Generating Interlanguage Syllabification in Optimality Theory*

Hyouk-Keun Kim
Department of Linguistics
Georgetown University
37th & O St. NW, Washington, DC 20057, USA
kimhk@epsilon3.georgetown.edu

Abstract

This paper proposes an Optimality Theory (Prince & Smolensky, 1993) [OT]-based generator of the Interlanguage [IL][1] syllabification of Korean speakers of English. Basically, I accept the ideas of 'cyclic CON-EVAL loop' and 'locally encoded finite candidate set' proposed by Hammond (1995, 1997b). However, in order to treat some features of Korean accented English such as vowel epenthesis, segment modification (stop voicing, devoicing, nasalization, etc.), and ambisyllabicity, I suppose that the input string of phonemes be linked to two templates of candidate grid sets; one for syllable position, and the other for (finite) possible output segments for each input segment. I will also use the ALIGN family of constraints to treat the Korean coda neutralization phenomena effectively.

1. Introduction

In a second language acquisition, one of the most important factors in learner errors is first language transfer (Kenworthy, 1987; Major, 1994). Many Korean speakers, for example, mispronounce *'pick me up'* as [pʰiŋmiʌp] (stop nasalization) and *'stop it'* as [s<u>ɯ</u>tʰa<u>b</u>it] (vowel epenthesis and voiceless stop voicing), since they tend to transfer Korean phonology to their English IL.

Ellison (1994), Tesar (1995), Eisner (1997), Hammond (1995, 1997b), etc. propose how to implement OT. However, their implementation is usually based on basic syllable structure constraints such as *PEAK/C, *MARGIN/V, PARSE, FILL, ONSET, NoCoda, *COMPLEX, etc. Accordingly, it is hard for such implementation to cover the syllabification of real language data properly, since most languages including IL of Korean learners of English are not simple enough to be governed by only such constraints.

In the next section, I will present the salient features of Korean phonology and IL phonology of Korean speakers of English [IL-K-E] in terms of OT. In Section 3, I will propose how to implement an OT-based generator of IL-K-E. In Sections 4, I will mention contribution and future work of this project. Section 5 is the conclusion.

---

* The generator in this paper is a subsystem of my Korean Accented English Pronunciation Simulator [KAEPS] system, which is part of the Ph.D. dissertation I am writing. The KAEPS system is implemented in PERL and deals with pronunciation not only of word-level but also of phrase-level English orthographic representations. The output of the KAEPS is three types of English pronunciations: 1) a phoneme-based English pronunciation, 2) a desirable allophone-based English pronunciation, and 3) one or some possible Korean accented English pronunciation(s). It is run on the Web and its URL is **http://epsilon3.georgetown.edu/~kimhk/cgi-bin/kaeps/.** I am grateful to Lisa Zsiga and an anonymous reader for their comments on an earlier version of this paper, and to Cathy Ball and Donna Lardiere for their help and encouragement which aided me in accomplishing this project, and also to Michael Hammond for allowing me to try to revise his parser (1995) in order to make it compatible with my KAEPS system. Of course, all errors or mistakes in this paper are mine.

1 Selinker (1969, 1972) proposed and elaborated the term "interlanguage" to explain the unique utterance of L2 learners. It is regarded as a separate linguistic system which results from a learner's attempts to produce a target language norm.

2. IL Phonology of Korean Speakers of English [IL-K-E]

Korean has three distinctive types of voiceless stops as phonemes: aspirated /p^h, t^h, k^h/, fortis /p*, t*, k*/, and lenis /p, t, k/. Voiced stops do not exist as phonemes but exist as allophones, because lenis stops become voiced between two (voiced) sonorant sounds. Thus, unlike English, aspiration is a phonemic feature and voicing in stops is an allophonic one in Korean. Many Koreans tend to accept English voiceless stops as Korean aspirated stops, so they pronounce *school* as [sɯk^hul]. In this project, the input of an English voiceless stop is regarded as an aspirated stop.

There are three salient features of Korean syllable structure. First, consonant clusters are not allowed. When speaking an English word with consonant clusters, many Koreans tend to insert a vowel as shown below:

(1) a. school [sɯ.k^hul]
b. mint [min.t^hɯ]

Let us see how such IL pronunciation is obtained in terms of OT. First of all, the following OT constraints are to be considered:

(2) *COMPLEX [*COM] (P&S)
No more than one C or V may associate to any syllable position node.

(3) MAX (McCarthy, 1995a)
Every element of S1 has a correspondent in S2. (no deletion of a segment)

(4) DEP (McCarthy, 1995a)
Every element of S2 has a correspondent in S1. (no insertion of a segment)

The constraint ranking seems to be {*COM, MAX} >> DEP as depicted in (5):

(5) *mint*

/mint^h/	*COM	MAX	DEP
min		*!	
mint^h	*!		
☞ min.t^hɯ			*

In English, *COM should be lower-ranked than MAX or DEP.

In IL-K-E, vowel epenthesis occurs not only in consonant clusters but also in syllable-final fricatives or affricates or even stops preceded by a diphthong (Ahn, 1991; H-B Park, 1992; Broselow & H-B Park, 1995; Tak, 1996) as shown below:

(6) a. *kiss* [k^hisɯ]
b. *push* [p^huʃi]
c. tight [t^hait^hɯ].

This is related to the second feature of Korean syllable structure that only seven consonants [p, t, k, m, n, ŋ, l] can occur in the coda position in Korean. Labial and velar stops are neutralized as a homorganic lenis stop (/p^h, p*, p/ → [p]; /k^h, k*, k/ → [k]), and all coronal obstruents as [t] (/t^h, t*, t, ʧ^h, ʧ*, ʧ, s, s*/ → [t]). To deal with this coda neutralization phenomenon, I propose the following Feature Alignment Constraints, revising Hong (1997):[2]

(7) a. Align-Left ([stiff vocal folds], σ)
→[A-L(svf, σ)]
b. Align-Left ([+continuant], σ)
→[A-L(cont, σ)]

These constraints force a segment with the corresponding feature in the syllable-initial position. Like Hong, I also adopt IDENT-IO [F] constraints, which also belong to the Correspondence Theory family.

---

2 Adopting McCarthy & Prince's (1993) General Alignment and Itô & Mester's (1994) concept of CodaCond, Hong (1997) proposes the following:

Align-Left ([larygeal], σ).

This is based on Lombardi's (1995a,b) proposal of laryngeal neutralization. That is, laryngeal features such as aspiration or voicing appears only in the syllable-initial position. However, the scope of Korean coda neutralization is not limited only to laryngeal neutralization, since alveolar fricatives and alveo-palatal affricates are also neutralized as a plain lenis alveolar stop /t/ as described above. Furthermore, a voiced stop can occur in an ambisyllabic coda position by means of Lenis Stop Voicing as shown in (19) later.

(8) IDENT-IO [F] (McCarthy, 1995a)
Correspondent elements in S1 and S2 have identical values for feature [F]. (no phonological affiliation)

(9) IDENT-IO constraints

a. IDENT-IO [stiff vocal folds]	ID[svf]
b. IDENT-IO [continuant]	ID[cont]
c. IDENT-IO [voiced]	ID[vd]
d. IDENT-IO [sonorant]	ID[son]
e. IDENT-IO [lateral]	ID[lat]
f. IDENT-IO [place]	ID[place]

In Korean, the constraints of (7) outrank those of (9) as depicted below:

(10) *os* 'clothes' in Korean

/os/	A-L(cont, σ)	DEP	ID[cont]
☞ ot			*
os	*!		
osɯ		*!	

Since [os] and [osɯ] violate higher-ranked A-L (cont, σ) and DEP, respectively, [ot] is judged optimal even if it violates ID[cont]. However, as shown in (6), English *kiss* is not pronounced as [kit] but [kisɯ], which means ID[cont] should outrank DEP in IL-K-E. On the other hand, *top* is usually pronounced as [tʰap] not as [tʰapʰɯ], which means DEP outranks ID[svf] as shown in (11). That is, the modification of [continuant] feature is severer than that of [svf] feature.

(11) *kiss* and *top* in IL-K-E

/kʰis/	A-L (cont,σ)	A-L (svf)	ID [cont]	DEP	ID [svf]
kʰit			*!		
kʰis	*!				
☞ kʰisɯ				*	
/tʰapʰ/					
☞ tʰap					*
tʰapʰ		*!			
tʰapʰɯ				*!	

The third feature of Korean syllable structure is that increasing sonority across the syllable boundaries is disfavored. An obstruent before a nasal, for example, cannot be preserved due to the increasing sonority, but changes into a homorganic nasal with the same sonority as shown below:

(12) Obstruent nasalization
a) *os.man* /os.man/ → *on.man* [on.man] 'clothes only'
b) *aph.ni* /apʰ.ni/ → *am.ni* [am.ni] 'front tooth'

To deal with this, the following Syllable Contact Constraint is proposed:

(13) Syllable Contact Constraint [SCC][3]
Avoid rising sonority across the syllable boundaries.

The selection of the output form in (12-a) can be depicted as follows:

(14) *os.man* 'clothes only' in Korean

/os.man/	SCC	A-L (cont, σ)	ID [cont]	ID [pl]	ID [lat]	ID [son]	ID [vd]
a) os.man	*!	*					
b) ot.man	*!		*				
☞ c) on.man			*			*	*
d) ol.man			*		*	*	*!
e) om.man			*	*		*	*!

Candidates (a, b) are eliminated due to the violation of higher-ranked SCC. Candidates (c, d, e) satisfy both SCC and A-L(cont, σ), and violate ID[cont], ID[son] and ID[vd]. However, candidates (d, e) violate ID[lat] and ID[place] respectively, which candidate (c) does not violate. So candidate (c) is selected.

When Koreans transfer this obstruent nasalization phenomenon to English, *pick me up* is pronounced as [pʰiŋ.mi.ʌp] and *big mouse* as [piŋ.ma.u.sɯ]. Resyllabification is also related to

[3] This corresponds to Murray & Vennemann (1983) and Vennemann's (1988) Syllable Contact Law, which was based on Hooper (1976). Davis & Shin (1997) propose such a constraint and Hong (1997) adopts it.

SCC.[4] If a nominal particle, '-*i*', is attached to *os* (10), the neutralization does not occur, since /s/ is resyllabified as the onset of the following syllable as in *os-i* [o.si].[5]

(15) *os-i* 'clothes-NOM' in Korean

/os/	SCC	A-L (cont, σ)	ID[cont]
a) os.i	*!	*	
b) ot.i	*!		*
☞c) o.si			
d) o.ti			*!

To satisfy a higher-ranked constraint SCC, the *C* in *VCV* sequences in a prosodic word must be syllabified as an onset of the second vowel as in (c) and (d). Between the two candidates, (c) is judged optimal since it satisfies the ID[cont] constraint, too, while (d) violates it.

However, a compound word *wus.os* /us.os/ 'upper garment' is not pronounced as [u.sot] but as [udot]. This indicates that the coda /s/ belonging to the first word *wus* is neutralized as /t/ and this lenis stop becomes voiced between two vowels. To avoid misjudging [u.sot] as optimal, Align-Left (7) is clarified as (16) and another Align constraint such as (17) is proposed, following Hong (1997).

(16) a. Crisp-Align-Left ([stiff vocal folds], σ)
→ CA-L(svf, σ)
b. Crisp-Align-Left([+continuant], σ)
→ CA-L(cont, σ)

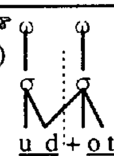
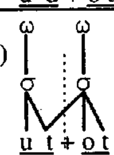
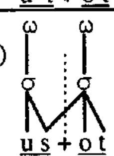
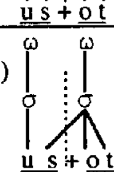
(17) Non-Crisp-Align-Right (Root$_0^{max}$, PrWd)
→ NCA-R(Rt, PW)

Crisp Alignment does not allow ambisyllabicity, while Non-Crisp Alignment allows it (Itô & Mester, 1994). Accordingly, (16) does not allow aspirated stops, fortis stops, fricatives, or affricates to occur in the coda position whether they are ambisyllabic or not. On the other hand, (17) allows the last element of the root word to become ambisyllabic, but does not allow it to be disconnected from the original word.

Proposing a Voice constraint (18) to deal with Korean Lenis Stop Voicing phenomenon, let us consider how to syllabify *wus.os*.

(18) VOICE [VCE]
Stops with a [−stiff vocal folds] (i.e., non-aspirated or non-fortis) feature are realized as voiced between two sonorant sounds within an accentual phrase, and as voiceless elsewhere.

(19) *wus.os* 'upper garment' in Korean

/us+os/	SCC	NCA-R (Rt, PW)	CA-L (cont, σ)	VCE	ID [cont]	ID [vd]
☞ a) 					**	*
b) 				*!	**	
c) 			*!		*	
d) 		*!			*	
e) u . d ot		*!			**	*
f) ud . ot	*!				**	*
g) us . ot	*!		*			

Candidate (a) is judged optimal since it violates only the lower-ranked ID[cont] and ID[vd], which are compelled to satisfy a higher-ranked CA-L (cont, σ) constraint. Candidate (b) is eliminated due to the violation of a higher-ranked VCE. Candidate (c) violates CA-L(cont, σ) since /s/

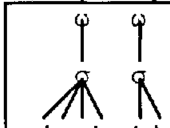
[4] There are other phonological phenomena related to the SCC such as lateralization, delateralization, /n/-insertion, /l/-insertion, etc. However, I will not deal with them in this paper.

[5] In fact, /s/ becomes palatalized before a high front vocoid. But I skip this phenomenon in this paper. The other issue is that ONSET can also play a role for triggering resyllabification. However, SCC covers the role of ONSET, i.e., ONSET can be regarded as a subset of SCC.

should not be a coda anyway, and is eliminated. Candidates (d, e) are eliminated due to the violation of higher-ranked NCA-R(Rt, PW). The word *wus* is a root and also a prosodic word by itself, so its final element /s/ should not belong to another word. Candidates (f, g) are eliminated, since they violate another highly-ranked SCC.

The transfer of the coda neutralization and lenis stop voicing phenomena to English may result in the pronunciation of *stop it* as [sɯtʰabit]. The following tableau shows how it works (Note: ambisyllabic C is represented as "·C·"):

(20) *pick up* in IL-K-E

 s t^h a p^h + i t^h	*COM	SCC	NCA-R (Rt, PW)	CA-L (svf, σ)	VCE	ID [vd]	DEP	ID [svf]
a. st^hap^h.it^h	*!	*		**				
b. sɯ.t^hap^h.it		*!		*			*	*
c. sɯ.t^hap.it		*!			*		*	**
d. sɯ.t^hab.it		*!				*	*	**
e. sɯ.t^ha.p^hit			*!				*	*
f. sɯ.t^ha.p it			*!		*		*	**
g. sɯ.t^ha.b it			*!			*	*	**
h.sɯ.t^ha·p^h·it				*!			*	*
i. sɯ.t^ha·p·it					*!		*	**
☞ j. sɯ.t^ha·b·it						*	*	**

Candidate (a) is eliminated due to the violation of a higher-ranked *COM. All other candidates violate a lower-ranked DEP to satisfy *COM. Candidates (b, c, d) and (e, f, g) are cast out due to the violation of higher-ranked SCC and NCA-R(Rt, PW), respectively. Candidates (h, i) are eliminated due to the violation of CA-L(svf, σ) and VCE, respectively. Candidate (j) is selected even if it violates ID[vd], DEP and ID[svf], which are lower-ranked. The reason why ID[vd] is considered to outrank DEP is due to the observation that Koreans tend to insert a vowel after a voiced stop even preceded by a lax vowel. That is, *sad* may be pronounced as [sɛdɯ] rather than as [sɛt], where the former violates DEP but the latter violates ID[vd].[6]

To sum up, the constraint ranking in IL-K-E considered up to now is as follows:

(21) Constraint Ranking in IL-K-E
{ *COM (2), MAX (3), SCC (13), CA-L(svf/cont, σ) (16), NCA-R(Rt, PW) (17) }
>> ID[cont, vd, son, lat, place] (9b,c,d,e,f)
>> DEP (4)
>> ID[svf] (9a)

3. Syllable Generation of Korean Accented English in OT

3.1 Problems

According to Hammond (1997b), the greatest problem in OT-based implementation is the possibility of the infinite candidate set when epenthesis (violation of DEP) or deletion (violation of MAX) are allowed, since **Gen** can produce infinitely any candidates. Even if there is no epenthesis or deletion, assuming that any segment can be syllabified as an onset, peak, coda or unparsed element, a word with *n* elements may have 4^n possible syllabifications, which is an exponential problem. In addition, each candidate has to be tested by each constraint. That is, the combination of the number of candidates times the number of constraints must be considered, which is an arithmetic but still nontrivial problem.

To solve these problems, he proposes: 1) implementation of syllabification is made by a form of a parser, which does not need to consider epenthesis nor deletion; 2) syllabification is encoded locally; and 3) a cyclic CON-EVAL loop is applied constraint by constraint.

The problems of implementation of IL-K-E in OT are more complicated than those raised by Hammond, since ambisyllabicity, epenthesis, and segment modification should be considered. The

[6] However, some words like *good* usually do not allow epenthesis, even if the final segment is a voiced stop. More experimental research is required on this issue, and I will skip this in this paper.

system dealing with such syllabification cannot be a parser but a generator.[7]

3.2 Korean accented English Generator

I assume that the initial candidate set produced by **Gen** can be predictable and finite, following the previous researches (Ellison, 1994; Tesar, 1995; Eisner, 1997; Hammond, 1995, 1997b). I adopt the concept of local encoding (Hammond, 1995, 1997b) developed from the concept of finite state automata (Ellison, 1994) and that of dynamic programming (Tesar, 1995). Unlike Hammond, however, since the role of this generator is not only syllabifying the input segments but also modifying them into suitable output segments, I suppose there are two templates of candidate grids: one representing syllable positions, and the other representing potential segment output forms.

Supposing the input is a phrase like *stop it*, whose string of phonemes is /s tʰ a pʰ # i tʰ/, the grids look like below:

(22) Grid A (for syllable position)

s	tʰ	a	pʰ	#	i	tʰ
o	o	o	o	o	o	o
n	n	n	n	n	n	n
nn	nn	nn	nn	nn	nn	nn
c	c	c	c	c	c	c
co	co	co	co	co	co	co
o n	o n	o n	o n	o n	o n	o n
u	u	u	u	u	u	u

(23) Grid B (for output segments[8])

s	tʰ	a	pʰ	#	i	tʰ
s	tʰ	a	pʰ	#	i	tʰ
	t		p			t
	d		b			d
	n		m			n
s ɯ	tʰ ɯ		pʰ ɯ			tʰ ɯ

As shown in (22), each segment can be an onset, a nucleus, two nuclei,[9] a coda, an ambisyllabic coda-onset, an onset + a nucleus (due to epenthesis), or an unparsed segment. Deletion is not considered in the current project. Constraints (cyclic CON-EVAL) prune away disfavored candidates cyclically. If there is only one member left in the candidate set, it should not be pruned away by any constraint. Grid A (22) is treated first, and then Grid B (23) is treated.

There are constraints such as NUC requiring a nucleus in a syllable, and *MARGIN/V saying a vowel cannot be an onset or a coda, and *PEAK/C saying a consonant cannot be a nucleus. So, if a segment is a vowel, all candidates but 'nn' (for a diphthong) or 'n' (for another vowel) are removed, and if a consonant, 'n' and 'nn' is removed as shown in (24) (the removable candidate is italicized and underlined):

(24) NUC, *MARGIN/V, *PEAK/C

s	tʰ	a	pʰ	#	i	tʰ
o	o	*o*	o	o	*o*	o
n	*n*	n	*n*	n	n	*n*
nn	*nn*	*nn*	*nn*	nn	*nn*	*nn*
c	c	*c*	c	c	*c*	c
co	co	*co*	co	co	*co*	co
o n	o n	*o n*	o n	o n	*o n*	o n
u	u	*u*	u	u	*u*	u

7 Hammond differentiates, "The generator would start with an input form, generate candidate syllabifications, and apply constraints to produce a syllabified output. A parser would start with an unsyllabified output, generate candidates, and produce a syllabified output" (p.6).

8 In producing candidates for segments, only those which may become optimal forms in some situations are considered. The number of candidates for segments depends on the types of segments. For example, a voiceless stop has five candidates: aspirated one, unaspirated one, voiced one, nasal one, and epenthesized one, while a nasal has only one candidate, itself.

9 Many Koreans tend to regard an English dipthong like /aɪ/ as two distinctive vowels like /a i/.

I adopt Hammond's idea of housekeeping, too, and propose the following cases:

(25) Housekeeping
- a. word-initial coda and coda-onset
- b. word-final onset
- c. phrase-final coda-onset
- d. word-final coda before another word starting with a vowel
- e. no parsing of word boundary

A word cannot start with 'c' or 'co' (a), and cannot end with 'o' (b). 'co' can occur in the word final position, but not in the phrase final position (c). A 'c' in a word-final position is deleted if it follows by '#' and 'n' (due to the SCC(13)) (d). A word boundary '#' has only 'u' (e).

(26) Housekeeping

s	t^h	a	p^h	#	i	t^h
o	o		$\underline{o}^{b}$	$\underline{o}^{e}$		$\underline{o}^{b}$
		n		$\underline{n}^{e}$	n	
$\underline{c}^{a}$	c		$\underline{c}^{d}$	$\underline{c}^{e}$		c
$\underline{co}^{a}$	co		co	$\underline{co}^{e}$		$\underline{co}^{c}$
o n	o n		o n	$\underline{o\ n}^{e}$		o n
u	u		u	u		u

Every segment is considered to be parsed, i.e., no deletion is considered in the current system. So 'u' is deleted in each set.

(27) PARSE

s	t^h	a	p^h	#	i	t^h
o	o					
		n			n	
	c					c
	co		co			
o n	o n		o n			o n
$\underline{u}$	$\underline{u}$	$\underline{u}$	$\underline{u}$	u	$\underline{u}$	$\underline{u}$

A consonant immediately before a vowel should be an onset, and has only 'o'.

(28) ONSET

s	t^h	a	p^h	#	i	t^h
o	o					
		n			n	
	$\underline{c}$					c
	$\underline{co}$		co			
o n	$\underline{o\ n}$		o n			o n
				u		

It is better to delete the portion of the word boundary symbol, before *COM starts to work. *COM does not allow a sequence of 'o + o', or 'c + c'. So delete 'o' or 'co' preceded by 'o', and 'c' or 'co' followed by 'c'.[10]

(29) *COM

s	t^h	a	p^h	i	t^h
$\underline{o}$	o				
		n		n	
					c
			co		
o n			o n		o n

Since a vowel is usually epenthesized after a fricative or an affricate in the coda position, 'c' or 'co' under such a segment should be deleted by means of CA-L(cont, σ). For the current example, however, this application is vacuous.

(30) CA-L (cont, σ)
the same as (29)

Now, DEP plays a role of pruning 'o n' candidate in a set containing more than one element.

[10] Not only is an English diphthong regarded as two vowels (cf. footnote 9), but also an obstruent followed by it tends to be epenthesized (cf. 6-c), so that 'c' and 'co' in this position may be deleted. I think this phenomenon is also related to *COM. For the current example, however, this is not applicable.

(31) DEP

s	t^h	a	p^h	i	t^h
	o				
		n		n	
					c
			co		
o n			o n		o n

At last, the optimal set of Grid A candidates is determined, which is dispatched to Grid B (23). Before applying constraints, one-to-one matching occurs. That is, if the syllable position needs two segments, i.e., if it is 'o n', candidates such as 't^h ɯ' with epenthesized vowel will be selected; if it is not 'o n', candidates such as 't^h ɯ' will be removed.

(32) One-to-one matching

s	t^h	a	p^h	i	t^h
o n	o	n	co	n	c
s	t^h	a	p^h	i	t^h
	t		p		t
	d		b		d
	n		m		n
s ɯ	t^hɯ		p^hɯ		t^hɯ

SCC checks the sequence of 'c' and 'o', and compare the sonority degree of the segments, usually for nasalization. Here, it is not applicable. CA-L(svf, σ) deletes an aspirated stop under 'c' or 'co' as shown below:

(33) SCC: vacuous, CA-L(svf, σ)

s	t^h	a	p^h	i	t^h
o n	o	n	co	n	c
	t^h	a	p^h	i	t^h
	t		p		t
	d		b		d
	n		m		n
s ɯ					

The conditions of voicing or devoicing will be checked. The voiceness of each segment is checked by using feature geometry.

(34) VOICE

s	t^h	a	p^h	i	t^h
o n	o	n	co	n	c
	t^h	a		i	
	t		p		t
	d		b		d
	n		m		n
s ɯ					

IDENT[F] constraints remove candidates which has a different [F] feature from that of input as shown in (33) and (34):

(35) IDENT[nas]

s	t^h	a	p^h	i	t^h
o n	o	n	co	n	c
	t^h	a		i	
					t
	d		b		
	n		m		n
s ɯ					

(36) IDENT[vd]

s	t^h	a	p^h	i	t^h
o n	o	n	co	n	c
	t^h	a		i	
					t
	d		b		
s ɯ					

Finally, the generation is done as follows:

(37) σ σ σ

s ɯ t^h a b i t [sɯ.t^ha·b·it].

Let us examine another example *big mouse* which is related to obstruent nasalization. I'll skip the syllabification of the syllable position grid.

(38) One-to-one matching

b	**i**	**g**	**m**	**ou**	**s**
o	**n**	**c**	**o**	**nn**	**o n**
b	i	g	m	ou	*s*
k		k			
m		ŋ			
b ɯ		*g ɯ*			s ɯ

(39) SCC

b	**i**	**g**	**m**	**ou**	**s**
o	**n**	**c**	**o**	**nn**	**o n**
b	i	*g*	m	ou	
k		*k*			
m		ŋ			
					s ɯ

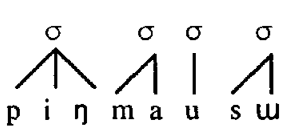
(40) VOICE

b	**i**	**g**	**m**	**ou**	**s**
o	**n**	**c**	**o**	**nn**	**o n**
b	i		m	ou	
p					
m		ŋ			
					s ɯ

(41) IDENT[nas]/[vd]/[svf]

b	**i**	**g**	**m**	**ou**	**s**
o	**n**	**c**	**o**	**nn**	**o n**
	i		m	ou	
p					
m		ŋ			
					s ɯ

The result is as follow:

(42) 

[piŋ.ma.u.sɯ]

3.4 Contribution and Future Work

The current work is significant in that it analyzes and implements the generation of the syllabification of an IL, IL-K-E, in OT. It tries to generate not only syllable positions but also modified output segments.

There are some Korean phonological phenomena, which are not considered at the current system. They are palatalization, /l/~/r/ alternation, etc. Next step is dealing with these phenomena.

IL among L2 learners must be different according to the learners, and it is not always the same even in the same person. The current system produces only one type of output based on the transfer of some Korean phonology. Further efforts will be made to generate several possible IL pronunciations according to the different levels of proficiency.

4. Conclusion

This paper deals with an OT-based generator for Interlanguage phonology of Korean speakers of English. Hammond (1997b) insists that OT-based syllabification be made as a parser which does not need to consider epenthesis or deletion. However, in order to syllabify Korean accented English, not only epenthesis but also segment modification such as stop voicing/devoicing or stop nasalization should be considered.

I adopted the basic ideas proposed by Hammond and others: finite candidate set, cyclic CON-EVAL and local encoding. I proposed two templates of candidate sets: one representing syllable positions (onset, nucleus, two nuclei, coda, ambisyllabic coda-onset, onset + epenthesized vowel, and unparsed position), the other representing potential output forms for each segment. Under the basic syllable constraints, the optimal candidates in the syllable position grid are selected. The result is used to generate the potential output segments which can be revised from the original input segments under the constraints basically applicable to Korean phonology.

References

Ahn, Sang-Cheol. 1991. Vowel Deletion and Epenthesis: the Vowel ɨ. *Studies in the Linguistic Sciences* 21, 2: 1-18.

Bird, Steven. 1994. Introduction to Computational Phonology. *Computational Linguistics* 20, 3: 1-8.

Bird, Steven. 1995. *Computational Phonology: A constraint-based approach.* Cambridge: Cambridge University Press.

Broselow, Ellen & Hye-Bae Park. 1995. Mora Conservation in Second Language Prosody. In John Archibald, ed., *Phonological Acquisition and Phonological Theory*, 151-168. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Davis, Stuart and Seung-Hoon Shin. 1997. Is there a syllable contact constraint? Poster presented at Hopkins Optimality Theory Workshop/University of Maryland Mayfest 1997, Maryland: Baltimore.

Ellison, T. Mark. 1994. *Phonological Derivation in Optimality Theory.* Ms., University of Edinburgh. [ROA 75]

Eisner, Jason. 1997. *Efficient Generation in Primitive Optimality Theory.* Ms., University of Pennsylvania. [ROA 206]

Hooper, Joan. B. 1976. *An Introductions to Natural Generative Phonology.* New York, NY: Academic Press.

Hammond, Michael. 1995. *Syllable parsing in English and French.* Ms., University of Arizona. [ROA-58].

Hammond, Michael. 1997a. Vowel quantity and syllabification in English. *Language, Vol. 73, No. 1*, 1-17.

Hammond, Michael. 1997b. *Parsing in OT.* Ms., University of Arizona. [cmp-lg/9710004].

Harris, John. 1994. *English Sound Structure.* Cambridge, MA: Blackwell Publishers.

Hong, Soonhyun. 1997. *Prosodic Domains and Ambisyllabicity in Optimality Theory.* Ph.D. dissertation. University of Pennsylvania. [ROA-209].

Itô, Junko and Armin Mester. 1994. Reflections on CodaCond and alignment. *Phonology at Santa Cruz, Vol. 3*, 27-46. [ROA-141]

Itô, Junko and Armin Mester and Jaye Padgett 1995.NC: licensing and underspecification in optimality theory. *Linguistic Inquiry, Vol. 26, No., 4*, 571-613. [ROA-38].

Kenstowicz, Michael. 1994. Syllabification in Chukchee: a Constraints-Based Analysis. *FLSM, Vol. 4*, 1-24. [ROA30].Kenworthy, Joanne. 1987. *Teaching English Pronunciation.* NY: Longman Inc.

Lombardi, Linda. 1995a. Laryngeal Neutralization and Syllable Well-formedness. *Natural Language and Linguistic Theory* 13: 39-74.

Lombardi, Linda. 1995b. Laryngeal Features and Privativity. *The Linguistic Review* 12: 35-59.

Major, Roy C. 1987. A Model for Interlanguage Phonology. In Georgette Ioup & Steven Weinberger, eds., *Interlanguage Phonology: the Acquisition of a Second Language Sound System*, 101-124. Cambridge, MA: Newbury House Publishers.

Major, Roy C. 1994. Current Trends in Interlanguage Phonology. In Mehmet Yavaş, ed., *First and Second Language Phonology*, 181-204. San Diego, CA: Singular Publishing Group, Inc.

McCarthy, John and Alan Prince. 1993. Generalized Alignment. In G. Booij and J. van Marie, eds., *Yearbook of Morphology.* Dordrecht: Kluwer, 79-153.

Mester, Armin and Jaye Padgett. 1994. *Directional syllabification in generalized alignment.* Ms. University of California, Santa Cruz. [ROA-1]

Murray, R. W. & Theo Vennemann. 1983. Sound change and syllable structure in Germanic phonology. *Language* 59: 514-528.

Park, Hye Bae. 1992. *External Evidence for Representations, Rules, and Constraints in Korean and Japanese.* Ph.D. dissertation. State University of New York at Stony Brook.

Prince, Alan and Paul Smolensky. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar.* Ms., Rutgers University and University of Colorado at Boulder.

Selinker, Larry. 1969. Language Transfer. *General Linguistics* 9: 67-92.

Selinker, Larry. 1972. Interlanguage. *IRAL* 10: 209-231. (Reprinted in Jack C. Richards, ed., *Error Analysis: Perspectives on Second Language Acquisition*, 1974, London: Longman, 31- 54.)

Tak, Jin-Young. 1996. Variable Vowel Epenthesis in Korean-accented English. *Proceedings of the Annual Boston University Conference on Language Development* 20, 2: 768-779.

Tesar, Bruce. 1995. *Computing Optimal Forms in Optimality Theory: Basic Syllabification.* Ms., University of Colorado at Boulder. [ROA-52].

Vennemann, Theo. 1988. *Preference Laws for Syllable Structure and the Explanation of Sound Change.* Berlin: Mouton de Gruyter.