[Review]

Doing Optimality Theory: Applying Theory to Data


KYOKO YAMAGUCHI
University of Tokyo*

Keywords: phonology, Optimality Theory, constraint

1. Introduction

Optimality Theory (OT) (Prince and Smolensky (1993, 2004)) has been developed mainly in the area of phonology since the 1990s. It is a constraint-based theory, making a sharp contrast with the previous rule-based theory. Although a large amount of research has been conducted based on OT, there have not been systematic instructions for how to conduct research within this framework. Doing Optimality Theory fills this gap, examining theoretical backgrounds and providing practical advice for readers.

This book has some noteworthy features. First, it contains many Exercises and Questions, which give readers hands-on opportunities to analyze data. Second, although data from various languages are discussed, this book provides an in-depth analysis of one language: Yawelmani (also Yowlumne: a nearly extinct dialect of the California Penutian language Yokuts). The author uses data from Yawelmani to argue for the inadequacy of rewrite rules throughout the book. This exposition enables readers to pursue the course of analysis step by step. The next section introduces the outline of this book, showing the clarity of the author’s exposition and the value of his advice.

2. An Overview

Doing Optimality Theory consists of six chapters. After an introduction to OT in Chapter 1, Chapter 2 shows how to construct an analysis in this

* I am greatly indebted to Shin-ichi Tanaka, Shigeto Kawahara, Kohei Nishimura, Bum-Ki Son, and two anonymous reviewers for their insightful comments. I would like to express my gratitude to Jennifer Hansen for proofreading this paper. Needless to say, all remaining errors and inadequacies are my own.
framework, and the next chapter gives advice about writing up the analysis. Chapter 4 introduces various types of constraints and examines the way to motivate new ones. In Chapter 5, language typology and universals are discussed in relation to constraints and their permutations. The last chapter discusses variation, acquisition, and other current research questions.

Chapter 1 is an introduction to OT, surveying the reasons why OT became the dominant framework in place of the traditional rule-based theory and explaining basic concepts of this new framework. First, the problem of rewrite rules (A→B/C_D) is illustrated with the data of Yawelmani. The language has phonological processes such as final vowel deletion (e.g. /taxa:-k^2a/-→[ta.xak^2] ‘bring!’) and vowel epenthesis (e.g. /ʔi.lik.hin/-→[ʔi.lik.hin] ‘sing (nonfuture)’). In the style of SPE (The Sound Pattern of English) (Chomsky and Halle (1968)), these processes are stated in rewrite rules: V→φ/VC_# and φ→i/C_CC. Although these two rules appear to be unrelated, final vowel deletion does not apply if the output would produce CVCC syllables (e.g. /xat-k^2a/-→[xat.k^2a]/*[xatk^2] ‘eat!’), which is the very same structure the vowel epenthesis rule repairs. This situation, where more than one process is blocked or triggered by the same output constraint, is called a conspiracy (Kisseberth (1970)). Rewrite rules, however, could not capture conspiracies formally, which led to the development of the constraint-based theory.

The author also argues in Chapter 1 that constraints should be violable and shows how the violability is essential in OT. For instance, three constraints are involved in Yawelmani: *C unsyll, *COMPLEX-SYLLABLE, and *V#. These constraints prohibit unsyllabified consonants, complex syllables, and word-final vowels respectively. Although [ta.xak^2] (‘bring!’) and [ʔi.lik.hin] (‘sing (nonfuture)’) obey these three constraints, [xat.k^2a] (‘eat!’) violates *V#. If constraints were inviolable, *[xa.tik^2], which satisfies all these constraints, would be produced. This example illustrates the violability of constraints and implies that some constraints are more likely to be violated than others. This difference in priority among constraints is expressed formally as constraint ranking. For example, [xat.k^2a] (‘eat!’) shows that *C unsyll and *COMPLEX-SYLLABLE dominate *V# ( *C unsyll, *COMPLEX-SYLLABLE >> *V#).

The latter half of Chapter 1 explains the basic framework of OT: /input/ → GEN (Generator) → {candidate_1, candidate_2, …} → EVAL (Evaluator) → [output]. Candidates are produced from an input in GEN, and the optimal one is selected as an output in EVAL through interaction of constraints. The author emphasizes that OT is a theory of constraint interac-
tion and offers a framework of selecting optimal outputs; it does not say anything about the nature of representations or constraints themselves. This point is also important in that it enables OT to be employed in other fields than phonology.

With regard to underlying representations, however, OT makes a proposal of its own: Lexicon Optimization, which requires an underlying representation to be identical to the surface representation without independent evidence (Prince and Smolensky (1993, 2004)). This principle, which was originally proposed for nonalternating morphemes, is extended to alternating morphemes and sheds light on the issue of underspecification in OT (Inkelas (1995)). Nevertheless, the author refers to Lexicon Optimization only in a footnote (Chapter 2, p. 136) and does not recommend employing the principle as an analytic tool because it may cause misunderstanding. However, more focus on Lexicon Optimization would be helpful for better understanding of OT as it is one of the important principles in this theory.

In Chapter 2, the author gives a detailed account of procedure for doing analysis in OT and provides practical advice for readers based on his own experience. He does so with the example of Yawelmani phonology and also presents the analysis of do-support in English, showing how to develop an OT analysis in syntax.

Constructing an analysis includes several preliminary steps. It starts from choosing a good research topic with an appropriate size. In OT, phenomena which involve systematic alternations are good research topics because they indicate the possibilities of constraint interaction. Second, a descriptive generalization is an important step which mediates between data and analysis. Good generalizations refer to target output configurations and relevant unfaithful mappings since the interaction of markedness constraints and faithfulness constraints is essential to OT. Especially, “except when” phrases indicate constraint interaction. For example, the generalization of Yawelmani vowel deletion is as follows: Words cannot end in vowels (=target output configuration), and this requirement is enforced by deletion (=unfaithful mapping), except when an unsyllabified consonant would result. Third, the author points out that faithfulness constraints, which are less varied and better understood than markedness constraints, are a good starting point for analysis. When a certain faithfulness constraint is violated, it is dominated by some markedness constraint. The author calls readers’ attention to the definition of markedness constraints: they should never mention the input or contain words that refer to the phonological system (e.g. “phoneme”) or processes (e.g. “delete”). He also suggests the phrase
“Assign one violation mark for every …” as a strategy for precise definition of constraints. For instance, the definition of *V# is as follows: “Assign one violation mark for every phonological word that ends in a vowel.” In sum, these helpful suggestions show the importance of preliminary steps.

The next stage is to rank constraints based on the comparison of candidates. In Section 2.2, the author points out the danger of asserting ranking with no evidence and shows how tableaux illustrate valid ranking arguments, where constraint conflict, a winner–loser comparison, and no ranking disjunction are essential factors. This section is also a good guide to different types of tableaux: the violation tableau, the comparative tableau, and the combination tableau. Furthermore, the author introduces a Hasse diagram to show the constraint ranking more accurately.

The following sections discuss the candidates, constraints, and inputs that need to be considered in one’s analysis. First, losing candidates which threaten to tie or beat the intended winner should not be overlooked. The author advises looking for other losers which satisfy the relevant markedness constraint by violating other faithfulness constraints. In particular, he points out that attention should be paid to the ranking where MAX is dominated because it is possible to vacuously satisfy markedness constraints by deletion, i.e. the violation of MAX. On the other hand, he points out that harmonically bounded candidates, which have a proper superset of another candidate’s violation marks, need not be included because they are not informative about ranking. As pointed out in the text, ERC (Elementary Ranking Condition) (Prince (2002)) entailment can tell how informative a winner–loser pair is. Second, loser-favoring constraints and winner-favoring constraints should be considered because the former may threaten the winner and the latter may undermine the rankings established up to that point. In contrast, constraints which favor neither losers nor winners can be safely ignored. Third, as there are no restrictions on the lexicon, an unrestricted range of inputs should be considered (i.e. Richness of the Base). The systematic differences between languages result from the differences in constraint ranking and cannot be attributed to the differences in inputs. Although Richness of the Base is a difficult concept to grasp and often produces misunderstanding, the author gives an excellent explanation

\footnote{MAX is a faithfulness constraint which prohibits deletion. It is defined as follows: “Let input = i_1i_2i_3...i_k and output = o_1o_2o_3...o_m. Assign one violation mark for every i_x if there is no o_y where i_x \not\in o_y.” (\mathcal{R} means the correspondence relation.)}
for this concept. In summary, it is important to consider various candidates, constraints, and inputs in order to construct solid ranking arguments.

The author also lays out cases where direct ranking arguments are impossible. First, unviolated constraints cannot be ranked with respect to one another due to the absence of conflict. Second, with regard to constraints in general-specific (stringency) relation, the author argues that it is mistaken to use Paninian ranking (Specific >> General) as a universal ranking, which has been commonly assumed in previous studies. As there is no conflict between constraints in the stringency relation, direct ranking arguments are impossible although they can be ranked by transitivity. Another common mistake is that tie-breaking constraints should be ranked low. Although this seems natural intuitively, the fact is that they cannot be ranked.

Arriving at the whole constraint hierarchy based on the procedure above, it should be verified by summary tableaux with all of the relevant inputs and candidates. RCD (Recursive Constraint Demotion) algorithm and ERC fusion can also be employed to detect inconsistency in the analysis.

Chapter 3 gives advice about how to write up an analysis. In this chapter, the author stresses clarity in writing, not only for the readers, but for the writers themselves. First, the author discusses how to organize papers. The author argues that writers should give readers an outline of the whole discussion at the beginning of a paper to get their attention. Therefore, Section 1 as the introduction needs to raise questions and answer them, though this may seem counterintuitive. Furthermore, Section 2 should introduce essential background assumptions and new ideas. Although Section 2 often reviews previous studies in many papers, the comparison of previous proposals with new ones is only possible after the latter are presented.

Second, the way to present an OT analysis is discussed. As OT analysis is complicated because of constraint interaction, the author emphasizes the importance of presenting analyses step by step; authors should avoid presenting big tableaux or asserting a total ranking of constraints without justification of each ranking. The first step to incremental presentation is to find the tableau of input-output mapping which has the most constraints that favor neither the winner nor the loser among the summary tableaux constructed in the process of analysis because those constraints can be safely ignored. Then, after the presentation of the data and descriptive generalization, authors should present ranking arguments. The winner and a single loser are compared in the tableau which includes all constraints introduced so far; therefore, the size of the tableau is $2 \times n$, unlike $2 \times 2$ as assumed in many previous studies. For example, (1) is one of the tableaux which are
required in the ranking argument of vowel epenthesis in Yawelmani.²

(1) *COMPLEX-SYLLABLE >> DEP

<table>
<thead>
<tr>
<th>/ʔi.lik.hin/</th>
<th>MAX-C</th>
<th>*COMP-SYLL.</th>
<th>*Cunsyll</th>
<th>DEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. →ʔi.lik.hin</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>b. ʔilk.hin</td>
<td>*W</td>
<td></td>
<td></td>
<td>L</td>
</tr>
</tbody>
</table>

This comparative tableau compares the winner [ʔi.lik.hin] and the loser *[ʔilk.hin]. The former candidate, which has an epenthetic vowel [i], violates DEP to satisfy *COMPLEX-SYLLABLE. On the other hand, the latter is faithful to the input but violates *COMPLEX-SYLLABLE. As the former is the winner, *COMPLEX-SYLLABLE dominates DEP.

After a summary tableau is produced for the first input-output mapping based on tableaux described above, the same procedure is repeated for another mapping where one constraint at a time is added or where more information is given about ranking. It is important that the order of presentation proceeds incrementally.

Chapter 3 also mentions the responsibilities of good scholarship, including proper citation and attitudes toward previous studies. In particular, the author stresses the importance of fairness, rationality, and self-control in producing and receiving criticism, which of course applies to any field. Furthermore, he gives many helpful and practical suggestions as to choosing good research topics in general. For instance, some methods he suggests to find a good topic are adjusting the balance between new discoveries and previous work, inverting conventional wisdom, or accomplishing the same or more with less.

As mentioned above, OT is not a theory of constraints but a theory of constraint interaction; therefore, constructing an analysis in OT often involves theorizing about the constraint component (CON). Chapter 4 shows how to introduce new constraints and gives a comprehensive explanation of common constraints in OT. First, a ranking paradox, a contradiction between different input-output mappings, is a representational case where a new constraint is required. The new constraint should favor only the winner, and a pair of similar winners of different mappings enables identification of the constraint. Although more than one constraint may satisfy the

² This is based on tableau (5) on p. 146, where the winner (a) lacks the violation mark for DEP. The violation mark for DEP of the winner needs to be added also in (4) on p. 144, (6) on p. 146, and (7) on p. 147.
condition, language typology often indicates which is appropriate. With regard to the definition of a new constraint, the author recommends the phrase “Assign one violation mark for every …” In this formulation, the role of constraints is very simple: to assign violation marks to a candidate. Constraints do not specify tendencies or processes, which are the result of constraint interaction.

The latter half of Chapter 4 explains markedness constraints and faithfulness constraints. The former is a constraint on the output form, such as features, syllable structures, and prosodic hierarchy. These constraints are classified in the last section of this chapter, giving a handy list to readers. On the other hand, faithfulness constraints militate against the disparity between the input and the output, including deletion, epenthesis, and featural changes.

Both types of constraints invoke various issues that call for in-depth research. With regard to markedness constraints, the author discusses several issues, including the top-down or bottom-up definition, gradient evaluation, harmonic alignment, and the issue of “fixed hierarchies vs. stringency.” The issues that the author discusses regarding faithfulness constraints include two models of faithfulness (i.e. containment and correspondence), faithfulness to features (IDENT (feature) according to the segmental theory vs. MAX (feature) according to the autosegmental theory), and positional faithfulness. Explaining these important issues lucidly is one of the remarkable characteristics of this book. In addition, the author also classifies common markedness constraints in OT and gives readers a useful list of them.

Another important theme of Chapter 4 is how to justify constraints. Doing so in a proper way is essential to any OT analysis because a constraint without justification is merely an ad hoc device. The author lists three ways of justifying constraints: formal, functional, and typological. First, formal justification is to organize constraints into families. In particular, constraints defined by a certain schema are good examples, such as alignment constraints and local conjunction. Second, functional justification is to argue that a constraint aims at reducing the burden on the speaker or hearer: the ease of articulation and enhancing perception in phonology and promoting the ease of comprehension in syntax. Typological justification is the most convincing one; if the introduction of a certain constraint predicts a plausible typology, the constraint has good justification.

Language typology is discussed in Chapter 5 in more detail. In OT, all constraints exist in the grammars of all languages (i.e. universality of CON),
and the difference between languages results from the different ranking of constraints. One of the important aspects of studying the permutation of CON is the explanation of language universals. If some property appears in every language, this language universal should be observed under any ranking of CON. Furthermore, under Richness of the Base, the universal should not derive from the distribution of the input. They rather derive from the content of CON. For example, the language universal “Every language has some syllables with onsets.” implies that CON does not include a constraint which prohibits onsets (i.e. No-Onset). In addition, the author also refers to harmonic improvement (Prince (1997), Moreton (2003)) as a universal which is independent of assumptions on CON. This is one of the basic characteristics of OT: unfaithful mapping is possible only when it leads to the improvement in markedness.

In the latter half of this chapter, research into factorial typology is illustrated with the example of Yawelmani. The permutation of CON predicts possible languages and impossible languages, and these predictions are tested against the facts. Factorial typology also examines new constraints. In view of the universality of CON, introducing a new constraint can affect all languages, so it is important to consider its typological consequence. In addition, the author indicates that typological predictions can be employed to identify implausible constraints. This chapter also refers to empirical research on language typology based on the example of the relationship between vowel height and rounding harmony in various languages. The crucial point is that differences between languages should be attributed to constraint interaction, not to differences in the constraints themselves.

The last chapter deals with five topics of current research in OT: language variation, language acquisition, derivation, absolute ungrammaticality, and the too-many-solutions problem. OT offers new perspectives on the first two issues. As the difference of systems is accounted for by the difference of constraint ranking in OT, language variation and language acquisition are explained in the same way as language typology is. For instance, with regard to acquisition, markedness constraints are ranked above faithfulness constraints in child grammars, while the relationship can be the opposite in adult grammars.

The other three issues present new problems. First, flat derivation with no intermediate steps is controversial although it is one of the basic characteristics of OT. Transderivational similarities and opacity have been argued to be evidence for derivation. Second, as Eval always selects some candidate as the winner, a phonologically motivated gap, which results from
ungrammaticality, requires explanation. In this case, to posit the candidate *null output* and the constraint MPARSE, which is violated by the null output, is a possible solution. Third, factorial typology often predicts unattested patterns, which is called the too-many-solutions problem. Several approaches to the problem are proposed, such as the revision of the theory of faithfulness and perceptual similarity.

Lastly, before concluding this review, I would like to make a brief mention of the errata, which are generally unavoidable for any book. A list of typographical errors I noticed is included in the footnotes.

3 Although many of them are insignificant, (vii), (ix), (x), and (xiv) are concerned with content.

3. Conclusion

In conclusion, *Doing Optimality Theory* offers a comprehensive and profound explanation of the mechanism of OT and gives readers practical advice for constructing analyses within this framework as the title shows. All the advice given in this book is based on the long experience of the author, who has made an important contribution to the progress of phonology. In particular, readers will feel fortunate to find many Exercises and Questions which will help them to acquire the skills he suggests. As this book includes highly advanced discussions in phonology, some readers may find it technical or difficult; it is probably most appropriate for researchers who specialize in phonology, including graduate students. However, I firmly be-

---

3 The errata I noticed are as follows.
(i) p. 5, line 16: yawelmani → Yawelmani
(ii) p. 52, Tableau 3/p. 53, Tableau 1, 2: Cand 3 should be candidate (c), not candidate (b).
(iii) p. 74, line 2 from bottom: *[xat.k^7] → *[xat.k^2]*
(iv) p. 109, 11 lines below (78): a argument → an argument
(v) p. 128, 2nd line above (101): marks and of → marks
(vi) p. 138, line 20: to to call → to call
(vii) p. 168, the last line: it must add → it must not add
(viii) p. 202, line 14: able move → able to move
(ix) p. 237, line 21: x over y → y over x
(x) p. 237, line 22: y over x → x over y
(xi) p. 243, Tableau (4): The winner should be [xat.k^2a?], not [xat.k^2a?].
(xii) p. 251, Tableau (13): MEP → DEP
(xiii) p. 256, 2nd line above (20): both high (b) → both high (a)
(xiv) p. 257, Tableau (22), candidate (b): gophon → gophun
lieve that this book provides all readers with an insight into doing research in the field of linguistics in general as well as in the specific field of phonology.

REFERENCES


[received July 18 2009, revised and accepted November 18 2009]

Department of Language and Information Sciences
Graduate School of Arts and Sciences
University of Tokyo
3–8–1 Komaba, Meguro-ku
Tokyo 153–8902
e-mail: takano@phiz.c.u-tokyo.ac.jp