# SOME SANDHI' RULES IN PORTUGUESE1

Joel Rotenberg, in his dissertation *The Syntax of Phonology*, deals with syntactic influence in phonology. According to the author, this influence is direct, not mediated by string boundaries. With data from different languages, he demonstrates that certain phonological phenomena show syntactic conditioning of various sorts, at distinct. levels of structure. Defining the notions of 'edge', 'juncture', 'constituency', he postulates a theory of hierarchical domains of application of phonological rules. These domains are: the sentence, the phrase, the 'clitic groups', the word, the morpheme and the syllable. Syllable-level rules have a different status, since they are *not* influenced by any kind of lexico-syntactic conditioning. The use of terminal boundary symbols in the formulation of rules that account for segmental phenomena is, then, proven to be both inadequate and inappropriate.

This paper is an attempt to see if Rotenberg's theory works for Portuguese. I will examine some phonological facts involving sequences of two vowels belonging to adjacent words, trying to postulate the rules that would account for them. The data will be analyzed following the general lines of Rotenberg's ideas, particularly with respect to his notion of domain of application of rules. In connection with the phenomena involving vowels, a rule of palatalization will also be investigated.

#### 1. Some facts about vowels in portuguese

In Portuguese, there are seven oral vowel phonemes<sup>2</sup>, distinguished in stressed position: /a, &, e, i, o, >, u/. Consider the following words:

(1) saco	[sáku]	'bag'
seco	[séku]	'l dry'
seco	[séku]	'dry'
cedo	[sédu]	'early'
sido	[sídu]	'been (p.p.)'
soco	[sɔ́ku]	'I punch'
soco	[sóku]	'blow, punch'
suco	[súku]	'juice'

In unstressed position, however, this system is reduced, some of the oppositions being neutralized. So it is that, in pretonic syllables for example, the distinction between /e,  $\epsilon$ /, on the one hand, and /o,  $\circ$ /, on the other, disappears. Then, we find:

(2)	cereja	[seréža] /	[sɛ̞réžə̞]	'berry'
	melado	[meládu] /	[m¿ládu]	'melada'
	dobrado	[dobrádu]/	[d <u>ɔ</u> brádu]	'folded'
	morada	[moráda]	[m_rádə]3	'abode'

In final unstressed position, a further reduction occurs and only /a, i, u/ appear, there being no distinction between /i, e/ and /o, u/:

(3) bebe	[béb <u>i</u> ]	1	[bébe]	'he drinks'
carro	[káhu]	1	[káho]	'car'

It is also worth mentioning that /a, i, u/, in this final unstressed position, have [a], [1] and [v], respectively, as allophones.<sup>4</sup>
Let us now consider sequences of vowels, as in:

(4) rua	[hú <i>ə</i> ]	'street'
lia	[lía]	'he/she read'
país	[país]	'country'
coar	[kuáh]	'to filter'

In (4), each word contains a sequence of two vowels, one of which is stressed. The pattern is either  $\dot{V}V$  or  $V\dot{V}$ . When this occurs, the vowels belong to different syllables, both being [+ syll ] . Now observe:

(5) pais	[p <u>á</u> ys]	'parents'
beiço	[béysu]	'lip (pop)'
quase	[k <u>wá</u> zi]	'almost'

Here, only one of the vowels is [+ syll]. The same element is also [+ stress]. So, the sequences of 'Vowel + Semivowel' and 'Semivowel + Vowel' belong to the same syllable and we have diphthongs. So far, in the examples given, one of the elements in the sequence is [+ stress]. And either we have two vowels in separate syllables, or they are together in the same syllable, one of them being a glide. Now suppose both are unstressed, as in:

(6) feitiço	[feyčisu]	'sorcery'
glória	[gl͡ɔśryə] <sup>5</sup>	'glory'
aumento	[awmetu]	'raise'

The data in (6) show that again we have diphthongs, one of the two elements being [+ syll]. And this situation is invariable, in the domain of the word: in fact, no two [+ syll] elements occur in sequence, unless one of them is also [+ stress]. Of course, the two vowels will, then, be part of separate syllables. The point now is to see what happens when the vowels are in adjacent words.

## 2. The sanahy rules in' V#V' sequences

Consider the following examples:

(7) nó apertado	[n <u>ɔ̀a</u> pehtádu]	'tight knot'
café escuro	[kaf¿iskúru]	'dark coffee'
esta hora	[Èstəɔ́rə]	'this hour'
carta útil	[kàhtəúčiw]	'useful letter'
está ótimo	[istàjɔ́cimu]	'it is very good'

In the above cases, one of the vowels is [+ stress] and there is no special phenomenon: the words are pronounced as when in isolation. The two adjacent vowels maintain their individuality and are assigned to different syllables. In the last example in (7), both vowels are [+ stress] and, again, nothing happens.

Cases of another sort can be seen in:

(8) mesa amarela	/méza#amar€la/	'yellow table'
atitude ilícita	/atitúdi#ilísita/	'illicit attitude'
caso utópico	/kázu#utɔ́piku/	'utopic case'

In (8), both underlined vowels are [- stress] and, furthermore, they are identical. And the corresponding phonetic forms are:

(9)[mèzamarélə] [ačitù]ilísitə] [kàzutɔ́piku]<sup>6</sup>

We need, then, a rule to account for the fact that the two identical vowels are reduced to just one. Before postulating such a rule, we should investigate further, to see if there is any syntactic condition that controls the phenomenon. In the three examples in (8), we have a sequence of 'Noun + Adjective' and the loss of the vowel could be restricted to this environment only. But then we find:

(10) a. 'Det + Noun' esta atividade [ Estačividáji ] 'this activity' [uzuráryu] 'the usurer' o usurário b.'Verb + Adverb' 'he speaks wisely' [fàlažuyzádu] fala ajuizado 'he drinks in secret' [bkbiskujídu] bebe escondido c. 'Det + Noun + Verb' o doce estragou [udòsistragów] 'the candy is spoiled' 'the guy used o moço usou. . . [umòsuzów] d. 'Prep + Noun' [pàrasésu] para acesso 'as access' [sòbristźrya] sobre história 'about history'

The cases in (10) are enough to show that the reduction of the two identical vowels to just one is *not* restricted to 'Noun + Adjective' sequences. In fact, the list of environments could be extended to indicate that the phenomenon occurs everywhere within the sentence, disregarding the nature or class of the constituents. We could, then, be tempted to say that the rule to be postulated will be sentence-level. But this is not quite correct, since it applies even across the abnormal 'junctures': traces, parentheticals and sentence boundaries. According to Rotenberg, these junctures do prevent the application of rules with syntactic conditioning. However, they do not block the reduction of vowels in the Portuguese following cases:

(11) a. A moça apesar de que odeio admiti-lo — armou-se de coragem.

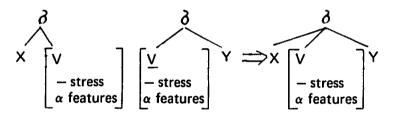
'The girl — although I hate to admit it — armed herself

with courage'.

b. Maria entrou no carro, usou a chave e partiu. 'Mary got in the car, used the key and left.'

In (11a.), we have parentheticals and in (11b.), sentence edges. And the reduction occurs in both cases. We are forced to recognize that the rule in question is *not* subject to any lexico-syntactic conditioning and has a special status. We cannot maintain anymore that it is sentence-level, since it goes beyond its limits. In fact, it behaves exactly like syllable-level rules, that are conditioned by prosodic structure instead. So it is that the reduction of the two vowels into one is blocked by pause, ignoring phrase structure —as long as it is not reflected by pause —, applies even across traces, parentheticals and sentence edges. Those are the characteristics of the palatalization rule in English, that, according to the author in study, Rotenberg, has the syllable as domain? We will call the rule of Portuguese VOWEL COALESCENCE and it will have the following form:

(12)



The rule is obligatory, since we do not have:

- (13) \*[mèzaamaréla]
  - \*[ačitù]ilísita]
  - \*[kàzuutɔ́piku]8

Once again, we see that there is no sequence of two [+ syll, - stress] elements in Portuguese. In the examples just seen, to avoid the occurrence of such sequence, a rule applies that converts the two vowels into one. Observe now the following data:

(14) casa escura /káza#iskúra/ 'dark house' cara otário /kára#otáryu/ 'a dupe' aquela eremita /akɛ̃la#eremíta/ 'that hermit (fem.)' problema uterino /problema#uterínu/ 'uterine problem'

In (14), we have a final unstressed /a/, followed by different vowels, also unstressed. The final phonetic forms are:

```
(15)[kàziskúra]
[kàrotáryu]
[akèlerēmítə]
[problěmuterinu]
```

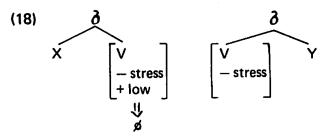
In all cases, the final /a/ is dropped and the initial vowel of the following word remains unchanged. We need another rule to account for the facts. We can see that it applies without any syntactic conditioning, disregarding the abnormal junctures, as in the case of VOWEL COALESCENCE (VC):

```
(16) a. 'PRO + Verb'
    ela estuda [ɛ̃listúdə] / [ɛ̃lestúdə] 'she studies'
    ela examina [¿lizamína] / [¿lezamína] 'she examines'
    b. 'Prep + Noun'
                     [pàroperáryus]
    para operários
                                          'for workmen'
                                         'against nettle'
    contra urtiga
                     [kõtruhčígə]
    c. Verb + Adverb'
    fala escondido [fàliskūjídu] / [fàleskõjídu] 'he speaks
                                                    in secret'
    canta humildemente [katumiwjiméči] 'he sings humbly'
    d. Across parentheticals
    A moça – e Maria viu tudo – foi atropelada pelo carro.
    'The girl - and Mary saw everything - was hit by the car'.
    e. Across sentence edges
    O homem chamou a moça, olhou em volta e correu.
    'The man called the girl, looked around and ran'.
```

We will, then, assume that the rule is also syllable-level. It is blocked by pause: the relevant parts in (16 d.) and (16 e.), above, are pronounced as in (17), when a pause separates the vowels under investigation:

(1,2) [amósa . imaría...] / [...amósa . ολόw...] <sup>9</sup>

The rule will be called VOWEL DELETION (VD):



It will have to be optional, since we have:

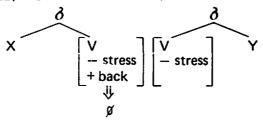
The examples in (19) are disturbing. They do not fit into the pattern found so far: they show a sequence of two[+syll, - stress] segments, that belong to different syllables. We would need to weaken the statement about sequences of vowels in Portuguese and permit two unstressed syllabic elements in a row, when the first one is /a/. And we should remember that this is only a partial conclusion, since we have not investigated, so far, sequences where the first vowel is /i/ or /u/. But it would be better if we could find another way out of the problem. And it seems to me that this is possible, if we assume that the forms in (19) are, in fact:

(20) does not disturb the pattern. It would be necessary to postulate a rule to make the second vowel in the sequence a semi-vowel: /i, e/ would be changed into[y] and /u, o/ into[w]. I will leave this as an open question at the moment.

Now consider:

From the examples in (21), we see that final unstressed /u/, followed by another unstressed vowel, is deleted in the same way as /a/a. We, then, modify VD as:

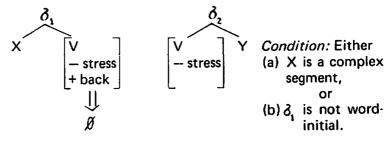
### (22) VOWEL DELETION (modified version)



Now VD drops final unstressed /a/ and /u/. When we investigate further, we find the following:

Since the forms in (23) are ungrammatical, we need to impose a condition in rule (22) - VD -, to prevent its application in the relevant cases. Observe that words like 'no', 'da', etc, are not tonic and, in fact, combine two morphemes: 'no' results from 'em + o', 'da' comes from 'de + a'. Let us, then, say that the condition in (22) will have to mention that the syllable in which the first vowel occurs is not word initial. We will also add a condition to allow for the dropping of final /a/ in the example where the preposition para assumes the reduced form 'pra'. So, we reformulate VD again, this time as:

## (24) VOWEL DELETION (final version) - Optional



As an illustration, some of the relevant syllables in the examples in (23) are:

(25) 
$$\stackrel{\#}{\downarrow}$$
  $\stackrel{\delta_1}{\downarrow}$   $\stackrel{\#}{\downarrow}$   $\stackrel{\delta_2}{\downarrow}$   $\stackrel{\#}{\downarrow}$   $\stackrel{\dagger}{\downarrow}$   $\stackrel{\dagger}{$ 

Condition (a) in (24) makes it possible to have [pris...]; condition (b) does not allow the deletion of the final /u/ in / $\mu$ du $\mu$ a.../, thus preventing the generation of ungrammatical \* [dapéhtu], for instance.

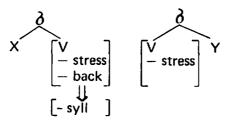
We pointed out before that only /a, i, u/ occur in final unstressed position, in Portuguese. So far, we have dealt with facts involving /a/ and /u/, postulating rule (24), to account for them. Let us now see what happens with /i/:

By the above examples we realize the need for yet another rule: final unstressed /i/ becomes a semivowel. The rule is also syllable-level; it is not subject to any syntactic conditioning, applying across abnormal junctures. It is an obligatory rule, as the following data show:

'The hare - although the dogs were loose - ran.'

Here we see again another way to avoid sequences of two [+ syll, - stress] segments: a rule will convert the first one into a [- syll] element, obligatorily:

# (28) VOWEL REDUCTION - Obligatory



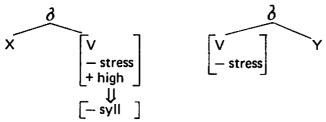
(28) will have no conditions attached to it, of the type VD has. See, for example:

Consider again the following data:

As mentioned before, VD is an optional rule. In this case, the forms in (30) should have a phonetic realization where the final /u/ is still present. But, instead, we find:

(31 a) shows that it is not grammatical to keep the final vowel unchanged; in fact, as (31b.) confirms, this vowel becomes the semi-vowel [w]. We can modify VOWEL REDUCTION (VR), so as to accompdate these facts:



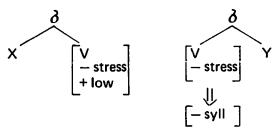


With this last modification of VR-I, sequences of two [+ syll, - stress] segments are, again, destroyed. In fact, the three rules postulated so far lead to this result. The only disturbing examples are brought about by the fact that VD is an optional rule and, when it does not apply, the following forms turn out:

As mentioned before, the examples in (19) could be interpreted as ungrammatical and the correct phonetic representations for them would be:

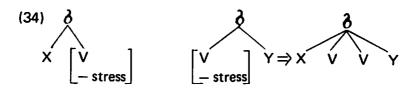
Suppose we accept that (20) is the actual pronunciation. Now, in the light of VR-I, that makes a semivowel out of /i/ and /u/, thus avoiding the sequence of two unstressed vowel in a row, it is not unreasonable to assume that another rule will take care of the cases where /a/ is the first member of the sequence. This new rule will operate on the second vowel, producing[y] or[w]. Recall that, since in pretonic position only /a, e, i, o, u/ occur, we will have to explain /e, i/ becoming [y] and /o, u/ becoming[w]. I will go further, however, and assume that the new VR rule will operate only on /i/ and /u/: there is an alternation between /i, e/ and /u, o/, when in initial unstressed position. We have, for example, [izaminə]/ [ezaminə]; [urísu] / [orísu]. Tentatively this rule is:

# (33) VOWEL REDUCTION II — Obligatory

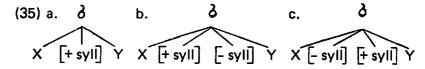


Adopting (33)<sup>14</sup>, we can assume that there are no exceptions to the fact that Portuguese does not have any sequence of two [+ syll, - stress] segments. This would be one statement about the segmental

phonology of the language. And, then, we could formulate:



The four rules postulated in this paper would aply to the outcome of (34) and would be simplified, as a result: no mention of the feature [+stress] is necessary. These rules will operate so as to modify the sequence 'XVVY' into one of the following syllable structures:



VC — rule (12), obligatory, that applies to identical vowels — will produce the structure in (35a.). The same configuration will come about by the use of VD, the optional rule (24), which drops the first vowel of the sequence, either /a/ or /u/.

Recall that VD is optional. So, when it does not apply, /a/ and /u/ will remain as the first member of the two syllabic elements in a row. Then, to take care of /a/, VR-II — rule (33), obligatory — acts and we get the structure in (35b.). And to account for the sequences of '/uV/' and '/iV/', VR-I — rule (32), also obligatory — changes /u/ and /i/ into [w] and [y], respectively, and the structure (35c.) is obtained.

I will now explore another rule of Portuguese phonology, to see its interaction with the ones postuled so far. Observe the examples bellow:

(36)	lata	/lá <u>ta</u> / [lá <u>ta]</u>	'can'
	tipo	/ <u>tí</u> pu/ [ <u>čí</u> pu]	'type'
	bondade	/bő <u>dádi</u> / [bő <u>dájí</u> ]	'goodness'
	chateado	/ša <u>tiádu</u> / [ša <u>čy</u> á <u>du]</u>	'troubled'
	telha	/té/a/ [té/a]	'tile'

To account for the data in (36), we need a rule that palatizes /t/ and /d/, when followed by [i]or[y]:

Palatalization is obligatory, since we do not have:

Now we must decide what is the domain of application of (37), Will it be free of any lexico-syntactic conditioning, as VC, VD, and VR? Can it apply across words, phrases and sentences? And the answer is that Palatalization is a word-level rule. It does not operate beyond the limit of the word. We verify this by looking at:

(39) shows that Palatalization is *not* allowed, outside the domain of the word. See that, after VD applies, we have the structural description adequate for (37). And, yet, [t] and [d] remain unchanged. We see here that these facts of Portuguese confirm Rotenberg's theory. Having different levels of rule, establishing domains of application for the rules, makes it possible to explain some phonological phenomena of Portuguese, without the use of boundary symbols. But observe also that we could obtain the same results in (39), if we should order Palatalization before any of the rules postulated in this paper.

As a last remark on the Palatalization rule, we should see if it applies within the word, across morphemes. According to Rotenberg, this should be expected. And, in fact, it is:

In the above examples, /- iña/ is the diminutive morpheme. And (37) is obligatory also across morphemes.

#### 3. Summary

Investigating the structure of isolated words and of sequences of words in Portuguese, we ended up by assuming that, in colloquial speech:

(41) a. Sequences of the form 
$$\begin{pmatrix} V \\ -stress \end{pmatrix} = \begin{bmatrix} V \\ -stress \end{bmatrix}$$
 are not possible

- b. Four different rules operate, at the level of the syllable, to avoid such sequences:
  - (i) VOWEL COALESCENCE Obligatory
    This takes care of two identical vowels in a row;
  - This takes care of two identical vowels in a row;
    (ii) VOWELL DELETION Optional

It accounts for the cases where /a/ and /u/ are the first vowel in the sequence;

- (iii) VOWEL REDUCTION I Obligatory It changes /i/ and /u/ into the corresponding semivowels. It applies after VD, to take care of the cases where /u/ is not dropped;
- (iv) VOWEL REDUCTION II Obligatory It makes /i, e/ and /u, o/ the semivowels [y ] and [w], respectively; it also operates after VD;
- c. Palatalization is a word-level rule in Portuguese; it could be ordered after VC, VD and VR-I and VR-II.

The facts in (41) are evidence in favor of Rotenberg's theory: syllable level rules are not subject to any lexico-syntactic conditioning and apply even across abnormal junctures. The use of terminal boundary symbols in the formulation of rules is desnecessary, if you adopt the notion of domain of application.

I would like to finish by saying that the results arrived at in this paper could be changed and are not, in any way, absolute. I have not investigated data with nasal vowels or sequences of more than two syllabic elements, for example. The rules I postulated lead to resyllabification and I have not treated this in any explicit way. All these fatcs could result in another paper.

#### NOTES

- 1. This article was originally presented as a final paper at the course 'Seminar in Phonology', ministered by Prof. John McCarthy, at The University of Texas at Austin, during the Fall of 1979. Some slight modifications were made, since then. I am indebted to Yara Goulart Liberato, who, indirectly, gave me the idea for this paper. She deals with the same phenomena, under a dif' ant point of view, in: LIBERATO, Yara Goulart. "Alterações valicas em Final de Palavra e a Regra de Palatalização". Ensaios de Lingüística 1:80-95, 1978.
- 2. According to some authors, there are also five distinct nasal vowel phonemes in Portuguese: /ã, ẽ, ĩ, õ, ũ/. There is some controversy as to their phonemic status, however. On this, see, for example: PONTES, Eunice. Estrutura do Verbo no Português Coloquial. Petrópolis, Vozes, 1973.
  I will not deal here with the nasal vowels. The phenomena to be investigated are to be restricted to sequences of oral vowels.
- 3. This difference in pronunciation is dialectal. In certain cases, depending, also, on the level of speech, there is still a third form for the words, where [i] alternates with [e, e] and [u] alternates with [o,o]. This is what happens with ferida 'wound', that is pronunced [firida], in Minas Gerais, [ferida], in São Paulo and careful speech in Minas Gerais and [ferida], in Bahia. For more details, see: CAMARA, Mattoso. Estrutura da Língua Portuguesa. Petrópolis, Vozes, 1970. The data under investigation, in this paper, will be from the dialect spoken in Minas Gerais. Furthermore, I will be discussing the colloquial, fast speech.
- 4. The forms in (3) with final [e] and [o] are not used in colloquial speech. But they would be considered Portuguese, if pronounced. Now, the Portuguese system of oral vowels and their allophones is, of course, more complex. I am not interested in giving here all the facts about it, since it will not be necessary for this paper.
- 5. This is not the only possible pronunciation for the final vocalic group in the word, here. It is, however, the one used in fast speech. If we should consider the sequence [yə] as two separate [+ syll] elements [iə] —, then the statements about (6) should be weakened. It would be possible to have two [+ syll, stress] segments in a row. But I am adopting the point of view that this is not the case in colloquial, fast speech.

- 6. As pointed out before, final unstressed /a, i, u/ have [ə], [ɪ] and [v], respectively, as allophones. In fact, these forms also occur in pretonic position. In this case, the sequences in (9) have at least one more pronunciation, where the relevant vowels are the lax ones. I will assume, however, that the rule that produces the lax allophones is a very late one, occuring after the dropping of the vowels in (9).
- 7. For details on this theory, see: ROTENBERG, Joel. *The Syntax of Phonology*. Unpublished doctoral dissertation, MIT, 1978, p. 201-206.
- 8. Considering the allophones of unstressed /a, i, u/, these forms could have the lax counterparts in them. But it would not make a difference, since they would still be ungrammatical. The rule in (12) does not allow for the reduction to occur, when the sequences are 'a+ a', 'I + i', for instance. I am assuming, however, that the lax vowels are produced later in the derivation.
- 9. The period (.) indicates the pause, in these examples.
- 10. The explanation for (17) assumes a different light now. Observe that, in (19), it is already the allophone [ə], that appears. It is, in fact, very difficult to hear it. I do not really know if these forms are possible in fast speech. They do not seem very natural.
- 11. This is the correct phonetic form for the expression 'de esquina'. Again I should say that I am not very certain about the status of the representations in (23). That is, are they really ungrammatical? In case they should be grammatical, we would drop the condition (b) in rule (24).
- 12. These examples were given before, in (21), p. 11 of this paper. I am not really very interested in the accuracy of the phonemic representations in here or in the other changes from (30) to (31).
- 13. Liberato (1978) assumes exactly the opposite. She would consider (19) the correct phonetic form for these words with [ə] the [- syll] element of the sequence —; and (20) would be ungrammatical. I do not think we can be absolutely certain about the facts here, without the use of a mechanical apparatus. It seems to me that, when /a/ is the first member of the sequence, this vowel is almost always dropped, in fast speech. In other words, (19) and (20) would be ungrammatical. Evidently,

this would lead to a reformulation of my proposal.

14. Note that, considering the way rule (33) is formulated, it does not really matter if it is /i/ or /i, e/, on the one hand and /u/ or /u, o/, on the other, that change into [y] and [w], respectively. The formulation of VR-II would be more complex, if we should consider the groups /i, e/ and /u, o/.

#### REFERENCES

- 1. CAMARA, Mattoso. Estrutura da Língua Portuguesa. Petrópolis, Vozes, 1970.
- 2. CUNHA, Celso. *Gramática do Português Contemporâneo*. Belo Horizonte, Bernardo Álvares, 1970.
- 3. LIBERATO, Yara Goulart. Alterações em Final de Palavra e a Regra de Palatalização. Ensaios de Linguística 1: 80-95, 1978.
- 4. PONTES, Eunice. Estrutura do Verbo no Português Coloquial. Petrópolis, Vozes, 1973.
- 5. ROTENBERG, Joel. *The Syntax of Phonology*. Unpublished doctoral dissertation, MIT, 1978.