages*. DOI: <http://dx.doi.org/10.1075/scl.68.website>. Accessed on 01/07/2013. (= KAB_AM)
- Muysken, Peter. 1995. Code-switching and grammatical theory. In *One Speaker, Two Languages: Cross-disciplinary Perspectives on Code-switching*, Lesley Milroy & Peter Muysken (eds), 177–198. Cambridge: CUP. DOI: 10.1017/CBO9780511620867.009
- Myers-Scotton, Carol. 1992. Comparing codeswitching and borrowing. *Journal of Multilingual and Multicultural Development* 13(1–2): 19–39. DOI: 10.1080/01434632.1992.9994481
- Myers-Scotton, Carol. 1993. *Duelling Languages: Grammatical Structure in Codeswitching*. Oxford: OUP.

- Myers-Scotton, Carol. 2001. The matrix language frame model: Developments and responses. In *Codeswitching Worldwide II*, Rodolfo Jacobson (ed.), 23–58. Berlin: Mouton de Gruyter.
- Myers-Scotton, Carol. 2002. *Language Contact: Bilingual Encounters and Grammatical Outcomes*. Oxford: OUP.
- Olson, Daniel & Ortega-Llebaria, Marta, 2010. The perceptual relevance of Code Switching and Intonation in creating narrow focus. In *Selected Proceedings of the 4th Conference on Laboratory Approaches to Spanish Phonology*, Marta Ortega-Llebaria (ed), 57–68. Somerville MA: Cascadilla Proceedings Project.
- Poplack, Shana. 1980. Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of codeswitching. *Linguistics* 18(7-8): 581–618.
DOI: 10.1515/ling.1980.18.7-8.581
- Romaine, Suzanne. 1989. *Bilingualism*. Oxford: Blackwell.
- Sankoff, David & Poplack, Shana. 1981. A formal grammar for codeswitching. *Papers in Linguistics: Journal of Human Communication* 14(1): 3–45. DOI: 10.1080/08351818109370523
- Shenk, Petra Scott. 2006. The interactional and syntactic importance of prosody in Spanish-English bilingual discourse. *International Journal of Bilingualism* 10(2): 179–205.
DOI: 10.1177/13670069060100020401
- Tadmor, Uri. 2009. Loanwords in the world's languages: Findings and results. In *Loanwords in the World's Languages. A Comparative Handbook*, Martin Haspelmath & Uri Tadmor (eds), 55–75. Berlin: Mouton de Gruyter.
- Tosco, Mauro and Stefano Manfredi. 2013. Pidgins and creoles. In *The Oxford Handbook of Arabic Linguistics*, Jonathan Owens (ed.), 495–519. Oxford: Oxford University Press.
- Vanhove, Martine. 2012. 'Beja Corpus'. Corpus recorded, transcribed and annotated by Martine Vanhove. In Amina Mettouchi & Christian Chanard (eds). *The CorpAfroAs Corpus of Spoken AfroAsiatic Languages*. DOI: <http://dx.doi.org/10.1075/scl.68.website>. Accessed on 01/07/2013. (= BEJ_MV)
- Vicente, Ángeles. 2012. 'Moroccan Arabic Corpus (Ceuta)'. Corpus recorded, transcribed and annotated by Ángeles Vicente. In Amina Mettouchi & Christian Chanard (eds). *The CorpAfroAs Corpus of Spoken AfroAsiatic Languages*. DOI: <http://dx.doi.org/10.1075/scl.68.website>. Accessed on 01/07/2013. (= ARY_AV)
- Watson, Duane, Breen, Mara & Gibson, Edward. 2006. The role of syntactic obligatoriness in the production of intonational boundaries. *Journal of Experimental Psychology: Learning Memory and Cognition* 32(5): 1045–1056. DOI: 10.1037/0278-7393.32.5.1045
- Zentella, Ana Celia. 1997. *Growing up bilingual*. Oxford: OUP.
- Ziamari, Karima. 2010. Moroccan Arabic-French codeswitching and information structure. In *Information Structure in Spoken Arabic*, Jonathan Owens & Alaa Elgibali (eds), 189–206. London: Routledge.