
Compensatory Lengthening as Coalescence: Analysis and Implications

MEGHAN SUMNER

State University of New York-Stony Brook

1 Introduction

In this paper, I argue for a new view of compensatory lengthening, hereafter CL, in which CL results from the coalescence rather than deletion of a segment. Coalescence results in lengthening because in CL languages, coalescence is not free. The cost of coalescing two segments is that both must be represented prosodically. I further argue that the generalization that only the deletion of weight-bearing codas triggers CL (Hayes 1989) is not consistent with all the CL data.

There are a number of phenomena within the CL data that raise questions about previous treatments. First, Modern Colloquial Tehrani Farsi exhibits CL triggered by the loss of only a subset of consonants (Darzi 1991). Also, CL triggered by the loss of consonants that are not possible codas, referred to as moraic mismatches by Hyman, is found in some Bantu languages (Hyman 1992). CL can also be triggered by the loss of onsets, as in Greek double flop cases. Additionally, there are data which pair CL with coalescence as in Luganda (Clements 1986), Old English (de Chene 1976), or Komi I_{ma}. Finally, there is an apparent incompatibility of CL with an Optimality Theory (OT) framework (McCarthy & Prince 1993). The main problem is the fact that OT lacks an intermediate stage in which codas can receive moras by a Weight-By-Position rule. However, this step is not desired in the cases of moraic mismatches because it forces us to make unwanted claims about the moraic structure of the languages in question.

* I would like to thank Mark Aronoff, Christina Bethin, Yiya Chen, and Lori Repetti for many valuable comments on various versions of this paper. I would also like to offer special thanks to Ellen Broselow whose time, patience, and guidance have been invaluable to me.

The primary goal of this paper is to present a new view of CL focusing on coalescence. I show that mora preservation is neither a necessary nor sufficient condition for CL. In this paper, I use data from Modern Colloquial Tehrani Farsi to address lengthening in OT and data from Greek to illustrate the behavior of onsets in CL languages.

2 Lengthening

In this section, I consider CL in Modern Colloquial Tehrani Farsi. I use the data from Farsi to illustrate past analyses, as well as the analysis I propose in this paper.

2.1 The data

Lengthening in Farsi occurs in a number of situations. This variation is useful in choosing an analysis for CL and therefore provides a nice starting point. In Farsi, lengthening occurs in the colloquial dialect but not the conservative dialect (Darzi 1991). The loss of glottal consonants leads to lengthening word finally, as in 1.

(1)	Conservative	Colloquial	Gloss
	a. robʔ	ro:b	'hand'
	b. sobh	so:b	'morning'
	c. suʔ	su:	'bad'
	d. kuh	ku:	'mountain'

Note that 1a,b exhibit non-local lengthening. There is a consonant between the glottal consonant that is lost and the vowel that lengthens. In the final two forms, lengthening occurs when the glottal consonants are the only codas. Thus, the presence of consonant clusters is not a factor in the loss of glottal consonants. Lengthening also results from the loss of glottal consonants when they are the initial segment of a consonant cluster, as below.

(2)	Conservative	Colloquial	Gloss
	a. qæʔr	qæ:r	'bottom'
	b. qæhr	qæ:r	'wrath'
	c. læʔn	læ:n	'cursing'
	d. læhn	læ:n	'language'

The loss of syllable-final glottal consonants induces lengthening as well. This is shown in 3.

(3)	Conservative	Colloquial	Gloss
	a. næʔna	næ:na	'mint'
	b. pæhna	pæ:na	'width'
	c. ʔeʔzam	ʔe:zam	'dispatch'

d. behzad be:zad proper noun

In order to illustrate that only the loss of glottal consonants induces lengthening, the data in 4 are needed.

- | | | | |
|-----|-----------|-------|------------|
| (4) | a. fekr | fek | 'thought' |
| | b. ʃokr | ʃok | 'thanks' |
| | c. qænd | qæn | 'sugar' |
| | d. gereft | gerɛf | '(he) got' |
| | e. ʔæz | ʔæ | 'from' |

In the examples above, final consonants are lost whether or not they constitute part of a cluster but the vowel does not lengthen. So, the forms in 4 involve the deletion of a consonant without lengthening. We have already seen in 1 that Farsi allows non-local lengthening, so the same would be predicted for the examples in 4. Any adopted analysis must account for this variation.

The facts about Farsi CL are summarized below (where G=glottal consonant & C=non-glottal).

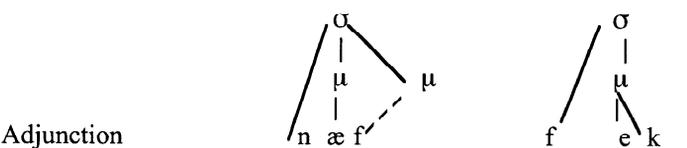
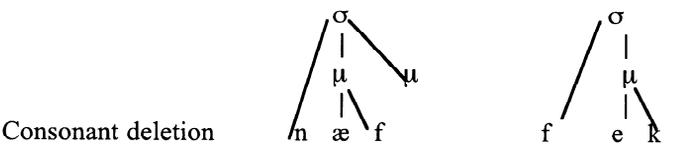
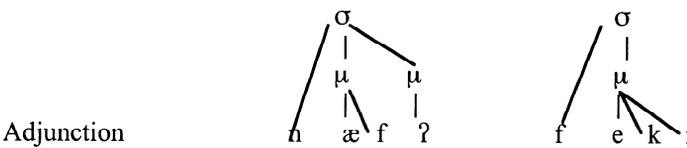
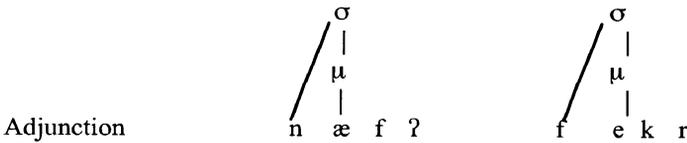
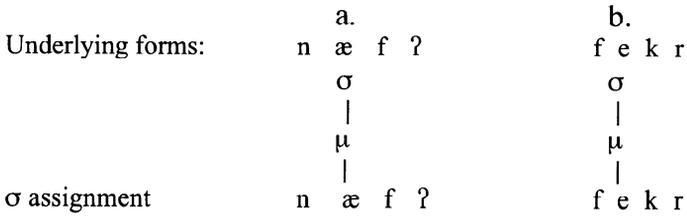
- | | | | |
|-----|---------------------------------|---|------------------|
| (5) | CVG | → | CV: |
| | CVGC | → | CV:C |
| | CVCG | → | CV:C |
| | CVC ₁ C ₂ | → | CVC ₁ |

It is not necessary for glottals to appear in clusters to trigger their deletion. In the colloquial dialect there appears to be a preference for long vowels over glottal consonants when possible. When non-glottals delete, no lengthening occurs.

2.2 Previous analyses

Darzi (1991) discusses the difficulty a moraic theory account of CL has with the Farsi data. Here, I briefly outline his arguments and raise questions about the analysis he provides. First, consider an analysis in the moraic framework.

(6)

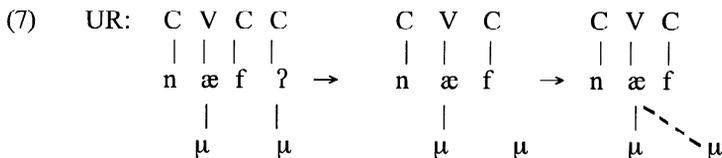


The derivation above is an example of a moraic theory account of CL, like Hayes 1989. Traditionally, syllable assignment takes place, followed by a Weight-By-Position rule that assigns moras to coda consonants. Subsequently, consonant deletion occurs and an adjacent segment associates to the stranded mora left by the deleted segment, resulting in lengthening.

Looking at 6a,b in more detail, it can be seen that the end result is not lengthening. Some distinction must be made to account for lengthening due to the loss of glottal consonants and no other consonants. This example is one in which the distinction is the ability to head a mora. The analysis above correctly predicts that the loss of final non-glottals will not result in lengthening (6b), but it does not account for non-local lengthening due to the loss of glottal consonants (6a). There are two possibilities not

shown. First, in 6a, [f] can be directly linked to the syllable, but if lengthening is attempted, association lines would cross and the derivation would then fail. Second, every consonant can be moraic, but this will predict lengthening when all consonants delete. The analysis then would be too weak to account for the strict conditions under which lengthening applies.

Darzi's analysis resolves the issues of crossing association lines. He does this by forming a CV tier and a moraic tier. By using two tiers, the vowel can freely associate to the stranded mora, as in 7.



In 7, like moraic theory, moras are assigned to eligible codas. The glottal consonant then deletes and since the CV tier is separate from the moraic tier, the vowel can freely associate with the mora left by the glottal consonant. A vowel would not lengthen in the case of non-glottal deletion because non-glottals would not be able to head their own moras, thus not benefiting from a Weight-By-Position rule. The deletion of non-glottals would then leave no stranded mora to which a vowel can associate.

This type of analysis, while solving some issues of a traditional moraic account raises issues of its own. First, this type of representation would allow unlimited long-distance CL that is not exhibited here. Second, In order for the analysis to succeed, it must be stipulated that the glottal consonants are moraic while other consonants are not. This may cause problems when considering the sonority of glottal consonants versus other consonants in the language. Problems will arise depending on the sonority of the glottal stop in Farsi because, as Zec argues, 'if obstruents appear in weight-bearing positions, this implies that all the remaining segments in the inventory are also members of the moraic set' (1995:107).

2.3 Main proposal

The main argument presented here is that CL languages lengthen the vowel because two segments coalesce and this lengthening results because coalescence is not free. In some languages, coalescence is represented by a change in vowel quality and in CL languages, the price is prosodic recognition. The core of CL is composed of two constraints forcing all segments to be acknowledged in the output and to avoid deletion of segments. The composition of these two constraints is the main analysis of CL. However, for each language, there are constraints that need to be used for a complete analysis. For example, in Farsi, glottal consonants delete everywhere except in word initial position. Therefore, a complete analysis will account

for this, and the interaction of this account with the two components of CL will result in lengthening. In this section, I provide an analysis of lengthening in Farsi that can be extended to the general domain of CL. First, as discussed above, some constraints are needed to account for the distribution of glottal consonants.

- (8) MAXWI: Every word-initial segment in the input must have a corresponding segment in the output (Casali 1997)
- (9) IDENTWI: Output correspondents of word-initial input segments with αF are also αF .
- (10) *GLOTTAL: No glottal consonants should be present in the output.

The three constraints above are used simply to ensure glottal deletion everywhere except word initially. Consider again the form 3c, $\text{ʔeʔzam} \rightarrow \text{ʔe:zam}$ ‘dispatch’. In this form, the glottal stop is lost at the end of the first syllable, but not at the beginning of the word. MAXWI, taken from Casali 1997, forces the preservation of word-initial segments. IDENTWI is an extension of that constraint, prohibiting the changing of that word-initial segment. These constraints are needed because of the preference Colloquial Farsi has for long vowels over glottal consonants. Without them, we would expect the initial glottal to delete or to change to a more acceptable consonant. This is illustrated in the tableau below.

(11) Input: /ʔeʔzam/

	MAXWI	IDENTWI	*GLOTTAL
→a. ʔezam			*
b. ʔeʔzam			**!
c. ezam	*!		
d. tezam		*!	

In the above example, 11b-d are ruled out for various reasons. First, candidate 11b has two glottal consonants in the output, making it a worse choice than 11a. As long as MAXWI and IDENTWI are ranked higher than *GLOTTAL, 11b,c will be ruled out because one deletes the word-initial segment and the other changes it. Therefore, the form that loses the medial glottal, but maintains the initial glottal is the optimal form.

Now consider lengthening in Farsi. As stated above, the lengthening analysis presented here has two conditions. They are presented below as constraints.

- (12) **BIPOSITION:** An output segment representing two input segments (denoted by subscripts) must be linked to two prosodic positions.
- (13) **MAX_{SEG}:** Maximize input segments.

BIPOSITION translates into a need to acknowledge all input segments in the output. If two segments coalesce, each segment must somehow be represented. In the case of CL, the representation of two coalesced segments surfaces as one segment linked to two prosodic positions, or as a lengthened segment. Consider the following example.

- (14)
- μ
 |
 Input: /r₁o₂ b₃?₄/

	BIPOSITION	MAX _{SEG}
μ μ / / V / → a.r o _{2,4} b ₃		
μ b. r o ₂ b ₃		*!
μ c . r o _{2,4} b ₃	*!	

There are three possibilities to consider. First is an output with two segments that have coalesced, lengthening the vowel (14a). Second, is a candidate that deletes the glottal stop (14b). Another option is an output like 14c that has coalesced two segments to avoid deletion, but does not acknowledge both segments prosodically. Of the three options, only one survives due to the two constraints used to account for CL. Candidate 14b is ruled out because it deletes a segment, violating MAX_{SEG}. The other possibility, 14c, violates BIPOSITION because it preserves all segments, but does not indicate prosodically that all segments are present. Because Farsi allows non-local lengthening, any type of contiguity constraint would be low ranking to allow for this. Clearly, this analysis is not yet enough for a complete analysis of Farsi, but the combination of BIPOSITION and MAX_{SEG} is the part of the analysis that can be extended to account for CL in general. The use of these two constraints will achieve lengthening in CL languages.

To continue with an analysis for Farsi, remember that CL only occurs with glottal consonants. As discussed earlier, there are potential problems with any analysis that bases CL in Farsi on the premise that glottal consonants are moraic, while other consonants are not. The analysis proposed

here focuses on a difference between glottal consonants and non-glottal consonants. Bessell & Czaykowska-Higgins (1991), following McCarthy, show that glottal consonants are placeless if they pattern differently than uvulars. This is the case in Farsi, as shown below.

- (15) Conservative Colloquial Gloss
 ʃahrox ʃa:rox (proper noun)

We have seen that word-final glottal consonants are lost and trigger lengthening. This is not the case for uvulars in Farsi, as shown above. I therefore consider glottals in Colloquial Farsi to be placeless. Thus, by the constraints in 16, the coalescence of glottal consonants and vowels is allowed, while the coalescence of non-glottal consonants and vowels is prohibited.

- (16) IDENT PLACE: Output correspondents of an input segment with α place are also α place.

The example below illustrates how IDENT PLACE affects non-glottal consonant coalescence, but has no effect on glottal-consonant coalescence.

(17)

Input: / fe ₂ k ₃ r ₄ /	IDENT PLACE
a. fe _{2,4} k ₃	*!
→b. fe ₂ k ₃	
Input: /n ₁ æ ₂ f ₃ ? ₄ /	IDENT PLACE
c. n ₁ æ ₂ f ₃	
→d. n ₁ æ _{2,4} f ₃	

The first form in the tableau is a form that deletes a non-glottal consonant. Deletion is preferred in this case because IDENTPLACE prohibits the coalescence of a non-glottal and a vowel because the consonant would then lose its place as in 17a. Examples 17c,d simply illustrate the fact that this constraint is not a deciding factor in the case of glottal consonants because they are placeless to begin with and whether they delete or coalesce with a vowel, they do not lose any place specifications.

In addition to IDENTPLACE, CL can be ruled out with the loss of non-glottal consonants with one final constraint.

- (18) DEP μ : Every mora in the output has a correspondent in the input.

DEP μ is used to rule out the arbitrary addition of a mora as in 19b below

(19)

μ
 |
 Input: /sa₂x₃t₄/

	IDENT PLACE	BIPOSITION	MAX _{SEG}	DEP μ
μ μ a.s a _{2.4} x ₃	*!			
μ μ b. s a ₂ x ₃			*	*!
μ → c. sa ₂ x ₃			*	
μ d. s a _{2.4} x ₃	*!	*		

The deciding factor between the surfacing form and the losing candidate is DEP μ , which rules out 19b because of the addition of a mora. DEP μ must also be low-ranking because in the cases of CL, an additional mora does appear in the output. However, as example 14 shows, other candidates are eliminated by BIPOSITION and MAX_{SEG}. The lengthened form will violate DEP μ but it will also be the only remaining candidate. Example 19 also shows the crucial ranking of IDENT PLACE above MAX_{SEG}, BIPOSITION above MAX_{SEG}, and DEP μ below MAX_{SEG}. Examples 19a,d are both ruled out because a non-glottal consonant ([t]) coalesces with a vowel and loses its place specification. This leaves us with a final ranking of IDENT PLACE >> BIPOSITION >> MAX_{SEG} >> DEP μ for Colloquial Farsi.

In this section, I hope to have accomplished a number of things. First, the ranking Bipoosition >> Max_{seg} and the use of these two constraints will achieve lengthening across languages. Second, independent constraints decide whether or not segments are deleted or coalesced, as Ident Place in Colloquial Farsi. Third, when no lengthening occurs, a segment has been deleted. Finally, for Farsi, lengthening occurs with the loss of a subset of consonants because glottal consonants are placeless and are able to coalesce, while non-glottal consonants have their own place specifications and must preserve it, thus prohibiting them from coalescing with a vowel.

3 Onsets

As Farsi, Greek also exhibits CL, but the analysis of this will not be covered in this paper. By using BIPOSITION and MAX_{SEG}, Greek CL, including double flop cases, are easily handled. In this section, I concentrate on the loss of onsets because at first glance they appear to be problematic for my analysis. Greek has both word-initial and word-medial onset loss. For this reason, I use data from Greek to illustrate my account of onset loss in CL languages.

3.1 Word-initial onsets

In Greek, initial segments are lost and there is no resulting lengthening, even though Greek is a CL language. This is shown below.

- (20) a. wanaks → anaks 'lord'
 b. werg → erg 'work'
 (Steriade 1982)

If lengthening occurs as the result of coalescence, when there is no lengthening due to the loss of onsets, the onsets must then be deleted leaving no trace of the segment. This can easily be explained for word-initial onsets. By the constraint IDENTWI, CL languages that lose initial onsets would always prefer to delete an initial onset rather than to change one. This idea can be carried over into an analysis of Greek.

(21)

μ
 \uparrow
 Input: /w₁e₂r₃g₄/

	IDENT WI	BIPOSITION	MAX SEG
<p>a. e_{1,2} r₃ g</p>	*!		
<p>→ b. e₂r₃g</p>			*

Candidate 21a follows the CL requirements perfectly, not deleting any segments and representing both prosodically. It is ruled out, though, because it fails to maintain the identity of the word-initial segment. The winning form is the one that deletes the initial segment. This is similar to Farsi. The main difference is that Greek prefers to delete rather than to

change an initial segment and Colloquial Farsi neither changes nor deletes an initial segment.

3.2 Word-medial onsets

Medial onsets are also lost in Greek. This is shown in 22.

- (22) a. elaiwon → elaion 'oil'
 b. newos → neos 'new'
 c. klewos → kleos 'renown'
 (Steriade 1982)

So far, the general trend of this analysis is that segments try to remain in the output and when that is not a possible option, the best result ends up being deletion. For medial onsets, however, there is no straightforward reason to delete them with the constraints used above. The result then, would be as in 23.

(23)

$\mu \quad \mu\mu \quad \mu$
 $\quad | \quad | \quad |$
 Input: /e₁l₂a₃i₄w₅o₆n₇/

	IDENT WI	BI- POSITION	MAX _{SEG}	WBP
<p>a. e₁ l₂ a₃ i₄ o₆ n₇</p>			*!	
<p>→ b. e₁ l₂ a₃ i_{4,5} o₆ n₇</p>				
<p>c. e₁ l₂ a₃ i_{4,5} o₆ n₇</p>		*!		

For Greek, I use a weight-by-position (WBP) constraint parallel to Hayes's rule to have weight-bearing codas in the output, as in all the forms above. This tableau illustrates the idea that medial onsets are not actually deleted, instead they are coalesced. In the surviving candidate, the vowel is acting as an onset to the following syllable. Candidate 23c does not obey BIPOSITION and candidate 23a deletes a segment. The only form left is the

one that has coalesced. While this form is predicted by my analysis, there is an important implication here that cannot go untouched. While there is no phonological lengthening of the vowel in this case, I do predict that there would be some phonetic lengthening because even though onsets are weightless, they do have a measurable length.

4 Conclusion

In this paper, I have addressed reasons for reconsidering previous analyses of CL based on the wide array of data found among CL languages. I have further shown that in Farsi, lengthening can be considered to be the result of coalescence and the lack of lengthening results from deletion. Overall, the ranking of BIPOSITION >> MAX_{SEG} will yield an end result of lengthened segments cross-linguistically. I have also shown that in languages that lose initial onsets without lengthening, those onsets are forced into deletion. As for medial onset loss, I propose that they are, in fact, the result of coalescence and they serve as onsets to the syllables they precede.

While I only have time to go into detail with Farsi lengthening and Greek onsets, this analysis can be extended to account for lengthening in languages that exhibit vowel quality changes along with lengthening, moraic mismatches (Hyman 1992), double flop cases as in Greek, and CL in general. For the Bantu cases of moraic mismatches, adopting this analysis would eliminate the stage in which nasal codas are assigned moras, which in turn eliminates the circularity of the analysis.

When considering compensatory lengthening in Optimality Theory, we have two options. First, we can hold tight to a traditional analysis as in Hayes 1989 and change the theory to fit the data or we can consider the possibility that the previous analyses can benefit from a few adjustments. I opt for the latter in this paper for two main reasons. First, the implications of manipulating the theory to account for CL are far reaching. Second, and finally, changing current theory to involve unnecessary intermediate steps only takes us back to the multi-stage derivational theory from which we came.

References

- Bessell, Nicola, and Ewa Czaykowska-Higgins. 1991. Interior Salish evidence for placeless laryngeals. *Proceedings of the North Eastern Linguistic Society* 21.
- Casali, Roderic F. 1997. Vowel elision in hiatus contexts: Which vowel goes? *Language* 73:493-533.
- Clements, G. N. 1986. Compensatory lengthening and consonant gemination in Luganda, in L. Wetzels & E. Sezer, eds.
- Darzi, Ali. 1991. Compensatory lengthening in modern colloquial Tehrani Farsi. *Studies in the Linguistic Sciences* 21.23-37.

- de Chene, Brent. 1975. The historical phonology of vowel length. UCLA dissertation.
- Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 15.33-74.
- Hyman, Larry. 1992. Moraic mismatches in Bantu. *Phonology* 9.255-65.
- McCarthy, John and Alan Prince. 1993. Prosodic morphology I: Constraint interaction and satisfaction. Technical report 3, Rutgers University Center for Cognitive Science.
- McCarthy, John, and Alan Prince. 1995. Faithfulness and reduplicative identity. *Papers in optimality theory*, ed. By Jill N. Beckman et al., 249-385. Amherst, MA:GrLSA.
- Steriade, Donca. 1982. Greek prosodies and the nature of syllabification. MIT PhD dissertation.
- Zec, Draga. 1995. Sonority constraints on syllable structure. *Phonology* 12.85-129.