Locating code-switching in the grammar: Role of postsyntax and morphological wordhood

The long term bi- and multi-lingualism in southeastern Turkey has yielded patterns of 'word'-internal CS in verbal and nominal domains among Turkish (TK), Anatolian Arabic (AA), and Northern Kurdish (NK). Informed by the very rarely-studied trilingual language-mixing, this paper first demonstrates the existence of code-switching (CS) within a complex head (*contra* MacSwan 1999; MacSwan & Colina 2014), and then argues that various formal approaches to CS, which rely on either a distinction between functional vs lexical categories (Poplack 1981; Belazi et al. 1994), or phasehood (López et al. 2017) as the defining constraint on CS are not tenable. Instead, we argue that such instances of CS are subject to a restriction called *No-Reversal Constraint*, that places a ban on reverting back to a previous language for morpheme insertion. This constraint applies in Morphology at the level of subwords constituting morphological words (MWds, Embick & Noyer 2001), thus it is not a CS-specific mechanism, but it applies to independently needed units.

CS in the verbal domain. In (1), various TK roots are used in AA, by first identifying and extracting a tri- or quadriradical root, e.g., $\sqrt{\text{KPT}}$ 'to close'. Then the root is assigned to one of the templates.

(1) Stem II *qappat – īqappət* 'to close' cf. Turkish *kapat-*
Stem III *ḍāyan – īḍāyən* 'to be patient, to bear up' cf. Turkish *dayan-* (Talay 2007)

CS in the nominal domain. A common situation involves addition of the Arabic regular plural morpheme *-(a)d* to Turkish roots (Lahdo 2009), as in (2). The reverse pattern is also found, (3): The Arabic plural *-ad* realized on the Num head can be replaced with its Turkish counterpart *-lAr*, in a way that obeys the vowel harmony of Turkish. This is indeed attested in both regular, (3a), and double plural forms, (3b) (We adopt the verbal/nominal syntax Arad 2003 proposes for Semitic; see also Tucker 2011; Kramer 2016. Also *subscript* notation in the glosses indicates which language a morpheme comes from).

(2) a. *mılyar-ad* 'billion$_{\text{TK}}$-PL$_{\text{AA}}$'
b. *soba-d* 'stove$_{\text{TK}}$-PL$_{\text{AA}}$'

(3) a. *xatan-lar* 'son-in-law$_{\text{AA}}$-PL$_{\text{TK}}$' cf. *xatan-ad*
b. *kıleb-ler* 'dog$_{\text{AA}}$.PL$_{\text{AA}}$-PL$_{\text{TK}}$' cf. *kıleb-ad*

The data thus far shows that the root (and maybe root + a higher functional head) may come from Language 1 (Lg1), whereas the top node, Num in (2), comes from Language 2 (Lg2). Attempts to add outer morphemes are informative wrt potential restrictions. E.g., adding the possessive marker to (2b) is allowed only if the possessive is AA, (4a), whereas the TK possessive leads to ungrammaticality, (4b). Same considerations hold for (3), where an additional possessive or another morpheme would be licit only if it is from TK and not AA.

(4) a. soba-d-i
stove$_{\text{TK}}$-PL$_{\text{AA}}$-1SG.POSS$_{\text{AA}}$
'my stoves'

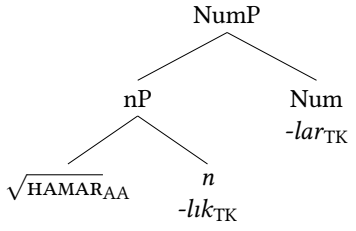
b. *soba-d-ım
stove$_{\text{TK}}$-PL$_{\text{AA}}$-1SG.POSS$_{\text{TK}}$
'my stoves'

One hypothesis is that whichever language the highest (functional) node is from, any additional morpheme(s) has to be from that language (plus the stipulation that roots do not enter into the calculus). This would also explain examples that involve 'derivational' affixes, (5)-(6). e.g., (5a) has the structure in (7), and is licit, while the ungrammatical (5b) would revert to AA for the Num after the TK morpheme on *n*.

(5) a. hamar-lık-lar
donkey$_{\text{AA}}$-DER$_{\text{TK}}$-PL$_{\text{TK}}$
'stupidities'
b. cf. *hamar$_{\text{AA}}$-lık$_{\text{TK}}$-ad$_{\text{AA}}$
c. hamar-lık-tan
donkey$_{\text{AA}}$-DER$_{\text{TK}}$-ABL$_{\text{TK}}$
'from the stupidity'

(6) a. salaq-tiy[e]-ad
stupid$_{\text{TK}}$-DER$_{\text{AA}}$-PL$_{\text{AA}}$
'stupidities'
b. cf. *salaq$_{\text{TK}}$-tiye$_{\text{AA}}$-ler$_{\text{TK}}$
c. *salaq-tiye-den
stupid$_{\text{TK}}$-DER$_{\text{AA}}$-ABL$_{\text{TK}}$
'from the stupidity'

(7)



Language-mixing patterns of trilingual speakers are crucial in demonstrating that the above hypothesis cannot be correct. In (8), the AA root and the plural morphemes are followed by the vocative morpheme of NK. This does not follow from the above hypothesis. Crucially, with the same meaning in (9), the AA root can be followed by the TK plural, which is then followed by the NK vocative. (10) is like (9) in that the AA root is followed by the TK plural, but the following morpheme is AA, which leads to ungrammaticality.

(8) xatan-ad-no
son-in-law$_{\text{AA}}$-PL$_{\text{AA}}$-VOC$_{\text{NK}}$
'Sons-in-law!'

(9) xatan-lar-no
son-in-law$_{\text{AA}}$-PL$_{\text{TK}}$-VOC$_{\text{NK}}$
'Sons-in-law!'

(10) *xatan-lar-i
son-in-law$_{\text{AA}}$-PL$_{\text{TK}}$-1POSS$_{\text{AA}}$
'my sons-in-law'

Longer sequence of morphemes, as in (11), are also acceptable.

(11) a. xatan-lar-ım-no
son-in-law$_{AA}$-PL$_{TK}$-1SG.POSS$_{TK}$-VOC$_{NK}$
'My sons-in-law!'

b. xatan-ad-ım-no
son-in-law$_{AA}$-PL$_{AA}$-1SG.POSS$_{TK}$-VOC$_{NK}$
'My sons-in-law!'

Generalization: The attested and unattested patterns are as in (12). CS allows various patterns of morpheme insertion in (12a) through (12c), but not the pattern in (12d). (12a) indicates that a M(orpheme) from Lg1 can be followed by a morpheme from Lg2, which in turn is followed by another morpheme from Lg2. In (12b) the first two morphemes are from Lg1 and they are followed by a morpheme from Lg2. (12c) shows that each M can belong to a different language, while in the ungrammatical (12d), the first morpheme is from Lg1, followed by the second morpheme from Lg2. Crucially, the third morpheme is drawn back from Lg1.

(12)

	M1	**M2**	**M3**
a.	Lg1	≺ Lg2	≺ Lg2
b.	Lg1	≺ Lg1	≺ Lg2
c.	Lg1	≺ Lg2	≺ Lg3
d.	*Lg1	≺ Lg2	≺ Lg1

This generalization captures all the observed cases of intra-word CS in bilingual settings that involve many other languages, compiled in Stefanich et al. 2019, as they form a small subset of (12). It also captures the trilingual pattern in Corfiot, (13), where the root is Hebrew, the verbalizer Greek and the infinitival Romance.

(13) dibur-efs-ár
talk$_{Hebrew}$-VRBZ$_{Greek}$-INF$_{Romance}$
'talk' (Corfiot; Vardakis 2023:6)

We call the pattern in (12) the *No-Reversal Constraint*. It prohibits switching *back* to a language that has already contributed an exponent (via Vocabulary Insertion) earlier in the derivation.

At what stage of the derivation does No-Reversal Constraint apply? We propose that this constraint applies to subwords in Morphology concomitant with linearization of MWds (morphosyntactic words), and before Vocabulary Insertion (Embick & Noyer 2001). The analysis further makes two correct predictions: (i) It imposes a restriction on MWds, but allows its obviation for larger structures. For example, it explains why unlike (6c), both (14) and (15) are grammatical (as the P head is a separate MWd). (ii) It also correctly predicts that nodes that do not form complex words via head-movement should not be subject to the No-Reversal Constraint. Indeed, the clitics *-(y)sA/ise* 'as for', (16), *mI* (interrogative), are apparent exceptions.

(14) mı salaq-tiye
from$_{AA}$ stupid$_{TK}$-DER$_{AA}$
'from the stupidity'

(15) salaq-tiye için
stupid$_{TK}$-DER$_{AA}$ for$_{TK}$
'for/because of the stupidity'

(16) soba-d=sa
stove$_{TK}$-PL$_{AA}$=as.for$_{TK}$
'as for the stoves'

This constraint also explains examples like (17), or like the Media Lengua example, (18): combination of Spanish roots/lexical items (italicized) plus Quechua inflection/grammatical properties (boldfaced).

(17) ma-co xatan-lar-ım.
NEG$_{AA}$-came$_{AA}$ son.in.law$_{AA}$-PL$_{TK}$-1POSS$_{TK}$
'My sons-in-law didn't come.'

(18) *unu fabur*-**ta** *pidi*-**nga**-**bu** *bini*-**xu**-**ni**.
one favor-ACC ask-NOM-BEN come-PROG-1SG
'I come to ask a favor.' (Muysken 1997)

Issues with prior analyses • PF Interface Condition of MacSwan (1999); MacSwan & Colina (2014) states that "switching within a complex head is prohibited" even when there is phonological integration. This constraint fails to capture the grammatical examples reported in this study. • the Functional Head Constraint of Belazi et al. (1994) states that CS may not occur between a functional head and its complement, while CS between a lexical head and its complement proceeds unimpeded. This incorrectly rules out examples like (8), (9), (11), (13). • López et al.'s (2017) analysis is built on the *phase*-theory and hypothesizes that "code-switching may take place at phase boundaries but not within the phase". In a nominal structure like [$\sqrt{}$- **n** - Num - D - **K**(ase)], López et al. (2017) adopt the standard view that n and K are phase heads. In this analysis, *n* and its complement $\sqrt{}$ belong to different phases in Spell-Out. Moreover, D would be transferred with Num and *n*, while K is transferred with the higher phase that contains it, i.e., *v*P or *p*P. Despite capturing many of the examples (e.g., (2), (4), (5), (6)), it incorrectly rules out many acceptable forms, such as (3) or (11b). E.g., in (11b), both the plural on Num head and the possessive on D/Poss are in the same Spell-out, and thus are transferred together. Thus, CS is not predicted to be possible, contrary to fact. Similarly, in (3), the categorizer n and the plural are within the same phase, thus CS between the two should be disallowed, also not correct.

Going forward. We also consider issues an alternative approach to Embick & Noyer 2001, e.g., spanning, faces. Moreover, various more examples will be adduced to highlight the need for also syntactic diacritic features, so that the system correctly linearizes an abstract morpheme as a prefix or a suffix in the first place.