Jeremy O'Brien

Foreign Consonants in Hindi

Introduction:

In the following analysis, we will be discussing foreign consonants in Hindi-Urdu—their status in the phonological inventory, the way they are nativized, and the relationship between the phonology and the orthography. We will be looking at consonants that come from words taken from Arabic, Persian, Turkish, and English. Hindi-Urdu borrows from other languages, the noteworthy ones being Portuguese and Sanskrit, but we will not focus on them in this analysis.

Hindi-Urdu is an Indo-Aryan language variety spoken primarily on the Indian subcontinent. It is spoken as a first and second language by a great number of people—estimates for native speakers are in the range of 180 million[1], and many more can understand Hindi (340 million in India, according to a 1991 India census[2]). Hindi is used as a lingua franca in India, and Urdu is used as a national language in Pakistan.

Hindi and Urdu are often socially considered distinct language varieties, but linguistically the division between the two varieties is complex. Masica 1993 explains that while they are different languages officially, they "are not even different dialects or subdialects" in a linguistic sense; rather, "they are different *literary styles* based on the *same* linguistically defined subdialect" (p. 27). In everyday interpersonal conversation, Hindi and Urdu are nearly identical. In the higher registers, using vocabulary pertaining to government, religion, academia, etc., the two language varieties diverge considerably, to a point of mutual unintelligibility. This is mostly due to the fact that Hindi borrows most of its high-register vocabulary from Sanskrit, while Urdu mostly borrows from Persian and Arabic. This is not to say that Hindi does not have words of Persian or Arabic origin—in fact, a great deal of everyday items comes from these languages, and this will be the main focus of our discussion.

The distinction between Hindi and Urdu is more salient in the written form. Hindi and Urdu are commonly written with different orthographies, Hindi being written in Devanagari (the script that Sanskrit was often written in), and Urdu being written in a modified Perso-Arabic script. In short, we are examining the entire Hindi-Urdu spectrum, but for convenience we will call the language 'Hindi', and in consideration of space we will only take a look at the orthographic effects of Devanagari.

[1] Ethnologue, http://www.ethnologue.com/show_language.asp?code=hin
[2] http://www.censusindia.net/cendat/language/lang_table5.PDF

The Turko-Arabic /q/:

We'll start our analysis with the glottal stop /q/. These can be found in Arabic loanwords such as *qatla* 'murder', and Turkish words such as *qa:bu:* 'control'. In the speech of highly educated individuals, especially those familiar with the languages of Arabic or Turkish, these loanwords are realized in a manner similar to the source language—namely, the /q/ is realized as a uvular stop. In the speech of other Hindi speakers, however, it is realized as a velar stop *k*, the place of articulation being fronted.

To account for this phenomenon in Optimality Theory, we need to posit a few constraints. The first of these constraints is *FOREIGN, which in essence is a constraint against certain phones being found in the output—in this case, it prohibits /q/ from being realized as the output [q].

1. *FOREIGN — Assign one violation for every output segment that is 'foreign', meaning it is *q, x, ɣ, f, z, t d, θ, ð,* or any other segment that is not part of the Hindi phoneme inventory.

As formalized above, *FOREIGN is a constraint placed on the output of the phonology, but it is acting as a constraint on the phoneme inventory. The constraint could be broken into different constraints for each proposed foreign phone (i.e. *Q, *X, etc.), but the shorter form in (1) is less cumbersome for the present analysis. In terms of support, constraints like (1) could come about as an after-effect of a system of constraints on phonological contrasts. Fleming 2004 develops a system of constraints on contrasts, Dispersion Theory, with the result being a constrained phonological inventory. Fleming deals with vowels, nasalization, and stop voicing contrasts, but does not deal with consonant place of articulation. It is not obvious, but there should be a way to use Dispersion Theory and related theories to explain why a foreign consonant would not be immediately adopted into the phoneme inventory of a language, and that would be the cause of a specific set of constraints like (1).

Furthermore, the definition of *FOREIGN is not perfect, because we also need to prevent segments from being realized as highly phonetically modified versions of native segments. For instance, in the case of an average speaker of Hindi who cannot pronounce *q*, he or she does not produce a fronted version of *q* or a back version of *k*—the stop is realized as /k/ would be realized in a native word.

Accepting *FOREIGN for what it is, we also have IDENTPLACE, IDENTVOICE, and IDENTASP, essentially three aspects of the same type of constraint. Finally, in order to prevent simple deletion of the foreign segment, we use MAX. These constraints are shown below in (2)

through (5).

2. IDENTPLACE — For every output segment that fails to match the place specification dictated by the corresponding input segment, assign one violation for every 'hop' between the two places of articulation. The order is labial > dental > alveolar (& postalv.) > retroflex > palatal > velar > uvular

3. IDENTVOICE — Assign one violation for every output segment that fails to match the voicing specification of its corresponding input segment.

4. IDENTASP — Assign one violation for every output segment that fails to match the aspiration/breathy voicing specification of its corresponding input segment.

5. MAX — Assign one violation for every output segment that fails to have exactly one corresponding output segment.

Notice the careful wording of MAX. It is worded such to prevent fusion from being possible in our framework. Because this constraint is delimiting our framework, we will posit that it is ranked very high, and never violated in the data.

The tableau in (6) shows these constraints in action. As can be seen, the lowest ranked constraint is IDENTPLACE. The other constraints are not ranked in any particular order with respect to each other.

6.

/qatla/	MAX	*FOREIGN	IDENTASP	IDENTVOICE	IDENTPLACE
☞a) katla					uvular→velar
b) qatla		q(!)			
c) kʰatla			q→ kʰ(!)		uvular→velar
d) gatla				q→g(!)	uvular→velar
e) atla	q!				

With speakers that are educated in Arabic or Turkish, or those who try to maintain a distinction between historical /q/ and /k/, the *FOREIGN constraint is ranked low, allowing /q/ to be realized as a uvular stop, as in (7).

7.

/qatla/	MAX	IDENTPLACE	*FOREIGN
☞a) qatla			q
b) katla		uvular→velar!	
c) atla	q!		

The Perso-Arabic /x/:

Now, we will look at the velar fricative /x/. This is found in words such as *xara:b* 'bad' or the Arabic *fa:xta:* 'dove'. The foreign phonemes /q/ and /x/ are nativized in different ways. As

we discussed, /q/ is nativized as a stop at a different place of articulation. On the other hand, /x/ does not change place; it changes manner of articulation. It becomes a voiceless aspirated stop kʰ in the speech of many Hindi speakers. This introduces a very interesting question—why is /x/ realized as a voiceless aspirated stop, and not just a plain voiceless stop? In order to account for this phenomenon, we must look at the essence of a stop.

In the accepted view of features, a stop is simply any segment that has the feature [–continuant]. A fricative, on the other hand, is [+continuant]. Affricates are a tricky matter, because they start as [–cont] and end [+cont]. In this view, aspiration is a completely distinct feature [+spread glottis], and it does not interact with continuity in any theory-internal way.

I propose a less traditional way of looking at this. In this different view, continuity consists of a continuum. At the extreme of [+cont], we have fricatives. A little less continuant are the affricates, and even less continuant are aspirated stops. The most [–cont] segments are unaspirated stops. This viewpoint is formalized in (8), and this constraint takes the place of our previous IDENTASP constraint.

8. IDENTCONT — Assign one violation for every discrepancy between the continuant quality of an input and an output, where a discrepancy is a 'hop' on the continuum of continuity, which is defined as:
fricative > affricate > aspirated stop > unaspirated stop

So, assuming that the input is a fricative /x/, and also assuming that *FOREIGN will not allow it to be realized as *x*, then IDENTCONT constrains the output to the next best allowable segment on the continuant scale. The fricative is not allowed, so it hops to the affricate. A velar affricate is also a 'foreign' segment, so it hops again to aspirated stop. The velar aspirated stop is anything but foreign in Hindi, so the constraint allows the segment to be realized as that. The tableau in (9) demonstrates this, focusing on the interaction between IDENTPLACE and IDENTCONT. Note that the ranking in the tableau is necessary to achieve the desired output form *kʰara:b*.

9.

/xara:b/	*FOREIGN	IDENTPLACE	IDENTCONT
☞a) kʰara:b			fri → affric → asp
d) kara:b			fric → affric → asp → unasp
e) hara:b		velar → uvular → glottal !	
f) ʃara:b		velar → … → postalv !	
b) xara:b	x!		
c) k͡xara:b	k͡x!		fric → affric

The Perso-Turko-Arabic /ɣ/:

The voiced velar fricative /ɣ/ is found in words from Persian, Arabic, and Turkish. One example is *ɣa:fil* 'negligent'. Although /ɣ/ is quite similar to its voiced equivalent /x/, it does not nativize in the same way.

Hindi, like many languages of the Indian subcontinent, has what some linguists call voiced aspirates. Phonetically speaking, these segments might be more appropriately described as creaky voiced or breathy voiced, but phonologically speaking it is sufficient to say that they pattern in very similar ways to the voiceless aspirates. Even though there exists a voiced velar aspirate in Hindi, the voiced fricative does not become this segment when nativized. Rather, it becomes a stop *g*. To explain this, we need two different continua when dealing with continuants—one for voiced segments, and one for voiceless. These are shown in (10).

10.	IDENTCONT (REVISED)	The same as (9), but follow different continua, depending on whether the corresponding input segment is voiced or voiceless (new) voiced: fricative > affricate > unaspirated stop (old) voiceless: fricative > affricate > aspirated stop > unaspirated stop

With this new continuum, it is easy to show how the voiced fricative becomes an unaspirated stop.

The Perso-Anglo-Arabic /f/ and /z/:

Next we have /f/, a segment contained in loanwords from Persian, Arabic, and English. Examples include *fila:sfi:* 'philosophy' and the Persian *fara:mo:ʃi* 'forgetful'. Like the nativization of /x/, the labial fricative /f/ becomes an aspirated stop *pʰ*. Presumably, this is because there is no native labial affricate *p͡f*. Of course, *f* and *p* are not at the exact same place of articulation—*f* is a labiodental consonant (with the feature [+strident]), while *p* is a bilabial consonant ([–strident]). The direct correspondent would be *ɸ*. A much more nuanced understanding of IDENTCONT and IDENTPLACE is necessary to account for the data to a very fine point, but our current understanding of these constraints is sufficient to explain why Hindi would nativize consonants the way it does. In later research I would like to uncover this nuanced relationship between place and manner identity constraints.

As you might have noticed this far in the analysis, the affricates have been included in the continua, but up until now they haven't come into action. They have just been convenient to complete the paradigm. Now we will see a useful reason for including affricates in the continua. The foreign consonant /z/, again found in Arabic, Persian, and English, is often nativized to an

affricate dʒ (which we will write as *j*, using the symbol *y* to represent the front glide if need be). Examples of /z/ are *zarda:* 'tobacco' and the Persian *na:z* 'flirtatious behavior'.

Following the example of the previous consonants, this situation relies upon the continuum of continuity to determine the appropriate native phone. The closest thing to a fricative is an affricate, so IDENTCONT forces the segment to be realized as *j*. Again, *z* and *j* do not have the exact same place of articulation, but they are very close, and in some ways can be said to be at (almost) the same place.

So, as a partial summary, we have used the continuum of continuity to explain why certain foreign sounds are nativized in a way that, on first glance, seems arbitrary. Through satisfying the IDENTPLACE and IDENTCONT constraints, the foreign input becomes a more native output in a regular manner. This system has external support as well. The consonants /q/ and /ɣ/ are often nativized as *k* and *g* in other languages, the most obvious of which is English (e.g. Qur'an as [kəran]). Furthermore, there are attested historical sound changes that follow this pattern. Ancient Greek φ used to be a /pʰ/, but in Modern Greek it is an /f/.

Foreign Consonants From English:

We will now focus our attention on consonants that are considered primarily in the loanwords of English. These are the alveolar stops /t/ and /d/, and the dental fricatives /θ/ and /ð/. These foreign phonemes behave in a way that is very different from our previous phonemes.

The alveolar stops are nativized as retroflex stops. So, /t/ becomes ʈ and /d/ becomes ɖ. The dental fricatives change manner of articulation, becoming t̪ʰ and d̪ʒ[3]. Unlike the other foreign consonants, this substitution does not seem to be optional—there is no literature on alteration between these sounds, and English loanwords are always transcribed as having retroflex and dental stops. It does not seem to be the case that those with extensive background in English, even the English of North America or Europe, try to keep the /t/ as an alveolar in English-origin words—everyone pronounces it as a retroflex, and the same with the dentals. These are informal observations—more research is necessary before stating this matter as a strong claim about the Hindi lexicon and phonology—but the observations point to an interesting view of these consonants.

As with the case of /q/ and *k*, the foreign segment /d/ is realized at the closest allowable place of articulation. But we defined our IP constraint as being bidirectional. Going backward or forward in terms of place is just as equally good (i.e. two steps back is the same as two steps

[3] This change is in accordance with the IDENTCONT constraint as defined in (10)

forward when calculating number of violations). Then, why aren't the alveolar stops realized as dental stops? Dental stops are allowable native consonants, so why not choose them?

The answer is confusing and difficult to put into Optimality Theory. The dental fricatives /θ/ and /ð/ both change manner of articulation, and become dental stops *t̪ʰ* and *d̪*. Intuitively, we can describe this with a constraint that wants to prevent convergence of phonemes in the non-native stratum. That is to say, Hindi doesn't want to neutralize the distinction between /d/ and /ð/, realizing them both as *d̪*. The most harmonious resolution is to realize /ð/ as a stop at the same place, and move /d/ back to the retroflex, since the retroflex, like the dental, is only one point of articulation away from the alveolar. /ð/ can only be realized as *d̪* because IDENTPLACE is ranked higher than IDENTCONT. Moving /d/ forward to the dental would violate the same number of constraints as moving it to ɖ, but it would also cause convergence with /ð/.

But what about /t/? It could be realized as a dental without any problems, because there would still be the aspiration distinction between *t̪ʰ* and *t̪*. There seems to be a constraint to make similar foreign phonemes be realized similarly—/t/ is just as happy as a dental or a retroflex, so all things being equal, it will try to be more like its voiced equivalent /d/, which means being realized as a retroflex.

A set of tableaux representing this phenomenon is nearly impossible, because it involves a sort of meta-constraint on the language—the constraint that forces there to be a distinction between /d/ and /ð/ (and, in a secondary effect, /t/ and /θ/). Again, these ideas are much in the same vein as Fleming's Dispersion Theory, and further investigation in this area might find more connections of this sort.

Orthographic Correlations:

The distinction between the Perso-(Turko-)Arabic loan-phonemes /q/, /x/, /ɤ/, /f/, /z/, and the English loan-phonemes /t/, /d/, /θ/, and /ð/ is more than just phonological. The distinction is also salient in the orthography. For the Perso-Arabic consonants, the Devanagari letter is formed with the corresponding nativized letter, with a diacritic dot to show that it is the foreign sound. So, the devanagari letter for *kə* is क, while the letter for *qə* is क़. Sometimes this alternation is optional orthographically as well as phonological—people will write the sound as a क but pronounce it as a क़, or vice versa, or other times the phoneme and grapheme will correspond.

For the English consonants, on the other hand, there are no characters corresponding to *d* or *ð*. Instead, they are always written as the nativized equivalents *ɖə* ड or *d̪ə* द.

Conclusion and Further Research:

We've only scratched the surface as to what's happening with these loan-phonemes. We

have proposed a departure from the traditional, feature-bound analysis, to a more gradual, articulatorily-grounded analysis. A summary of the consonant alternations/variations is shown in (11), making it easy to see why a feature-bound analysis is so difficult to maintain while analyzing the data.

11.

foreign — native	foreign features	native features	violations of our constraints
q – k क़ - क	–high (dorsal)	+high	1 * for IDENTPLACE
x – kʰ ख़ - ख	–spread glottis +continuant	+spread glottis –continuant	2 * for IDENTCONT
ɣ – g ग़ - ग	+continuant	–continuant	2* for IDENTCONT
f – pʰ फ़ - फ	–spread glottis –continuant +strident	+spread glottis +continuant –strident	2* for IDENTCONT
z – j ज़ - ज	+anterior (coronal) –distributive –dorsal –high +continuant	–anterior +distributive +dorsal +high +/–continuant	1 * for IDENTCONT
t – ʈ Ø - ट	+anterior (coronal)	–anterior	1* for IDENTPLACE
θ – t̪ʰ Ø - थ	–spread glottis +continuant	+spread glottis –continuant	2* for IDENTCONT

There are a few more generalizations that are worth noting. For one, nativization always makes segments *less* continuous—we have not seen an input stop that becomes a fricative, only the opposite. This might be worth exploring, especially considering that the input /t/ never spirantizes to an *s* or *ʃ*.

Also notice that the phonology seems to match the orthography pretty closely. Those phonemes that alternate have different ways of representing the foreign sound, while the phonemes that fail to alternate—always being realized as native—do not have orthographic representation. It is too early to tell whether there is a causal relationship between the orthography and the phonology, but in my opinion, this matter deserves a great deal of research.

We have analyzed the repair strategies of Hindi, giving theoretical reasons for why certain segments are a better repair than other segments. For instance, *kʰ* is a better substitution for /x/ than any other allowable segment is for /x/. This parallels the work of Steriade 2001, but that particular research project (P-Map) is on a larger, more cross-linguistic scale than the present analysis. How much generalizations like the continuum of continuity are cross-linguistic is still an

open question, and if they are universal[4], then they might be built into the P-Map of human language.

References:

"Arabic transliteration", Wikipedia, accessed on 12/13/2005. http://en.wikipedia.org/wiki/Arabic_transliteration

Fleming, Edward. (2004). "Contrast and perceptual distinctiveness (pre-publication version)". B.Hayes, R.Kirchner, and D.Steriade (eds.) *Phonetically-Based Phonology*. Cambridge University Press, 232-276. http://web.mit.edu/~flemming/www/paper/CandP13.pdf

"Hindi", Wikipedia, accessed on 12/13/2005. http://en.wikipedia.org/wiki/Hindi

"Hindi-Urdu Phrasebook", WikiTravel, accessed on 12/13/2005. http://wikitravel.org/en/Hindi-Urdu_phrasebook

Jain, Usha R. (1995) *Introduction to Hindi Grammar*. University of California at Berkeley. ISBN 094461325X

Kelkar, Ashok R. (1968) *Studies in Hindi-Urdu I: Introduction and Word Phonology*. Deccan College, Pune, India.

Masica, Colin P. (1991). *The Indo-Aryan languages*. Cambridge: Cambridge University Press. ISBN 0-5212-3420-4.

"Persian language", Wikipedia, accessed on 12/13/2005. http://en.wikipedia.org/wiki/Persian_language

Snell, Rupert and Simon Weightman. (2003) *Teach Yourself Hindi*. Hodder & Stoughton Educational. ISBN 0-07-142012-6

Steriade, Donca. (2001). "The Phonology of Perceptibility Effects: the P-map and its consequences for constraint organization". http://www.linguistics.ucla.edu/people/steriade/papers/P-map_for_phonology.doc

[4] 'Universal' might be too strong a term. Steriade herself points out that perception of similarity between sounds could be governed by the phonological systems they are imbedded in.