Compensatory lengthening in Samothraki Greek: some problems and theoretical consequences

Nina Topintzi, UCL
i.topintzi@ucl.ac.uk --- www.phon.ucl.ac.uk/home/nina

1 Facts and the problems

Samothraki Greek (SamG) compensatory lengthening (CL) occurs after /\-/loss in an onset position (Hayes 1989, Katsanis 1996).

Word-initially, singleton /\-/ deletion leads to vowel lengthening [(1)], but word-medially, no lengthening occurs [(2)].

(1) róγa > óoγa riz > ïz réma > éema rúxa > úuxa rafts > áafts
(2) θaró > θaό kséru > kséu lurí > luí

/\-/ deletion within an onset cluster always happens with subsequent lengthening [(3)].

(3) krató > kaató ádras > ádaas prótos > póotos téreo > téexu mávros > máavus

N.B: This is the core of the data; for an additional pattern where the cluster is followed by a vowel in which case metathesis applies and no r-deletion occurs, see Topintzi (2005, in prep.).

However, /\-/ does not delete when it belongs to a coda.

(4) fanár arpázu karpós

Classic analysis of CL (Hayes 1989)

CL as mora preservation: the total number of moras stays constant, i.e. segments delete, moras survive, e.g. kan → .ka\textsuperscript{m}n → .ka\textsuperscript{m}∅ → .ka\textsuperscript{m}m

- two defining features:
  - the process occurs serially - existence of intermediate stages
  - the application of Weight-by-Position (WBYP): a rule that assigns moras on coda consonants

Problems for an OT analysis of CL:

i) general problems:
- Lack of intermediate stages during which syllabification, moraification and subsequent deletion occur

ii) specific to SamG:
- CL occurs after onset loss and not after coda loss
- Seems to contradict Hayes’ claim (1989) that CL may only occur in languages where there is a syllable weight contrast (cf. Piro for a similar example, Lin (1997))

Current Proposal:

CL is not about mora preservation, but about position preservation through a mora.
2 OT analyses of CL

2.1 Intermediate representations

- An attempt for a direct translation of the Hayesian model in OT either implicitly or explicitly. Such analyses are also designed to handle opacity more generally. Major proponents include:
  - Sprouse (1997): intermediate representations in the form of enriched inputs, i.e. inputs which exhibit syllabification and moraification.
  - McCarthy’s (2003) sympathy theory: covert intermediate level called a sympathetic candidate, which never emerges, but interacts with other candidates and eventually determines the correct output.

- Problems: i) complexity of systems, ii) learnability, iii) failure to produce the correct results in certain occasions (cf. Bermúdez-Otero 2001), iv) risk of re-introducing serialism in the model and thus reverting to older theories with all the problems these faced.

2.2 Underlying moraicity and Richness of the Base (ROTB)

- CL can be generated depending on the moraicity of inputs in some cases (5), but not in others (6).

<table>
<thead>
<tr>
<th>(5)</th>
<th>moraic input → CL</th>
</tr>
</thead>
<tbody>
<tr>
<td>/CV^μC^μ/</td>
<td>NOCODA</td>
</tr>
<tr>
<td>a. CV^μ</td>
<td></td>
</tr>
<tr>
<td>b. CV^μμ</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(6)</th>
<th>non-moraic input → no CL</th>
</tr>
</thead>
<tbody>
<tr>
<td>/CV^μC/</td>
<td>NOCODA</td>
</tr>
<tr>
<td>a. CV^μ</td>
<td></td>
</tr>
<tr>
<td>b. CV^μμ</td>
<td></td>
</tr>
</tbody>
</table>

- Possible solution: inputs are moraically specified like in (5).

- Problem: Richness of the Base violation (Prince and Smolensky 1993/2004). ROTB requires that there are no restrictions placed on inputs.

- Is ROTB needed? A debatable issue, not thoroughly examined so far, see van Oostendorp (2000) or Zamma (1998): English cluster simplification: [n] ~ [gn] as in sign ~ signature. A single input has to be selected, namely that with /gn/, because an input with /n/ cannot generate the surface [gn].

- Despite certain problems ROTB faces, underlying moraicity is not promising (cf. Hermans 2001).

2.3 Segmental faithfulness

- Lee (1996) sees CL as the preservation of the “numerical integrity of segments of a morpheme”. Thus, a form such as /C_V2C_C_V5/ becomes [C_V2C_C_V5]. Several problems: i) no explanation why an output [C_V2C_C_V5], i.e. without lengthening, should be excluded, ii)
although the use of moras is initially deserted, it is re-introduced in CL cases such as double 
flop, e.g. Ancient Greek /ødws/ → [ɔːdɔs] or Bantu prenasalization /muntu/ → [muːtu], iii) 
underlying moras (cf. p. 21-26) and input syllabification (cf. p. 4, 22-24) are utilized, both of 
which are problematic.

2.4 CL as coalescence

* Sumner (1999) treats CL as segmental preservation through the use of numerical indices. The 
advantage of this approach is that it does not refer to input moraicity. We will see later on how 
this analysis fails with respect to SamG.

3 An alternative: CL as position preservation

* Basic idea: CL is not mora preservation, but position preservation via a mora.

* An input segment needs to have an output correspondent, by either: i) surviving in the output 
intact (or with some of its identity features altered) or ii) having a prosodic correspondent in 
the guise of a mora. This can be coded in the constraint termed Position Correspondence 
(Poscorr).

(7) Poscorr: An input segment must have an output correspondent either segmentally by means 
of a root node or prosodically by means of a mora

Implementation: hypothetical example of synchronic CL -- the root /kan/ may either surface on its 
own or be followed by vowel-initial (-a) or consonant-initial (-ta) suffixes.

a. /kan/ → [ka:]

b. /kan-ta/ → [ka:ta]

c. /kan-a/ → [kana]

N.B: Indices are used to identify corresponding segments. To avoid too many indices, moras of 
input segments will have no indices. Only the inserted mora [in bold] due to CL will have an index to 
mark the mapping to the deleted segment.

(9) /kan/ → [ka:]

<table>
<thead>
<tr>
<th>/ka\t\i\t2\i\h\n3/</th>
<th>NOCODA</th>
<th>POSCORR</th>
<th>DEP-μ</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ...a\i\t\i\h\n2\i\h\n3\i\h\n3</td>
<td>*!</td>
<td>...</td>
<td>*</td>
</tr>
<tr>
<td>b. ...a\i\t\i\h\n2\i\h\n3\i\h\n3</td>
<td>*!</td>
<td>...</td>
<td>*</td>
</tr>
<tr>
<td>c. ...a\i\t\i\h\n2\i\h\n3\i\h\n3</td>
<td>...</td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td>d. ...a\i\t\i\h\n2\i\h\n3\i\h\n3</td>
<td>...</td>
<td>...</td>
<td>*</td>
</tr>
</tbody>
</table>

But how about ka\i\t\i\h\n2\i\i\h\n3\i\i\h\n3\i\h\n3? This has no lengthening, but presents coalescence of the deleted segment 
with the previous vowel. Ranking UNIFORMITY (McCarthy and Prince 1995) highly enough, this 
candidate will be easily excluded as in (11a).

(10) UNIFORMITY: No element of the output has more than one correspondent in the input, i.e. no 
coalescence

(11) /kan/ → [ka:]

<table>
<thead>
<tr>
<th>/ka\t\i\t2\i\h\n3/</th>
<th>UNIFORMITY</th>
<th>POSCORR</th>
<th>DEP-μ</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ...a\i\t\i\h\n2\i\h\n3\i\h\n3</td>
<td>*!</td>
<td>...</td>
<td>*</td>
</tr>
<tr>
<td>b. ...a\i\t\i\h\n2\i\h\n3\i\h\n3</td>
<td>...</td>
<td>...</td>
<td>*</td>
</tr>
</tbody>
</table>
Same analysis works for /kan-ta/ → [ka:ta]. How about /kan-a/ → [kana]?

(12) \[
\begin{array}{c|cccc}
/k_a a_2 u_n_3 a_4 u/ & \text{UNIFORMITY} & \text{NOCODA} & \text{POSCORR} & \text{DEP-\(\mu\)} \\
\hline
\text{a. }..a_2^u n_3 a_4^u & & & & \\
\text{b. }..a_2^u p n_3 a_4^u & & & *! & \\
\text{c. }..a_2^u a_4^u & & *! & & \\
\end{array}
\]

(13) **General pattern for CL**
- a trigger for deletion of segments in particular syllabic positions must exist
- PosCORR >> DEP-\(\mu\) ensures that CL occurs, i.e. insertion of a mora identified with the deleted input segment
- UNIFORMITY >> DEP-\(\mu\): prosodic rather than segmental identification is preferred

This system can also easily generate the absence of CL, as shown in (14).

(14) **Deletion of a segment not followed by CL:**
- requires minimal re-ranking: DEP-\(\mu\) >> PosCORR implying that no trace (segmental or prosodic) of the deleted segment is left behind

- Advantages of this proposal:
  - it does not require intermediate levels
  - it makes no reference to underlying moraicity
  - it does not violate ROTB
  - it treats segments equally, thus it is anticipated that deletion of all segments (under certain conditions, cf. Topintzi in prep.) can lead to CL
  - thus it can naturally account for cases such as SamG CL after onset loss, or Piro CL after deletion of unsyllabified consonants (Topintzi in prep.)

4 Samothraci Greek analysis

Summarizing the patterns:

(15) SamG /r/ patterns

\[
\begin{array}{c|cc}
\text{Coda /r/} & \text{Deletion} & \text{Lengthening} \\
\hline
\text{NO} & \text{NO} \\
\hline
\text{Singleton onset /r/} & \text{word-initially} & \text{YES} & \text{YES} \\
\text{word-medially} & \text{YES} & \text{NO} \\
\hline
\text{Complex onset /r/} & \text{Cf+V+C} & \text{YES} & \text{YES} \\
\end{array}
\]

**Assumption:** SamG /r/ is placeless [for discussion see Topintzi 2005]. This has two effects: a) it can survive as a coda, since placeless codas are accepted, but will delete from an onset, b) it allows a particular instance of V-spreading (for more on this see Topintzi 2005, in prep.). This is encoded in:

(16) *ONSET/r: /r/ is disallowed in onset position

Additional constraints required [some extra will be added in due course]:

(17) **UNIFORMITY:** No coalescence
**POSCORR:** An input segment must have an output correspondent either segmentally by means of a root node or prosodically by means of a mora
**DEP-\(\mu\):** Do not insert moras in the output
4.1 Coda /r/

In codas, /r/ does not delete (18), thus any lengthening is gratuitous.

(18) fanár arpázu karpós

(19) no CL when /r/ in coda: UNIFORMITY, *Onset/r, PosCorr >> DEP-µ

<table>
<thead>
<tr>
<th>/kᵢ₉ᵣ₃pᵢ₉₀s₈/</th>
<th>UNIFORMITY</th>
<th>*Onset/r</th>
<th>PosCorr</th>
<th>Dep-µ</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ...aᵢ₉ r₃·</td>
<td></td>
<td></td>
<td></td>
<td>*!</td>
</tr>
<tr>
<td>b. ...aᵢ₉[rᵢ₉]·</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. ...aᵢ₉[rᵢ₉]·</td>
<td></td>
<td></td>
<td>*!</td>
<td></td>
</tr>
</tbody>
</table>

4.2 Singleton /r/ word-initially

/r/ deletion is enforced by *Onset/r, which now becomes relevant. Lengthening also occurs.

(20) róya > óoya riz > íiz réma > éema rúxa > úuxa rafts > áafts

The ranking UNIFORMITY, *Onset/r, PosCorr >> DEP-µ is justified.

(21) CL of singleton /r/ word-initially: UNIFORMITY, *Onset/r, PosCorr >> DEP-µ

<table>
<thead>
<tr>
<th>/rᵢ₉aᵢ₉[sᵢ₉uᵢ₉]·</th>
<th>UNIFORMITY</th>
<th>*Onset/r</th>
<th>PosCorr</th>
<th>Dep-µ</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. aᵢ₉·</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. rᵢ₉aᵢ₉·</td>
<td></td>
<td></td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>c. aᵢ₉·</td>
<td></td>
<td></td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>d. aᵢ₉[rᵢ₉]·</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

4.3 Singleton /r/ word-medially

Word-medially no lengthening occurs.

(22) əaró > əaó kséu > kséu luri > lúi

* Is this puzzling? Imagine if an input like /lᵢ₉uᵢ₉[rᵢ₉iᵢ₉]/ had lengthening. We would either get: i) lu:i [lᵢ₉uᵢ₉[rᵢ₉iᵢ₉]] or ii) lui: [lᵢ₉uᵢ₉[rᵢ₉iᵢ₉]].

* A sequence VːV (or VVV) would be created, but this tends to be avoided in languages, cf. Kavitskaya (2002: 43) who claims that the ban on VːV “... can be motivated by perceptual properties of VV sequences: since vowel-to-vowel transitions are always very long, a two-vowel sequence is not likely to be re-interpreted as a three-vowel one”.

* Proposal: *S-L VT markedness constraint which is very highly-ranked in SamG.

(23) *Super-Long Vocalic Transitions (*S-L VT) / *VːV


<table>
<thead>
<tr>
<th>/lᵢ₉uᵢ₉[rᵢ₉iᵢ₉]·</th>
<th>*Onset/r</th>
<th>*S-L VT</th>
<th>PosCorr</th>
<th>Dep-µ</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ...uᵢ₉[rᵢ₉iᵢ₉]·</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. ...uᵢ₉[rᵢ₉iᵢ₉]·</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. ...uᵢ₉[rᵢ₉iᵢ₉]·</td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. ...uᵢ₉[rᵢ₉iᵢ₉]·</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
This time the winning candidate is one where PosCorr cannot be satisfied, as in (d), so the winner shows no lengthening. The relationship between (a) and (d) offers the ranking argument *Onset/r >> PosCorr, yielding the adapted ranking:


### 4.4 Complex clusters of the type Cr+V+C

Output of clusters of this type involve /r/ deletion and lengthening as in (26).

(26) krató > kaató ádras > ádaas prótos > póotos tréxo > téexu mávros > máavus

Given the grammar employed so far, the right results obtain in a manner by now familiar.

(27) CL of /r/ in Cr+V+C: Uniformity, Ons/r >> PosCorr >> Dep-μ

<table>
<thead>
<tr>
<th>/m₁u₂₅₄u₅₆</th>
<th>Uniformity</th>
<th>*Onset/r</th>
<th>PosCorr</th>
<th>Dep-μ</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ...₃₃₄₅₆</td>
<td>*!</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. ...₄₃₄₅₆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. ...₃₃₄₅₆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. ...₃₃₄₅₆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

N.B: A candidate like mi₃₄₅r₆₇₈₉, ie. [₃₄₅₆₇₈₉] is excluded because it would entail that the onset is moraic. Assuming that *Moraic Onset is high-ranked, this competitor is avoided. It is interesting however to note that while the current proposal allows Cl because of onset loss, this does not have to relate to moraic onsets. As is evident, SamG CL presents the proposed pattern that CL needs to exhibit, namely: Markedness, Uniformity >> PosCorr >> Dep-μ.

### 5 SamG and the other analyses

- All the alternatives as we have seen fail with respect to basic principles of OT and thus are untenable
- Moreover, none of them is designed to handle CL after onset loss, since they either involve mora preservation (and they assume that onsets never have moras) or explicitly deny the existence of such a possibility
- Only Sumner (1999) seems to allow such a case, which is why I return to her proposal in more detail.
- Sumner sees CL not as the product of mora preservation, but as the result of:

(28) Biposition >> Max-Seg

The former constraint is defined in the following way.

(29) Biposition: An output segment representing two input segments (denoted by subscripts) must be linked to two prosodic positions. [Sumner 1999: 538]

- Reference to ‘two prosodic positions’, does not necessarily impose linking to two moras. Let us see how this approach fares when considering the SamG data.
- Consider a case, such as /krató/ becoming [kaató].
(30) Input: /kr_{23}a_{34}o_{56}/

(a) fails BIPOSITION since the coalesced vowel only links to one prosodic position; (c) has two
links, but exhibits a strange structure which could be acceptable for high vowels as glides, but is
highly unlikely one for other vowels. (b) satisfies BIPOSITION by virtue of the extra mora the
coalesced vowel is linked to. Under this approach then, CL indeed has to do with segmental
preservation, and the ranking in (31).

(31) *Onset/r, BIPOSITION >> Max-Seg

* Problem for Sumner and SamG:
  o cases where the segment deletes without any trace of lengthening, cf. medial /r/-
deletion in SamG. Consider /lur/ → [lui]. The only structure that adheres to
  Sumner’s proposal is the following

(32) /l_{12}u_{23}r_{34}/ → [l_{12}u_{23}.r_{34}]

* This captures the lack of vowel lengthening after /r/-loss word-medially, but has the
implication that the high back vowel acts as an onset of the following syllable, i.e. effectively
something like lu.wi. The same problem arises even if take [l_{12}u_{23}.r_{34}] as the output, i.e lu.ji.
These are clearly wrong since the actual output is lui.

6 Concluding remarks

* All previous OT analyses fail either because they assume intermediate levels or because they
assume underlying moraification. Both assumptions are problematic.
* They also fail to account for CL in SamG because they do not allow CL after onset loss.
* Only Sumner (1999) seems more promising, but this permits other structures that do not
  present any CL, so it is inadequate.

General pattern for CL

* General pattern for CL: Markedness, Uniformity >> PosCorr >> Dep-μ; while Dep-μ >>
  PosCorr produces the lack of it.
* Several advantages:
  o it does not require intermediate levels
  o it makes no reference to underlying moraicity
  o it does not violate ROTB
  o it treats segments equally, thus it is anticipated that deletion of all segments (under
certain conditions, cf. Topintzi in prep.) can lead to CL
Thus it can naturally account for cases such as SamG CL after onset loss, or Piro CL after deletion of unsyllabified consonants (Topintzi in prep.)

References

Bermúdez-Otero, Ricardo (2001) Underlyingly nonmoraic coda consonants, Faithfulness and Sympathy. ms, University of Manchester


Hermans, Ben (2001) CL in OT. Linguistics in the Netherlands **18**: 139-150


Oostendorp, Marc van (2000) Back to the Base. On the Richness of the Base Hypothesis. ms, Meertens Institute, Amsterdam


Topintzi, Nina (in prep.) Onset, stress and weight interactions [provisional title]. Doctoral Dissertation, UCL
