Malayalam-Kannada Code-mixing

PAROMA SANYAL, *Indian Institute of Technology Delhi*
THERESE LIAM TOM, *Indian Institute of Technology Delhi*

ABSTRACT

This paper presents the results of a pilot study undertaken to see the possible limits of code-mixing among Malayalam-Kannada bilinguals in a syntactic context of featural mismatch. The results of the study reiterate that the fact that despite being influenced by psycholinguistic factors like being primed with the language of the task paragraph, simultaneous (balanced) bilinguals can make the choice of strategy based on structural factors, but sequential bilinguals do not access syntactic structure while making these decisions. The latter's choice of strategy is based either on the acquisitional factor of MT/non-MT or on psycholinguistic factors from the task design.

1 Introduction

A bilingual person can mix two languages within a single syntactic construction. This shows that the syntactic machinery offers the possibility of combining the syntactic frames from different languages. (1) below shows such a possibility where the English and Malayalam frames have been mixed.

(22) [I told John [that *avande co:r njaan kazhichu*]
I tell.PST John COMPL he-GEN rice I eat.PAST
'I told John that I ate his rice'

Here the matrix clause is in English, and the embedded proposition (*His rice, I ate*) is in Malayalam. Note that the Complementizer embedding the Malayalam clause, is in English, the language of the matrix sentence.

In addition to clausal level of mixing, bilinguals also insert lexical items (Muysken 2000) from one language into the syntactic frame of the other language as seen in (2)

(23) I eat *co:r* in the morning.

Here, the Malayalam lexical item, *co:r*, 'rice' replaces the English word, *rice* in an English frame sentence. Henceforth, in this paper we refer to the process in (1) as code-switching, and the one in (2) as code-mixing[1].

Neither (1) nor (2) require interactions among the functional features of the two languages since there are no overt functional features from English, the frame language, that are syntactically dependent on the Malayalam part. However, if the code-mixed lexical

[1] The terms code-mixing and code-switching have been used in literature (MacSwann 2014, Muyksen 2000, Grimstad et al. 2018, Myers-Scotton 1993) to mean a number of different kinds of mixing patterns between languages. Sometimes they include loan-words and sometimes they do not. For clarity and consistency, we are making our use of the terms transparent.

item needs to syntactically interact with a functional head from the other language there is a possibility for feature mismatch between them. This potential mismatch between the functional structures of two different languages is the scope of enquiry for this paper.

In naturalized environments, balanced bilinguals, who are proficient in both languages, are likely to avoid code-mixing in circumstances where there is a potential mismatch in syntactic characteristics between the two languages. So instead of looking for well-formed cases of naturalized code-mixing within a bilingual corpus, we are eliciting code-mixed data from bilingual participants in our study. This is a pilot study designed to understand what people would intuitively do if they had to code-mix in a certain way and had no time restriction to produce their response.

1.1 The participants in the code-mixing study

For this study on code-mixing, we have selected two Dravidian languages, Malayalam and Kannada, because they are lexically and syntactically similar, and have some structural distinctions that could produce mismatch of functional features in code-mixing contexts. Since the two languages are spoken in neighboring geographical regions, one would have expected balanced bilinguals among both Malayalam and Kannada speakers. However, that is not the case since the border region between the states of Kerala (where Malayalam is the official language) and Karnataka (where Kannada is the official language) has several other Dravidian languages such as Tulu, Kodagu etc. Residents of this region speak both Kannada and Malayalam but neither of them natively. Consequently, we have not tried the elicitation with people from this region.

All the participants in our study are Malayalam speakers who are long-time residents of Karnataka, Bangalore city to be specific, and use Kannada as a lingua franca. Ideally, we should also have had another set of participants who are Kannada speakers living in Kerala for a long time. However, due to the directionality of population migration, such a demographic is not readily available. Consequently, we have created two subsets, shown in (3), within the Malayalam-Kannada bilingual population available to us.

(3) Sequential and Simultaneous Bilinguals

Sequential Bilinguals	Participants who acquired L1 in childhood, and later acquired L2 as an adult.
Simultaneous Bilinguals	Participants who acquired both L1 and L2 in childhood

Both the groups of Malayalam-Kannada bilingual participants were presented with the same elicitation task. In the section 4 of the paper, where we analyze the sentences produced by them, we will see if the two groups are using divergent strategies to handle feature mismatch.

1.2 The mismatch contexts in the code-mixing study

Ordinarily, inflectional morphemes do not attach to borrowed words directly. Instead, the borrowed word is embedded within another lexical item from the frame language. This is observed in (4) where the English verb form 'buy' combines with the Malayalam tensed "do" verb *cejjudu* to fit into the syntactic frame of the Malayalam tensed clause.

(4) ɲa:n dosha **bɑɪ** cejjudu
I dosa buy do.PST
'I bought dosa'

Here, the English verb form 'buy' cannot attach directly with the tense morphology of Malayalam, and thus requires 'do support'. Will the same kind of restriction hold for code mixing between Malayalam and Kannada which share lexical and syntactic similarities?

Theoretically, one prominent distinction between the verbal inflections of Kannada and Malayalam is that the former has subject agreement for person and number while the latter has no agreement marking. Examples (5) and (6) demonstrate this.

(5) Kannada

a. naanu mane-ge hogid-**e**
1Sg home-DAT go.PST.1Sg
'I went home.'

b. avalu mane-ge hogid-**alu**
3SgF home-DAT go.PST.3Sg.F
'She went home.'

(6) Malayalam

a. njaan viitil-ottu pooji
1Sg home-DAT go.PST
'I went home.'

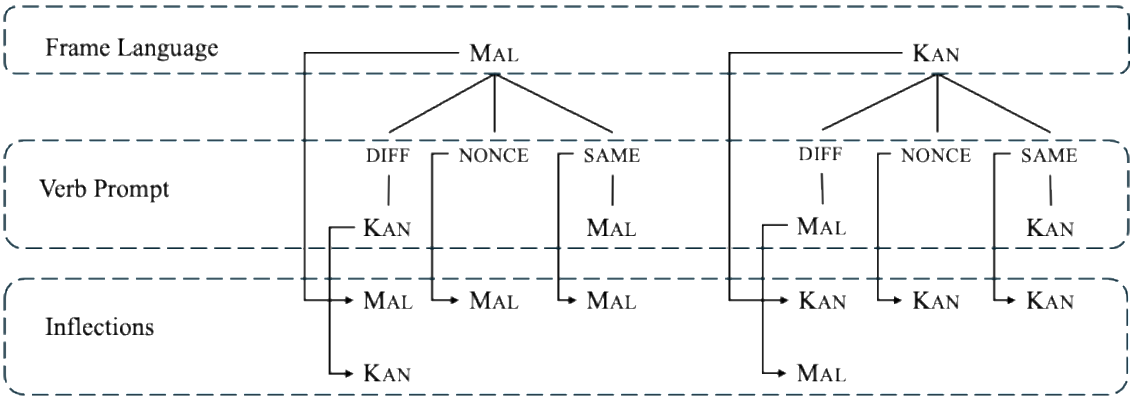
b. aval viitil-ottu pooji
3SgF home-DAT go.PST
'She went home.'

We designed our elicitation paradigm using this distinction between the two languages by inserting a blank in the place of specific verbs which the participant had to inflect and use. The language of the frame sentence was distinct from the language of the lexical verb prompts. For example, a specific verb is replaced with blanks within a Kannada text paragraph and the participant is prompted with the Malayalam lexical verb root rather than the Kannada one. While reading the text, the participant has a free choice to use 'do support' from Kannada, or inflect the Malayalam verb with either Kannada or Malayalam inflections. The same process is done with Kannada verb prompts and Malayalam Text.

Further, to determine whether the strategy used results from a preference for a particular option or a dispreference for the other, we have used a control set of nonce

prompts for both Kannada and Malayalam text contexts. If there was no interference from the inflectional paradigm of the language of the verb prompt, there should be no distinction between the choice of strategy between the nonce and verb prompt contexts. This entire paradigm along with the potential outcomes has been schematically presented in (7).

(7) Schematic presentation of the code-mixing contexts



The use of Kannada inflection in the DIFF verb prompt context has the further matter of whether the Kannada agreement markers have been used along with TAM inflections or if they have been left out, similar to Malayalam.

In the following section, we elaborate on theoretical reasons to distinguish between TAM inflections and Agreement markers. Following that is a section elaborating on the specifics of the elicitation task design and the results obtained.

2 The syntax of verb inflections

The theory of generative syntax (Chomsky 1970, 1981, 1995) has always noted the distinction between lexical and functional items in the vocabulary. While lexical items are independent meaningful words with corresponding phonological strings that are stored in the lexicon, functional items may or may not have phonological strings associated with them and have "meaning" only in the context of the syntactic arrangement of which they are a part. Within generative syntax, there are two broad ways in which these distinctions are analyzed. The two approaches are commonly referred to using the terms Lexicalist (Chomsky & Lasnik 1993) and the Non-lexicalist (Halle & Marantz 1993, Borer 2003, 2015, Embick & Noyer 2007) and hereafter we will use the same.

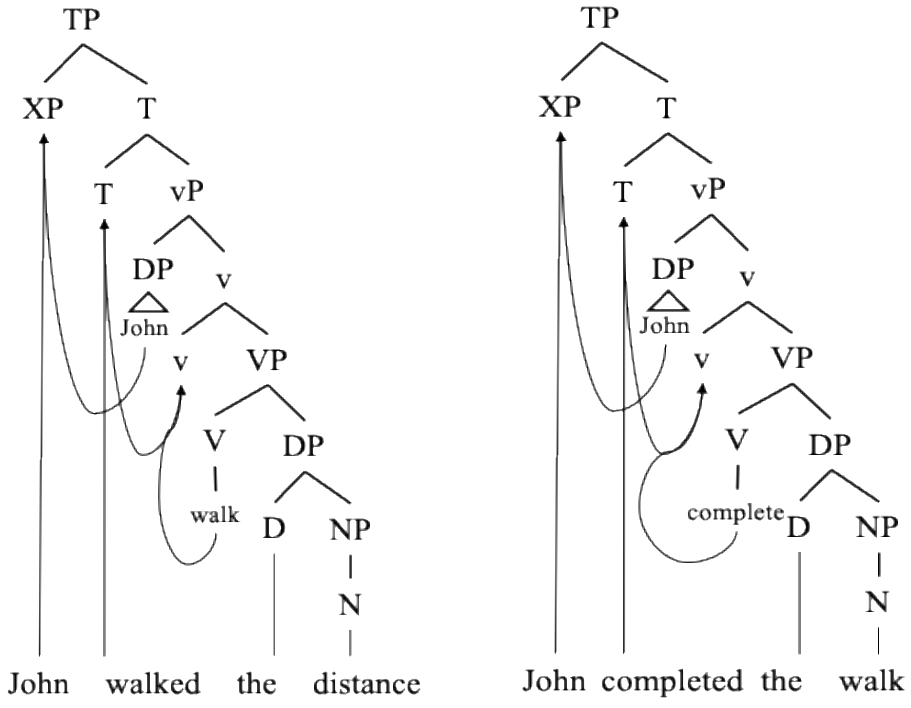
In the Lexicalist approach syntactically categorized lexical items are stored in the lexicon and enter the syntactic module by projecting a syntactic structure headed by the lexical item. Each such structure then MERGEs with further functional heads that host the syntactic and semantic features associated with that lexical head. For example, a Noun would further project functional heads corresponding to Number, Gender, Definiteness etc and a Verb would project functional heads corresponding first to the event structure and

then the tense, aspect, mood (TAM). This latter T(ense) head is also argued to be the functional head that contains uninterpretable ϕ features corresponding to person, number and gender. These are cancelled by matching them with the corresponding interpretable feature values present within a nominal projection via the operation AGREE. So, the difference between a language with overt ϕ agreement, like Kannada, and a language without ϕ agreement, like Malayalam, rests in the nature of the functional head T that is projected from the lexical verb of the language. A Kannada T has uninterpretable ϕ features that need agreement, while a Malayalam T does not have uninterpretable ϕ features and therefore does not need agreement.

In the non-lexicalist approach the general structure within the syntax and the mechanisms of structure building through MERGE and ϕ agreement via AGREE are the same as in lexicalist approach. The difference lies in the way the syntactic module interacts with the lexicon. First, the lexical items are syntactically distinct from the computational structure generated in narrow syntax. The entire syntactic hierarchy is generated on the basis of the syntactic features present on functional heads within the syntactic module. The lexicon contains category-less √Roots that gain their syntactic category as well as specific contextual meaning with respect to the syntactic structure they MERGE under. This fundamental difference between the two theoretical approaches is shown in (8) and (9).

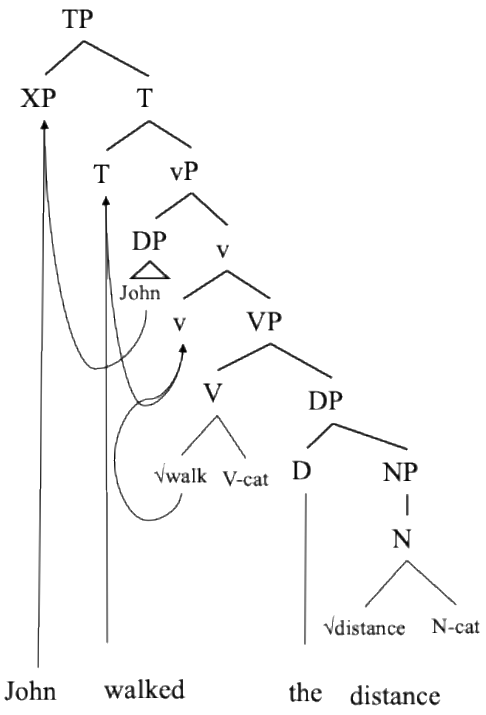
(8) Schematic presentation syntactic structure in the lexicalist approach

a. $[walk]_{verb}$ b. $[walk]_{noun}$

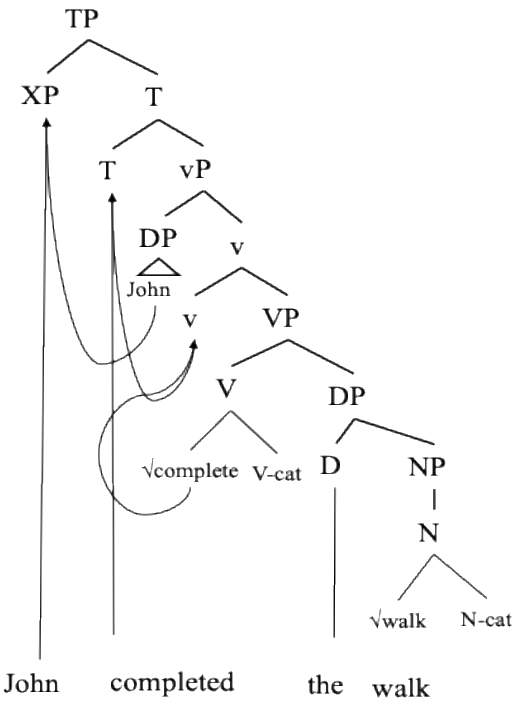


(9) Schematic presentation syntactic structure in the non-lexicalist approach

a. [√walk, v-cat]



b. [√walk, n-cat]



In (8) and (9) we see the first major distinction between these two theoretical approaches. In the lexicalist framework there are two different lexical entries corresponding to the phonological string [walk], one corresponding to verbal meaning and the other to the nominal one. In contrast, the lexicon contains just one single root √walk that corresponds to the both the nominal and the verbal outputs in the non-lexicalist framework. When the root merges with a n-categorizer under a NP-DP frame it gets a nominal meaning, and when it merges with a v-categorizer under a VP-TP frame it gets a verbal meaning.

These two theoretical approaches would predict different outcomes in the code-mixing context. If the syntactic structure is projected based on the features of the lexical item, then the Kannada verb will project a T with uninterpretable ϕ features while the Malayalam verb will project a T without it. The probe from this Kannada T will look for ϕ agreement with functional features within the Malayalam DP. The Malayalam DP will not have those functional heads since there is no ϕ probe in Malayalam. The structure will crash. On the other hand, if the syntactic structure is independent of the lexical choice the Kannada lexical item can nest under a Malayalam frame with Malayalam inflections and vice versa.

A second distinction between these two theoretical approaches also has crucial implication for our study. In the lexicalist approach all morphological derivations happen before the syntax and these lexical items already contain the phonological material

associated with inflections. In the non-lexicalist approach insertion of phonological material correlating to functional heads (non-roots) happens post-syntactically through an operation called Vocabulary Insertion (Embick, 2015). Apart from the un-categorized roots the lexicon also contains another set of phonological strings called vocabulary items (VI). Each VI corresponds to a syntactic substring containing a set of syntactic features. The same feature could be part of the defining correspondence of more than one VI. When that happens the Vocabulary Insertion progresses by disjunctive rule ordering, i.e. the more specific rule gets ordered above the less specific one. This has been demonstrated in (10).

(10) Past tense allomorphy in English

Suppletive past	Regular past
[sat]↔[past]/√sit [went]↔[past]/√go (…)	[-ed]]↔[past]

Each of the suppletive [past] VI insertion rule is specified for a particular set of roots and is therefore more specific than the regular past tense VI insertion rule. Like English, Malayalam also has multiple allomorphs corresponding to the syntactic feature [past]. There are two past tense markers [-i] and [-u] and both are specified for two distinct sets of roots.

There is a potential phonological pattern to the morphological contexts for marking the past tense with [-i] or [-u] in Malayalam. Malayalam has a phonemic contrast between voiceless geminate stops and voiced singleton stops in word medial contexts. There is a general tendency to geminate the stops before attaching the [-u] past marker and voice the stop consonants before attaching [-i]. However, this is not a clear case of complementary distribution since gemination can be seen in the context of past marking with [-i] as well. The corresponding data is shown in (11) (p.c and insights from Greeshma Joseph).

(11) Phonological context of past tense allomorphy in Malayalam

[past] marking with [-i]		[past] marking with [-u]	
paaḍ-i	sing	irunn-u	sit
caaḍ-i	jump	paranj-u	say
karakk-i	rotate	tot̪t̪-u	touch
uruṭ̪t̪-i	roll	koduṭ̪t̪-u	give

However, since voicing does not seem to accompany [-u] marking, we make the hypothesis that [-i] is the specific rule and [-u] the general one. This will be borne out in our elicitation task where the participants are theoretically predicted to use the general rule in case of nonce or novel conjugation. The disjunctive ordering for these is shown in (12).

(12) Past tense allomorphy in Malayalam

Suppletive past	Regular past
[-i]↔[past]/(√sing, √jump, √go, …)	[-u]]↔[past]

To summarize the discussion in this section so far, the lexicalist framework would predict that the TAM inflections would match the language of the verb prompt rather than the language of the remaining frame. In case of Kannada, the structure will crash since the uninterpretable ϕ features on the Kannada T would not be valued. But since the syntactic structure is literally generated out of the lexical verb given, the numeration of the participants will fail to generate an output with the Kannada verb prompt in the mismatch context, without "do support" from Malayalam. In contrast, the Malayalam verb prompt will not have this issue due to the absence of ϕ agreement. Participants will be able to produce the inflected Malayalam verbs in the Kannada frame.

The predictions from the non-lexicalist frameworks like the Exoskeletal approach (Grimstad et al., 2018) would differ. The functional projection of a Kannada V is compatible with a functional projection of D that has a Malayalam noun at its base. This is because the syntactic frame into which the Malayalam root is merged to form a noun is not generated out of the lexical item. It could contain the functional projections for person and number that could AGREE with the ϕ probe from the T. Similarly, the functional projection of a Kannada V need not have a Kannada lexical item at its base. So, a Malayalam verb prompt could end up with Kannada inflections including ϕ agreement in a Kannada frame paragraph.

Since the study design is open-ended and could result in varied outcomes, we also need an evaluative framework to analyze the responses. For this we are using a constraint-based evaluative framework. In the following section we elaborate of this evaluative framework.

3 The evaluative framework for analysis

Constraint based evaluative frameworks like Harmonic Grammar (Legendre et al., 2022) and Optimality Theory (Prince and Smolensky 1993) are used to analyze contexts where multiple factors are at play, to determine an outcome. These factors are termed as constraints. In Optimality Theory, the constraints have three core properties: violability, rankability, and universality. Of these we are going to use only the first two for our evaluative purpose. Since we are evaluating a performance output, all our constraints are not derived out of the formal aspects of the language module. In such a context, universality would be an inappropriate claim.

Violability refers to the fact that all constraints are in principle violable, and rankability means that all constraints are freely rankable with respect to all other constraints. To elaborate, suppose we are evaluating three candidate outputs A, B and C with respect to the constraints P, Q and R. Given that A violates P and Q, B violates Q and C violates P and R, the only possible outcomes are B and C and these are shown in (13).

(13) Evaluation Table for the possible outputs

a. Equal weights

	Con P	Con Q	Con R	Harmonic
	w=1	*w=1*	*w=1*	Value
Cand A	-1	-1		-2
☞Cand B		-1		-1
Cand C	-1		-1	-2

b. Q weighed up

	Con Q	Con P	Con R	Harmonic
	w=3	*w=1*	*w=1*	Value
Cand A	-1	-1		-4
Cand B	-1			-3
☞Cand C		-1	-1	-2

In (13) all the constraints are assumed to weights (*w=x*). Given the violations where each violation is denoted [-1], the harmonic value for each candidate, summation of violation and weight, is calculated for each candidate. The one with the highest value is the optimal output for that table. In (13a) where all the constraints are assumed to have equal weights, B surfaces since it has the highest harmonic value on account of minimum number of violations. Note that the violations of B being a proper subset of the violations of A, A is harmonically bound by B and will always be blocked by it from surfacing. On the other hand, C which is not harmonically bound, and also incurs two violations, can surface as the output if the weight of constraint Q is greater than the sum of the weights of P and R. This is seen in (13b).

The possible output candidates in our Kannada-Malayalam code-mixing set up are listed in (14).

(14) Candidate outputs

Denotation	DP Frame	Lexical Verb	TAM inflection	Agreement
a. MKM	M	K	M	
b. KMKK	K	M	K	K
c. MKKK	M	K	K	K
d. KMM	K	M	M	
e. MKK	M	K	K	
f. KMK	K	M	K	

(14a) is a context where a Kannada root is merged under a Malayalam VP frame and is comparable to (14b) where a Malayalam root is merged under a Kannada VP frame. Similarly, in (14c) and (14d) a Malayalam and Kannada root is merged under a Kannada and Malayalam DP frame. Technically, these four should have been the only contexts we should have expected. However, since our participants are native Malayalam speakers with

Kannada as a L2 there is a possibility that some of them would not have acquired the agreement paradigm of Kannada very well. Thus (14e) and (14f) are output candidates with a Kannada VP frame where the T does not have ϕ features.

Determining the nature of the potential constraints that influence the outcome is more challenging than determining the possible output candidates. We have divided these constraints into three sub-types based on our experimental paradigm. The first set of constraints are based on structure-based predictions from syntax and morphology, the second set from acquisition related factors, and the third considers psycholinguistics factors that might be induced by the elicitation task design. These have been described in (15).

(15) Potential evaluative constraints

a. Structural Constraints

Max F	*Project the Syntactic frame with maximal functional features* This constraint states that, given an option between two syntactic structures the one with the greater number of functional heads will be preferred than the one with the lesser one. This means both DP and VP frame choice will prefer Kannada over Malayalam.
Min F	*Project the Syntactic frame with minimal functional features* This constraint states that, given an option between two syntactic structures the one with the lesser number of functional heads will be preferred than the one with more features. This means both DP and VP frame choice will prefer Malayalam over Kannada.

b. Acquisition Factors

MT Faith	*Project the Syntactic frame of MT* This constraint states that, given an option between two syntactic structures where one of them is the mother tongue of the participant, there is a positive bias towards the MT frame. This means both DP and VP frame choice will prefer Malayalam over Kannada.
MT Anti-faith	*Do not project the Syntactic frame of MT with non-MT verb prompt* This constraint states that, given an option between two syntactic structures where one of them is the mother tongue of the participant, there is a negative bias towards the MT frame. This means both DP and VP frame choice will prefer non-Malayalam over Malayalam.

c. Psycholinguistic Factors

Match I-FT	*The language of the Inflection must match the language of the frame text*

This constraint states that given an option between two syntactic structures where one of them matches the context, there is a positive bias towards that frame. This means the VP frame choice will prefer the language of the context.

Match I-VP *The language of the Inflection must match that of the verb prompt*
This constraint states that given an option between two syntactic structures where one of them matches the root of the verb prompt, there is a positive bias towards that. This means the VP frame choice will prefer the language of the verb prompt.

While the constraints in (15b) and (15c) are self-explanatory, those proposed in (15a) warrant some additional discussion. The syntactic features of case and agreement have historically been linked to other syntactic features such as EPP. In the light of empirical data from languages with Multiple Agree, Addressee Agreement, and Split Agreement, agreement as a phenomenon is no longer linked with any particular syntactic position for the DP with the interpretable features. Further, Adger (2003) proposes a further distinction between these set of features. The semantically interpretable features such as number and definiteness do not get deleted after the match, but semantically vacuous purely syntactic features such as case and gender get deleted after the match. This [+/- interpretable] factor has been used by L2 acquisition researchers (Bel 2003, Díaz et al. 2008, Hulk & Müller 2000, Sorace 2003, Tsimpli 2001, Tsimpli & Dimitrakopoulou 2007, Tsimpli & Mastropavlou 2008) to observe that [-interpretable] features like gender and case tend to be problematic for L2 acquisition in a way that [+interpretable] features are not.

In the context of our study, since the agreement in Kannada involves only the [+interpretable] features that need to be acquired in L2, we are presuming that Max F will add to the depth of semantic information encoding, while Min F will prefer minimal semantic depth.

4 The elicitation task

In the code-mixing data elicitation task, the participant had to read out a paragraph of text from the screen. The sentences in the paragraph had blanks in place of the inflected verbs. The verb root was provided in brackets next to the blank (Figure 1) and participants were asked to inflect the verbs appropriately while reading out.

ನಮ್ಮ ಶರೀರವನ್ನು ಉತ್ತಮ ಅರೋಗ್ಯವಂತವಾಗಿ ಇಡಲು ನಾವು ವ್ಯಾಯಾಮ ಮಾಡಬೇಕು. ನಾನು ಪ್ರತಿದಿನ ಸಾಯಿಂಕಾಲ ನನ್ನ ಮನೆಯ ಪಕ್ಕ ಇರುವ ಪಾರ್ಕಿನಲ್ಲಿ ___ [nadi] ಹೋಗುತ್ತೀನಿ. ನನ್ನ ಗೆಳೆಯರು ನನ್ನ ಜೊತೆ ಪ್ರತೀ ದಿನ ಅಲ್ಲಿಯೇ ____[nadi]. ಹಿಂದಿನ ವಾರ ನನಗೆ ಜಾಸ್ತಿ ಕೆಲಸ ಇರಲಿಲ್ಲ ಅದಿಕೆ ಇನ್ನು ಸಮಯ ಸಿಕ್ತು. ಅದಿಕೆ ಆ ದಿನಗಳಲ್ಲಿ ನಾನು ದಿನ ಎರಡು ಸತಿ ____[nadi]

Figure 1. Reading Illustration.

Each participant read out six such paragraphs. While three of these paragraphs had Kannada as the context Language (Set A), three of them has Malayalam (Set B). For each set, the first paragraph has a verb prompt from the same language, the second from the other language and the third has a nonce prompt. Each paragraphs had three blanks that corresponded to an infinitive, present-habitual and a past tense marker. This task design is schematically shown in (16).

(16) Elicitation Task Design

Set	A			B		
Language of the Paragraph	K			M		
Language of Verb prompt	K	M	N	M	K	N
TAM context for Verb Prompt	NF	NF	NF	NF	NF	NF
	Hab	Hab	Hab	Hab	Hab	Hab
	Past	Past	Past	Past	Past	Past

The participant set had 10 Malayalam-Kannada bilinguals, 5 of whom were simultaneous bilinguals (SM), and the other 5 sequential bilinguals (SQ). The SM participants belonged to the age group 18-30, and the SQ participants belonged to the age group 25-50. The SM group includes some children of the migrant SQ population. Hence the age-groups intersect, but do not match.

The participants were first screened for fluency in both Malayalam and Kannada. Further, they were also asked about the age of acquisition of Kannada. For the reading elicitation task, they could choose to read the script in Kannada, Malayalam, or Roman. Accordingly, they were shown the task paragraphs in the script of their choice. The verb prompt was always in the roman script to maintain uniformity.

While most SQ bilinguals chose to read all the paragraphs in the Malayalam script, most SM bilinguals chose the Roman script for the Malayalam and the Kannada script for the Kannada paragraphs. This asymmetry shows that most participants chose a script that they learnt as part of their school education. Most SM bilinguals studied Kannada as part of their schooling in Bangalore. One SM bilingual chose to read both languages in Roman script. In the following section, as we present the results of the elicitation choice, we will see if the choice of script had some impact in priming for the syntactic frame of the language. Further, it might also have an impact on the priming for the vocabulary item irrespective of the frame.

5 Results and analysis

The entire set of 60 responses from the 10 Malayalam-Kannada participants in the code-mixing contexts KM_ and MK_ has been presented in (17). In this section two sets of participants SQ and SM are not differentiated. Once the overall analysis of the inflection

choice has been presented, we will discuss the differences in choice between the two populations in section 6.

(17) Results

Set	A			B		
Language of the Paragraph	K			M		
Language of Verb prompt	M *paDi* 'study'			K *kari* 'call'		
TAM context for Verb Prompt	NF	Hab	Past	NF	Hab	Past
Suffix K	*-yakke*	*-yutta:re*	*-ide*	*-yakke*	*-yutta:re*	*-daru*
Suffix M	*-kkya:n*	*-kkyum*	*-ccu*	*-kkya:n*	*-kkyum*	*-ccu*
SQ1	M	M	M	M	M	M
SM2	M	M	M	M	M	M
SM1	M	M	M	K	K	K
SQ4	M	M+K	~K	K	K+M	K+M
SM3	K	K	K	~M	~M	M
SM4	K	K	M	~M	~M	M
SQ5	K	T	~K	#M	#M	#M
SQ2	K	~K	K	M/K	M/K	M
SQ3	K	~K	~K	K	~M	K
SM5	K	K	~K	K	K	K

The results from the 10 participants in the study can be broadly categorized into four types. The diacritic (~) is used K to indicate absence of or incorrect agreement marking. With M the diacritic (~) shows that the affix initial [k] was not geminated. The diacritic (#) with M indicates that the participant reinterpreted the Kannada verb prompt as a similar sounding but contextually inappropriate verb from Malayalam and added Malayalam inflection to it.

The first group (SQ1 and SM2) consistently used the Malayalam inflectional form with both Kannada and Malayalam verbs ignoring the language of the paragraph altogether. We refer to them as the "pakka mallus". By ignoring the language of the paragraph, they bypassed the problem of ϕ agreement altogether. The evaluation table in (18) shows that either Min F or MT Faith weighed up by a factor *(+x)* would result in such an output. Weighing up Min F would indicate that the choice is syntactically driven while weighing up MT Faith would indicate that the choice is an acquisition related phenomenon. Since one of the participants is a sequential bilingual and the other a simultaneous bilingual, it is possible that both of them conflated at the same result due to different factors.

(18) Evaluation Table for uniform M inflection

KM_	Min F	MT Faith	Max F	MT anti-Faith	Match FT	Match VP	HV
	w=1(+x)	*w=1(+x)*	*w=1*	*w=1*	*w=1*	*w=1*	
☞KMM			-1	-1	-1		-3
KMK	-1	-1				-1	-3+*(-x)*

MK_	Min F	MT Faith	Max F	MT anti-Faith	Match FT	Match VP	HV
	w=1(+x)	*w=1(+x)*	*w=1*	*w=1*	*w=1*	*w=1*	
☞MKM			-1	-1		-1	-3
MKK	-1	-1			-1		-3+*(-x)*

The second group (SM1 and SQ4) show a slight inter-participant variation. While both prefer to inflect in the language of the verb prompt SQ4 shows a further distinction between non-finite and finite contexts. This distinction, indicates a clear impact of the syntactic frame since ϕ agreement needs to appear only in finite contexts. In such scenario, SQ4 shifts to a "do support" structure. The evaluation tables for these two participants are shown in (19) and (20) respectively.

(19) Evaluation Table for SM1

KM_	Match VP	Min F	Max F	MT Faith	MT anti-Faith	Match FT	HV
	w=1(+x)	*w=1*	*w=1*	*w=1*	*w=1*	*w=1*	
☞KMM			-1		-1	-1	-3
KMK	-1	-1		-1			-3+*(-x)*
MK_	Match VP	Min F	Max F	MT Faith	MT anti-Faith	Match FT	HV
	w=1(+x)	*w=1*	*w=1*	*w=1*	*w=1*	*w=1*	
MKM	-1		-1		-1		-3+*(-x)*
☞MKK		-1		-1		-1	-3

(20) Evaluation Table for SQ4

a. Non-finite context

KM_	Match VP	Min F	Max F	MT Faith	MT anti-Faith	Match FT	HV
	w=1(+x)	*w=1*	*w=1*	*w=1*	*w=1*	*w=1*	
☞KMM					-1	-1	-2
KMK	-1			-1			-2+*(-x)*
MK_	Match VP	Min F	Max F	MT Faith	MT anti-Faith	Match FT	HV
	w=1(+x)	*w=1*	*w=1*	*w=1*	*w=1*	*w=1*	
MKM	-1				-1		-2+*(-x)*
☞MKK				-1		-1	-2

b. Finite context

KM_	Match FT	Match VP	Min F	Max F	MT Faith	MT anti-Faith	HV
	w=1(+x)	*w=1*	*w=1*	*w=1*	*w=1*	*w=1*	
KMM	-1			-1		-1	-3+*(-x)*
☞KMK		-1	-1		-1		-3

MK_	Match FT	Match VP	Min F	Max F	MT Faith	MT anti-Faith	HV
	w=1(+x)	*w=1*	*w=1*	*w=1*	*w=1*	*w=1*	
☞MKM		-1		-1		-1	-3
MKK	-1		-1		-1		-3+*(-x)*

There is no distinction in functional features on the Malayalam and Kannada in the non-finite context, so Min F and Max F are not evaluated in (20a).

The third set (SM3, SM4, SQ5 and SQ2) includes participants who have preferred to inflect in the language of the text paragraph in both finite and non-finite contexts. Of these, the outputs of SM3 and SM4 are near identical, with just one variance. SM4 chose to uniformly use Malayalam past inflection, even when the text paragraph was Kannada. Given our earlier discussion of the Malayalam past suppletive morphology (see 11 and 12 for reference), we find that the general rule for Malayalam past vocabulary insertion is ordered over the regular past in Kannada. The Evaluation table for SM3 and SM4 is shown in (21).

(21) Evaluation Table for SM3 and SM4

KM_	Match FT	Match VP	Min F	Max F	MT Faith	MT anti-Faith	HV
	w=1(+x)	*w=1*	*w=1*	*w=1*	*w=1*	*w=1*	
KMM	-1			-1		-1	-3+*(-x)*
☞KMK		-1	-1		-1		-3
MK_	Match FT	Match VP	Min F	Max F	MT Faith	MT anti-Faith	HV
	w=1(+x)	*w=1*	*w=1*	*w=1*	*w=1*	*w=1*	
☞MKM		-1		-1		-1	-3
MKK	-1		-1		-1		-3+*(-x)*

The evaluation table in (21) is also applicable to the SQ5 and SQ2. SQ5 in fact seems to distinguish Malayalam from Non-Malayalam, so the Non-Malayalam set includes both Kannada past marking without agreement as well as Tamil for the present-habitual. SQ2 differed from (21) in producing a second alternative output MKK for the MK_ context. By doing so, we see that the participant varys between weighing up Match FT and Match VP. The latter option is however superseded by the suppletive morphology of Malayalam in the Past context, just like SM4.

The fourth and final group (SQ3 and SM5), have chosen Kannada inflections in both KM_ and MK_ contexts. Once again, like the first group, the context has been ignored and the choice is either based on syntactic structure (Max F) or MT (MT anti-Faith). The evaluation table for this group is shown in (22).

(22) Evaluation Table for uniform K inflection

KM_	Max F	MT anti-Faith	Min F	MT Faith	Match FT	Match VP	HV
	w=1(+x)	*w=1(+x)*	*w=1*	*w=1*	*w=1*	*w=1*	

KMM	-1	-1			-1		-3+*(-x)*
☞KMK			-1	-1		-1	-3
MK_	Max F	MT anti-Faith	Min F	MT Faith	Match FT	Match VP	HV
	w=1(+x)	*w=1(+x)*	*w=1*	*w=1*	*w=1*	*w=1*	
MKM	-1	-1				-1	-3+*(-x)*
☞MKK			-1	-1	-1		-3

While consistently favoring MT anti-Faith, the sequential bilingual SQ3 went with the geminate-less Malayalam inflection in the habitual MK_ context instead of the agreement-less Kannada inflection that they used in the KM_ context. This small anomaly aside, our evaluation framework has successfully analyzed every piece of code-mixing data that was elicited in the task.

6 Discussion

In (23) we re-present the results table from (17) sorted according to SQ and SM.

(23) Results table sorted by SQ and SM

Set	A			B		
Language of the Paragraph	K			M		
Language of Verb prompt	M *paDi* ‘study’			K *kari* ‘call’		
TAM context for Verb Prompt	NF	Hab	Past	NF	Hab	Past
Suffix K	*-yakke*	*-yutta:re*	*-ide*	*-yakke*	*-yutta:re*	*-daru*
Suffix M	*-kkya:n*	*-kkyum*	*-ccu*	*-kkya:n*	*-kkyum*	*-ccu*
SM1	M	M	M	K	K	K
SM3	K	K	K	~M	~M	M
SM4	K	K	M	~M	~M	M
SM2	M	M	M	M	M	M
SM5	K	K	~K	K	K	K
SQ2	K	~K	K	M/K	M/K	M
SQ4	M	M+K	~K	K	K+M	K+M
SQ5	K	T	~K	#M	#M	#M
SQ1	M	M	M	M	M	M
SQ3	K	~K	~K	K	~M	K

The simultaneous bilinguals are expected to be more balanced in the relative proficiency between the two languages, than sequential bilinguals. The higher proficiency in Kannada, the L2, is clearly apparent in the fact that there is only one instance of (~K) among in the SM responses, whereas five out of seven cases of Kannada inflection usage in the SQ responses are instances of (~K). In contrast, the SM group has four instances of (~M) as opposed to a single instance in SQ. Considering that Malayalam is the L1 for all the

participants and nobody made any (~M) error in the control contexts with nonce, it is likely that all the (~M) errors are the effect of MT anti-Faith, where the participants produce an output that is morpho-syntactically M, but phonologically unlike M. If this is an innovation, then it is interesting to note that such innovation is more common among the SM group than the SQ.

The "do support" strategy is also conspicuous by its total absence from the SM group. Although the sample size of this pilot study is very small, it indicates that the SM group is comfortable to code-mix in the context of featural mismatch between the two languages. They have used three different strategies to code-mix. For the first and second strategy, they weighed up the Match VP and Match FT constraints respectively. In the third case, they went with Min F and Max F producing uniform M and K inflections. Thus, the MT-based constraints did not have any significant influence on the SM participants.

The SQ participants were a lot more varied in their outputs, and the cases of dual response, "do support", and complete lexical change in the verb (#M) responses indicate that this group is not so comfortable with code-mixing in the featural mismatch context. Their responses can be grouped into two sets based on whether they weigh up the
Match FT constraint, or one of the MT-based constraints. Evidence for the argument that uniform K inflection in SQ3 is a consequence of weighing up MT-anti faith rather than Max F, comes from the fact that the output lacks agreement marking.

References

Adger, D. (2003). Core syntax: A minimalist approach. Oxford, UK: Oxford University Press.

Alexiadou, A. & Anagnostopoulou, E. (1998). Parametrizing AGR: word order, V-movement and EPP-checking. Natural Language and Linguistic Theory, 16, 491-539.

Bel, A. (2003). The syntax of subjects in the acquisition of Spanish and Catalan.

Borer, H. (2003). Exo-skeletal vs. Endo-skeletal explanations: syntactic projections and the lexicon. In M. Polinsky and J. Moore (Eds.), The Nature of Explanation (pp.31-67). Chicago: Chicago University Press, (CSLI).

Borer, H. (2015). The category of roots. In A. Alexiadou, H. Borer and F. Schäfer (Eds.), The Roots of Syntax, the Syntax of Roots (pp.).Oxford: Oxford University Press.

Borer, H. (2015). Wherefore roots? Theoretical Linguistics 2014; 40(3/4), 343 – 359.

Chomsky, N. (1970). Remarks on nominalization. In R. Jacobs & P. Rosenbaum (Eds.), Readings in English transformational grammar (pp. 184–221). Waltham, MA: Ginn-Blaisdell.

Chomsky, N. (1981). Lectures on government and binding. Dordrecht, Netherlands: Foris.

Chomsky, N. (1995). The minimalist program. Cambridge, MA: MIT Press.

Chomsky, N., & Lasnik, H. (1993). The theory of principles and parameters. In A. Jacobs, W. von Stechow, & T. Vennemann (Eds.), Syntax. An international handbook of contemporary research (pp. 506–569) Berlin: de Gruyter.

Díaz, L., Bel, A., & Bekiou, K. (2017). Interpretable and uninterpretable features in the acquisition of Spanish past tenses. In *The role of formal features in second language acquisition* (pp. 484-512). Routledge.

Embick, D. (2015). The Morpheme: A Theoretical Introduction (Vol. 31). Walter de Gruyter GmbH & Co KG.
Embick, D. and Noyer, R. (2007). Distributed morphology and the syntax/morphology interface. In Oxford handbook of linguistic interfaces, chapter 9. Oxford : Oxford University Press
Grimstad, M. B., Riksem, B. R., Lohndal, T., & Åfarli, T. A. (2018). Lexicalist vs. exoskeletal approaches to language mixing. The Linguistic Review, 35(2), 187-218.
Halle, M. & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale & S.J. Keyser (eds.), The View from Building 20 (pp.111–176). Cambridge, Mass.: MIT Press.
Harley, H. (2014). On the identity of roots. Theoretical linguistics, 40(3-4), 225-276.
Hulk, A., & Müller, N. (2000). Cross-linguistic influence at the interface between syntax and pragmatics. Bilingualism: Language and Cognition, 3(III), 227–244
Legendre, G., Miyata, Y., & Smolensky, P. (2022). Harmonic Grammar–A formal multi-level connectionist theory of linguistic well-formedness: An application. In *12th Annual Conference. CSS Pod* (pp. 884-891). Psychology Press.
Liceras, J. M., Zobl, H. & Goodluck, H. (2008). Introduction. In J. M. Liceras, H. Zobl & H. Goodluck (Eds.), The role of formal features in second language acquisition (pp.1-19). New York: Lawrence Erlbaum Associates. Taylor & Francis Group.
MacSwan, J. (2014). A minimalist approach to intrasentential code switching. Routledge.
Mahootian, S. (1993). A null theory of codeswitching. Northwestern University.
Muysken, P. (2000). Bilingual speech: A typology of code-mixing. Cambridge University Press.
Myers-Scotton, C. M. (1993). Duelling languages: Grammatical structure in codeswitching. Oxford: Oxford University Press. Myers-Scotton, C. M. (2001). The matrix language frame model: Developments and responses. In R. Jacobson (ed.), Codeswitching worldwide II, pp. 23–58. Berlin: Mouton de Gruyte
Prince, A., & Smolensky, P. (2003). Optimality Theory in phonology. *International encyclopedia of linguistics*, *3*, 212-22.
Riksem, B. R. (2018). Language mixing in American Norwegian noun phrases. Journal of Language Contact, 11(3), 481-524.
Sorace, A. (2003). Near-nativeness. In M. Long & C. Doughty (Eds.), Handbook of second language acquisition (pp. 130–152). Oxford, UK: Blackwell.
Tsimpli, I.-M. (2001). LF-interpretability and language development: A study of verbal and nominal features in Greek normally developing and SLI children. Brain and Language, 77, 432–448.
Tsimpli, I.-M. & Dimitrakopoulou, M. (2007). The Interpretability Hypothesis: evidence from wh-interrogatives in second language acquisition. Second Language Research 23, 215-242.
Tsimpli, I-M. & Mastropavlou, M. (2008). Feature interpretability in L2 acquisition and SLI: Greek clitics and determiners. In J. M. Liceras, H. Zobl & H. Goodluck (Eds.), The role of formal features in second language acquisition (pp. 142-183). New York: Lawrence Erlbaum Associates. Taylor & Francis Group.