

FROM A RULE-BASED MODEL TO A CONSTRAINT-BASED MODEL

จากแนวคิดกฎเกณฑ์สู่แนวคิดแบบข้อจำกัด

Dr. Apichai Rungruang

Abstract

In the late 1950's, Noam Chomsky and Moris Halle broke new ground by introducing the Generative School of Phonology. This school believes that phonological structure echoes the linguistic competence of the speakers of particular languages. Speakers are able to compute a phonetic representation for the infinite number of sentences generated by the grammar. Generative phonology consists of five crucial components-levels of phonological representation (underlying form and surface form), phonological rules, derivations, distinctive features, and linearity. In this paper, two generative phonological models are examined. Even though a constraint-based model or Optimality Theory (OT) can solve the duplication problem, a rule-based model still exists for some reasons. This study does not intend to identify which model is the better between the two. Rather, both models account for phonological alternations in different ways.

Key words : *phonology, Optimality Theory, rules-based model, phonological alternations, English plural suffixes*

บทคัดย่อ

ในปลายทศวรรษที่ 1950 นอม ชอมสกีและมอริส แฮลลี่ ได้นำแนวการศึกษาสัทวิทยาหรือศาสตร์แบบเพิ่มพูนเข้ามา แนวคิดนี้เชื่อว่าโครงสร้างของรูปแบบเสียงสะท้อนถึงความสามารถทางภาษาศาสตร์ของผู้พูดในภาษานั้นๆ ผู้พูดสามารถสร้างเสียงที่ใช้สื่อความหมายเป็นประโยคที่หลากหลายได้นับไม่ถ้วนจากไวยากรณ์ในสมอง กล่าวโดยทั่วไปว่าสัทวิทยาแบบเพิ่มพูนมีองค์ประกอบหลักๆ 5 ประการคือ ระดับของการ

ทดแทนเสียง กฎระบบเสียงต่างๆ ลักษณะเด่นจำแนก การแปลงรูปของเสียง และแนวเชิงเส้น บทความนี้ ศึกษา ลัทธิตฤษฎี 2 แนวคิดซึ่งถือว่าเป็นส่วนหนึ่งของลัทธิตฤษฎีแบบเพิ่มพูน แม้ว่าแนวคิดแบบข้อจำกัดต่างๆ (Constraint-based model) หรือ ทฤษฎีจุดมผล (Optimality Theory) จะแก้ปัญหาการทำซ้ำได้ แต่แนวคิดแบบกฎเกณฑ์ต่างๆ (Rule-based model) ยังปรากฏอยู่ อนึ่งการศึกษานี้มิได้มุ่งหวังเพื่อตัดสินว่าแนวคิดใดดีกว่ากัน แต่ต้องการชี้ให้เห็นว่าทั้งสองแนวคิด มีมุมมองในการอธิบายการเปลี่ยนแปลงของเสียงที่แตกต่างกัน

คำสำคัญ: วรรณคดี ทฤษฎีจุดมผล แนวคิดแบบกฎเกณฑ์ การเปลี่ยนแปลงทางด้านวรรณคดี คำลงท้าย ภาษาอังกฤษในรูปพหูพจน์

Introduction

Before the 1990s, a rule-based model (RBM) played a significant role in employing rules to account for natural languages' alternations. However, nothing lasts forever. Since the 1990s, as a new model of phonological derivation, constraint-based model, or Optimality Theory (OT), has come into prominence. Rules fall from grace and the explanatory burden is placed entirely on constraints of Universal Grammar (UG). This article demonstrates two models in which phonological alternations have been made. First, a concept of rule-based model, which reveals changes in the representation so as to make it conform to the requirements of the language, is discussed. Then, the concept of OT, which focuses on a ranking of a set of violable and universal output constraints, is investigated with related examples. The final part wraps up the striking points of the two models.

1. Rule-Based Model

Phonological Rules: Why do they matter?

Rule-based model (RBM) ¹ uses phonological rules to account for the patterns of speech sounds in our mind. The rules help us track down what happens in our mind (the underlying forms) before we hear the words or sentences (the surface forms). For instance, the word 'link' is pronounced [lɪŋk] rather than [lɪnk]. Why is the former, not the latter? Why and how does an [ŋ] come out? It looks like something must occur in our brain before we hear it. Therefore, what is going on in our

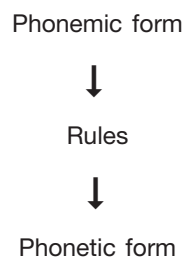
¹Rule-based model (RBM) can be called in different names such as a derivational approach, a traditional approach, a traditional model, a rule-based theory, a rule-based approach, etc.

brain requires phonological rules to account for. This traditional approach is related to the Sound Pattern of English, Chomsky and Halle (1968) Called a derivational approach, RBP states that rules must be ordered except the first rule; that is, an output of the previous rule serves as an input of the following rule. After all the appropriate changes have been made, the correct form (surface representation) surfaces. In other words, the surface form of the expression is derived in a series of structural changing operations from the underlying form.

Below is a general form:

$$A \rightarrow B / C_D$$

This form shows that A sound (the focus of the change) changes into B sound (the structural change), if A is between C and D. A change from the phonemic underlying form to the actual phonetic form of a word by means of phonological rules can be represented in the following diagram:



To put it simply, a phonological rule is formulated to connect the mind and the mouth.

Underlying Forms	Distribution Statements	Surface Forms
/phonemic level/	phonological rules	[phonetic level]

Consider the following data set from American English.

/ˈbɪtəɹ/	→	[ˈbɪɾəɹ]	‘bitter’
/ˈleɪtəɹ/	→	[ˈleɪɾəɹ]	‘letter’
/ˈraɪdəɹ/	→	[ˈraɪɾəɹ]	‘rider’
/ˈraɪtəɹ/	→	[ˈraɪɾəɹ]	‘writer’

How do phonologists account for this phenomenon? Obviously a /t/ and a /d / become a flap [ɾ]. The next step is to find out the environment in which the flap occurs. What we see is that the flap occurs between vowels (or an intervocalic position) in an unstressed syllable. Now, use a general form to formulate a rule.

[t, d] → /r/ [-cons] _____ $\left[\begin{array}{c} \text{V} \\ \text{-stress} \end{array} \right]$

(Gussenhoven; & Jacobs. 2003: 97)

Another easy way to formulate the rule is following:

Alveolar stop → flap / v _____ v

unstressed

(Finegan. 1999: 119)

Below is how to apply a flapping rule.

Underlying forms (UF)	'bɪt̩əɪ	'ɹaɪt̩əɪ
Flapping rule	ɾ	ɾ
Surface forms (SF)	'bɪr̩əɪ	'ɹaɪr̩əɪ

RBP also requires a constraint². In this case, the constraint provides a very specific environment in which a flap occurs in both an intervocalic position and an unstressed syllable. When a constraint and a rule appear to do the same work, this inherent drawback of the traditional model is known as the duplication problem (Kenstowicz; & Kissenberth. 1977: 136). An interesting question is: if RBP has some problems, why does it still exist? For one important reason, phonological rules are applied in sequence, and they could express generalizations in simple ways.

Next, consider another example of the English plural suffixes³. Then, use RBM to account for the following set of data.

[-s]	[-əz]	[-z]
[k ^h æ-s] cat <u>s</u>	[bʊʃ-əz] bush <u>es</u>	[k ^h ii-z] key <u>s</u>
[bəlɪf-s] belie <u>fs</u>	[rɒʊz-əz] rose <u>s</u>	[sliiv-z] sleeve <u>s</u>
[k ^h ʌp-s] cup <u>s</u>	[k ^h ɪz-əz] kiss <u>es</u>	[k ^h æb-z] cab <u>s</u>
[t ^h æk-s] tack <u>s</u>	[tʃəɪʃ-əz] church <u>es</u>	[bɔɪ-z] boy <u>s</u>
[p ^h æθ-s] path <u>s</u>	[dʒʌdʒ-əz] judg <u>es</u>	[dɒg-z] dog <u>s</u>

² A constraint refers to a condition which restricts the application of a rule to ensure that the well formed structure is established. It is an argument adduced in favor of OT over RBP in that one should not have both constraints and rules if constraints alone can account for everything (McCarthy 2002: 243). What can be found in OT is a markedness constraint, namely *t/*dinter. This point will be addressed later in the discussion of OT constraints.

³ This part does not consider irregular plural formation such as man-men, child-children, sheep-sheep, crisis-crises, foot-feet, ox-oxen, etc. In addition, since this paper focuses on phonology, the interaction between phonology and morphology or morphophonemics is beyond the scope of this study.

	hats	bushes	dogs
UF	/hæt+z/	/bʊf+z/	/dɒg-z/
Schwa insertion rule	-	bʊf+əz	-
Devoicing rule	hæt+s	-	-

SF	[hæts]	[bʊfəz]	[dɒgz]
----	--------	---------	--------

What happened if the second rule (devoicing rule) would be applied first?

	rats	places	jobs
UF	/ɹæt+z/	/pleis+z/	/dʒʌb+z/
Devoicing rule	ɹæt+s	pleis+s	-
Schwa insertion rule	-	-	-

SF	[ɹæts]	*[pleiss]	[dʒʌbz]
----	--------	-----------	---------

UF: underlying form (phonemic level); SF: surfaceform (phonetic level)

Obviously, the order of phonological rules is crucial. Devoicing rule must follow schwa insertion rule. Notice that devoicing rule reflects an assimilation process, specifically a regressive assimilation process, in which the following segment influences the previous one. The rule appears to be significantly constrained to allow only a certain feature to occur in a particular environment. Once again, RBP seems to have both rule and constraint at the same time. Rules are ordered serially or sequentially⁵. It can be concluded that RBP has two major basic elements. One is there are two main levels of representation: input representation and output representation. Second, the mapping from both levels is guided, and constrained, by rules. Surface forms result from rule application.

2. Constraint-Based Model or OT

Constraint-based model or Optimality Theory is a theory of constraint interaction. It was first proposed at the start of the 1990s by Alan Prince & Paul Smolensky, but it has become widely

⁵ Note that even though English plural forms and past forms have two similar rules, the forms of phonological rules are different, as in: Assimilation rule: /d/ → voiceless / voiceless + _____ #

Insertion rule: /d/ → [əd] / alveolar stop + _____ # (Finegan. 1999: 128)

Since rules are language-specific and descriptive, an infinite number of rule types are possible.

known through the work of John McCarty. As a new model of phonology, rules are abandoned and the explanatory burden is placed entirely on constraints. The focus is placed on the output. The theory proposes that the grammars of all languages have a set of constraints which are a part of Universal Grammar or the innate language knowledge that humans have. Constraints are universal and present in all grammars. Kager (1999) claims that OT recognizes two types of constraints: markedness and faithfulness. The markedness constraint requires a structure to be unmarked; it prohibits any marked forms from appearing on the surface. The major force or constraint counterbalancing markedness is faithfulness to lexical contrasts. Faithfulness refers to the relation between the surface form and their lexical representation or the underlying form. That is, the input forms should match the output form. In other words, the input form must be faithful to the output form. No changes are allowed. There are three basic components of the theory:

1. GEN generates the list of possible outputs, or candidates.
2. CON provides the criteria, violable constraints (markedness and faithfulness constraints), used to decide between candidates.
3. EVAL chooses the optimal candidate based on the constraints.

In OT, an evaluation of output candidates by a set of ranked constraints can be shown by a tableau. To draw a clearer picture, below is an example.

C ₁	C ₂	C ₃	_A
*	*	***	B
*	**!		

The tableau above shows that there are three main constraints C₁, C₂, and C₃, where C₁ dominates C₂, which dominates C₃ (C₁ >> C₂ >> C₃). A is optimal if it does better than B on the highest ranking constraint which assigns them a different number of violations. If A and B tie on C₁, but A does better than B on C₂, A is optimal, even if A has many more violations (with a number of asterisks) of C₃ than B. A '!' indicates the fatal violation for the candidate (so the winning candidate should not have a '!' in its violation profile)

The one which survived is called the optimal output candidate, which is most harmonic with respect to the set of ranked constraints, and it is marked by the pointing finger . It is also a type of well-formedness. All less harmonic candidates are ill-formed. In this theory, perfect output forms

cannot exist since every output form violates at least one constraint. In other words, only one form wins, optimally but never perfectly. The next question is: how can OT deal with a flap segment? Four constraints are formulated. Three faithfulness constraints are:

MAX-IO: Input segments must have output correspondent (no deletion)

DEP-IO: Output segments must have input correspondent (no insertion)

IDENT-IO: The specification for the feature of an input segment must be preserved in its output correspondent (no replacement)

(Kager. 1999: 67)

A relevant markedness constraint I propose is below. $*t/*d^{inter}$: t and d cannot occur both in an intervocalic position and an unstressed syllable.

Tableau I MAX-IO, DEP-IO, $*t/*d^{inter}$ >> IDENT-IO

/bitəɪ/ bitter	MAX-IO	DEP-IO	$*t/*d^{inter}$	IDENT-IO
(a).[bɪɾəɪ]				*(r)
(b).[bɪəɪ]	*!(t)			
(c).[bitəɪ]			*! (t)	
(d).[bɪɾəɪə]		*!(ə)		* (r)
(e).[bitəɪə]		*!(ə)	* (t)	

The ranking of the first three constraints (MAX-IO, DEP-IO, $*t/*d^{inter}$) does not make any difference as long as IDENT-IO is in the lowest position. Otherwise, candidates (b) and (c) can be the winner. In OT, there is only one winner. Candidate (b) is fatal because the [t] is deleted, which violates MAX-IO.

A fatal candidate can be identified by *!. Candidate (c) is ill-formed since the [t] occurs in an intervocalic position, which violates $*t/*d^{inter}$. Like candidate (a), candidate (d) violates IDENT-IO. Worse, it is out because it violates a higher ranking constraint or DEP-IO. Like candidate (d), candidate (e) violates not only DEP-IO, but a markedness constraint or $*t/*d^{inter}$. Clearly, the optimal candidate is the one which violates the lowest constraint. Notice that dotted lines show that constraints are not ranked relative to each other; solid lines means constraints are ranked relative to each other. Now, employ OT to the English plural suffix. It turns out that five constraints are formulated. Below are two markedness constraints⁶.

⁶ Gussenhoven, Carlos and Jacobs, Haïke (2003) propose the following two markedness constraints in the English plural suffix.

- *ss: Sibilant clusters are not allowed.
- *AGREE [voice]: Adjacent obstruents must match for voicing.

(Wikipedia. 2012)

Notice that *ss and *AGREE [voice] are equivalent to schwa insertion rule and devoicing rule, respectively. Three faithfulness constraints are:

- MAX-IO : Input segments must have output correspondent.
- DEP-IO : Output segments must have input correspondent.
- IDENT-IO : The specification for the feature of an input segment must be preserved in its output correspondent.

Tableau II *ss, *AGREE [voice], MAX-IO >> DEP-IO, IDENT-IO

Tableau II *ss, *AGREE [voice], MAX-IO >> DEP-IO, IDENT-IO

/dɪf+z/ dishes	*ss	*AGREE [voice]	MAX-IO	DEP-IO	IDENT-IO
☞ (a) [dɪfəz]				*	
(b) [dɪfəs]				*	*!
(c) [dɪfz]	*!	*			
(d) [dɪf]			*!		
(e) [dɪfs]	*!				*

Candidate (d) plays a crucial role to have a strict ranking between DEP-IO and MAX-IO. Otherwise, there is no optimal form. It does not matter if DEP-IO is outranked IDENT-IO since both candidates (b) and (e) violate more constraints than candidate (a). This shows that lower-ranked constraint can be violated by the optimal output, but its violation must be minimal. Unlike a rule-based model, OT concerns how constraints are ranked. We can conclude that the ranking in Tableau II is: *ss, *AGREE [voice], MAX-IO >> DEP-IO, IDENT-IO.

Consider the word ‘cat’ by using the above constraint ranking.

Tableau III *ss, *AGREE [voice], MAX-IO >> DEP-IO, IDENT-IO

Tableau III *ss, *AGREE [voice], MAX-IO >> DEP-IO, IDENT-IO

/kætz/ cats	*ss	*AGREE [voice]	MAX-IO	DEP-IO	IDENT-IO
☞ (a) [kæts]					*
(b) [kætəz]				*!	
(c) [kætz]		*			
(d) [kæt]			*!		
(e) [kætzs]	*!	*		*	

Tableau III reveals that IDENT-IO must be ranked in the lowest position. Otherwise candidate (b) will be another winner, which contradicts the theory. That is, there is only one winner. So, the dotted line must be changed into a solid line between the two constraints. Before making any changes, consider another word to assure that DEP-IP must be ranked in a higher position than IDENT-IO.

Tableau IV *ss, *AGREE [voice], MAX-IO >> DEP-IO, IDENT-IO

Tableau IV *ss, *AGREE [voice], MAX-IO >> DEP-IO, IDENT-IO

/dɒgz/ dogs	*ss	*AGREE [voice]	MAX-IO	DEP-IO	IDENT-IO
☞ (a) [dɒgz]					
(b) [dɒgəz]				*!	
(c) [dɒgs]		*!			*
(d) [dɒg]			*!		
(e) [dɒgz̩s]	*!	*		*	

Tableau IV shows that candidate (a) is a clean winner. There is no violation. Thus, Tableau III proves that a solid line is required to make DEP-IO outranked IDENT-IO. As a result, a new tableau or Tableau V is needed.

Tableau V *ss, *AGREE [voice], MAX-IO >> DEP-IO >> IDENT-IO

/dɒgz/ dogs	*ss	*AGREE [voice]	MAX-IO	DEP-IO	IDENT-IO
(a) [dɒgz]					
(b) [dɒgəz]				*!	
(c) [dɒgs]		*!			*
(d) [dɒg]			*!		
(e) [dɒgz̥s]	*!	*		*	

A new ranking is: *ss, *AGREE [voice], MAX-IO>> DEP-IO >> IDENT-IO. The ranking of the first three constraints does not matter. On the other hand, the ranking of the last two constraints matter. In sum, five tableaux show that in OT, languages strive for well-formedness, and phonology operates so as to increase well-formedness without phonological rules. Instead, a set of ranked constraints examines the set of all possible output representations for a given input, and assigns degrees of well-formedness to these; the optimal member of this set is chosen as the preferred, or optimal, output. A highly ranked constraint is the one that must be satisfied. It cannot be violated. A lower ranked constraint is satisfied if possible, but not at the expense of a higher-ranked one.

Conclusion

The role of phonological rules is to express phonological structure of a particular language. Rules are also ordinarily taken to represent specific derivational instructions, or steps along the path relating underlying or deep to surface representations. In OT, the list of markedness and faithfulness constraints evaluates a set of possible output candidates to find the optimal candidate which violates

the fewest number of constraints. A shift from ordered rules to well-formedness constraints, giving rise to Optimality Theory does not mean that one is better than the other. RBM and OT provide different ways to satisfy syllable structure target. Sharing certain aspects together, both models explicitly encode markedness directly in the grammar. For instance, *ss and *AGREE [voice] are equivalent to a schwa insertion rule and a devoicing rule, respectively. Like Yoshimoto (2002), I agree that OT analysis provides more generalization than the old model. *AGREE [voice] can be employed to account for English plural forms and English past forms. RBM needs to state devoicing or assimilation rule separately although both English plural and past forms share the same name or assimilation rule. However, even though optimalists claim that OT focuses on the output, I argue that the theory does not obviously ignore the input in the spite the fact that it has Richness of the Base (ROTB), which refers that there are no restrictions on the input for the grammar. Anything that is a logically possible linguistic representation should also be a possible input for the grammar of any language. A good example can be seen in faithfulness constraints matching the features between input and output representation. As a result, both RBM and OT pay attention to underlying and surface forms. Since there is no massive pile of severe problems with RBM that everyone agrees to abandon, RBM is still taught and does exist. Below is a comparison between the two models.

A Comparison of RBM and OT

RBM	OT
- RBM shows that one input leads to one output	- OT has one input; it has an infinite number of output candidates. However, there is only one optimal output candidate or the winner.
- Rules are inviolable	- Constraints are violable. It is fatal if the candidate violates the highest constraint. The optimal candidate might violate the lowest constraint.
- Rules are ordered serially; they are applied one at a time in sequence. However, rules can be applied simultaneously.	- Constraints are ranked in parallel.
- A surface form results from rule application.	- A surface form is optimal if it incurs the least serious violations
- RBM reflects multilevel derivation. One rule's output is the next rule's input.	- There are no intermediate stages.
- Rules reveal language-specific. Rules reflect some aspects of phonological behavior in a particular language.	- Constraints are universal; they are part of UG.

References

- Chomsky, Noam; & Halle, Morris. (1968). **The Sound Pattern of English**. New York: Harper & Row.
- Finegan, Edward. (1999). **Language: Its Structure and Use**. 3rd ed. Fort Worth: Harcourt Barace college Publishers.
- Gussenhoven, Carlos; & Jacobs, Haike. (2003). **Understanding Phonology**. India: Arnold.
- Kager, Rene. (1999). **Optimality Theory**. New York: Cambridge University Press.
- Kenstowicz, Michael; & Kisseberth, C. (1977). **Topics in Phonological Theory**. New York: Academic Press.
- Wikipedia**. (2012). Retrieved November 12, 2012, from http://en.wikipedia.org/wiki/Optimality__theory