Reducing the Number of Farsi Epenthetic consonants

Navid Naderi, Marc van Oostendorp

1.Introduction

Consonant epenthesis is the most productive means of hiatus resolution in the formal register of Farsi, which is the register described in this paper. The number and variety of the epenthetic consonants, in addition to their morphological specificity, has led to substantial controversy in the literature. According to a common view, there are nine epenthetic consonants in the language; i.e. [ʔ],[1] [h], [j], [g], [dʒ], [t], [d], [w] and [v] (see SADEGHI 2002; MAJIDI 1990:27-45; BIJANKHAN 2006:12-15; KAMBUZIYA 2007:277-306).[2] In this paper we will argue that it is possible to reduce this number to three, i.e. the glottal stop and the two glides [j] and [w]. Glottal stop epenthesis will be analyzed as emergence of the unmarked. Based on typological studies of consonant epenthesis, we argue that the glottal stop is the pure phonological, syllabification-driven epenthetic consonant in Farsi. Moreover, we argue that by assuming a specific process of feature spreading, it is possible to drive the insertion of glides on a purely phonological basis. Under under this assumption, glide epenthesis incurs exactly the same number of faithfulness violations as glottal stop epenthesis.

Apart from these three segments, we argue that [g], and [j] before the Ezafe-marker, are better analyzed as "latent segments" which have segmental features but lack a root node and we can derive them under the assumption of subsegmental specification. We also argue that the [j] after certain back vowels is a plain segment not parsed in isolation due to a markedness constraint, but it will be parsed in hiatus to avoid more faithfulness violations. As for the other above-mentioned consonants, we show that they are either semantically contrastive and hence not epenthetic ([t] and [dʒ]); or fossilized and non-productive ([h], [d] and [v]).

2.Theoretical Considerations

In our analysis we use two theories developed by LOMBARDI (1997/2002) and ZOLL (1996; 2001), whose main lines of argumentation are briefly sketched in this section.

---

[1] We use phonetic brackets [ ] for all epenthetic consonants–which lack a root node, or for fully phonetically realized forms involving such segments; for the sake of coherence we have used phonetic brackets also for cases where a segment is described as epenthetic in the literature, although we might have rejected that interpretation later. Phonemic brackets / / are reserved either for the forms before their phonetic realization, e.g. /sepɒhi+ɒn/, or when we suggest that certain segments are neither epenthetic nor latent, but have a root node in the underlying phonemic representation, e.g. the case of /j/ in the coda.

[2] LAZARD (1957:26-32) believes that in many cases vowels can remain in hiatus in Farsi, however, this is a rather contested issue. Moreover, he believes that in some cases, the first vowel in hiatus is followed by a "facultative" glide which cannot be considered a phoneme (LAZARD 1957:28). MAHOOTIAN (1997:339) also believes that "when a mid vowel is followed by any other vowel, nothing can insert" and the vowels remain in hiatus; however her examples, *to-i* "it's you" and *zendeam* "I'm alive", are quite dubious. Concerning the cases of [g] epenthesis, it seems that she takes the [g] to be a part of the suffixes with which it appears (see MAHOOTIAN 1997:273-274, 276, 279, inter alia) , in which case it should be explained in terms of allomorphy.

2.1. Glottal stop as the unmarked epenthetic consonant

By definition, any epenthetic consonant should be specified as [+cons]; so, in a constraint-based analysis, no form with an epenthetic consonant will be completely faithful to the input since it necessarily violates a DEP constraint. However, the quality of the epenthetic consonant is subject to a hierarchy of markedness constraints which guarantees that the consonant which emerges in this position will be as unmarked as possible.

Based on typological data, LOMBARDI (1997/2002) argues that the "occurrence [of coronals and glottals] as epenthetic segments shows different patterns;" and demonstrates that coronals are never the "general, purely syllabification-driven epenthetic consonant of a language" and that "they appear when additional factors force the use of more marked coronals instead of less marked glottal stops" (LOMBARDI 2002:1-2).[3] LOMBARDI (2002:4) assumes that the glottal stop has a pharyngeal place specification and suggests that the hierarchy of ranked markedness constraints according to the place of articulation as proposed by PRINCE / SMOLENSKY (1993) should be modified to include *Phar in the rightmost position. This hierarchy is reproduced in (1) below.

(1) *Lab, *Dor >> *Cor >> *Phar

With this hierarchy "/ʔ/ will be [the] optimal epenthetic consonant [and] its place markedness violation is even lower than that of the relatively unmarked /t/" (LOMBARDI 2002:4). Moreover, she assumes the markedness constraint ranking in (2) to avoid the false prediction of highly marked pharyngeals like /ʕ/ in epenthesis; also to show that their markedness is not due to their primary pharyngeal place (LOMBARDI 2002:5-6).

(2) *[-glottal] >> *[+glottal]

With these rankings LOMBARDI (2002:7) concludes correctly that "coronals only occur in epenthesis in certain specific types of situations, while glottal stop is seen in more general situations when pure markedness is permitted to reign over other considerations." She also considers that when a language has corresponding glides which agree in some features with the vowels in hiatus, their insertion violate the markedness constraints less, so the language may choose them where possible (LOMBARDI 2002:10).

2.2. Latent segments as floating features

An important representational distinction in autosegmental phonology is the one between full and latent segments. Latent segments are parsed only when necessary to facilitate syllabification and traditionally are represented diacritically to ensure that they are "either defective or extramaterial. Syllabification ignores elements thus marked, resulting in the distinction between full and latent segments" (ZOLL 2001:47). ZOLL (1996; 2001) argues against the traditional representation of latent segments and in favor of representing them as subsegments. She shows that of all the traditional "ostensible differences between segments

[3] This is necessarily a very brief and sketchy discussion of an issue which has been widely discussed in the literature; we can only refer here to literature such as BLEVINS 2008; HUME 2003; ORGUN 2001; RICE 1996; STERIADE 2009; TRIGO 1988; UFFMAN 2002; VAUX 2001; and ZYGIS 2010.

and subsegments," i.e. visibility to the syllable, dependency of the surface form, and mobility, the only one which holds consistently is (in)visibility to the syllable (ZOLL 2001:49); and "if this is the case then all latent segments, quasi-segments and dependent features which share some kind of invisibility to syllabification, should be represented uniformly as floating features that lack a root node" (ZOLL 2001:55).

Explaining the manner in which floating features surface is then a matter of specific grammars, and representational distinction is not necessary (ZOLL 2001:55). When floating features and latent segments are represented uniformly as subsegments, "the differences among subsegments derive from the source of the inserted root node:" while surface dependant features anchor to an existing segment, "for latent segments a new root node is epenthesized to host the feature" (ZOLL 2001:48). When a grammar includes a high-ranking constraint which prohibits a feature to anchor on an existing segment in the inventory of the language, there is no other way than epenthezing a root node which consequently will yield the manifestation of the floating feature as a full segment (see ZOLL 1996:184-186).

In what follows we shall analyze consonantal epenthesis as well as the conditions for manifestation of latent segments in Farsi with a constraint-based approach and with the aid of the theories suggested by LOMBARDI (1997/2002) and ZOLL (1996; 2001).

3. [ʔ]-epenthesis

3.1. Word-initial [ʔ]-epenthesis

The status of the glottal stop in Farsi has always been a source of controversy among linguists. Some believe that it is always epenthetic regardless of its position and even in Arabic loanwords such as *tabi'at* 'nature' (LAZARD 1989:266), while others are of the opinion that in word-initial position the glottal stop is not epenthetic but phonemic (e.g. WINDFUHR / PERRY 2009:427), and also in word-initial position the glottal stop should be described either as phonemic (WINDFUHR 1979:140) or as an "inherited automatic feature," the latter being the right description of the hiatus resolving [ʔ] as well (WINDFUHR / PERRY 2009:427). In this paper we follow the analysis suggested by KAMBUZIYA (2007) that words starting with glottal stops in Farsi should be divided into two groups of which one has phonemic /ʔ/ and the other epenthetic [ʔ]. Comparing the data in (3) and (4) sheds light on this issue (data taken from KAMBUZIYA 2007:278-280):

(3)	a.	[ʔaɢide]	"belief"		b.	[ʔaknun]	"now"
		[ʔahd]	"promise"			[ʔandɒze]	"size"
		[ʔasr]	"era"			[ʔandiʃe]	"thought"

(4)	a.	[hamʔaɢide]	"fellow-believer"		b.	[hamaknun]	"just now"
		[hamʔahd]	"alien"			[hamandɒze]	"same size"
		[hamʔasr]	"contemporary"			[hamandiʃi]	"conference"

Comparing the data in (3) and (4) shows that the distribution of the glottal stop in the words of group (a) is complete, but the words in group (b) lose the initial glottal stop in affixation or any case of compounding. So in the morphemes in (3b.) the glottal stop is only inserted

to fulfill a necessary condition of Farsi syllable structure. The morphemes belonging to group (a) are Arabic loans, while group (b) members are of Iranian origin.

It is not surprising to see the glottal stop as the epenthetic consonant in word-initial position. Cross-linguistically it is common to see the glottal stop, due to its unmarked Place, satisfying purely phonological requirements such as ONSET (LOMBARDI 2002:39).[4] Given the hierarchy of marked constraints in (1) and the constraints (5) to (7) ranked as in (8), we explain the word-initial [ʔ]-epenthesis in Farsi as illustrated in tableau (9).[5]

(5) ONSET assign one violation mark for every syllable that begins with a vowel
(6) DEP Let *input*=$i_1 i_2 i_3 \ldots i_n$ & *output*=$o_1\ o_2\ o_3 \ldots o_m$,
Assign one violation mark for every o_y
if there is no i_x K o_y.
(7) MAX Let *input*=$i_1 i_2 i_3 \ldots i_n$ & *output*=$o_1\ o_2\ o_3 \ldots o_m$,
Assign one violation mark for every i_x if there is no o_y where i_x K o_y.
(8) ONSET, MAX >> *Cor >> *Phar >> DEP

(9)

/aknun/	ONSET	MAX	*Cor	*Phar	DEP
☞ ʔaknun			*	*	*
aknun	*!W		*	L	L
knun		*!W	*	L	L
taknun			**!W	L	L

The ranking in (8) plus a constraint *CODA (MCCARTHY 2008:177) can also show why under prefixation the morphemes belonging to the group (a) of examples (3) and (4) do not lose their glottal stop while the members of group (b) do. This is illustrated in tableaux (11) and (12).

(10) *CODA assign one violation mark for every consonant in the coda of a syllable

(11)

/ham+aknun/	ONSET	MAX	*Phar	DEP	*CODA
☞ ha.mak.nun			**		**
ham.ʔak.nun			***!W	*W	***W
ham.knun		*!W	**		**

(12)

/ham+ʔaGide/	ONSET	MAX	*Phar	DEP	*CODA
☞ ham.ʔa.Gi.de			**		*
ha.ma.Gi.de		*!W	*L		L

Tableau (11) shows that [ha.mak.nun] is more harmonic than the candidate with an

[4] It is also instructive that most cases of syllabification-driven coronal epenthesis are in Coda position. This is due to the fact that coronals are more sonorant and make better codas, so a high-ranked markedness constraint like SonCoda can explain why glottals have less occurrence in this position (see LOMBARDI 2002:19-27).

[5] Constraint definitions are taken from MCCARTHY (2008:176, 197, 204), respectively; the constraints *Cor, *Phar, etc. are assumed to assign a violation mark for every occurrence of their respective features on the surface.

epenthetic consonant. But in tableau (12), where the glottal stop is not epenthetic, violating *CODA is preferable, because the deletion of /ʔ/ is blocked by the high-ranked MAX.[6]

3.2. Morpheme-medial [ʔ]-epenthesis

Morpheme-medial hiatus exists only in loanwords in Farsi. (13) provides some examples:

(13) [videʔo] "video"
[teʔori] "theory"
[seʔul] "Seoul"
[tɒʔɒtr] "theatre"
[mɒʔo] Mao (proper name)

In all the examples of morpheme-medial epenthesis, the epenthetic segment only serves to avoid an ONSET violation; hence, it is purely phonological. The hierarchy of ranked place markedness constraints in (1) suffices to explain why the optimal epenthetic consonant should be [ʔ]. Tableau (14) illustrates the optimal candidate.

(14)

/video/	ONSET	MAX	*Cor	*Phar	DEP
vi.de.ʔo			*	*	*
vi.de.o	*!W		*	L	L
vi.de.to			**!W	L	*
vi.do		*!W	*	L	L

3.3 [ʔ]-Epenthesis at morpheme boundaries

At morpheme boundaries, when there are no specific morphological considerations, the glottal stop resolves all the hiatuses, except when the first vowels in hiatus are /i/ or /u/, where hiatus is resolved by an agreeing glide in the same way as in morpheme-medial epenthesis. This is in turn due to the fact that there are no agreeing glides for other vowels in the Farsi inventory, i.e. /e/, /a/, /o/ and /ɒ/. (15) provides an example for any possible hiatus **not** starting with a high vowel:[7]

(15)

[mohɒvereʔi]	"colloquial"	[rɒdijoʔi]	"relating to radio"
[xɒneʔemɒn]	"our home"	[rɒdijoʔemɒn]	"our radio"
[xɒneʔat]	"your home"	[rɒdijoʔam]	"my radio"
[gerjeʔu]	"one who cries a lot"	[toʔoman]	"you and me"
[xɒneʔokɒʃɒne]	"abode"	[bɒlɒʔi]	"the one above"
[sofreʔɒrɒ]	"table designer"	[dɒnɒʔand]	"they are wise"
[minɒʔɒrɒ]	"flower designer"	[zɒʔu]	"pregnant"
[ʔaʔaʔi]	"pee-pee"	[mɒʔoʃomɒ]	"we and you"
[na:naʔu]	"one who always disagrees"	[bɒlɒʔe]	"s/he is upstairs"
[naʔonu]	"rambling"		

[6] Note that the candidates **knun* in (9) and **ham.knun* in (11) also occur fatal *CCV constraints which are not represented in the tableaux above.

[7] NPs, VPs and ConjPs are here represented as Phonological Phrases which have access to syllabification in the prosodic hierarchy. It seems that there are no examples for /a-e/, /a-ɒ/, /o-u/, /o-ɒ/.

It is easy to observe that in all the examples above, a DEP violation is incurred to avoid the fatal ONSET violation. Since there are no higher ranked constraints or additional factors to force the use of a more marked segment, the hiatus in all the examples is resolved by the emergence of glottal stop as the optimal epenthetic consonant according to the markedness hierarchy in (1).

4. Glide epenthesis as feature spreading

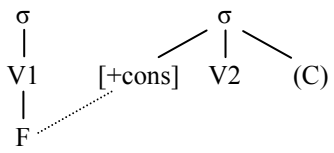
A common observation is that "in languages which resolve hiatus via epenthesis, [a] common approach is to epenthesize a glide that agrees in features with an adjacent vowel, often a high vowel" (LOMBARDI 2002:9). This is exactly what happens in Farsi as well. Farsi has only two glides [j] and [w] in its inventory, which agree with /i/ and /u/ respectively. Hence, should the left vowel in hiatus be /i/ or /u/, the agreeing glide inserts. (16) and (17) exemplify hiatus with /i/ or /u/ as the left vowel:[8]

(16) [bɒzɒri**j**i] "a businessman"
[sini**j**emɒn] "our tray"
[kasifi**j**aʃ] "its dirtiness"
[ʔenɢelɒbi**j**un] "revolutionaries"
[pari**j**o nimɒ] "Pari and Nima" (proper names)
[sepɒhi**j**ɒn] "soldiers"

(17) [ʔɒhu**w**i] "a deer"
[tarsu**w**e] "the chickenhearted"
[ʔɒhu**w**ast] "it's a deer"
[kɒhu**w**o kalam] "salad and cabbage"
[bɒzu**w**ɒn] "arms"

We will show in this section that assuming a certain process of forward feature spreading, i.e. from a high vowel to the epenthetic root node on its immediate left as in (18), glide epenthesis incurs the same number of faithfulness violations as glottal stop epenthesis while the output will be more harmonious with an epenthetic glide than with the glottal stop. Under this assumption the glottal stop emerges only when spreading is not possible.

(18)



We assume that [j] is an /i/ in consonantal position, and [w] an /u/ in consonantal position. Although in some other languages also other vowels (e.g. /e/ and /o/ can lead to glides) we assume that this comes at a cost and Farsi is not willing to pay that cost. Furthermore, we also assume that spreading can only be progressive, i.e., it comes from V1 rather than from V2. Languages as diverse as Mandarin Chinese and Dutch avoid [ji] and [wu] sequences (VAN OOSTENDORP 2000); whatever is responsible for this avoidance may play a role here as well.

[8] It seems there are no examples for /u-u/ hiatus.

The feature spreading process illustrated in (18) is the consequence of a conflict between a faithfulness constraint *F and a dominated markedness constraint *EMPTY (VAN OOSTENDORP 2000). These constraints are defined in (19) and (20) respectively and their ranking with respect to the other constraints discussed so far is given in (21).

(19) *F assign one violation mark for any feature in the output which does not correspond to a feature in the input.

(20) *EMPTY assign one violation mark for any consonant without a place specification.

(21) ONSET, MAX >> *F >> *EMPTY >> DEP

The decision making process for the quality of the epenthetic consonant in these cases is illustrated in tableaux (22) and (23).

(22)

/sepɒh+i+ɒn/	ONSET	MAX	*F	*EMPTY	DEP
☞ se.pɒ.hi.**j**ɒn					*
se.pa.hi.wɒn			*!W		*
se.pɒ.hi.ʔɒn				*!W	*
se.pɒ.hi.ɒn	*!W			*W	L
se.pɒ.hɒn		*!W			L
se.pa.hin		*!W			L

(23)

/bɒzu+ɒn/	ONSET	MAX	*F	*EMPTY	DEP
☞ bɒ.zu.**w**ɒn					*
bɒ.zu.jɒn			*!W		*
bɒ.zu.ʔɒn				*!W	*
bɒ.zu.ɒn	*!W			*W	L
bɒ.zɒn		*!W			L

5. Latent segments and hiatus resolution

5.1. Latent [g]

[g] resolves hiatus only in the following three morphophonologic contexts (SADEGHI 2002:34-35; BIJANKHAN 2006:14-15; KAMBUZIYA 2007:298-300; inter alia)
i. after /e/ and before the nominalizer/adjectivizer suffix *-i*
ii. after /e/ and before the adverbializer suffix *-ɒne*
iii. after /e/ and before the plural suffix *-ɒn*
The use of hiatus-resolving [g] is illustrated by the following examples.

(24) a. [xaste] "tired"
[xaste**ʔ**i] "you're tired"
[xaste**j**e kɒr] "tired of working"

b. [xaste**gi**] "fatigue"
[xaste**gɒn**] "tired ones"
[xaste**gɒne**] "with tiredness"[9]

[9] This word, admittedly, sounds odd in Farsi, but it is not an impossible word. However, it is used here only for the sake of coherence and one can easily replace it with another example like [batʃegɒne].

Analyzing [g] as an epenthetic consonant is certainly problematic for a diversity of reasons. First and foremost, [g] is a highly marked consonant in the hierarchy of place markedness (see no. 1 above), hence highly improbable to occur as an epenthetic consonant. Though it is not impossible to have a phonologically-driven epenthetic velar plosive cross-linguistically, if a language has such an epenthetic consonant, it should actually be enforced by higher ranking constraints which strictly prohibit labials, coronals and pharyngeals, which is quite improbable. Further, while [g] has an obvious phonological function here, i.e. hiatus resolution and ONSET-satisfaction, it is clearly morphologically restricted, so there should be a morphophonological process which enforces its emergence and not a pure phonological one. ZOLL's theory of subsegmental representation of latent segments, discussed in Section 2.2 above, provides a good framework for analyzing this issue. [g] has all the characteristics of a "subsegment" according to ZOLL's (2001:46-56) definition. It surfaces only when a high-ranked constraint enforces it.

Moreover, there is also no natural articulatory explanation for an agreement between /e/ and [g]. The crucial question, however, is whether we should associate this subsegment with above-mentioned suffixes or with the stems to which they attach? Since the so-called epenthetic [g] emerges only before these three suffixes, it is quite tempting to assume that it is there in their underlying representation.

However, we know that in Middle Persian [g] was the final segment of the majority of words which end in /e/ in today's Farsi (SADEGHI 2002:39). Hence we assume that [g] is still present in the underlying representation of these words but syllabification only parses it when it is forced to do so.[10] Considering the morphoogically restricted occurrences of the hiatus-resolving [g], there should be morpheme-specific constraints which allow [g] to be parsed only in the contexts mentioned above. This is guaranteed by the Alignment constraints defined in (25) to (27).

(25)	ALIGNR(*-i*, stem)	assign one violation mark for every segment intervening between the nominalizer/adjectivizer *–i* and the right edge of the stem
(26)	ALIGNR(*-ɒne*, stem)	assign one violation mark for every segment intervening between the adverbializer *–ɒne* and the right edge of the stem
(27)	ALIGNR(*-ɒn*, stem)	assign one violation mark for every segment intervening between the plural morpheme *-ɒn* and the right edge of the stem

Then we need two constraints whose conflict determines whether the subsegment should be materialized or not. We follow ZOLL (1996/2001) in solving this problem with a combination of Maximality and Dependency constraints as defined below (from ZOLL 2001:60, 66 respectively).

[10] One might argue that [g] in these cases does not function as a hiatus-resolving segment at all, but rather the forms with [g] had a separate line of development and have entered Modern Persian independently. This might sound plausible in the case of the nouns, adjectives or adverbs made with -i and -*ɒ*ne, but it would be highly improbable to imagine that all the plural forms with *-ɒn* have also entered Modern Persian independent of their singular counterparts (thanks are due to Geoffrey Haig for this argument). The privilege of our theoretical approach is that it explains all the cases in a unified manner.

(28) MAX(SUBSEG) every subsegment in the input has a correspondence in the output
(29) DEP(Root) an output root node has an input correspondent

MAX(SUBSEG) is the faithfulness constraint which encourages the realization of subsegments and is violated by any subsegment in the input which does not have a correspondent in the output and should be crucially dominated in Farsi to ensure that [g] does not surface unless a higher ranked constraint forces it. The rationale for DEP(Root) is that in order for a subsegment to surface as a full segment, a root node should epenthesize to host it (ZOLL 2001:48). The proper ranking of the constraints for materialization of latent segments in Farsi would be as follows:

(30) ALIGNR, DEP(Root) >> MAX(SUBSEG)

Given the fact that MAX(SUBSEG) is crucially dominated, the latent [g] never surfaces when morphemes containing it are parsed in isolation. Moreover, the domination of ALIGNR over MAX(SUBSEG) ensures the edge-boundedness of [g]; and with DEP(Root) dominating the MAX(SUBSEG) we can be sure that [g] surfaces only when its absence would lead to a fatal violation of the morpheme-specific ALIGNR active in each specific context. The following Tableaux illustrate the working of these constraints.

(31)

/xaste[g]/	ALIGNR	DEP(Root)	MAX(SUBSEG)
☞ xas.te			*
xas.teg		*!W	L

(32)

/xaste[g]+i/	ALIGNR(-*i*, stem)	DEP(Root)	MAX(SUBSEG)
☞ xas.te.gi		*	
xas.te.ʔi	*!W	L	*W

(33)

/bɒzu+ɒn/	ONS	MAX	ALIGNR (-*ɒn*, stem)	DEP(Root)	*F	*EMPTY	MAX(SUBSEG)
☞ bɒ.zu.**w**ɒn			*	*			
bɒ.zu.**j**ɒn			*	*	*!W		
bɒ.zu.ʔɒn			*	*		*! W	
bɒ.zu.ɒn	*!W		L	L			
bɒ.zɒn		*!W	L	L			

(31) shows that when morphemes containing the velar plosive subsegment are parsed in isolation, the Alignment constraint becomes inactive, thus DEP(Root) decides about the optimal output. (32) exemplifies a case where the morpheme-specific Alignment constraint and the DEP(Root) conflict over the optimal output. Finally, (33) demonstrates that in the case of a morpheme which lacks the subsegment, the morpheme-specific Alignment constraint and DEP(Root) cannot make any decision because those candidates preferred by them are already out of competition by higher-ranking ONSET or MAX and all the rest are equally bad for them, so they become inactive.

There are also a few loanwords in Farsi which show a [g] insertion before the same suffixes discussed above, in spite of the fact that they do not really have a latent [g]. Here are some examples:

(34) [sarjuxegɒn] "lieutenants"
[seflegɒn] "poor ones"
[talabegi] "studentship (esp. in a religious school)"
[ʔamalegi] "workmanship (pejorative) "

As we can see, all of these words end in /e/, which is also the final vowel of the forms with latent [g]. This can obviously be explained in terms of analogy, i.e. the presence of [g] in the words in (34) is due to a process of historical reanalysis. SADEGHI (2002:39-44) mentions many examples which show that in the early periods of New Persian, after the decline of Middle Persian, the hiatus resolving /g/ occurred in various morphophonological contexts, but in the course of several centuries it has become morphologically restricted.

5.2 Latent [j]

The second latent segment in Farsi is [j]. However one should be wary not to take the latent [j] mistaken for its homophonous glide discussed in Section 4..; and also for the plain segment /j/ which we will discuss in Section 6. below. Here again a morphological observation can draw a clear line between the latent [j] and its homophonous glide. In Section 4. we discussed situations where the glide inserts to avoid hiatus, but it seems that the following data oppose our analysis:

(35) [xɒn**eje** nimɒ] "Nima's home" [hard**oje** ʃomɒ] "both of you"
[n**aje** to] "your negative answer" [ɒG**ɒje** zamɒni] "Mr. Zamani"
[pat**uje** sabz] "green blanket"

In all these examples, [j] occurs in the same morphological context, i.e. before the morpheme *-e* which is the marker of the so-called Ezafe-construction. As an enclitic, *-e* always attaches to its preceding morpheme. Should the morpheme end in a consonant, [j] does not surface:

(36) [ketɒb-e nimɒ] "Nima's book"
[tʃaman-e sabz] "green grass"
[daftar-e to] "your notebook"

But as we can see in (35) when it attaches to a codaless morpheme, [j] resolves the hiatus regardless of the quality of the preceding vowel, and there is no exception to this rule. We assume that in this case [j] is actually a subsegment belonging to -e which surfaces only when forced by a high-ranked constraint.

As we have seen above, constraints which conflict on the materialization of latent segments are MAX(SUBSEG) and DEP(Root) as defined in (28) and (29) in combination with a morpheme-specific Alignment constraint, ranked as (30) above. We suggest the following morpheme-specific Alignment constraint to predict the occurrence of the latent [j].

(37) ALIGNR(-^{j}e, stem) assign one violation mark for every segment intervening between the Ezafe marker -^{j}e, and the right edge of the stem

Tableau (38) illustrates the constraints conflicting over materialization of [j].

(38)

/patu+-^{j}e /	ONS	MAX	ALIGNR (-^{j}e, stem)	DEP(Root)	*F	*EMPTY	MAX(SUBSEG)
pa.tu.je				*			
pa.tu.we			*!W	*	*W		*W
pa.tu.ʔe			*!W	*		*W	*W
pa.tu.e	*!W			L		*W	*W
pa.te		*!W		L			*W

6. Final /j/ after /ɒ / and /u/ in Coda

There are quite a lot of words in Farsi ending in the two sequences /ɒj/ and /uj/. However the final /j/ in these morphemes is not parsed unless a high-ranked constraint forces it. Since the occurrence of /j/ after these morphemes is not restricted to any specific morphological context, there is no need to describe /j/ as a latent segment. /j/ in this case is just a normal plain segment which is present in the underlying representation, but it is not parsed due to the markedness constraint defined in (39).

(39) *$V_{[+back]}J_{CODA}$ assign one violation mark for every /j/ in coda after back vowels

With this constraint ranked above MAX, no morpheme which ends in /ɒj/ or /uj/ can be fully parsed in isolation, but when it encounters a vowel-initial morpheme to its right, its final /j/ should be parsed because syllabification necessarily puts it in the Onset position. On the other hand, when /j/ is parsed there is no hiatus anymore, and MAX, which definitely favors its parsing, automatically outranks the other candidates. Tableau (40) shows a case where *$V_{[+back]}J_{CODA}$ outranks a faithful candidate (*begu* "say"), and (41) illustrates a competition where the final /j/ is parsed (*gujef* "dialect").[11]

(40)

/be+guj/	ONSET	*$V_{[+back]}J_{CODA}$	MAX
☞ begu			*
beguj		*!W	L

(41)

/guj+eʃ/	ONSET	*$V_{[+back]}J_{CODA}$	MAX	*F	*EMPTY
☞ gu.jeʃ					
gu.weʃ			*!W		
gu.ʔeʃ			*!W		*W
gu.eʃ	*!W		*W		*W

[11] A reviewer points out to us that given our constraints, a form *[beguw] might be expected to win in (40), assuming that the [w] would derive from the [j] by spreading: such a form would not violate MAX and thus beat the winner in (40).

7. Other segments

As mentioned in Section 1. there are five other consonants described as epenthetic in the literature, viz. [h], [d], [v], [t] and [dʒ]. In this section we will briefly argue against the epenthetic nature of these segments.

Of these segments, [h], [d], and [v] are neither productive nor frequent enough to be considered epenthetic. They only occur in a set of frozen or fossilized phrases, in a fixed situation, and show no variation or alternation. [h] occurs only in twelve fixed forms, i.e. when pronominal enclitics attach to the propositions *be* "to" and *bɒ* "with," as exemplified in (42).

(42)	[behem]	"to me"	[bɒhɒm]	"with me"
	[behet]	"to you"	[bɒhɒt]	"with you"
	[beheʃ]	"to her/him"	[bɒhɒʃ]	"with her/him"
	[behemun]	"to us"	[bɒhɒmun]	"with us"
	[behetun]	"to you (Pl.)"	[bɒhɒtun]	"with you (Pl.) "
	[beheʃun]	"to them"	[bɒhɒʃun]	"with them"

[d] is attested only in the following four archaic forms which even have their own entries in the most recent Farsi dictionary (see ANVARI 2002:846, 871, 874, 875).

(43)	[bedu]	"to him/her"	[bedin]	"with this/for this"
	[bediʃɒn]	"to them"	[bedɒn]	"with that/for that"

And finally [v] is attested only in a few fossilized forms before the attributive suffix *-i* (mostly with the name of cities).

(44)	[gandʒavi]	"from Ganja
	[sɒravi]	"from Sari
	[Gaznavi]	"from Ghazni
	[koravi]	"spherical

The so-called epenthetic [t] occurs only in a verb-initial construction with a pronominal enclitic as its direct object. The [t] is always situated between the verb and the enclitic.[12] Moreover the so-called epenthetic [t] emerges only with 3rd person singular verbs as exemplified below:

(45)	[didateʃ]	"he's seen him"
	[koʃtatet]	"he'll definitely kill you (lit. he has killed you)"
	[bordatemun]	"he's taken us"

The semantics of this construction is the key to the origin of [t]. As SADEGHI (2002:37) also observes, [t] in this construction is indeed a remainder of the partially deleted 3rd person singular perfect marker, i.e. cliticised *-ast*; but it is not really clear why he then describes it as an epenthetic consonant. The presence of /t/ in these examples is a result of deletion and not epenthesis.

[12] This construction is used only in the informal register.

The last segment labeled epenthetic in the phonological literature on Farsi is the voiced affricate [dʒ] which occurs only before the loan plural morpheme *-ɒt*. But there is both semantic and phonologic evidence that *-dʒɒt* is an independent morpheme in Farsi. An epenthetic segment cannot cause any semantic difference by definition, but *-dʒɒt* and *-ɒt* show a semantic difference: *-ɒt* is only a plural morpheme and synonymous to the other more frequent plural suffix *-hɒ* while *-dʒɒt* adds both senses of plurality and variety to the stem to which it attaches. Compare the following data:

(46)	a.	[mosɒbeɢ**ɒt**]	"competitions"	c.	[mosɒbeɢe**hɒ**]	"competitions"
		[monɒzer**ɒt**]	"debates"		[monɒzere**hɒ**]	"debates"
	b.	[kɒrxɒne**dʒɒt**]	"different factories"	d.	[kɒrxɒne**hɒ**]	"factories"
		[ʃirini**dʒɒt**]	"a variety of pastries"		[ʃirini**hɒ**]	"pastries"
		[dɒru**dʒɒt**]	"a variety of medicines"		[dɒru**hɒ**]	"medicines"

The phonological evidence for the independence of *-dʒɒt* comes from the fact that it can productively attach to consonant-final morphemes (specially in youth speech). The data in (47) are taken from the Internet.

(47)	[xafandʒɒt]	"awesome things"
	[ʔesemesdʒɒt]	"different (funny) text messages"
	[seksdʒɒt]	"different things related to sex"
	[alkoldʒɒt]	"different kinds of alcoholic drinks"

8. Summary

According to some of the phonological literature on Farsi there are nine epenthetic consonants in this language. We argued in this paper, on the basis of an Optimality-Theoretical analysis, that this number can be reduced to three. The inventory of true epenthetic consonants in Farsi consists of the glottal stop and the two glides [j] and [w]. In Section 3. and 4. we argued that these consonants are the only pure phonological, syllabification-driven epenthetic consonants in Farsi and their emergence could be explained in terms of place markedness and feature spreading. Then in Section 5. we showed that the highly morphologically restricted hiatus-resolving [g] and the palatal glide before the Ezafe-marker are better analyzed as latent segments which lack a root node and their occurrence could be predicted under the assumption of subsegmental specification. In Section 6. we argued that the palatal glide after back vowels in coda is neither epenthetic nor latent, but rather a plain segment which, due to a high-ranked markedness constraint, is not parsed unless syllabification renders its parsing necessary. Finally in Section 7. we briefly argued against the epenthetic nature of other segments described as such in the literature.

References

ANVARI, Hasan 2002: *Farhang-e Soxan*. Tehran: Sokhan 1381 h.š.

BIJANKHAN, Mahmoud 2006: *Vājšenāsi: nazariye-ye behinegi*. Tehran: Samt 1384 h.š.

BLEVINS, Juliette 2008: "Consonant epenthesis: natural and unnatural histories." In: Jeff GOOD (ed.): *Language Universals and Language Change*. Oxford: Oxford University Press, pp. 79-107

HUME, Elizabeth 2003: "Language specific markedness: The case of place of articulation." In: *Studies in Phonetics, Phonology and Morphology* 9:2, pp. 295-310

KAMBUZIYA, Aliye K. Z. 2007: *Vājšenāsi: ruykard-hā-ye qā'ede-bonyād*. Tehran: Samt 1385 h.š.

LAZARD, Gilbert 1957: *Grammaire du Persan Contemporain*. Paris: Klincksieck

—— 1989: "Le persan." In: Rüdiger SCHMITT (ed.): Compendium Linguarum Iranicarum. Wiesbaden: Reichert, pp. 263-293

LOMBARDI, Linda 1997: "Coronal Epenthesis and Markedness." In: *Phonology* 19, pp. 219-251 [The modified version of 2002 is used as a reference here; it is available on Rutgers Optimality Archive, ROA-245]

MAHOOTIAN, Shahrzad 1997: *Persian* [*Descriptive Grammars Series*]. London: Routledge

MAJIDI, Mohammad Reza 1990: *Strukturelle Grammatik des Neupersischen (Farsi), Band II: Morphologie*. Hamburg: Helmut Buske Verlag

MCCARTHY, John J. 2008: *Doing Optimality Theory*. Oxford: Blackwell

VAN OOSTENDORP, Marc 2000: *Phonological Projection*. Berlin: Mouton De Gruyter

ORGUN, Cemil O. 2001: "English r-insertion in optimality theory." In: *Natural Language & Linguistic Theory* 19, pp. 737–749

PRINCE, Alan, and Paul SMOLENSKY 1993: "Optimality Theory. Constraint Interaction and Satisfaction in Generative Grammar." Manuscript

RICE, Keren 1996: "Default variability: The coronal-velar relationship." In: *Natural Language & Linguistic Theory* 14.3, pp. 493-543

SADEGHI, Ali Ashraf 2002: "Elteqā-ye mosavvet-hā va mas'ale-ye sāmet-e miyānji." In: Ali Ashraf SADEGHI: *Masā'el-e tārixi-ye zabān-e fārsi*. Tehran: Sokhan, 1381 h.š., pp. 25-50

STERIADE, Donca 2009: "The phonology of perceptibility effects: The P-map and its consequences for constraint organization." In: Sharon INKELAS, Kristin HANSON (eds.): *On the Nature of the Word*. Cambridge: MIT Press

TRIGO, Lorenza R. 1988: *On the phonological behavior and derivation of nasal glides*. PhD dissertation, MIT

UFFMAN, Christian 2002: "Intrusive [r] and optimal epenthetic consonants." In: *Language Sciences* 29 (2007), pp. 451-476

VAUX, Bert 2001: "Consonant insertion and hypercorrection." LSA Annual Meeting, Washington DC. Ms. available online at <people.pwf.cam.ac.uk/bv230/OT/epenthesis-lsa.pdf>

WINDFUHR, Gernot L. 1979: *Persian grammar: history and state of its study*. The Hague: Mouton

WINDFUHR, Gernot, and John PERRY 2009: "Persian and Tajik." In: Gernot WINDFUHR (ed.): *The Iranian Languages*. London, New York: Routledge, pp. 416-544

ZOLL, Cheryl C. 1996: *Parsing Below the Segment in a Constraint Based Framework*. Doctoral Dissertation. University of California, Berkeley. Berkeley, CA [Available on Rutgers Optimality Archive, ROA-143]

—— 2001: "Constraints and Representation in Subsegmental Phonology." In: Linda LOMBARDI (ed.): *Segmental Phonology in Optimality Theory: Constraints and Representations*. Cambridge: Cambridge University Press, pp. 46-78

ZYGIS, Marzena 2010: "Typology of consonantal insertions." In Melanie WEIRICH, Stefanie JANNEDY (eds.): *Papers from the Linguistics Laboratory. ZAS Papers in Linguistics* 52, pp. 111-140