TREND
May 3, 1997

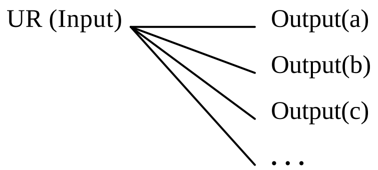
Ronald Sprouse
ronald@uclink.berkeley.edu

A case for Enriched Inputs

1. Standard model of OT:

In the Standard model of Optimality Theory (OT; P&S 1993, M&P 1993), the underlying representation, or Input, maps to a set of candidate Outputs created by the special function GEN. No intermediate representations are allowed.

(1)



2. Some challenges for Standard OT:

Sometimes an intermediate representation is required – compensatory lengthening and some facts about epenthesis and deletion in Turkish are difficult to analyze in Standard OT.

2.1 Compensatory lengthening

The fact that compensatory lengthening (CL) is triggered by deletion of coda consonants but not onsets is difficult to express in Standard OT (Zec 1994). The segment that deletes is not in the coda in UR, nor can it be in the coda in the Output. How can we express the fact that the deleted segment corresponds to a position in the coda at some level of representation without appealing to an intermediate representation?

(2) Syllable-based approach to CL

```
               Metrification (1)   Deletion       Metrification (2)
                μμ  μμ              μμ  μμ          μμ  μμ
               /|| /||             /|  /||         /|/ /||
/CVC-CVC/ →    CVC.CVC →           CV .CVC →       CV .CVC
```

2.1.1 Oromo (Lloret 1988, 1991)

/ɗ/ deletes before C-initial suffixes, triggering lengthening of the preceding vowel

(3)

(a)	/feɗ-na/	[feena]	'we wish'
(b)	/feɗ-sisa/	[feesisa]	'I make wish'
(c)	/feɗ-aɗɗa/	[feɗaɗɗa]	'wish for self'
(d)	/feɗ-umsa/	[feɗumsa]	'wishing' (n)
(e)	/feɗ-a/	[feɗa]	'I wish'

2.1.2 Komi (Harms 1968)

/l/-deletion triggers CL (104-105) (Ižma dialect) (for more discussion see Baker 1985: 98 and De Chene and Anderson 1979: 524-25)

(4)

	stem	1sg. past	infinitive	
(a)	/kɨl/	kɨl-i	kɨ:-nɨ	'to hear'
(b)	/sulal/	sulal-i	sulo:-nɨ	'to stand'
(c)	/lɨy/	lɨy-i	lɨy-nɨ	'to shoot'
(d)	/mun/	mun-i	mun-nɨ	'to go'

Word-final /l/-deletion triggers CL

(5)

		definite	indefinite	
(a)	/pil/	pilɨs	pi:	'the cloud'
(b)	/pi/	piyɨs	pi	'the son'
(c)	/vəl/	vəlɨs	və:	'the horse'

These are cases where Weight-by-Position would be said to apply (i.e., μ is not in UR). How can we handle this in standard OT? Paradox (Zec 1994)

Is it possible that μ is really in UR? – No, μ cannot be in UR:

compare Oromo /feɗ-umsa/ → [feɗumsa] (*[feɗɗumsa] or *[feeɗumsa])

and /fed-na/ → [feena]

we don't get V-lengthening, even though moraic preservation predicts we should get it, as we do when /ɗ/ deletes

Could we use segment correspondence only? (Lee 1996)

2.2 Turkish epenthesis and deletion

In Turkish an epenthetic vowel is required to break up certain disallowed coda consonants clusters. Yet this vowel sometimes triggers the deletion of one of the offending consonants. How can this be expressed without an intermediate representation?

(6) **2sg possessive suffix *-n***

(a)	bebek	'baby'	
	bebe-[i]n	'your baby'	Epenthesis and Deletion
	cf. bebekler	'baby-PL'	
(b)	temel	'foundation'	
	temel-[i]n	'your foundation'	Epenthesis
(c)	araba	'car'	
	araba-n	'your car'	
	(cf. /araba-Im/ → arabayım 'I am a car')		

<u>Question:</u> The epenthetic vowel can't be in UR. Why is it needed in *bebein* where there is no consonant cluster to break up? Alternatively, since deletion of *k* is allowed in *bebein*, why isn't **beben* the optimal candidate?

(7)

	/bebek-n/	*VkV	*CC]$_\sigma$	Dep	*VV
a.	bebekin	*!		*	
b.	bebein			*!	*
c.	beben				
d.	bebekn		*!		

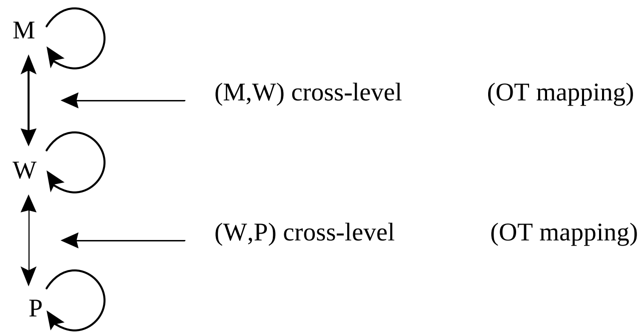
(8) Rule-based approach to Turkish

```
                Metrification (1)    k-deletion
                 μ  μ  μμ             μ  μ μμ
                /| /| /||            /| /| ||
/bebek-n/ →     be.be.kin →          be.be.in
```

3. A possible approach – a 3-level theory

In rule-based generative phonology some linguists have advocated 3-level models of phonology (Goldsmith 1993, Lakoff 1993). Incorporation of such an approach into OT might allow a straightforward analysis of the above cases.

(9) **M/W/P model** (Goldsmith 1993:33)



Problem with this approach: No real reason to stop at three levels.

4. Another approach – 2-level constraints

Can we refer to both Input and Output simultaneously (McCarthy 1995, Cho 1995)

(10) μProjection: project μ for every input C followed by a C (Zec 1994)
(11) K-vocalization: Input /k/ corresponds to an output [i] before C]$_\sigma$?

(12)

a) Theoretical problem: CL constraint is a bad constraint – it's more of a rule, combining a well-formedness constraint (how syllables should be organized) with its repair (project a mora) (see Paradis 1988 for TCRS) even when the well-formedness can't be evaluated directly since the mora and segment that projects it are never present at the same level; the 2-level constraint for Turkish is similarly rule-like
b) Typological problem: this morification algorithm for CL won't work for a language with complex onsets/codas
c) Empirical problem: doesn't make the right predictions about final Cs in Komi

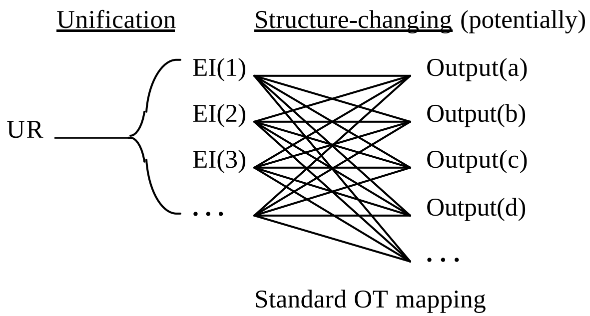
5. Proposal – Enriched Inputs

Let us admit that we need an intermediate representation to handle CL and Turkish epenthesis/deletion. How can we introduce this representation without creating a 3-level theory and opening the door to unlimited levels of representation?

5.1 Enriched Input model

UR — underlying representation – the basis of (but not necessarily the same as) candidate inputs to GEN
U-GEN — 'Unification GEN', a weaker form of GEN that provides an Enriched Input set of inputs that unify with UR
EI — Enriched Input, a set of candidate inputs provided by U-GEN
Output — a set of candidate outputs provided by GEN

(13) Enriched Input model



(14) Intrinsic properties of the UR, EI relationship (Unification; same principle as in syntactic theories, e.g. HPSG)

i) structure-filling relationship only (information in UR must be a subset of information in each EI candidate)
  ii) NO deletion, NO feature changing
  iii) NO Correspondence constraints (Max is irrelevant anyway since deletion not allowed)
iv) well-formedness constraints may apply to Enriched
v) No constraints ever refer to UR

(15) EI → Output mapping (same as Standard OT)

i) structure-filling and structure-changing
ii) deletion, insertion allowed
iii) Correspondence constraints (MAX and DEP) well-formedness constraints on Output

The EI consists of a candidate set just like the Output consists of a candidate set, although the class of potential members of the set is more restricted. The EI must contain all information present in UR, but the Output may delete information present in EI. The UR-EI mapping and EI-Output mapping are evaluated simultaneously, as in standard OT.

Example candidate inputs for UR /feɗna/: {fe ɗna}, {feɗna}, {feɗinata} are possible EIs of UR /feɗna/, but {feena} is not since it doesn't unify with UR.

5.3 Analysis of CL

(16) Oromo; UR /feɗ-na/

<table>
<tr><th></th><th>EI</th><th>Output</th><th>Syll$_{Input}$</th><th>Syll$_{Output}$</th><th>*ɗ]$_\sigma$</th><th>Max</th><th>Dep</th></tr>
<tr><td>1a.</td><td>σ σ
/\ |
μμ μ
/|| /|
feɗ na</td><td>μμ μ
/|//|
fe na</td><td></td><td></td><td></td><td>ɗ</td><td></td></tr>
<tr><td>1b.</td><td></td><td>μ μ
/|/|
fena</td><td></td><td></td><td></td><td>ɗμ!</td><td></td></tr>
<tr><td>1c.</td><td></td><td>μμ μ
/|| /|
feɗ na</td><td></td><td></td><td>*!</td><td></td><td></td></tr>
<tr><td>2a.</td><td>σ σ
| |
μ μ
/|\ /|
feɗ na</td><td>μμ μ
/|//|
fe na</td><td>*!</td><td></td><td></td><td>ɗ</td><td>*</td></tr>
<tr><td>2b.</td><td></td><td>μ μ
/|/|
fena</td><td>*!</td><td></td><td></td><td>ɗ</td><td></td></tr>
<tr><td>2c.</td><td></td><td>μμ μ
/|| /|
feɗ na</td><td>*!</td><td></td><td>*</td><td></td><td></td></tr>
</table>

*ɗ]$_\sigma$ [ɗ] not allowed in the coda
Max Don't delete
Dep Don't insert
Syll Have well-formed syllables (in Oromo, ɗ must be moraic)

5.4 Analysis of Turkish

(17) Turkish epenthesis and deletion; UR /bebek-n/

	EI	Output	*CC]$_\sigma$ Input	*CC]$_\sigma$ Output	*VkV	Dep	Max
1a.	[be]$_\sigma$[bekn]$_\sigma$	[be]$_\sigma$[bekn]$_\sigma$	*!	*			
1b.		[be]$_\sigma$[be]$_\sigma$[kin]$_\sigma$	*!		*	*	
1c.		[be]$_\sigma$[be]$_\sigma$[in]$_\sigma$	*!			*	*
1d.		[be]$_\sigma$[ben]$_\sigma$	*!				*
2a.	[be]$_\sigma$[be]$_\sigma$[kin]$_\sigma$	[be]$_\sigma$[bekn]$_\sigma$		*!			i
2b.		[be]$_\sigma$[be]$_\sigma$[kin]$_\sigma$			*!		
2c.		[be]$_\sigma$[be]$_\sigma$[in]$_\sigma$					k
2d.		[be]$_\sigma$[ben]$_\sigma$					ki!

This analysis also accounts for the allomorph distribution of the 3pers. poss. suffix /-I/ and /-sI/:

(18) (a) /bebek-I/ → [bebei] 'his baby'
(b) /temel-I/ → [temeli] 'his foundation'
(c) /araba-sI/ →[arabasi] 'his car'

Allomorphy selection in OT has been motivated by output well-formedness (Dolbey 1996, Kager 1995). This is a case of allomorph selection based on *input* well-formedness. A Standard OT account would need to use subcategorization frames to select the appropriate allomorph.

(19) Turkish allmorph selection; UR /bebek-{-I,-sI}/

	EI	Output	NoCoda Input	NoCoda Output	*VkV	Dep	Max
1a.	[be]$_\sigma$[be]$_\sigma$[ki]$_\sigma$	[be]$_\sigma$[be]$_\sigma$[ki]$_\sigma$			*!		
1b.		[be]$_\sigma$[be]$_\sigma$[i]$_\sigma$					
2a.	[be]$_\sigma$[bek]$_\sigma$[si]$_\sigma$	[be]$_\sigma$[bek]$_\sigma$[si]$_\sigma$	*!	*			

5.5 Why Outputs and Inputs with superfluous structure are not optimal

Given UR /paka/, why isn't an EI {pakata} or {pakatata} as optimal as {paka}? {pakata} is equally well-formed, and there are no Correspondence violations for adding structure in EI.

Reason: *struc Don't have structure.

(20) Hypothetical UR /paka/

	EI	Output	Syll	Dep	*struc
1a.	pakata	pakata			pakat!a
2a.	pakatata	pakatata			pakat!ata
3a.	paka	paka			paka

Important point: *struc can never be ranked so low as to be irrelevant. Given equally well-formed candidates, *struc will always favor the one with less structure (i.e. the one most closely resembling UR).

Duke of York gambit – In a rule-based theory with intermediate representations it is possible to introduce a segment that is ill-formed in the grammar (never surfaces) solely for the purpose of influencing a later rule. It is not possible to introduce such a segment in the EI model. Superfluous structure can't be added, as in (20). Only segments that increase harmony can be added.

5.6 Lexicon Optimization

The idea that segmental and syllable inventories can be judged as more or less harmonic by working backward from constraint rankings is known as the principle of Lexicon Optimization (P&S 1993). The additional insight that Lexicon Optimization can be used to find the most harmonic UR for a given morpheme that displays some phonological alternation is found in Inkelas (1994):

(21) Lexicon Optimization (Inkelas 1994)

> Given a grammar G and a set S = {S_1, S_2, ... S_i} of surface phonetic forms for a morpheme M, suppose that there is a set of inputs I = {I_1, I_2, ... I_j}, each of whose members has a set of surface realizations equivalent to S.

There is some $I_i \in I$ such that the mapping between I_i and the members of S is the most harmonic with respect to G, i.e. incurs the fewest marks for the highest ranked constraints. The learner should choose I_i as the underlying representation for M.

The EI model provides the optimal input already. In cases where a morpheme never undergoes alternations, this optimal input can be taken to be the UR. For a morpheme that alternates, such as /feɗ/ in Oromo, the UR would be the intersection of all the optimal inputs corresponding to /feɗ/:

(22) Optimal Inputs for /feɗ/:

(a) With a C-initial suffix

```
  σ
 /\
 μμ
/||
feɗ
```

(b) With a V-initial suffix

```
 σ
 |
 μ
/|
feɗ
```

(23) UR = intersection of Optimal Inputs

```
 σ
 |
 μ
/|
feɗ
```

This approach to Lexicon Optimization reduces the number of possible EI candidates.

6. Yowlumne

Can the EI model handle other types of opacity?

Harmony = spread [rnd] to the right to vowels of the same height

u…i → u…u   o…i → o…i
u…a → u…a   o…e → o…o

Lowering = VV[high] → VV

(data from Newman 1944 via Goldsmith 1993:33) (see also Archangeli 1988, Cole & Kisseberth 1996, Kisseberth 1969)

Future passive	Aorist passive	Gerundive	Gloss
a. CVC – short vowels only			
xil-nit	xil-it	xil-ʔas	‘confuse (a situation)’
hud-nut	hud-ut	hud-ʔas	‘recognize’
gop-nit	gop-it	gop-ʔos	‘take care (of a child)’
b. CVVC			
mek^{ʔ}-nit	meek^{ʔ}-it	mek^{ʔ}-ʔas	‘swallow’
sog-nut	**soog-ut**	**sogʔ-as**	**‘unpack’**
dos-nit	doos-it	dosʔ-os	‘recount’
c. CVCVVC			
hiwet-nit	hiweet-it	hiwet-ʔas	‘walk’
sudok^{ʔ}-nut	sudook^{ʔ}-ut	sudok^{ʔ}-ʔas	‘remove’
ʔopot-nit	ʔopoot-it	ʔopot-ʔos	‘get up’
d. CVCVV			
ʔilee-nit	ʔile-t	ilee-ʔas	‘expose to wind’
cuyoo-nut	cuyo-t	cuyoo-ʔas	‘urinate’
hoyoo-nit	hoyo-t	hoyoo-ʔos	‘name’
e. Epenthesis CVCC			
ʔilik-nit	ʔilk-it	ʔilik-ʔas	‘sing’
ʔugun-nut	**ʔugn-ut**	**ʔugun-ʔas**	**‘drink’**
logiw-nit	logw-it	logiw-ʔas	‘pulverize’

Rule ordering:

Epenthesis >> Harmony (feeding)	/ʔugn-t/ → [ʔugnut]
Harmony >> Lowering (counterbleeding)	/suugnIt/ → [sognut]
Lowering >> Closed-σ Shortening (counterbleeding)	/suugnIt/ → [sognut]

Constraints for Enriched Input approach:

Ident[rnd],[high] (round is privative)

Syll outlaws CVVC, CCC (i.e., drives closed σ shortening and epenthesis)

Agree spread [rnd] to right edge of word if vowels are of the same height (see Cole & Kisseberth for a formulation of these constraints)

*VV[high] no [high] on a V linked to two μ in Input or Output

Epenthesis and harmony in an example without lowering – no rankings crucial since winner has no violations

	/ʔugn-t/		$Syll_{Input}$	$Syll_{Output}$	Ident[rnd]	$Agree_{Input}$
1a.	ʔugnut	ʔugnit			**	
1b.		ʔugnut				
2a.	ʔugnit	ʔugnit				*
2b.		ʔugnut			*	*
3a.	ʔugnt	ʔugnit	*			
3b.		ʔugnut	*			

harmony, lowering and shortening – crucial rankings, exemplified with a form that undergoes high vowel shortening:

$Syll_{Output}$ >> Maxμ (shortening)

$Syll_{Input}$ violated in all candidates (shortening not possible in EI because of unification requirement)

*VV[high] >> Ident[high] (lowering)

	/suug-nIt/		$Syll_{Input}$	$Syll_{Output}$	Ident [rnd]	*VV [hi]	Ident [hi]	$Agree_{Input}$	Max μ
1a.	suug.nit	suug.nit	*	*!		*		*	
1b.		soog.nit	*	*!			*	*	
1c.		suug.nut	*	*!	*	*		*	
1d.		soog.nut	*	*!	*		*	*	
1e.		sug.nut	*		*!	*		*	*
1f.		sug.nit	*			*!		*	*
1g.		sog.nut	*		*!		*	*	*
1h.		sog.nit	*				*	*!	*
2a.	suug.nut	suug.nit	*	*!	*	*			
2b.		soog.nit	*	*!	*		*		
2c.		suug.nut	*	*!		*			
2d.		soog.nut	*	*!			*		
2e.		sug.nut	*			*!			*
2f.		sug.nit	*		*!	*			*
2g.		sog.nut	*				*		*
2h.		sog.nit	*		*!		*		*

References

Archangeli, Diana. 1988. *Underspecification in Yawelmani phonology and morphology*. Garland: New York.

Broselow, Ellen, Su-I Chen, Marie Huffman. 1996. Syllable weight: Convergence of phonology and phonetics. ms.

Cho, Young-mee. 1995. *Rule ordering, and constraint interaction in OT*. Handout from Phonology Workshop. Stanford University.

Cole, Jennifer and Charles Kisseberth. 1996. Restricting multi-level constraint evaluation: Opaque rule interaction in Yawelmani vowel harmony. ms.

Dolbey, Andrew. 1996. Output optimization and cyclic allomorph selection. WCCFL.

Goldsmith, John. 1993. 'Harmonic phonology' in *The last phonological rule*. John Goldsmith, ed. 21-60.

Gragg, Gene B. 1982. Oromo dictionary. East Lansing, MI: African Studies Center, Michigan State University.

Harms, Robert. 1968. *Introduction to phonological theory*. Prentice-Hall. New Jersey.

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20:253-306.

Hayes, Bruce. 1995. *Metrical stress theory*. Chicago: University of Chicago.

Hyman, Larry. 1985. *A theory of phonological weight*. Dordrecht: Foris.

Inkelas, Sharon. 1994. *The consequences of optimization for underspecification*. ROA 40-1294.

Kager, René. 1995. On affix allormorphy and syllable counting. ROA 88-00002.

Kisseberth, Charles. 1969. On the abstractness of phonology: The evidence from Yawelmani. *Papers in linguistics* 1. 248-82.

Lakoff, George. 1993. 'Cognitive Phonology' in *The last phonological rule*. John Goldsmith, ed. 117-45.

Lee, Byung-Gun. 1996. The emergence of the faithful. ms. Seoul University/University of Massachusetts, Amherst. ROA 167-1296.

Lloret-Romanyach, Maria-Rosa. 1988. *Gemination and vowel length in Oromo morphophonology*. PhD dissertation. Indiana University.

Lloret, Maria-Rosa. 1991. Moras or skeletal units? A question of parametric variation. *Catalan Working Papers in Linguistics*. Universitat Autònoma de Barcelona. 149-165.

McCarthy, John. 1994. *Remarks on phonological opacity in Optimality Theory*. Prepared for Jacqueline Lecarme, Jean Lowenstamm, & Ur Sholonsky, eds., Proceedings of the Second Colloquium on Afro-Asiatic Linguistics (Sophia Antipolis, June 1994)

McCarthy, John and Alan Prince. 1993. *Prosodic morphology 1: Constraint interaction and satisfaction*. Ms. University of Massachusetts, Amherst and Rutgers University. [To appear, MIT Press]

Newman, Stanley. 1944. *The Yokuts language of California*. The Viking Fund Publications in Anthropology #2. New York.

Paradis, Carole. 1988. On constraint repairs and strategies. *Linguistic Review* 6. 71-99.

Prince, Alan and Paul Smolensky. 1993. *Optimality Theory: Constraint interaction in generative grammar*. Ms. Rutgers University and University of Colorado, Boulder. [To appear, MIT Press]

Selkirk, Elisabeth. 1990. A two root theory of length. *University of Massachusetts Occasional Papers* 14. 123-171.

Zec, Draga. 1988. *Sonority constraints on prosodic structure.* PhD dissertation. Stanford University.

Zec, Draga. 1994. Coda constraints and conditions on syllable weight. ms.

Zec, Draga. 1995. Sonority constraints on syllable structure. *Phonology* 12: 85-129.