Stress placement in English loanwords by speakers of Mirpur Pahari in the UK

Stress placement in English loanwords into Mirpur Pahari (MP) is used to explore whether a usage-based approach can inform incorporation of external factors (e.g. exposure to a donor language - here, English) into formal phonological analysis of loanword adaptation alongside internal factors (e.g. phonology of the recipient language - here, MP). Stress placement in English loanwords into MP shows across- and within-speaker variation between conformity to MP stress rules (formalized in classical Optimality Theory) and retention of stress on the syllable that is stressed in English; this is a challenge for most theories of loanword adaptation. In our hybrid approach, variable adaptation patterns in individual speakers' loanword realizations in production data from 12 MP speakers in the UK are correlated with degree of exposure to English operationalized as vocabulary size, and a unified formal account is sketched through usage-based weighting of constraints in Stochastic OT.

Keywords: loanword phonology; Mirpur Pahari; UK English; usage-based phonology

1. Introduction

A key issue in the loanword adaptation literature is whether the phonological properties of loanwords can be accounted for within the phonology of the borrowing language or not (Silverman, 1992). This paper investigates loanword adaptation patterns which address this issue directly. We will show that some speakers' patterns reflect the constraints observed in the phonology of the borrowing language, but other speakers' patterns do not; indeed, for the latter group, we see variable adaptation whereby some lexical items reflect the phonology of the borrowing language, but other lexical items match stress placement in the source language word. This paper explores a unified analysis of internal and external factors governing these patterns of across- and within-speaker variation in adaptation of stress placement in English loanwords into Mirpur Pahari (MP).

Prior cross-linguistic work on loanword stress shows competing effects of faithfulness to stress position in the source language word and of the markedness constraints of the stress grammar of the borrowing language (Alber, 1998; Davidson & Noyer, 1997; Kager, 2000). Faithfulness to the surface form of the source word would seem – necessarily – to require a loanword specific grammar, since by definition a rule or constraint requiring a match to a word in another language cannot be part of the autonomous phonology of the borrowing language (see Kang, 2010 for an overview). A number of accounts have sought to bring such patterns under the umbrella of the phonology of the borrowing language by appeal to the role of L1 perception in the adaptation process (Broselow, 2009; Hamann & Boersma, 2009). In this paper we contribute new evidence of stress adaptation from novel data in an under-researched language (MP) as spoken by heritage speakers in the UK. Our data reveals patterns of variable adaptation in stress position which cannot be accounted for within a single grammar. We thus sketch an account in terms of externally driven change to an individual speaker's grammar, formalized as weighted constraints, building on Message Oriented Phonology (Hall, Hume, Jaeger, & Wedel, 2016). We test the validity of the approach in a small-scale production study.

In section 2 we provide background information about MP, then present key generalizations regarding stress placement stress in English loanwords into MP as spoken in Pakistan, and a formalization in classical Optimality Theory which demonstrates that only a variable grammar can account for the data. We then outline how a usage-based approach might offer a principled way of predicting variable adaptation. In sections 3-4 we present methods and results of a small empirical study designed to test validity of the approach and conclude in section 5 by sketching a unified analysis.

2. Background to the Study

2.1 Mirpur Pahari

Mirpur Pahari (MP), also known as Mirpuri, is a Western Punjabi language spoken by approximately 4 million in Pakistan and 500K in the UK, with no written form (Stow, Pert, & Khattab, 2012). There are several Pahari dialects, but limited research on the phonology of any of them, and Pahari is under-represented in research on UK minority languages (Hussain, 2015). Khan (2012) studied English loanwords into Poonch Pahari and Tabbassam (2003) provides a description of some aspects of MP. The present work is part of a wider study of the phonetics and phonology of MP.

Many regional languages are spoken in Pakistan, including MP, alongside Urdu as the high-status standard language. English is widely spoken as an additional language and used in most official correspondence. In Pakistan, then, both Urdu and English carry prestige, and are used by educated speakers in preference to their regional language in many contexts, and especially in urban settings. In villages, MP speakers are more likely to retain use of MP in all contexts but will have regular contact with speakers from Mirpur City and/or the UK. As a result, the MP lexicon includes many loanwords from both Urdu and English. Some loanwords in our text corpus (see section 2) may have been borrowed into MP from English via Urdu, and in some cases the indirect route is clear from e.g. influence of Urdu orthography. Such cases are excluded here, to focus on tokens which are either clearly borrowed directly from English or which show no signs of borrowing via Urdu.

A large number of Mirpuris settled in the UK from the 1960s onwards, and there is ongoing primary immigration to the UK. British Asian speakers of Mirpuri heritage who live in cities with a large Mirpuri population (such as Bradford and Leeds) will tend to speak their heritage language, MP, at home. UK-based Mirpuris typically make regular visits to Pakistan, and a constant flow of new English words thus reaches even older speakers residing in rural areas in Mirpur. The language contact situation is thus complex, and the resulting variable loanword adaptation patterns are our focus here.

MP has a large consonant inventory, but the question of whether it retains four-way laryngeal contrast in plosives (voiced/voiceless/voiceless aspirated/voiced aspirated) is debated (Khan, 2012;

Tabbassam, 2003). Our data are broadly consistent with the vowel and consonant inventory proposed by Khan (2012) for Poonch Pahari (see Table 1), which we adopt for consistency with the literature. In our data, we observe the following syllable types: CV, CVV, CVC, CVCC, CVVC. Complex onsets are not permitted, and the only complex codas observed are nasal-obstruent sequences.

Table 1. Consonant and vowel inventory of MP (based on Khan, 2012 for Poonch Pahari).

Consonants	Bilabial	Labiodental	Dental	Alveolar	Retroflex	Palatal	Velar	Uvular	Glottal
Plosives	p b pʰ		t̪ d̪ t̪ʰ	t d tʰ			k g kʰ		
Nasals	m		n		ɽ		ŋ		
Fricatives		f v	s z			ʃ	x ɣ		ɦ
Affricates				tʃ dʒ tʃʰ					
Lateral			l						
Trill			r						
Glides						j			

Oral Vowels	Front	Central	Back
Close	i: ɪ		ʊ u:
Mid	e e:	ə	o o:
Open	æ æ:	a:[1]	

Nasal Vowels	Front	Central	Back
Close	ĩ:		ũ:
Mid	ẽ:		
Open		ã:	

The eastern Punjabi language has word-level lexical tone, alongside word-level stress (Bhatia, 1993), and presence of tone is correlated with loss of voicing contrast in plosives. Our investigations in MP to date suggest that tone is not productively contrastive, though acoustic differences are present in MP productions of cognate words which are tonally contrastive in Punjabi (Khan, 2017; Khan, Xu, & Sohail, 2020). The status of these tonal contrasts is tangential to the goals of this study, however, which relate solely to position of word-level stress, for which there is clear evidence (Shafi, 2019).

Primary stress is assigned in MP according to the algorithm in (1), as illustrated in (2).

(1) a) stress a final superheavy syllable,

[1] Khan (2012) used the symbol /a/ to represent the central vowel /ɐ/ for typographic convenience, and we adopt the same approach. This inventory does not include marginal consonants such as /ʒ/ which appear only in loanwords.

b) else the penultimate syllable (‘penult’)

(2)	a)	pə.ˈsənɖ	choice	dər.ˈbaːr	shrine
	b)	ˈɖər.zən	seamstress	ˈtʃaː.vəl	rice
		ˈxəs.ra	measles	ˈsoː.ti	cane
		bə.ˈraɖ.rɪ	caste	tʃəp.ˈɽaː.si	assistant

We have not found evidence in our data of secondary stress and thus assume that a single metrical foot is built per word. Stress never falls on an open/light CV syllable and the maximal monomorphemic word size appears to be trisyllabic (since in our data all words of four syllables or more are compounds). In the context of cross-linguistic stress typology, in Metrical Stress Theory (Hayes, 1995), the MP stress system would be analyzed by proposing bimoraic trochees built right to left, with main stress on the rightmost foot; consonant extrametricality applies, as well as a complete ban on degenerate feet. We formalize the MP stress grammar in classical Optimality Theory (Prince & Smolensky, 2004), based on Shafi (2019), in section 2.3.

2.2 Stress Placement in English loanwords in Mirpur Pahari (as spoken in Pakistan)

A text corpus of 1219 loanwords was created by the first author, who is a first language speaker of MP (Shafi, 2019). The corpus is based on grammaticality judgements obtained from family members in Pakistan of different generations, without recording audio production data, though pronunciation details were checked in informal phone elicitation sessions. A broad IPA transcription indicates the typical/expected pronunciation of each loanword, including stress position.

The loanword adaptation patterns in the text corpus fall into two categories, roughly matching varying exposure to English and educational background of the speakers’ whose judgements were obtained. One group of loanword patterns can be categorized as that of *older speakers*. Typically,

speakers who would produce this pattern are aged 45-75 years, with little or no education and almost no exposure to English. The second group of loanword patterns can be categorized as that of *younger speakers*, who acquired English as an additional language after MP as first language; speakers who would produce these patterns are aged 18-45 years, with diverse educational backgrounds ranging from 8-14 years of education, and with more exposure to English than older speakers. These two categories align broadly with Haugen's (1950) categories of loanword adaptation patterns arising in the context of pre-bilingualism (for patterns reflecting usage of older speakers) and adult bilingualism (for patterns reflecting usage of younger speakers).

In this text corpus of MP as spoken in Pakistan, Shafi (2019) reports four adaptation patterns, as in (3).[2]

(3)

	gloss	*English*	*MP loanword*
Pattern A	refuse	rɪ.ˈfjuːz	rəf.ˈjuːz
	public	ˈpʌb.lɪk	ˈpəb.lək
	inspector	ɪn.ˈspɛk.tə	əns.ˈpæk.tər
Pattern B	appendix	ə.ˈpɛn.dɪks	ˈpæn.dəs
	decision	dɪ.ˈsɪ.ʒən	də.ˈsiː.ʒən
	lettuce	ˈlɛ.tɪs	ˈlæː.təs
Pattern C	stadium	ˈsteɪ.dɪəm	sə.ˈteː.dɪəm
	ambulance	ˈæm.bju.ləns	ˈæm.bʊ.læns
	vaccine	ˈvæk.siːn	ˈvæk.siːn
Pattern D	shampoo	ʃæm.ˈpuː	ˈʃæm.puː

[2] We do not consider segmental adaptation patterns such as deletion, substitution or adaptation of individual phonemes, and these have not been investigated in this dataset as it is a text-only corpus. We discuss adaptations involving changes to syllable structure where relevant to analysis of stress (see Shafi, 2019 on variable adaptation of phonotactics). For MP speakers in Bradford the English input form would be Northern British English not Southern, thus [ˈpʊb.lɪk] for 'public' not [ˈpʌb.lɪk]; however, as we do not analyse segmental adaptation, we set aside this aspect of input variation for now.

impact (noun)	ˈɪm.pækt	əm.ˈpækt
glucose	ˈgluː.kəʊz	gəl.ˈkoːz

In pattern A, stress in English falls in a position that happens already to meet the MP stress rules (see section 2.1), that is, on a final superheavy syllable (‘refuse’ [rɪ.ˈfjuːz]$_{Eng}$) or on a heavy penult syllable (‘public’ [ˈpʌb.lɪk]$_{Eng}$).

In pattern B, stress is maintained on the syllable which is stressed in English at the cost of some kind of structural change. For example, in ‘decision’ the vowel of the stressed penult is lengthened [dɪ.ˈsɪ.ʒən]$_{Eng}$ →[də.ˈsiː.ʒən]$_{MPL}$) creating a suitably heavy penult syllable to bear stress. Similarly, in ‘appendix’, the word-final cluster [ks] is simplified, avoiding attraction of stress to a final superheavy and away from the penult which bears stress in English ([ə.ˈpɛn.dɪks]$_{Eng}$ → [ˈpæn.dəs]$_{MPL}$). The result is a word which retains stress on the syllable it is on in English, while still meeting the MP stress rules.

In patterns C and D, the differing demands of keeping stress on the syllable it is on in English versus meeting the MP stress rules are not both met. In pattern C, stress placement matches the source word at the expense of meeting the MP stress rules; for example, in ‘ambulance’ the final superheavy is retained but does not attract stress: [ˈæm.bju.ləns]$_{Eng}$ → [ˈæm.bʊ.læns]$_{MPL}$. In contrast, in pattern D, stress is placed according to MP stress rules, resulting in stress on a different syllable to that stressed in English; in ‘shampoo’ stress falls on the heavy penult rather the (non-superheavy) final syllable ([ʃæm.ˈpuː]$_{Eng}$ → [ˈʃæm.puː]$_{MPL}$), and in the noun ‘impact’ stress is attracted to a final superheavy, rather than falling on the penult as it does in English ([ˈɪm.pækt]$_{Eng}$ → [əm.ˈpækt]$_{MPL}$).

All four patterns are observed in the portion of the text corpus which reflects the production of younger speakers of MP in Pakistan, who are educated to high school level or higher and thus learn English as a foreign language at school. In contrast, pattern C is *not* found in the text corpus data which

reflects production of loanwords by older MP speakers, who generally have a lower level of education, and who have not learned English. Older speakers are expected to produce these words as in (4).

(4)

gloss	*English*	*MP loanword young speakers*	*MP loanword older speakers*
stadium	ˈsteɪ.dɪəm	sə.ˈteː.dɪəm	əs.tə.ˈdɪəm
ambulance	ˈæm.bju.ləns	ˈæm.bʊ.læns	æm.ˈboː.lə̃s
vaccine	ˈvæk.siːn	ˈvæk.siːn	vək.ˈsiːn

We see 'macro' variation between younger and older speakers, therefore, in that younger speakers' judgements permit 'pattern C', but older speakers' judgements do not. Within the younger speakers' patterns, we also see variable adaptation of *structurally parallel words*: some are adapted with stress on the same syllable stressed in English (pattern C) but others reflect the MP stress rules (pattern D). An example is the structural minimal pair in (3), 'vaccine'~'glucose', both of which involve a heavy penult followed by a final superheavy syllable [CVC.CVVC][3], but which are adapted by younger speakers in two different ways.

In the next section we set out the formalization of MP stress in Optimality Theory offered by Shafi (2019), to identify the different and/or variable grammars which yield the surface loanword adaptation patterns in (3) and (4).

2.3 Modelling variable adaptation

2.3.1 A formal account within Optimality Theory

[3] As noted in section 2.1, complex onsets such as *[gl] are not permitted in MP, thus the initial syllable adjusts from [glu] to [gəl] (see Shafi, 2019 for further details).

In classical Optimality Theory (OT) surface forms arise from the interaction of ranked constraints (McCarthy, 2008). Markedness constraints refer only to properties of candidate surface forms, but faithfulness constraints refer also to a putative input form. Differences between grammars (and thus languages) arise in the theory from differences in constraint rankings alone, regardless of the properties of the putative input form (Prince & Smolensky, 2004). In our analysis of English loanwords into MP, we illustrate the analysis using surface English forms as input, but in principle the grammar should successfully predict observed surface form(s) whatever the input form.

The stress patterns of MP in (2), can be analyzed in an OT grammar using the constraints in (5).

(5) a) FT-BIN: Feet are binary (under moraic analysis).

b) PARSE-σ: Syllables are parsed into feet.

c) ALLFTR (Ft, R, PWd, R): Every foot is at the right edge of the Prosodic Word.

d) STRESS-TO-WEIGHT ('SWP'): If stressed, then heavy (Crosswhite, 1998).

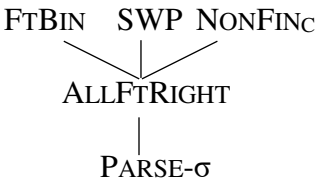
e) NON-FINALITY[C,w]: No mora-level gridmark occurs over the final consonant of a prosodic word (Hyde, 2011).

The first three – FT-BIN, PARSE-σ, and ALLFTR – are standard constraints indicated in the cross-linguistic typology of stress (Kager, 2007). We see the effect of Stress-to-Weight ('SWP') in the fact that all stressed syllables in MP are heavy, without exception. Finally, we use Hyde's (2011) grid-based definition of Non-Finality to capture robustly observed consonant extrametricality in MP.[4]

We motivate the MP stress grammar constraint ranking in (6) through worked examples in (7)-(8). Foot structure is marked with parentheses '( )', and extrametricality with angled brackets '< >'.

[4] Since stress is only ever realized on heavy syllables in MP, it is possible that, in order to satisfy SWP, an input short vowel may be lengthened in violation of a faithfulness constraint $\text{IDENT}_{[\text{long-V}]}$. In interests of brevity we set this issue aside here, but an analysis along these lines is explored in detail in Shafi (2019).

(6)



(7)

	/tʃaːvəl/ ‘rice’	FTBIN	SWP	NONFINC	ALLFTR	PARSE-σ
☞	a. (ˈtʃaː).vəl				*	*
	b. tʃaː.(ˈvəl)			*!		*
	c. tʃaː.(ˈvə)<l>	*!	*!			*
	d. (tʃaː).(ˈvəl)			*!	*	
	e. (tʃaː).(ˈvə)<l>	*!	*!		*	

(8)

	/dərbaːr/ ‘shrine’	FTBIN	SWP	NONFINC	ALLFTR	PARSE-σ
☞	a. dər.(ˈbaː)<r>					*
	b. (dər).(ˈbaː)<r>				*!	

In (7), the observed surface form is candidate a., which violates ALLFTR and PARSE-σ. Candidates b. and c. demonstrate ranking of FTBIN, SWP and NONFINC over ALLFTR; candidates d. and e. demonstrate ranking of FTBIN, SWP and NONFINC over PARSE-σ. Ungrammaticality of candidates d. and e. indicates that only one foot is formed in MP. The example in (8) shows ranking of ALLFTR over PARSE-σ.

Turning now to analysis of loanwords, the three adaptation patterns in the portion of the corpus reflecting typical productions of *older speakers* – patterns A, B and D – can all be analysed within the MP grammar, using the same constraints and ranking in (6). This is demonstrated by a worked example of a pattern B case, in (9).

(9)

	[ˈlɛ.tɪs]$_{Eng}$ 'lettuce'	FTBIN	SWP	NONFINC	ALLFTR	PARSE-σ
☞	a. (ˈlæː).təs				*	*
	b. (ˈlɛ).tɪs	*!	*!		*	*
	c. læː.(ˈtə)<s>	*!	*!			*
	d. læː.(ˈtəs)			*!		*
	e. (ˈlɛ.tɪ)<s>		*!			
	f. (ˈlɛ.tɪs)	*!	*!	*!		
	g. (lɛ).(ˈtɪs)	*!		*!	*	
	h. (ˈlɛ).(tɪs)	*!	*!	*!	*	

In (9), the surface form typically used by older speakers is candidate a. Candidates c-f. demonstrate ranking of FTBIN, SWP and NONFINC over ALLFTR in loanwords, as in home-grown MP words; candidates e-h. demonstrate ranking of FTBIN, SWP and NONFINC over PARSE-σ.

A further worked example in (10) confirms that pattern D loanwords can also be analysed within the autonomous MP grammar (setting aside segmental changes to resolve the complex onset).

(10)

	[ˈgluː.kəʊz]$_{Eng}$ 'glucose'	FTBIN	SWP	NONFINC	ALLFTR	PARSE-σ
☞	a. gəl.(ˈkoː)<z>					*!
	b. (ˈgəl).koːz				*!	*!

c. gəl.(ˈkoːz)			*!		*
d. (gəl).(ˈkoː)<z>				*!	

Pattern C loanwords are found only in the portion of the text corpus reflecting typical productions of younger speakers and are modelled by adopting the (previously proposed) loan-phonology specific constraint , shown in (11), MATCHSTRESS (Davidson & Noyer, 1997), which yields parallel surface effects to the Output-Output constraint MAXSTRESS proposed by Kenstowicz (2007).

(11) MATCHSTRESS: Stress falls on the same vowel in the source word as in the loan word.

(12)

[ˈvæk.siːn]$_{Eng}$ ‘vaccine’	MATCH	FTBIN	SWP	NONFIN$_C$	ALLFTR	PARSE-σ
☞ a. (ˈvæk).siːn					*	*
b. væk.(ˈsiː)<n>	*!					*

(13)

[ˈgluː.kəʊz]$_{Eng}$ ‘glucose’	FTBIN	SWP	NONFIN$_C$	ALLFTR	PARSE-σ	MATCH
☞ a. gəl.(ˈkoː)<z>					*!	*
b. (ˈgəl).koːz				*!	*!	

The example in (12) demonstrates that, in pattern C, MATCHSTRESS is highly ranked. It is not possible to derive patterns C and D within the same grammar, because, in the typical patterns of younger speakers, structurally parallel cases yield conflicting outputs: [ˈvæk.siːn] ~ [gəl.ˈkoːz]. Pattern

C requires a grammar in which MATCHSTRESS outranks ALLFTR, as in (12), but pattern D requires a grammar in which ALLFTR outranks MATCHSTRESS, as in (13).

Patterns A and B both involve stress falling on the vowel that is stressed in the English source word, so the relative ranking of MATCHSTRESS does not impact the analysis. Patterns A and B are thus also compatible with a grammar containing the loan-specific constraint MATCHSTRESS, as in (14); this is the same pattern B case shown in (9) for older speakers, without MATCHSTRESS.

(14)

	[ˈlɛ.tɪs]$_{Eng}$ 'lettuce'	MATCH	FTBIN	SWP	NONFIN$_C$	ALLFTR	PARSE-σ
☞	a. (ˈlæː).təs					*	*
	b. (ˈlɛ).tɪs		*!	*!		*	*
	c. læː.(ˈtə)<s>	*!	*!	*!			*
	d. læː.(ˈtəs)	*!			*!		*
	e. (ˈlɛ.tɪ)<s>			*!			
	f. (ˈlɛ.tɪs)		*!	*!	*!		
	g. (lɛ).(ˈtɪs)	*!	*!		*!		
	h. (ˈlɛ).(tɪs)		*!	*!	*!	*	

Overall, then, we can model macro-variation between younger and older speakers as a change in the grammar, brought about by introduction of the loanword-specific MATCHSTRESS constraint. However, it is not possible to model intra-speaker variable adaptation between patterns C and D by younger speakers within a single grammar; external factors must be involved. Given the definition of MATCHSTRESS, a unifying feature in younger speakers' pattern D cases might be frequency. Specifically, if speakers have not had sufficient exposure to the English source word they have nothing

to 'match' to, so MATCHSTRESS is vacuously satisfied and we only see the effects of autonomous MP rules in these words.

The role of external factors could be incorporated into a phonological analysis using a model such as Stochastic OT (Boersma & Hayes, 2001), which assigns variable weightings to constraints to predict which grammar will obtain in a particular case. Such an approach requires a principled way of determining the appropriate weightings to assign. In the next section, we outline an approach inspired by usage-based theory, based on Message Oriented Phonology (Hall et al., 2016), to see if it can offer a way to operationalise 'degree of exposure to English' as part of a variable grammar. The validity of the approach is tested in a small-scale empirical study in sections 3-4.

2.4 The Present Study

As shown above, it is not possible to explain the variable adaptation of English loanwords into MP both within and between speakers, without accepting a change in the grammar triggered by one or more external factors. Specifically, in an OT account of the type illustrated here, some external factor is needed to trigger re-ranking of MATCHSTRESS in the grammar of younger speakers.

In this section we outline an approach which could equip OT with weighted rankings to generate testable predictions. We propose modelling variable adaptation in terms of degree of 'mere exposure', in this case to English. In Message Oriented Phonology (MOP), the contents of each speaker's lexicon form the primary data over which statistical learning applies, thus the lexicon contributes directly to the emergence of phonological categories (Hall et al., 2016). We apply this model to the language contact situation in which lexical items are borrowed from one language into another. We assume a shared lexical space for each speaker between source and target languages: as the proportion of source language lexical items in the lexicon grows (here, English words), so the influence of the stress patterns in those words on the stress grammar increases.

This approach predicts inter-speaker variation in loanword adaptation as the norm, since individual speakers do not share an identical lexicon, but may add new lexical items in a different order or speed. The model also predicts intra-speaker variation since changes to the grammar take place gradually over time as the balance of word types in the lexicon changes. This situation, with both inter- and intra- speaker variation, is exactly the pattern we see in MP loanwords, with overall macro-variation between younger and older speakers, requiring *different* grammars, as well as variable adaptation patterns within the productions of younger speakers, requiring a *variable* grammar.

If this usage-based approach is valid, the amount and type of variable adaptation observed in an individual speaker's productions of English loanwords into MP should correlate with the degree of exposure of each person to English. In the remainder of the chapter, we explore the potential validity of this approach using English vocabulary size as a proxy for degree of exposure to English, in new data from heritage speakers of MP in the UK. We address two research questions: i) what stress patterns do MP speakers in the UK display in MP and in English loanwords into MP? and ii) does any variable adaptation of loanwords correlate with exposure to English measured in English vocabulary size?

3. Methodology

3.1 Materials

A picture naming task was designed to elicit 52 target loanwords from the corpus described in section 2 (see Appendix), selected to yield tokens across adaptation types, but choosing words amenable to visualisation. Picture naming was used because there is no agreed written form for MP; although it is possible to represent MP in Urdu script we expected that some of our UK based participants would not be confident reading MP written in this way. Each word was elicited in isolation in response to a single image. We also used picture naming to elicit a control set of 19 target words in MP, to establish

whether the stress patterns used in MP by the Bradford speakers matched the MP patterns described above for speakers in Pakistan.

To obtain a measure of vocabulary size across participants, we used the Vocabulary Size Test (VST, Nation & Beglar, 2007). VST is a measure of written receptive vocabulary size in English and has undergone validation (Beglar, 2010; Gyllstad, Vilkaite, & Schmitt, 2015). We wanted a test that could be administered on paper, since we would not always have internet access in testing sites, which ruled out an alternative online instrument, the V_YesNo test (Meara & Miralpeix, 2016). VST also allows us to use the same test for participants across a wide range of levels of proficiency in English, including early bilinguals, in contrast to e.g. the earlier Vocabulary Levels Test (Laufer & Nation, 1999). We used the original 140-word version of VST, which tests the first 14000 words of English at a 1 in 100 sampling rate. We favoured this over the 20000 words VST, which, although shorter, uses a lower sampling rate (1 in 200), since it has been found that, even at a 1 in 100 sampling rate, the VST test slightly overestimates actual vocabulary size (Gyllstad et al., 2015). The VST is nevertheless easy to administer and interpret and is a well-established test.

3.2 Participants

First language speakers of MP living in Bradford were recruited through contacts of the first author and the *Born in Bradford* community research team (https://borninbradford.nhs.uk/). Criteria for inclusion were being a first language speaker of MP resident in Bradford, aged 18 or over. A language background questionnaire gives a picture of the language use of each participant, but information provided on that questionnaire was not used as criteria for exclusion. No participants reported any experience with languages other than English, MP or Urdu. We report here responses from 12 participants (f1-f6 female/m1-m6 male). Details of their language background, age at onset of acquisition of English (AOA) and length of residence (LOR) in the UK are in Table 1.

Table 1. Demographic and language profile of the Bradford participants.

	age	*birthplace*	*AOA*	*LOR*	*highest education*	*language use at work*
f1	25	Mirpur	22	3	high school in Pakistan	MP/English
f2	47	Mirpur	6	41	high school in UK	English
f3	45	Mirpur	21	24	primary school in Pakistan	English/Urdu
f4	24	Bradford	n/a	24	undergraduate degree in UK	MP/English
f5	38	Mirpur	26	12	high school in Pakistan	MP/English
f6	36	Mirpur	4	32	high school in UK	MP/English
m1	34	Mirpur	25	9	undergraduate degree in Pakistan	English
m2	44	Bradford[5]	5	39	high school in UK	MP/English
m3	38	Bradford	n/a	38	high school in UK	MP/English
m4	31	Bradford	n/a	31	high school in UK	MP/English
m5	45	Mirpur	28	17	undergraduate degree in Pakistan	MP/English
m6	41	Mirpur	1	40	high school in UK	MP/English

The participant sample is by design heterogenous as to length of time in the UK or language used at work, so has the potential to reveal whether differences in exposure to English impact loanword adaptation.

3.4 Procedure

Data collection sessions were conducted in MP by the first author, who is a first language speaker of MP, and took place in participants' homes or suitable public places such as an office or community

[5] Although born in the UK, m2 lived in Mirpur as a child between the ages of 5-10 years.

centre. Recordings were made in .wav format at 44.1kHz 16 bit on a Marantz PMD620 recorder via a head-mounted Shure SM10A microphone. After provision of informed consent, each participant answered the language background questionnaire, before starting recordings. The loanword picture naming task was one of several elicitation tasks undertaken in MP for a wider study. After finishing all recording tasks in MP, each participant was asked to complete the Vocabulary Size Test in English on paper; the test comprises 140 multiple choice questions in which the participant must choose the correct meaning for a word in English.

3.4 Analysis

The production of each target loanword was phonetically transcribed by the first author. These transcriptions were checked, then coded for adaptation strategy by the second author. Each token was thus classified into adaptation strategies A, B, C or D, described in section 2. The VST answer key was used by the first author to generate a score for each participant. The adaptation strategy category of each token serves as dependent variable, with VST scores as potential predictor. Tests for correlation were run by the second author in R (Core Team, 2014) using Spearman's Rank and Kendall's Tau.

4. Results

All 12 Bradford speakers produced the control set of MP words with the MP stress patterns described in section 2.1, and thus have the same stress grammar in MP as MP speakers in Pakistan.

Figure 1 shows a count of observed tokens of each adaptation pattern type (A/B/C/D) in the 52 elicited loanwords for each speaker; the number of missing items is reported because in some cases the picture naming task failed to elicit the desired target word.

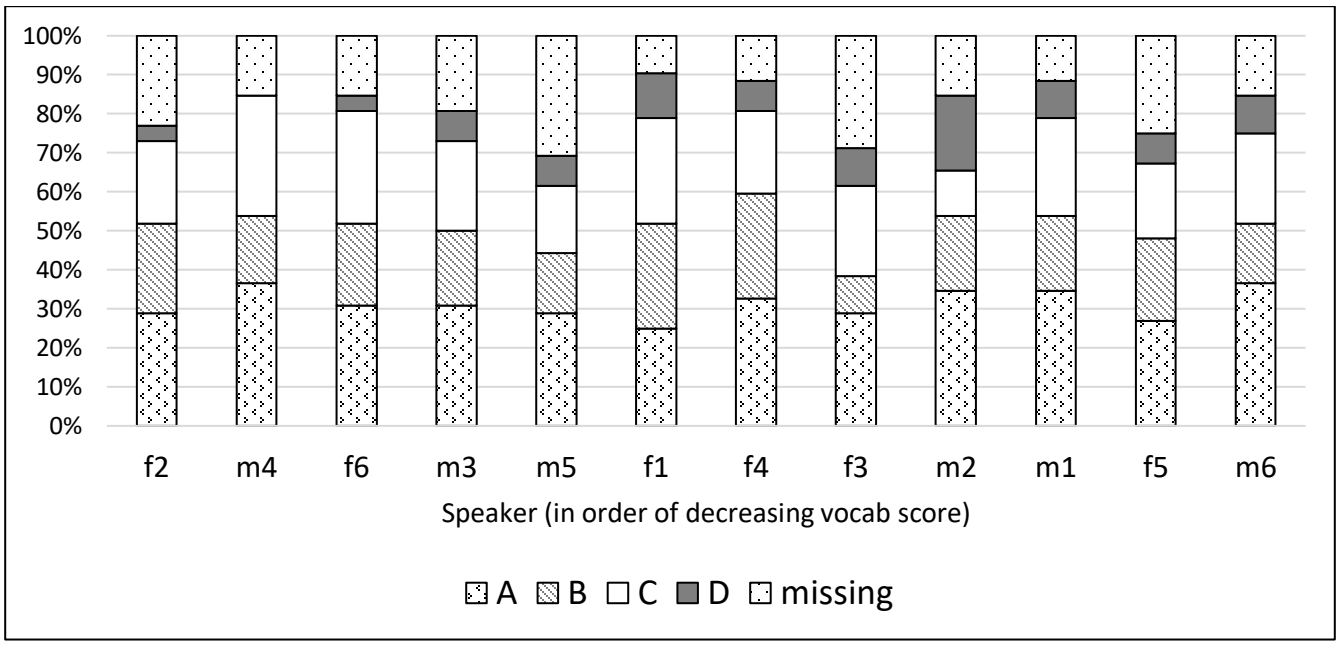


Figure 1. Count of each loanword adaptation pattern type (A/B/C/D/missing) by speaker.

All but one (m4) of the Bradford speakers show all four adaptation patterns, matching the range of variation observed in the text corpus for typical realisations of younger speakers in Pakistan. The age range of the Bradford speakers (24-47 years) roughly matches the age range of younger speakers in Pakistan (18-45 years) whose grammaticality judgements informed the loanword text corpus. The production data from heritage speakers in Bradford allows us to track variation between individuals and thus investigate the potential relationship between degree of exposure to English and use of pattern C versus D. Figure 2 shows examples of the word 'elephant' realised in pattern C (f2) and pattern D (f3).

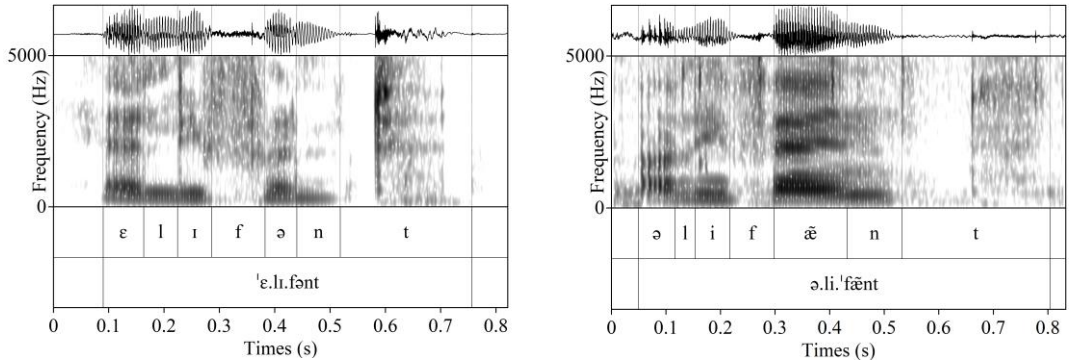


Figure 2. Waveform and spectrogram of 'elephant' produced by speaker f2 (left) and f3 (right).

In Figure 1 the Bradford speakers are ordered left to right by VST score, from highest to lowest. Speaker m6 did not complete the VST and is excluded from further analysis. Speakers' individual VST scores are shown in Figure 3. All speakers have a VST score consistent with being a second language speaker of English. Nation (2012) notes that a VST score of 50-60 (indicating receptive knowledge of the 5000-6000 most frequent word families) is similar to that of an undergraduate student for whom English is an additional language; an L2 English speaking PhD student would typically score 90, approaching the typical vocabulary size of a 13 year old L1 English speaker (10,000 words). In our sample (excluding m6), scores ranged from 49-94 (mean 69.2, SD 22.5).

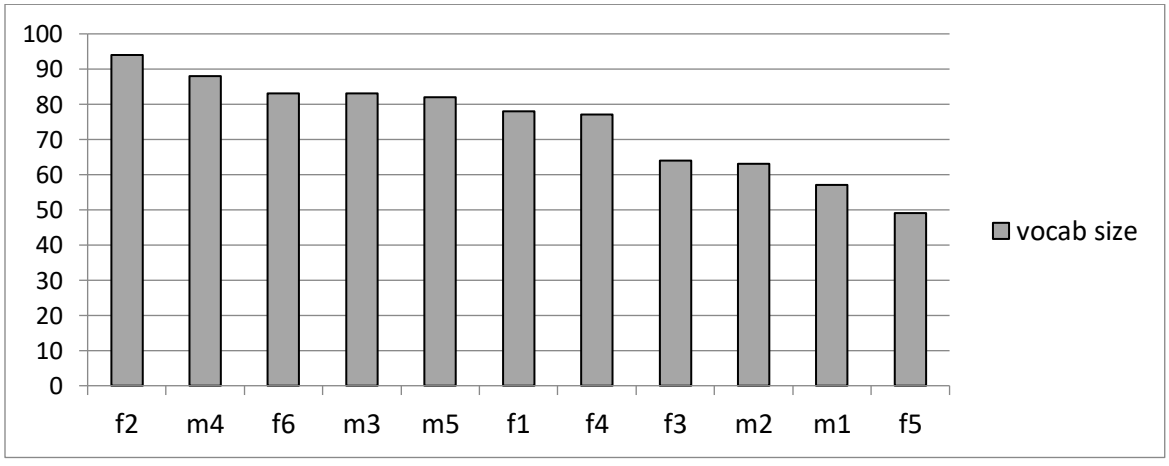


Figure 3. Vocabulary size test score, by speaker.

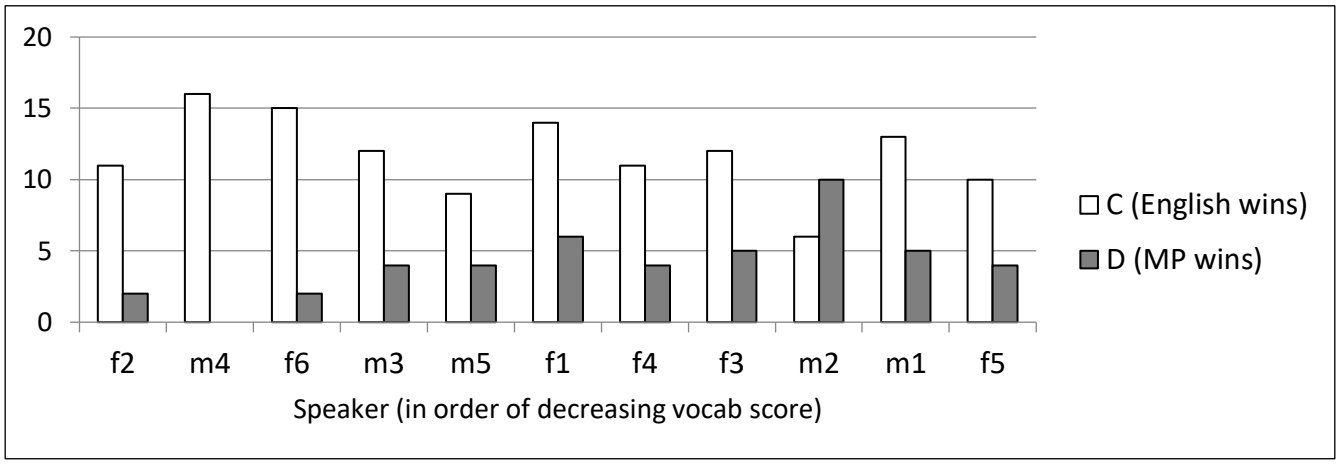


Figure 4. Count of pattern C and D tokens, by speaker.

Figure 4 shows the count of pattern C and D tokens, by speaker. Visual comparison of Figures 3 and 4 suggests an inverse relationship between number of D tokens and a speaker's VST score. Correlation tests confirm an inverse monotonic relationship between vocabulary size and number of D cases (Spearman's rho = -0.7028; p = .0159; Kendall's tau = -0.5557; p = .0238), illustrated in Figure 5, but no such relationship between vocabulary size and number of C cases (Spearman's rho = 0.3867; p = .24; Kendall's tau = 0.299; p = .2088).

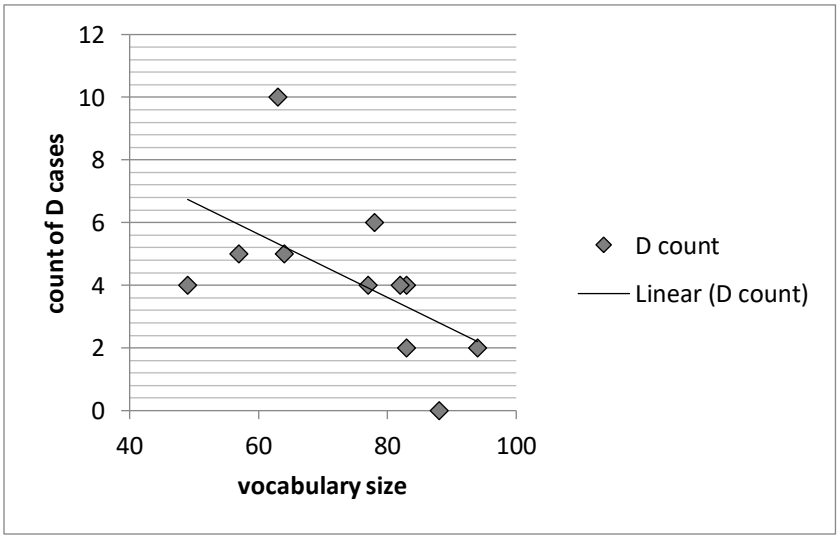


Figure 5. Scatter plot of vocabulary size and count of pattern D tokens, by speaker.

Although the sample size is small, this result is consistent with the hypothesis that exposure to English is inversely related to the number of D cases (which follow MP rules), rather than to the number of C cases (which have stress on the syllable stressed in English). We interpret this as an indication that all the Bradford speakers have a sufficient number of English words in their lexicon to generate an effect of faithfulness to stress placement in the English source word, as would be expected. In OT terms, all the speakers have had sufficient exposure to English to have 'switched' to a grammar in which the effects of MATCHSTRESS are seen. Furthermore, observed intra-speaker variation in residual use of MP-like pattern D co-varies with vocabulary size to a significant extent, with increased exposure to English reflecting reduced tendency to 'ignore' the English source word stress placement in favour of adhering to the rules of MP.

5. Discussion and Conclusions

We have presented patterns observed in a text corpus of English loanwords into MP, based on grammaticality judgements from MP speakers in Pakistan (Shafi, 2019); these indicate variation between younger and older speakers in Pakistan in stress placement in a subset of words (pattern C~D), and variable adaptation by younger speakers of otherwise structurally parallel words. Variation between and within speakers can be formalised in an OT grammar, where the patterns of adaptation favoured by older speakers are modelled with the same constraints and ranking that account for stress assignment in MP itself. A different and/or variable grammar, with re-ranking of MATCHSTRESS, is needed to account for adaptation patterns of younger speakers in Pakistan, but the OT analysis does not offer a way to explain what triggers re-ranking of constraints. A usage-based approach taking degree of exposure to English as a potential external trigger for change was outlined, and its validity tested in

new data collected from 12 MP speakers in the UK. The adaptation patterns of UK heritage speakers of MP proved to be correlated with an objective measure of receptive vocabulary size in English, supporting an account of observed across- and within-speaker variation in terms of exposure to English.

In Stochastic OT (Boersma & Hayes, 2001) the constraint ranking in an individual grammar is implemented as a continuous ranking scale, where higher values represent higher ranking. The distance between any pair of constraints along the scale reflects how fixed their relative ranking is, due to the manner in which candidates are evaluated: in a computational implementation, the position of each constraint (i.e. its numerical value on the scale) is multiplied by a randomly generated value within a standardised probability distribution; constraints whose ranking values are close on the scale will have overlapping probability distributions, yielding a variable ranking effect in the evaluation of candidates. Our analysis follows Feldhausen (2016) who shows how Stochastic OT can model variation between and within the grammars of individual speakers. Figure 6 sketches the relevant part of the stress grammar of MP – relative ranking of MATCHSTRESS and ALLFTRIGHT – for speakers with varying exposure to English, in a model where position of MATCHSTRESS is operationalised as a speaker's VST score: the higher the VST score, the higher MATCHSTRESS is ranked (moving leftwards in the diagram).

Figure 6: Overlapping constraints on a continuous ranking scale in three different scenarios.

Fig. 6 a) represents the ranking for an older speaker of MP in Pakistan, where MATCHSTRESS is low-ranked and its effects are not observed. Fig. 6 b) represents the ranking for a younger speaker of MP in Pakistan and/or a heritage speaker in the UK (e.g. m2), where exposure to English has moved MATCHSTRESS to a higher ranked position; its distribution overlaps with ALLFTRIGHT, yielding a

variable adaptation pattern in which pattern D (favoured by ALLFTRIGHT) is more common than pattern C (favoured by MATCHSTRESS). Fig. 6 c) illustrates the ranking for a speaker with greater exposure to English (e.g. f2), where MATCHSTRESS is now higher-ranked than ALLFTRIGHT; the overlapping distributions still result in variable adaptation but with fewer pattern D tokens. Speaker f2 has a higher VST score than speaker m4 who produced no pattern D, but scenario c) accounts for this outcome for speaker m4 due to the probabilistic manner in which rankings are generated for any one candidate, which could in principle yield a scenario c) sample with no surface pattern D. A speaker with an even higher VST score would have MATCHSTRESS ranked far enough away on the scale above ALLFTRIGHT that their distributions no longer overlap, resulting in no pattern C~D variable adaptation.

Exposure to English is, however, only one external factor likely to affect realisation of loanwords into MP or other languages, and a key limitation of this study is that it is based solely on single word elicitation. Future research using more naturalistic tasks would facilitate investigation of the effects on realization of loanwords of other contextual factors, such as those conducive to code-switching, alongside exposure to English, the role of which is demonstrated and formalized here.

Acknowledgements

Thanks to the Born in Bradford community engagement team for help with recruitment of participants.

Appendix

Table A: Target words elicited from UK participants, with expected realisation in MP by younger speakers, split by potential for adaptation in loanword patterns A, B or C~D (see section 2.2).

A		**B**		**C/D**	
English	**MP**	**English**	**MP**	**English**	**MP**
baby	ˈbeː.bi	alarm	ˈlaː.rəm	ambulance	æm.ˈboː.lə̃s

balloon	bæ.ˈluːn	almond	ˈal.mə̃d	cucumber	kə.ˈkəm.bər
banana	bəˈnaː.nə	appendix	ˈpæn.dəs	cylinder	sə.ˈlæn.dər
basket	ˈbas.kɪt	blender	bə.ˈlæn.dər	elephant	æ.lɪ.ˈfænt
burger	ˈbər.gər	bracelet	bə.ˈræs.lət	helicopter	hæ.lɪ.ˈkap.tər
carpet	ˈkar.pət	broccoli	bə.ˈrok.li	hospital	həs.pəˈt̪aːl
cartoon	kar.ˈtuːn	camera	ˈkæmb.rə	library	læb.ˈreː.ri
curtain	ˈkər.tən	computer	kəmˈpuː.tər	mobile	mə.ˈbaɪl
doctor	ˈdak.tər	decision	də.ˈsiː.ʒən	perfume	pər.ˈfjuːm
giraffe	dʒə.ˈraː.f	engagement	əŋ.ˈgeːdʒ.mɪt	salad	sə.ˈlaːd̪
injection	ən.ˈdʒæk.ʃən	introduce	ɪn.tər.ˈduːs	shampoo	ˈʃæm.puː
inspector	əns.ˈpæk.tər	lettuce	ˈlæː.təs	stadium	sə.teː.ˈdɪəm
monkey	ˈmaŋ.ki	mechanic	məˈkæː.nək	terrorist	tæ.ra.ˈrɪst
onion	ˈon.jən	mushroom	ˈmaʃ.rum		
paper	ˈpeː.pər	patient	ˈpeː.ʃə̃t		
pilot	ˈpaɪ.lət	pepper	ˈpeː.pər		
potato	pə.ˈtæː.to	printer	pə.ˈrĩn.tər		
teacher	ˈtiː.tʃər	student	sə.ˈtuː.də̃t		
tomato	tə.ˈmaː.to	umbrella	əm.ˈreː.la		
zebra	ˈzeb.ra				

References

Alber, B. (1998). Stress preservation in German loan-words. In W. Kehrein & R. Wiese (Eds.), *Phonology and morphology of the Germanic languages* (pp. 113-114). Tübingen: Niemeyer.

Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language Testing, 27*(1), 101-118.

Bhatia, T. K. (1993). *Punjabi: A Cognitive Descriptive Grammar*. London: Routledge.

Boersma, P., & Hayes, B. (2001). Empirical tests of the gradual learning algorithm. *Linguistic Inquiry, 32*(1), 45-86.

Broselow, E. (2009). Stress adaptation in loanword phonology: Perception and learnability. In P. Boersma & S. Hamann (Eds.), (pp. 191-234). Berlin: Mouton de Gruyter.

Core Team, R. (2014). *R: A language and environment for statistical computing*. Retrieved from Vienna, Austria: http://www.R-project.org/

Crosswhite, K. (1998). Segmental vs. prosodic correspondence in Chamorro. *Phonology, 15*(03), 281-316.

Davidson, L., & Noyer, R. (1997). Loan phonology in Huave: nativization and the ranking of faithfulness constraints. *Proceedings of the West Coast Conference on Formal Linguistics, 15*, 65-79.

Feldhausen, I. (2016). Inter-speaker variation, Optimality theory, and the prosody of clitic left-dislocations in Spanish. *Probus, 28*(2), 293-333.

Gyllstad, H., Vilkaite, L., & Schmitt, N. (2015). Assessing vocabulary size through multiple-choice formats: Issues with guessing and sampling rates. *International Journal of Applied Linguistics, 166*(2), 278-306.

Hall, K. C., Hume, E., Jaeger, T. F., & Wedel, A. (2016). *The message shapes phonology*. Ms. UBC, University of Canterbury, University of Rochester, University of Arizona.

Hamann, S., & Boersma, P. (2009). Loanword adaptation as first-language phonological perception. In A. Calabrese & W. L. Wetzels (Eds.), *Loan phonology* (pp. 11-58). Amsterdam: John Benjamins.

Haugen, E. (1950). The analysis of linguistic borrowing. *Language, 26*(2), 210-231.

Hayes, B. (1995). *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press.

Hussain, S. (2015). Missing from the 'minority mainstream': Pahari-speaking diaspora in Britain. *Journal of Multilingual and Multicultural Development, 36*(5), 483-497.

Hyde, B. (2011). Extrametricality and non-finality. In M. Oostendorp, C. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology* (pp. 1027-1051). Oxford: Blackwell.

Kager, R. (2000). Stem stress and peak correspondence in Dutch. In J. Dekkers, F. Reinoud, H. van der Leeuw, & J. M. van de Weijer (Eds.), *Optimality Theory: phonology, syntax, and acquisition* (pp. 121-150). Oxford: Oxford University Press.

Kager, R. (2007). Feet and metrical stress. In P. De Lacy (Ed.), *The Cambridge Handbook of Phonology* (pp. 195-227).

Kang, Y. (2010). Tutorial overview: Suprasegmental adaptation in loanwords. *Lingua: International Review of General Linguistics, 120*(9), 2295-2310.

Kenstowicz, M. (2007). Salience and similarity in loanword adaptation: a case study from Fijian. *Language Sciences, 29*(2-3), 316-340.

Khan, A. Q. (2012). *Phonology of Pahari: A Study of Segmental and Suprasegmental Features of Poonch Dialect.* University of Azad Jammu & Kashmir, Muzaffarabad.

Khan, A. Q. (2017). The tonal system of Pahari. *Acta Linguistica Academica, 64*(2), 313-324.

Khan, A. Q., Xu, Y., & Sohail, A. (2020). Multidimensionality of tone in Pahari. *Lingua, 245*, Online First 102923.

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing, 16*(1), 33-51.

McCarthy, J. (2008). *Doing Optimality Theory: Applying Theory to Data*. Oxford: Blackwell.

Meara, P., & Miralpeix, I. (2016). *Tools for Researching Vocabulary* (Vol. 105): Multilingual Matters.

Nation, P. (2012). Vocabulary Size in a Second Language. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics*. Oxford: Wiley.

Nation, P., & Beglar, D. (2007). A vocabulary size test. *The language teacher, 31*(7), 9-13.

Prince, A., & Smolensky, P. (2004). *Optimality Theory: Constraint Interaction in Generative Grammar. [Revision of Prince & Smolensky (1993).]*. Oxford Blackwell.

Shafi, S. (2019). *A Phonological Analysis of English Loanwords in Mirpur Pahari: Exploring Variable Adaptation in Optimality Theory.* University of York, York.

Silverman, D. (1992). Multiple scansions in loanword phonology: evidence from Cantonese. *Phonology, 9*(2), 289-328.

Stow, C., Pert, S., & Khattab, G. (2012). Translation to practice: Sociolinguistic and cultural considerations when working with the Pakistani heritage community in England, UK. In S. McLeod & B. A. Goldstein (Eds.), *Multilingual Aspects of Speech Sound Disorders in Children* (pp. 24-27). Bristol: Multilingual Matters.

Tabbassam, N. (2003). Phonological survey of Pahari Language. *Saroosh - Journal of Mirpur College, 11*(2), p380-397.