Variation and optionality in serial interaction*

Minkyung Lee
(Daegu University)

Lee, Minkyung. 2011. Variation and optionality in serial interaction. *Studies in Phonetics, Phonology and Morphology* 17.3, 445-463. The architecture of Harmonic Serialism (HS), a derivational version of OT, hinges upon harmonic improvement of parallel OT and a serial derivation of rule-based grammar. Under gradualness, Gen is forced to produce a limited candidate set with a single modification at a time. Through the multiple passes of Gen and Eval, optimality in HS is always determined in a gradual, local, and consistent mode. Given this HS tenet, (dialectal) variation and (intra-speaker) optionality reflect HS's local optimality. Both variation and optionality cause multiple optima on the surface but they adopt different harmonic improvement paths. Optimality in variation is entirely local and harmony ascent is gradually achieved via an invariant hierarchy but a different grammar from dialect to dialect. On the other hand, multiple optima from optionality are half local with a gradual harmony ascent and the other half global with ultimate harmony at once. **(Daegu University)**

Keywords**:** serial OT, local optimality, gradualness, harmonic improvement, multiple optima, variation, optionality

1. Introduction

This paper focuses on the major issue how a gradual version of OT called Harmonic Serialism (henceforth HS, McCarthy 2008a, b, 2009) distinguishes and analyzes the difference of dialectal variation from intra-speaker optionality.[1] Serial OT critically differs from parallel OT in two major aspects: Gen and the Gen→Eval→Gen ... loop (McCarthy 2008a: 273). Unlike in parallel OT, Gen is entirely constrained to produce a limited candidate set in which each candidate makes a single change at a time from its predecessor. This is due to the HS's unique property of

* This research was partially supported by the Daegu University Research Grant, 2011. I would like to thank three anonymous reviewers for their valuable and insightful comments and suggestions. All remaining errors are solely my own.

[1] As argued in Lee (2001), pre-OT grammar with no specific or formal distinction to distinguish intra-language variation from intra-speaker optionality treats both in the same manner via rules and their relative orderings. Multiple outputs from dialectal variation result from rule reordering or the presence or absence of certain rules while those from intra-speaker optionality come from the presence or absence of optional rules. In this respect, variation is used as a cover term ranging from region, speech group, speech style, or tempo to individual speaker difference. However, with the emergence of the input-output correspondence theory (McCarthy and Prince 1995), OT-grammar can distinguish variation from optionality. Variation results from differences due to region, speech group, and the like while optionality comes from an individual speaker's free choice with or without speech style or tempo influence. In this paper, given the input-output correspondence schema, optionality is one to many mapping in a speaker while variation is one to one mapping from dialect to dialect.

gradualness. Furthermore, HS allows a step-wise derivation via the multiple passes of Gen and Eval loop, so the output chosen at each Eval is locally optimal, not the ultimate output though, and more harmonic than its predecessor due to another property of harmonic improvement. This serial derivation continues up to the point of convergence where the most recent output of Eval is merged into the latest input to Gen, i.e., there is no further harmonic improvement possible. At this moment, the whole derivation is complete and the latest output becomes the ultimate output.

As argued in previous literature (Halle 1962, Kiparsky 1968, Schane 1972, Vennemann 1972, Calabrese 1989, Lee 2001, and others), dialectal variation is attributed to a dialect-specific rule ordering. Even though a single input corresponds to multiple outputs, with a closer look, each dialect has its own single output. In a serial OT perspective, dialectal variation arises due to the fact that related dialects have their own specific rankings though they share the same constraints. In each dialect, each pass through Gen and Eval is always determined by the same grammar, thus harmonic improvement is always consistent unlike in stratal OT (McCarthy 2008b:502).[2] In addition, as argued in previous literature (Zwicky 1972, Bolozky 1977, Hasegawa 1979, Kim-Renaud 1987, Lee 2001, and others), intra-speaker optionality is speech-style or tempo-driven. As argued in Bolozky (1977: 218), fast speech processes are optional, thus both slow and fast speech forms are acceptable as possible outputs.[3]

To examine the difference of (dialectal) variation from (intra-speaker) optionality in a gradual version of OT and provide an HS account as well, this paper targets the data of northern Greek where four related dialects show variation and the data of Modern Hebrew involving optionality. Section 2 introduces and discusses the basic architecture of serial OT in which gradualness and harmonic improvement are mainly highlighted. Section 3 observes and discusses the data of northern Greek for variation and Modern Hebrew for optionality. For the former, four related rules and their different orderings cause a dialectal difference. For the latter, the presence or absence of optionality incurs multiple optima on the surface. Section 4 provides an HS account. It will be shown that variation and optionality result from local optimality with a gradual harmony ascent. For the former, harmony ascent is fulfilled by an invariant hierarchy on the course of derivation but the grammar each dialect adopts is different, thus

[2] As McCarthy (2008b: 502) criticizes, unlike serial OT, stratal OT (Rubach 1997, Kiparsky 2000 and Ito and Mester 2003 and many others) adopts different grammars for different strata. In fact, what harmonic improvement regards differs from grammar to grammar, thus harmonic improvement in stratal OT is not consistent at all. However, serial OT uses the same grammar on each pass through Gen and Eval, thus harmonic improvement is always consistent. This is a critical difference of HS from stratal versions of OT.

[3] Note that his paper does not consider difference of speech style from speech tempo since, as Lee (2001: 22) pinpoints, casualness of speech varies from speaker to speaker and it is closely related to the rates of speech or complex sociolinguistic factors, thus it is not easy to distinguish a casual speech from a careful speech.

the ultimate output differs from dialect to dialect as well. For the latter, multiple optima come from the different harmonic improvement requirements, one locally and the other globally with the presence or absence of a gradual path to harmonic improvement. Section 5 summaries and concludes the paper.

2. Serial OT architecture

Serial OT (McCarthy 2008a, b, 2009, originally proposed by Prince and Smolensky 1993, 2004, but just briefly sketched) mainly differs in two respects from parallel OT. One is Gen and the other is the multiple passes of Gen and Eval loop. Under gradualness, Gen produces a limited candidate set in which each candidate makes a single difference at a time on each pass through Gen and Eval.[4] Therefore, unlike in parallel OT, HS does not permit any candidate with multiple changes in a single step. This guarantees a gradual harmonic improvement whereby each candidate is more harmonic than its predecessor.

Moreover, HS allows a step-wise derivation through the multiple passes of Gen and Eval loop. This implies that each pass of Gen and Eval has its own local optimum, though it is not the ultimate output (McCarthy 2008b: 503). Rather, it stands as a new input for the next pass of derivation in which the newly-born local optimum is more harmonic than its predecessor. This serial derivation continues until the latest output of Eval is identical to the most recent input to Gen, i.e., they are convergent. At this point, the latest output becomes the ultimate output and the whole derivation terminates since no further harmonic improvement is possible. Therefore, harmonic improvement in HS is entirely local. However, in parallel OT disallowing serialism, the most harmonic output is immediately created at once. This implies that harmonic improvement is always global.

The harmonic improvement tableau in (1) indicates how each candidate in a set gains more harmony than its predecessor.

---

[4] Given McCarthy (2008b: 501), the notion of 'a single change at a time' needs to be separately defined by basic vs. non-basic faithfulness constraint. For instance, voicing of [p] to [b] in <pap, pa.pi, pa.bi> as in (1) does not trespass the gradualness requirement since it violates just one basic faithfulness constraint Ident(voice) though it additionally violates the non-basic faithfulness constraint Ident-Onset(voice). This implies that a single step of an HS derivation may violate one or more non-basic faithfulness constraints at a time though never more than one basic faithfulness constraint such as Max(=No deletion), Dep (=No insertion) and Ident (=No change of a feature value). Here note that the direct mapping of pap→pa.bi is invalid in HS since it violates two different basic faithfulness constraints Dep and Ident in a single step, though it is entirely tolerant in parallel OT.

(1) Harmonic improvement tableau for <pap, pa.pi, pa.bi> (McCarthy 2008b: 503)

/pap/	NoCoda	*VC$_{voiceless}$V	Dep	Ident(voi)
a. pap *is less harmonic than*	*!			
b. pa.pi *is less harmonic than*		*!	*	
c. pa.bi			*	*

The tableau in (1) tells us that, given the constraint hierarchy, /pap/ in (1a) is mapping to the most harmonic candidate [pa.bi] in (1c), but not immediately from /pap/ to [pa.bi] at a time, instead gradually through the intermediate step of [pa.pi] in (1b). The crucial reason that the candidate in (1b) is more harmonic than (1a) is that the former successfully removes the violation mark that its predecessor in (1a) holds. Likewise, the candidate in (1c) is more harmonic than (1b) since the former eliminates the violation mark of *VC$_{voiceless}$V that its predecessor in (1b) possesses. Therefore, in the harmonic improvement tableau, the exclamation marker plays a key role as a signal that there will be more harmony if the exclamation marker of the preceding pass is successfully removed in the following step.[5]

In more details, given the step-wise derivation, the output of the /pap/'s first pass is [pa.pi] in (1b), which is more harmonic than its predecessor in (1a). Thus, [pa.pi] is locally chosen as optimal, but not the ultimate output. Rather, [pa.pi], as a new input, is submitted to the next pass. The output of [pa.pi]'s pass through Gen and Eval is [pa.bi] in (1c), which is more harmonic than its predecessor [pa.pi]. For the next derivation, the locally-born [pa.bi] is submitted to Gen, but it is yielded as optimal once again, i.e., convergence. With no further harmony ascent, [pa.bi] in (1c) becomes the ultimate output and the whole derivation is complete.

As observed in the tableau (1), in HS, harmonic improvement is always satisfied, first, locally since each pass through Gen and Eval gives birth to its own local optimum (but not the ultimate output until it is convergent to the latest input to Gen), second, gradually since each local output differs from its predecessor with a single change at a time, to put it another way, there is more harmony than its predecessor, and finally, consistently as well since an invariant constraint hierarchy on each pass of Gen and Eval always determines harmonic improvement.

[5] Unlike in the tableau (1) where each pass through Gen and Eval is consistently assessed by the same grammar, Kimper (2008), in line with Pater (2007), adopts the Partially Ordered Constraint (Anttila 1997, 2007) model to deal with three different optional ways of MiP (=Minor Phrase) parsing in Bengali. Given his HS analysis, each step posits a different grammar, thus harmonic improvement differs from pass to pass. See McCarthy (2008b) for the case where a single phonological process can be applied to multiple loci in a form. There is no need to resort to stochastic OT (Boersma 1997), floating constraint (Nagy and Reynolds 1997) or other similar multiple- rankings model.

Constraining Gen and allowing the multiple passes of derivation through Gen and Eval loop, serial OT well captures the characteristics of variation and optionality.[6] As will be discussed later in detail, given the major properties of gradualness and harmonic improvement in the HS architecture, each dialect adopts its own specific grammar whereby a local optimum is newly born on each pass of Gen and Eval up to the point of convergence. Thus, harmonic improvement in each dialect reflects HS's local optimality. For optionality, however, multiple optima come from different harmonic improvement routes, one from local optimality and the other from global optimality just like in parallel OT (Kimper 2008). For the former, harmonic improvement is gradually fulfilled but for the latter, it is global with no gradual path to harmony ascent, thus the ultimate output is immediately created at once.[7]

Taken together, unlike in parallel OT, HS, a serial version of OT, pursues local optimality based upon a gradual harmonic improvement via an invariant grammar on each pass of Gen and Eval. The ultimate output born at the convergent step of an HS derivation takes a gradual path to ultimate harmony, which is the unique asset of HS, not found in other OT grammars.

3. Data of variation and optionality

Focusing on the issue how surface optima of variation differ from those of optionality, this section examines and discusses the data of variation found in northern Greek (Chambers and Trudgill 1980: 47) and optionality from Modern Hebrew (Bolozky 1977: 219). Let us first look at the data of dialectal variation as laid out in (2).

(2) Dialectal variation of /ðikósmu/ ‘my own’

	Macedonia	Thessaly	Epirus	Euboea
UR SR	/ðikosmu/ [ɵkozim]	/ðikosmu/ [ɵkozum]	/ðikosmu/ [ɵkosim]	/ðikosmu/ [ɵkosum]

Given general dialectology (Chambers and Trudgill 1980: 45), a single

[6] As Lee (2001) argues, multiple optima, no matter where they come from, variation or optionality, arise due to two relevant constraints in conflict. Under parallelism, given the different mapping schema, variation is from re-ranking of two conflicting constraints while optionality results from their reversibility. Under serialism, however, variation and optionality are demarcated not by ranking strategy but by the presence or absence of a gradual path to harmonic improvement.

[7] As will be discussed later, in HS, McCarthy (2008b: 502) states that a singleton derivation in which an input-faithful candidate itself becomes the ultimate output at once is trivially gradual and harmonically improving. However, in this paper, along the lines of Pater (2007) and Kimper (2008), the derivation with no gradual path to harmony ascent is viewed as global as in parallel OT.

underlying representation is given to the related dialects as illustrated in (2). Even though a single input corresponds to multiple outputs on the surface, each dialect has its own single output since a rule ordering differs from dialect to dialect. The relevant rules are given in (3).

(3) Rules relevant to a dialectal variation (Chambers and Trudgill 1980: 47)[8]
Rule (A): High vowel loss
An unstressed high vowels /i/ and /u/ are banned.
Rule (B): Voicing assimilation
Voiceless stops are voiced before voiced stops while voiced ones become voiceless before voiceless stops.
Rule (C): Vowel epenthesis
When the second member of a word-final consonant cluster is nasal, an /i/ is inserted before the nasal.
Rule (D): Vowel rounding
/i/ becomes /u/ before a following labial consonant.

Rule (A) bars the appearance of an unstressed high vowel on the surface, no matter what it is front or back, if it is not stressed. Here, notice that, given the data, the stress falls on the penultimate position and that Rule (A) is prior applied to other rules.[9] Rule (B) requires the regressive voicing assimilation between a trigger and a target, thus two adjacent consonants share the same value in voicing. Rule (C) militates against any consonant cluster in coda where the second member is nasal. Thus, a high front vowel /i/ is inserted to split up a complex coda. Rule (D) demands the inserted high front vowel /i/ get rounded before a labial consonant.

Difference in rule ordering gives rise to a dialectal variation in northern Greek though each dialect shares the same rules displayed in (3). This is well organized in (4).

---

[8] Here note that, as Rule (B) in (3) indicates, voicing assimilation in northern Greek also arises before a voiced stop, i.e., a nasal stop, as well as a voiceless stop. Also, in Rule (D), as an alternative idea, vowel rounding may be affected by feature spreading of the preceding round vowel, not by the following labial consonant. However, putting this argument aside, this paper adopts the phonological rules set up by Brian Newton whereby Chambers and Trudgill (1980: 46) demonstrate the dialectal variation of northern Greek.

[9] Here note that metrical structure must be first assigned before the application of Rule (A), though Chambers and Trudgill (1980: 47) do not focus on this issue. In rule-based phonology, metrically-conditioned syncope (MCS) is analyzed by ordering metrical-structure assignment before deletion of unstressed vowels. In parallel OT (Prince and Smolensky 1993, 2004), however, metrical-structure assignment and syncope are both evaluated at the same time. In HS, as McCarthy (2008b: 500) emphasizes, metrical-structure assignment is always followed by MCS. That is, if no metrical structure, there are no metrically weak positions. Therefore, the ordering of metrical-structure assignment followed by MCS is intrinsic. For more detailed discussion on this, see McCarthy (2008b: 505-509).

(4) Dialectal variation via a different rule ordering

	Macedonia	Thessaly	Epirus	Euboea
UR	/ðikosmu/	/ðikosmu/	/ðikosmu/	/ðikosmu/
Rules	(A) ðkosm (B) θkozm (C) θkozim	(A) ðkosm (B) θkozm (C) θkozim (D) θkozum	(A) ðkosm (C) ðkosim (B) θkosim	(A) ðkosm (C) ðkosim (B) θkosim (D) θkosum
SR	[θkozim]	[θkozum]	[θkosim]	[θkosum]

In more details, given the Chambers and Trudgill's (1980: 47) data observations, for Macedonia, high vowel deletion rule (A) first applies to the input form, so the unstressed high vowels /i/ and /u/ are both deleted. Next, voicing assimilation rule (B) forces a target to assimilate to either a plus or minus value of a laryngeal feature that a trigger possesses. So the trigger regressively spreads its voicing value to the target. Finally, to avoid a consonant cluster in coda, vowel epenthesis rule (C) forces a high front vowel /i/ inserted before a nasal consonant. With the rule orderings of A, B, and C, the ultimate output [θkozim] occurs. For Thessaly, however, the rules given in (3) are all applied in that order. The difference of Thessaly from Macedonia is related to the presence or absence of rule (D) application. In Thessaly where vowel rounding rule (D) is additionally applied, the different output [θkozum] surfaces.

The dialects of Epirus and Euboea adopt the different rule orderings from Macedonia and Thessaly. A crucial difference between the two dialectal groups falls on the ordering of Rule (B) and Rule (C). For the latter group, Rule (B) precedes Rule (C) in their application order while for the former, their ordering is entirely opposite. Moreover, within each dialectal group, the presence or absence of Rule (D) application is the key criterion to demarcate Macedonia from Thessaly on the one hand and Epirus from Euboea on the other hand. As in Thessaly, Euboea also has the additional Rule (D) application. Epirus and Euboea have the same rule ordering but only Euboea goes one step further to Rule (D) where the inserted high front vowel turns into a round vowel.

Now let us move onto the data of optionality found in Modern Hebrew (Bolozky 1977: 219). Voicing assimilation optionally arises, which causes multiple optima on the surface as laid out in (5).

(5) Optional voicing assimilation in Modern Hebrew

yidfok	~	yitfok	he will knock
zkenim	~	skenim	old ones (pl.)
pzila	~	bzila	squinting
yiʃbor	~	yiʒbor	he will break
yifgoʃ	~	yivgoʃ	he will meet

As argued in previous literature (Bolozky 1977: 220, Kim-Renaud 1987: 343), speech tempo (fast vs. slow) plays a key role as a major factor to cause occurrence of multiple optima on the surface in a speaker. In fast speech, voicing assimilation regressively arises between obstruent consonants as the second column in (5) indicates. Therefore, a trigger shares its [voice] value with a target.

As such, given the data observations thus far, both variation and optionality involve multiple optima where a single input corresponds to multiple outputs on the surface. For variation, multiple optima are not all used in a single dialect. This implies that each dialect has its own single output, which differs from other related dialects. For optionality, however, multiple optima are freely mixed in a speaker.

4. An HS account on multiple optima

In this section, it will be highlighted that an HS account well captures the fact that variation and optionality reflect local optimality, an HS's inherent property.[10] For the former, each dialect adopts a single but different grammar from others whereby harmonic improvement is locally, gradually and consistently fulfilled through all iterations of Gen and Eval. For the latter, however, surface optima come from the different harmonic improvement paths, one is locally-born in a gradual mode and the other globally assessed at once.

4.1 Dialectal variation of northern Greek

To diagnose how serial OT deals with the dialectal difference of northern Greek as schematized in (4) and further provide an HS account, let us first posit some relevant constraints and discuss their interaction. As witnessed earlier, all dialects disfavor an unstressed high vowel, whatever it is front or back, due to the top-ranked markedness constraint $*\breve{v}_{[high]}$ in (6a) though it incurs the violation of Max-Seg in (6e).[11] Agree(voice) in (6b) militates

---

[10] Here I will clarify that the HS model this paper employs is not an alternative to parallel OT in that parallel OT can also treat the target data of this paper. As argued in McCarthy (2008b: 504), HS is a variant implementation of OT's basic ideas, just like parallel OT. However, serial OT, unlike parallel OT, stands on the entirely different architecture of gradualness. In the mapping of A→C, serial OT requires the intermediate stage of B prior to the step of C, but in parallel OT, A directly maps to C without B. In this respect, this paper makes an attempt to provide a new insight how HS distinguishes the difference of variation from optionality under gradualness.

[11] From the case study of Aguaruna (a Jivaroan language spoken in Peru and with left-to-right iambic feet) where MCS arises, McCarthy (2008b: 516) proposes a two-step derivation, stress assignment (with Ident[stress] violation) followed by syncope (with Max violation). As discussed earlier, in HS, violations of different basic faithfulness constraints require different derivational steps (McCarthy 2008b: 510). Here, the stressed vowels are boldfaced and the tableaux below are somewhat simplified from McCarthy's (2008b: 516) full-fledged tableaux.

against the disagreement of [voice] value between a trigger and a target, which sacrifices Ident(voice) in (6f). *Complex(Coda) in (6c) which disfavors a consonant cluster in coda requires vowel insertion, resulting in the violation of Dep-Seg in (6g). One step further, the inserted high front vowel turns into a round vowel due to the demand of Agree(round) in (6d), which compels the violation of Ident-OO(round) in (6h).

(6) Constraints related to the dialectal difference of northern Greek

a. *$\breve{V}_{[high]}$: Unstressed high vowels in the input are banned in the output.

b. Agree(voice) (=Agr(vo)) (Beckman 1999, Bakovic 2000, Bakovic and Wilson 2000): Consonant clusters agree in [voice].

c. *Complex(Coda) (=*Comp(Co)) (McCarthy 2007): Avoid a complex coda.

d. Agree(round) (=Agr(ro)): A [labial] trigger must share its [round] feature with its preceding high vowel.

e. Max-Seg (=Max) (McCarthy and Prince 1995): Realize all the segments from the input.

f. Ident(voice) (=Id(voi)) (McCarthy and Prince 1995): The corresponding segment must preserve the feature [voice].

g. Dep-Seg (=Dep) (McCarthy and Prince 1995): Output segments have input correspondents.

h. Ident-OO(round) (=Id-OO(ro))[12]: The corresponding segments

---

(1) Input: /itʃinakaŋumina/

Candidates for stress step	Non-Finality($^{'}\sigma_{light}$)	*V-PLACE$_{weak}$
→a.(itʃ**i**)(nak**a**)(ŋum**i**)na		4
b. (itʃ**i**)(nak**a**)(ŋum**i**)(n**a**)	1 W	3 L

(2) Result of stress step: [(itʃ**i**)(nak**a**)(ŋum**i**)na]

Candidates for syncope step	*V-PLACE$_{weak}$	Max
→a.(itʃ**i**n)(k**a**ŋ)(m**i**n)	1	3
b.(itʃ**i**)(nak**a**)(ŋum**i**)na	4 W	L

The tableau in (1) shows the result of each iterative stress assignment through Gen and Eval where one stress at a time is assigned as in <(itʃ**i**)nakaŋumina> and then another stress is assigned as in <(itʃ**i**)(nak**a**)ŋumina>… and so on (McCarthy 2008b: 520). For the syncope step in (2), given McCarthy (2008b: 516), three unstressed vowels are syncopated with the violations of a single basic faithfulness constraint Max. Note that McCarthy (2007: 61-61, 77-79, 2008a: 276) states that Gen can add violations of only one basic faithfulness constraint at a time, here Max. In addition, McCarthy (2008b: 507) assumes that vowel reduction (and deletion as well) involves loss of a vowel's place feature due to the demand of *V-PLACE$_{weak}$ compelling syncope of unstressed vowels. The choice between deletion and reduction is determined by the relative ranking of Ident(V-place) and Max. Here, for Aguaruna favoring deletion, Ident(V-place) is ranked over Max. Therefore, as verified in (2), gradualness is well preserved since there are no two different basic faithfulness constraints, Ident(V-place) and Max, violated at a time. For consonant deletion, however, McCarthy (2008a: 280) proposes a two-step process under the Max-Feature regime, that is, /t/→[H] (with Max[Place] violation) and [H]→ø (with Max violation) as in <pat.ka, paH.ka, pa.ka>.

[12] Note that the constraint posited here is indebted to Output-Output or OO faithfulness based on the output-to-output correspondence theory with respect to faithfulness relations (Benua 1997, Kager 1999, and Pater 2000). Given McCarthy (2008b: 503), faithfulness violation is calculated compared to the original underlying representation, not to the input of the latest

between outputs must share the same value for the feature [round].

Given the constraints postulated in (6), each dialect shares the same constraints but the hierarchy it prefers is different from each other. As discussed in (4), in all related dialects, the constraint in (6a) ranked on the top is satisfied though Max ranked at the bottom is at no cost though violated. The ranking difference between the constraints in (6b) and (6c) plays a vital role in dividing two dialectal groups, Macedonia and Thessaly versus Epirus and Euboea. Furthermore, the crucial role of the constraint in (6d) makes a difference within each dialectal group.

The constraint hierarchy for Macedonia is arranged in (7) and the harmonic improvement tableau is illustrated in (8).[13]

(7) The whole ranking of Macedonia
*v̆$_{[high]}$ >> Agr(vo) >> *Comp(Co) >> Id-OO(ro) >> Agr(ro), Dep

---

pass through Gen. However, for northern Greek where a high front vowel /i/ is inserted and rounded as well, the constraint Id-OO(ro) (or alternatively OO-Id(ro) (McCarthy 2007: 45)) evaluates its violation in comparison with the latest input though Gen. Note that OO faithfulness is originally proposed as a theory of opacity, but McCarthy (2007: 45) criticizes that it does not suffice to encompass opacity issue. Rather, he proposes OT with candidate chains and precedence constraint (OT-CC with Prec), which is beyond the scope of our major concern. This paper simply adopts the OO constraint and its basic idea for the purpose of this paper. Also note that, due to space limit, other faithfulness constraints except Id-OO(ro) and Dep are omitted from the tableaux.

[13] On the analogy of McCarthy's (2008b: 516) MCS analysis for Aguaruna (in footnote (11)), for deletion of unstressed vowels in northern Greek, this paper also supports a two-step derivation under gradualness, i.e., the step of stress assignment in (1a)(=(8a)) and the step of syncope in (2a)(=(8b)) as the rough tableaux below show. Note that creating a metrical structure causes *v̆[high] violated. In northern Greek, the foot form is assumed as trochee as in Modern Greek (Hayes 1995: 204) and the vowels stressed are boldfaced.

(1) Input: /ðikosmu/

Candidates for stress step	Foot=Trochee	*v̆[high]
→ a. ði(k**o**s)mu		2
b. (ð**i**)(k**o**s)(m**u**)	2 W	L

(2) Result of stress step: [ði(k**o**s)mu]

Candidates for syncope step	*v̆[high]	Max
→ a. (ðk**o**sm)		2
b. ði(k**o**s)mu	2 W	L

For ease of exposition and to focus the discussion on the main point of dialectal variation, this paper does not show the stress assignment process in the tableaux, but the tableaux given here show how stress assignment in (1a)(=(8a)) and syncope in (2a)(=(8b)) occur under gradualness. From the tableau in (2), the candidate in (2a) fares better in *v̆[high] in the sense that it eliminates two unstressed vowels at the expense of violating Max due to the deletions. After the step of syncope in (2a)(=(8b)), the following steps of an HS's derivation sequentially iterate up to the point of convergence.

(8) Harmonic improvement tableau for <ðikosmu, ðkosm, ɵkozm, ɵkozim>

/ð**i**kosm**u**/	*v̆[high]	Agr (vo)	*Comp (Co)	Id-OO(ro)	Agr (ro)	Dep
a. ðikosmu *is less harmonic than*	**!					
b. ðkosm *is less harmonic than*		**!	*			
c. ɵkozm *is less harmonic than*			*!			
d. ɵkozim					*	*

Here note that, to help in identifying the vowels that are unstressed and thus deleted, they are boldfaced in the underlying representation. From the harmonic improvement tableau in (8), how do we know the candidate in (8b) is more harmonic than (8a) and (8c) is more harmonic than (8b) and so forth? To be more harmonic, each local optimum should not violate the constraint that its predecessor violates (as well as the constraints ranked higher than the constraint that each predecessor violates). That is, removal of the violation mark that each predecessor has derives more harmony ascent in the following step. Therefore, (8b) is more harmonic than (8a) since it satisfies the top-ranked constraint that its predecessor in (8a) violates. Likewise, (8c) gains more harmony than its predecessor in (8b) since it successfully removes the violation mark of Agr(vo) that (8b) holds. (8d) fares better in harmony ascent than (8c) since it satisfies *Comp(Co) that its predecessor in (8c) violates.

Given the step-wise derivation, at the first pass of /ð**i**kosm**u**/, the output in (8b) is locally chosen as optimal but it is not the ultimate output since there is a room for more harmony ascent (McCarthy 2008a: 275). Therefore, as a new input, it is submitted to the next pass where, this time, (8c) is locally chosen as optimal, but still it is not the ultimate output, thus fed back into Gen. Now (8d) is locally-born as optimal and submitted to the next pass, but it is chosen as optimal once again, that is, convergence. The entire derivation terminates here and (8d) becomes the ultimate output.[14]

The dialectal difference of Thessaly from Macedonia is attributed to the crucial role of the constraint Agr(ro) in (6d) that requires the inserted high

[14] Given the hierarchy in (7), suppose that the derivation goes one step further from (8d) to [ɵkozum] with the inserted high vowel rounded. The latter never wins for Macedonia since it is less harmonic than (8d). Note that [ɵkozum] violates Id-OO(ro) that (8d) satisfies. Therefore, the pass of (8d) to [ɵkozum] is invalid in Macedonia, though it is legal in Thessaly, but by the different grammar from Macedonia. Also note that the inserted vowel in (8d), though unstressed, has nothing to do with the violation of *v̆[high] since it is present in the output, not in the input.

front vowel get rounded. The harmonic improvement tableau in (10) is based upon the whole ranking arranged in (9).

(9) The whole ranking of Thessaly
*V̆$_{[high]}$ >> Agr(vo) >> *Comp(Co) >> Agr(ro) >> Id-OO(ro), Dep

(10) Harmonic improvement tableau for <ðikosmu, ðkosm, ɵkozm, ɵkozim, ɵkozum>

/ð**i**kosm**u**/	*V̆$_{[high]}$	Agr (vo)	*Comp (Co)	Agr (ro)	Id-OO(ro)	Dep
a. ðikosmu *is less harmonic than*	**!					
b. ðkosm *is less harmonic than*		**!	*			
c. ɵkozm *is less harmonic than*			*!			
d. ɵkozim *is less harmonic than*				*!		*
e. ɵkozum					*	*

The tableau is self-explanatory. Each local optimum on each pass of Gen and Eval obtains more harmony than its predecessor with the successful removal of the constraint violation from the preceding pass. At the point of convergence, the whole derivation terminates and the latest output in (10e) becomes the ultimate. In Thessaly, harmonic improvement is locally pursued under gradualness and also consistent under an invariant but different grammar from Macedonia, resulting in a dialectal gap. Note that (10e) violates Id-OO(ro) requiring the faithful correspondence between output and output. The inserted high front vowel in (10d) gets rounded in the ultimate output in (10e) since, in Thessaly, the satisfaction of Agr(ro) is more imminent than that of the OO faithfulness constraint.

Now let us take a look at the dialect of Epirus. Here, note that *Comp(Co) is ranked over Agr(vo), which is opposite to Macedonia and Thessaly. Also, Id-OO(ro) sitting over Agr(ro), which is the crucial cue to make a dialectal difference from Euboea, plays a vital role as illustrated in the tableau (12). The whole ranking of Epirus is laid out in (11).

(11) The whole ranking of Epirus
*V̆$_{[high]}$ >> *Comp(Co) >> Agr(vo) >> Id-OO(ro) >> Agr(ro), Dep

(12) Harmonic improvement tableau for <ðikosmu, ðkosm, ðkosim, ɵkosim>

/ð**i**kosm**u**/	*V̆[high]	*Comp (Co)	Agr (vo)	Id-OO (ro)	Agr (ro)	Dep
a. ðikosmu *is less harmonic than*	**!					
b. ðkosm *is less harmonic than*		*!	**			
c. ðkosim *is less harmonic than*			*!		*	*
d. ɵkosim					*	*

Given the HS's major premises of gradualness and harmonic improvement, a single modification at a time is added on each pass of Gen and Eval. From the input in (12a) to the ultimate output in (12d), three intermediate steps are indispensable in HS, though the direct mapping of (12a) to (12d) at a time is always valid in parallel OT. Each pass of Gen and Eval improves harmony in a gradual, local and consistent manner.[15]

For Euboea, Agr(ro) outranks Id-OO(ro), which leads to the crucial difference from Epirus. The tableau in (14) favors the ultimate output in (14e) given the constraint hierarchy laid out in (13).

(13) The whole ranking of Euboea
*V̆[high] >> *Comp(Co) >> Agr(vo) >> Agr(ro) >> Id-OO(ro), Dep

(14) Harmonic improvement tableau for <ðikosmu, ðkosm, ðkosim, ɵkosim, ɵkosum>

/ð**i**kosm**u**/	*V̆[high]	*Comp (Co)	Agr (vo)	Agr (ro)	Id-OO (ro)	Dep
a. ðikosmu *is less harmonic than*	**!					
b. ðkosm *is less harmonic than*		*!	**			
c. ðkosim *is less harmonic than*			*!	*		*
d. ɵkosim *is less harmonic than*				*!		*
e. ɵkosum					*	*

[15] As discussed in footnote (14), with the same logic, the constraint hierarchy given in (11) for Epirus filters out the illegal pass of [ɵkosim] in (12d) to [ɵkosum] (which is the ultimate output for Euboea) since harmonic improvement is not gradually achieved. Note that (12d) is more harmonic than [ɵkosum] since the latter violates Id-OO(ro) that (12d) satisfies. Therefore, the surface-unattested [ɵkosum] cannot emerge as the ultimate output for Epirus at all.

The tableau in (14) tells us that the harmonic improvement requirement is gradually satisfied via removal of the constraint violation exclamation-marked as each pass of Gen and Eval iterates. At the point of convergence, the whole derivation is complete and the most recent output in (14e) becomes the ultimate.

As such, as in rule-based grammar, in HS, dialectal variation results from the fact that each dialect adopts its own specific grammar though they share the same constraints. Harmonic improvement in each dialect is consistently determined by an invariant hierarchy, which is different from other related dialects. Thus, a single but different ultimate output is created from dialect to dialect. Harmonic improvement in each dialect is also locally evaluated at each Eval where each local optimum arises and its harmony is gradually ascending over the course of derivation. Moreover, the difference between the two dialectal groups, Epirus and Euboea vs. Macedonia and Thessaly, comes from the relative ranking of *Comp(Co) and Agr(vo) while the difference within each dialectal group is due to the relative hierarchy of Agr(ro) and Id-OO(ro). In essence, each of multiple optima in dialectal variation results from local optimality with a gradual harmony ascent.

4.2 Intra-speaker optionality of Modern Hebrew

Optionality also reflects the spirit of local optimality that serial OT inherently entails. One big difference of optionality from variation is that surface optima come from the different harmonic improvement processes, one from local optimality with a gradual harmony ascent while the other from global optimality with no gradual path to harmonic improvement.

Given the data of Modern Hebrew observed in (5), here repeated but simplified in (15), optionality of voicing assimilation incurs multiple optima on the surface.

(15) Optionality of voicing assimilation in Modern Hebrew (Bolozky 1977: 219)

yidfok ~	yitfok	he will knock
zkenim ~	skenim	old ones (pl.)

In slow speech, the contrast of the feature [voice] in a cluster is well preserved since Ident(voice) in (16b) is ranked over Agree(voice) in (16a). In fast speech, however, it is neutralized with their opposite order. Meanwhile, a consonant cluster is tolerant in Modern Hebrew since *Complex in (16c) ranked at the bottom does not play a key role.[16]

[16] The data in (15) tell us that Max-seg and Dep-seg are ranked on the top from the fact that any segmental insertion or deletion is not witnessed to split up a consonant cluster in Modern Hebrew, though these constraints are not shown from the tableaux.

(16) Constraints relevant to the optional voicing assimilation
a. Agree(voice):
Consonant clusters agree in [voice].
b. Ident(voice):
Match the value of the feature [voice] between input and output.
c. *Complex: Avoid a consonant cluster.

Given the constraints laid out in (16), the presence or absence of optionality is attributed to the fact that the two constraints posited in (16a) and (16b) are in conflict. Thus, in fast speech, the presence of voicing assimilation leads to a gradual harmonic achievement via local optimality while, in slow speech, the absence of voicing assimilation induces global optimality with no gradual harmony ascent as organized in (17) in which two different grammars render multiple optima on the surface.

(17) Constraint hierarchy relevant to multiple optima
a. Local optimality
Agree(voice) >> Ident(voice) >> *Complex
b. Global optimality
Ident(voice) >> Agree(voice) >> *Complex

Given the different constraint hierarchy in (17), let us first consider the case where optimality in optionality is local as in (18).[17]

(18) Local optimality via the hierarchy in (17a)
Harmonic improvement tableau for <yidfok, yitfok>

/yidfok/	Agree(voice)	Ident(voice)	*Complex
a. yidfok *is less harmonic than*	*!		*
b. yitfok		*	*

At the first pass of derivation through Gen and Eval, (18b) is more harmonic than (18a) since it successfully removes the Agree(voice)

[17] As argued in McCarthy (2008b: 507, 508), in comparison with the tableau in (18), if the hierarchy in (17b) is adopted, the tableau below is illegal, thus ruled out in HS since harmonic improvement is not gradual.
Wrong harmonic improvement tableau for <yidfok, yitfok> via the hierarchy in (17b)

/yidfok/	Ident(voice)	Agree(voice)	*Complex
a. yidfok *is more harmonic than*		*	*
b. yitfok	*!		*

Under gradualness, (b) should be more harmonic than its predecessor in (a). Note that (b) must not violate the constraint ranked higher than Agree(voice) that its predecessor in (a) violates if it emerges as the ultimate output. Therefore, as vindicated in (18), the gradual path to harmonic improvement is guaranteed only from the hierarchy given in (17a).

violation mark that its predecessor in (18a) has. The local optimum in (18b) is fed back into Gen, but it is selected as optimal once again since there is no more harmony ascent. Therefore, the whole derivation stops here and (18b) becomes the ultimate output.

In slow speech, however, there is no gradual path to harmonic improvement as the tableau in (19) indicates.

(19) Global optimality via the hierarchy in (17b)

/yidfok/	Ident(voice)	Agree(voice)	*Complex
yidfok		*	*

Compared to the tableau in (18) where gradual harmony achievement is witnessed from the input to the ultimate output, in the tableau (19), the optimal candidate is immediately created at once as in parallel OT. No gradual path to harmony ascent is guaranteed here.[18]

Taken together, as evidenced in Modern Hebrew, optimality in optionality reflects local optimality with a gradual harmony ascent (as shown in (18)) and multiple optima resulting from optionality are the combination of a locally-born variant (as in (18)) and a globally-born variant (as in (19)).

5. Conclusion

Targeting the data of dialectal variation of northern Greek and intra-speaker optionality of Modern Hebrew, which involve multiple optima on the surface, this paper has focused on the issue how the serial OT model deals with variation and optionality and further how surface optima differ from each other under the HS's basic spirits of gradualness and harmonic improvement.

As evidenced in northern Greek, though each related dialect shares the same constraints, it adopts its own specific hierarchy, resulting in a dialectal difference. Thus, the ultimate output born at the convergent step through Gen and Eval is different from dialect to dialect. In this respect, though there appear multiple optima on the surface, each dialect has only one ultimate output according to its own preferred grammar. Note that multiple optima are not mixed in a single dialect. Also, as witnessed, in

[18] Throughout the paper, local optimality takes a gradual path to the ultimate output of the grammar (McCarthy 2008a: 272). However, global optimality, just like in parallel OT, takes an immediate path to the ultimate output at once. In fact, no crucial difference of HS from parallel OT is witnessed. Given McCarthy (2008b: 502), a longer derivation of <yidfok, yitfok> as exemplified in (18) is gradual and harmonically improving (cf. Riggle and Wilson (2005), Kimper (2008)) while a singleton derivation of <yidfok> as shown in (19) is trivially gradual and harmonically improving. However, as noted in footnote (7), this paper assumes that the latter globally pursues harmonic improvement as in parallel OT due to the lack of a gradual path to harmony ascent.

each dialect, optimality is always realized in a gradual, local and consistent pattern.

For the case of Modern Hebrew with optional voicing assimilation, it has been shown that surface optima result from different harmonic improvement routes. One is from local optimality whereby harmonic improvement is gradually fulfilled on each pass through Gen and Eval while the other comes from global optimality with no gradual harmony achievement. Therefore, surface optima in optionality are half local and the other half global, with a gradual harmony ascent and an immediate harmony improvement, respectively.

As such, serial OT restricting Gen to produce a limited candidate set with a single change at a time and allowing the multiple passes of derivation through Gen and Eval loop encompasses both variation and optionality which are well compatible with the HS's basic spirits of local optimality and gradualness in harmonic improvement.

REFERENCES

ANTTILA, ARTO. 1997. Deriving variation from grammar. *Variation, Change and Phonological Theory*. Amsterdam: John Benjamins.

__________. 2007. Variation and optionality. In Paul de Lacy (ed.). *The Cambridge Handbook of Phonology*, 519-536. Cambridge: Cambridge University Press.

BAKOVIC, ERIC. 2000. Nasal place neutralization in Spanish. *U. Penn Working Papers in Linguistics* 7.1, 1-13.

BAKOVIC, ERIC and COLIN WILSON. 2000. Transparency, strict locality, and targeted constraints. *West Coast Conference on Formal Linguistics* 19, 43-56.

BECKMAN, JILL. 1999. *Positional Faithfulness: An Optimality Theoretic Treatment of Phonological Asymmetries*. New York and London: Garland Press.

BENUA, LAURA. 1997. *Transderivational Identity: Phonological Relations between Words*. PhD Dissertation. University of Massachusetts, Amherst.

BOERSMA, PAUL. 1997. How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetics Sciences of the University of Amsterdam* 21, 43-58.

BOLOZKY, SHUMEL. 1977. Fast speech as a function of tempo in natural generative phonology. *Journal of Linguistics* 13, 153-368.

CALABRESE, ANDREA. 1989. Phonological variation. In Paola Beninca (ed.). *Dialect Variation and the Theory of Grammar: Proceedings of the GLOW Workshop in Venice*, 9-39. Dordrecht: Foris.

CHAMBERS, JOHN and PETER TRUDGILL. 1980. *Dialectology*. Cambridge: Cambridge University Press.

HALLE, MORRIS. 1962. Phonology in a generative grammar. *Word* 18, 54-72.

HASEGAWA, NOBUKO. 1979. Casual speech vs. fast speech. *Chicago Linguistic Society* 15, 126-137.

HAYES, BRUCE. 1995. *Metrical Stress Theory: Principles and Case Studies.* Chicago: The University of Chicago Press.

ITO, JUNKO and ARMIN MESTER. 2003. Lexical and postlexical phonology in optimality theory: Evidence from Japanese. *Linguistische Berichte* 11, 183-207.

KAGER, RENE. 1999. Surface opacity of metrical structure in optimality theory. In Ben Hermans and Marc van Oostendorp (eds.). *The Derivational Residue in Phonological Optimality Theory*, 207-245. Amsterdam: John Benjamins.

KIMPER, WENDELL. 2008. Local optimality and harmonic serialism. Ms. University of Massachusetts, Amherst.

KIM-RENAUD, YOUNG-KEY. 1987. Fast speech, casual speech, and restructuring. *Harvard Studies in Korean Linguistics* II, 341-359.

KIPARSKY, PAUL. 1968. Linguistic universals and linguistic change. In Emmon Bach and Robert Harms (eds.). *Universals in Linguistic Theory*, 171-204. New York: Holt, Rinehart and Winston.

__________. 2000. Opacity and cyclicity. *The Linguistic Review* 17, 351-367.

LEE, MINKYUNG. 2001. *Optionality and Variation in Optimality Theory: Focus on Korean Phonology*. PhD Dissertation. Indiana University, Bloomington.

MCCARTHY, JOHN. 2007. *Hidden Generalizations*: *Phonological Opacity in Optimality Theory*. London: Equinox Publishing.

__________. 2008a. The gradual path to cluster simplification. *Phonology* 25, 271-319.

__________. 2008b. The serial interaction of stress and syncope. *Natural Language and Linguistic Theory* 26, 499-546.

__________. 2009. Harmony in harmonic serialism. ROA 1009-0109.

MCCARTHY, JOHN and ALAN PRINCE. 1995. Faithfulness and reduplicative identity. *University of Massachusetts Occasional Papers in Linguistics* 18, 249-384.

NAGY, NAOMI and BILL REYNOLDS. 1997. Optimality theory and variable word-final deletion in Faetar. *Language Variation and Change* 9, 37-55.

PATER, JOE. 2000. Nonuniformity in English secondary stress: The role of ranked and lexically specific constraints. *Phonology* 17, 237-274.

__________. 2007. Local harmonic serialism. Handout: Revised excerpts for workshop presentation at CASTL, Tromso.

PRINCE, ALAN and PAUL SMOLENSKY. 1993. Optimality theory: Constraint interaction in generative grammar. Ms. Rutgers University and University of Colorado.

__________. 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden: Blackwell.

RIGGLE, JASON and COLIN WILSON. 2005. Local optionality. *Proceedings of the North East Linguistic Society* 35, 539-550. Amherst, MA: GLSA.

RUBACH, JERZY. 1997. Extrasyllabic consonants in Polish: Derivational optimality theory. In Iggy Roca (ed.). *Derivations and Constraints in Phonology*, 551-582. Oxford: Oxford University Press.

SCHANE, SANFORD. 1972. Natural rules in phonology. *Linguistic Change and Generative Theory*, 199-229. Bloomington: Indiana University Press.

VENNEMANN, THEO. 1972. Rule inversion. *Lingua* 29, 209-242.

ZWICKY, ARNOLD. 1972. On casual speech. *Chicago Linguistic Society* 8, 607-615.

Minkyung Lee
Department of English Education
Daegu University
15, Nari-ri, Jillyang-eup, Gyeongsan-si, Gyeongbuk
Tel: (053) 850-4122
e-mail: milee@daegu.ac.kr

received: May 9, 2011
accepted: December 15, 2011