

Optimality Theory

A framework in theoretical linguistics, used to formalise analyses in phonology, and less frequently other areas of linguistics. Its core is the assumption that linguistic generalisations should be described using a set of violable constraints on surface representations which are ranked in terms of their importance.

See also: Generative Phonology; Universal Grammar.

Optimality Theory ('OT') took centre stage in theoretical linguistics during the 1990s. Its basic tenets were both familiar and revolutionary, and this has doubtless contributed to its success: it developed ideas long present in linguistics, but gave them new characteristics, considerably changing the understanding of the grammar. OT provided linguists with new ways to work and new theoretical problems to crack, although some have rejected it as fundamentally misguided. It was introduced by Alan Prince and Paul Smolensky and has been developed by many others, most notably John McCarthy. OT was created with phonological problems in mind, and is still most popular among phonologists, but is also used in syntax and other areas. After early presentations in 1991, Prince & Smolensky distributed the manuscript *Optimality Theory: Constraint Interaction in Generative Grammar* in 1993. Although not published until 2004, this was highly influential, setting out the OT approach (along with work by McCarthy & Prince). Despite some influence from neural networks, OT was essentially conceived as a development of *Generative Phonology, so it maintains the basic competence/performance distinction, and that between underlying and surface levels of representation, although these are reinterpreted. One central OT assumption is that only constraints should be used to characterise linguistic generalisations. There are no substantive rules or transformations. A set of potential surface forms ('candidates') are evaluated by the grammar and one is chosen as the optimal candidate (the 'output') because it violates a language's constraints in the least bad way. This is possible thanks to another key characteristic: all constraints are violable, and languages rank them in order of importance, so it is worse to violate a high-ranked constraint than a low-ranked one. As the standard assumption is that the set of constraints (known as CON) is universal, language learners must work out how constraints are ranked in the languages they acquire.

OT analyses are formulated in a 'tableau', exemplified below for German Final Obstruent Devoicing (FOD) in its standard description: syllable-final voiced obstruents are devoiced, as in *Bund* [bunt] 'federation' (a contentious but standard assumption here is that [d] in German is specified for [voice], while [t] is unspecified). The tableau's top row shows the input (the underlying representation), and then the constraints, in ranked order from left to right. The first column shows the set of candidates, one of which is chosen as the output (indicated by a pointing finger). The other columns show constraint violations, each receiving one asterisk. The 'fatal' violation (which rules candidates out) is indicated by an exclamation mark. There are two basic types of constraint: markedness and faithfulness constraints. Markedness constraints penalise candidates which contain marked structures: the analysis of FOD uses *FINALOBSTRUENT/VOICE (*FOV), which dictates that final obstruents may not be specified for [voice]. Faithfulness is a crucial innovation of OT: markedness constraints can exert pressure for an output to differ from its input, but faithfulness constraints do the opposite, requiring identity between input and output. Individual constraints regulate particular aspects of faithfulness: IDENT(voice) requires that the value of [voice] be the same in input and output, and MAX requires everything in the input to have some correspondent in the output, ruling out deletion.

/bund/	*FOV	MAX	IDENT(voice)
bund	!* -----		
☞ bunt			*
bun		!* -----	
bu		!** -----	

Fully faithful [bund] is rejected because it violates high-ranked *FOV. An unlimited number of candidates are in fact produced (by the function GEN, which simultaneously performs every possible process on the input to derive the candidates). In practice only a few ‘reasonable’ candidates are considered in analyses (outlandish candidates are assumed to be ruled out by high-ranked faithfulness constraints). The candidates include [bunt], which satisfies *FOV, with no voicing in its final obstruent. [bunt] violates IDENT(voice), but is still the output because other candidates violate higher ranked constraints. The other two candidates do not violate *FOV, as they have no final obstruent (having undergone deletion), but they do violate MAX, also ranked above IDENT(voice). This ranking is shown by the unbroken line between MAX and IDENT(voice). The broken line between *FOV and MAX shows that their mutual ranking is irrelevant here. The above ranking is specific to languages with FOD. In English, with no FOD, the constraint *FOV must be ranked below IDENT(voice).

It is worth noting that the right candidate would still be selected if the input were /bunt/. In fact, in some cases, the precise nature of the input is indistinct: thus it doesn’t matter whether stops are aspirated underlyingly or not in English, as these segments do not contrast and, assuming that aspiration is enforced by a high-ranked ASPIRATION constraint (‘initial stops must be aspirated’), then [p^hɪk] could be derived from either /pɪk/ or /p^hɪk/. Which of these is the underlying form is simply not an important (or relevant) question. This characteristic is known as the Richness of the Base, and has some theoretical benefits. In the German case, however, alternations (such as *Bund~Bundes* [bunt]~[bʊndəs] ‘federation’~‘federation-GENITIVE’) show that the underlying segment must be voiced.

OT analyses can often be linked to pre- or non-OT work by implementing as a violable constraint what such work sees as a universal principle. This has likely contributed to OT’s success because it expresses what phonologists have wanted to be able to say: certain principles often play a role in the phonology of languages, but not always. For example, the Obligatory Contour Principle (OCP), was introduced in the 1970s to forbid adjacent identical aspects of structure, such as adjacent high tones or specifications for place. By the late 1980s, it had been shown that the OCP is not absolute, but is best understood as a tendency in languages, elevated to the level of an inviolable principle in some. This is exactly what OT predicts, and OT thus allows phonologists to formalise tendencies.

Theoretical development in OT has largely involved either work on the types of constraints allowed, driven by attempts to find new, better ways to express linguistic generalisations, or work on additions to the basic theoretical machinery, often to remedy perceived short-comings in the model, such as its inability to account for opacity. The former includes work on ‘positional faithfulness’, which preferentially preserves input specifications in strong positions, such as the onset. FOD can also fit into a positional faithfulness model: a positional IDENTONSET(voice) can require that the value of [voice] be the same in input and

output in onsets, and conflict with a general *OBSTRUENT/VOICE, which requires that no obstruent be specified for [voice]. This achieves the same result, but also opens up new 'positional' analytical options.

Opponents of OT often argue that its central tenets are mistaken. Critics have claimed that the status of inputs is problematic. The Richness of the Base has been criticised on both psycholinguistic and theoretical grounds (much previous theoretical success was due to the ability to specify the characteristics of underlying forms). The fact that OT has largely led its practitioners to abandon work on phonological representation (especially at the segmental level) has been condemned. The contents of CON have also proved controversial. As we have seen, the constraints needed to analyse simple FOD are subject to debate: there is not necessarily one straightforward way to analyse any phenomenon, because there is no clear theory of CON. Some see this as an opportunity for debate, others as a problem: how can we know which constraints exist? A more fundamental criticism questions whether we really should formalise all tendencies as cognitive constraints on surface forms. Should 'ease of articulation' (formalised in OT as LAZY) be seen as the same kind of thing as constraints on the faithfulness of features?

Despite such criticism, OT has a firm place in theoretical linguistics, as the framework for most analysis and debate in phonology, and considerable work in other areas of linguistics. It provides a novel set of principles to guide analysis and has allowed new answers to old problems. For those who follow its direction, it has focused work on the status of constraints and moved attention away from representations and underlying levels. The framework for constraint interaction has also been adopted by linguists who reject generative assumptions. Its influence is felt widely in linguistic theory.

Primary works

Prince, Alan & Smolensky, Paul (1993/2004). *Optimality Theory: Constraint Interaction in Generative Grammar*. Rutgers University & University of Colorado at Boulder. Published 2004, Oxford: Blackwell.

McCarthy, John & Prince, Alan (1993). *Prosodic Morphology: Constraint Interaction and Satisfaction*. University of Massachusetts, Amherst & Rutgers University.

McCarthy, John & Prince, Alan (1995). Faithfulness and Reduplicative Identity. In Jill Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *University of Massachusetts Occasional Papers in Linguistics* 18, 249-384.

Further reading

McCarthy, John (2002). *A Thematic Guide to Optimality Theory*. Cambridge: Cambridge University Press.

Patrick Honeybone
Linguistics and English Language, School of Philosophy, Psychology and Language Sciences
University of Edinburgh
patrick.honeybone@ed.ac.uk

NB: The (*See Also's*) at the start, and asterisked phrases in the text refer only to other entries which feature in *Key Ideas in Linguistics and the Philosophy of Language*. The 'KT' at the start refers to entries in the companion volume *Key Thinkers in Linguistics and the Philosophy of Language*. It's a good book – buy it!

Further further reading and notes

(Titles and things which I could not include in the published version due to constraints on space: 1500 words in total...)

- The set of potential output forms (the candidates) are evaluated by the grammar through a function called 'EVAL'. The name typically given to a language-specific ranking of constraints is a 'constraint hierarchy'.
- With their plausibility as potentially universal in mind, constraints are typically more prized if they are general (that is, are not formulated to directly account for the specifics of a particular process in one language), or can be shown to be 'phonetically grounded' (that is, they directly reflect a universal phonetic pressure). Others have argued that the 'phonetically grounded' criterion confounds the distinction between phonological and phonetics.
- The fact that OT allows for the *formalisation of tendencies* can be seen as beneficial because phonologists have long talked of tendencies (such as the historical tendency towards consonantal lenition) or for stress to fall on heavy syllables. Others argue that 'tendencies' have no place in phonology, which should deal with the possible and the impossible.
- Generative Phonology of the 1970s and 80s had increasingly developed a 'mixed' model which used both rules and constraints. Proponents of OT argued that if phonology can be done with only one theoretical entity (ie, constraints), this is better than a model which needs two types of entities (constraints and rules).
- Developments in the basic OT framework include a number of attempts to account for *opacity*, including Sympathy Theory (where failed candidates are allowed to influence the successful candidate) and Stratal OT, which introduces lexical strata, and hence a minimal serialism into the otherwise parallel OT, rescuing much of what was proposed in the model of Lexical Phonology (which was central in phonology in the 1980s).
- Opponents of OT, who argue that its central tenets are mistaken, prefer to carry on using pre-OT models (for example, by retaining the use of phonological rules), or to develop alternative models (for example, by retaining a focus on phonological representations).
- Lombardi (1999) analyses the case of Final Obstruent Devoicing discussed above quite differently, on a *positional faithfulness* model. This preferentially preserves input specifications in strong positions, such as the onset. She uses the positional constraint IDENTONSET(voice), which requires that the value of the feature voice may not be different between input and output in onsets, and a general constraint *OBSTRUENT/VOICE, which requires that obstruents may not be specified for [voice]. This achieves the same result as the analysis given above, but also opens up options of analysis in other cases of positional faithfulness.
 - The German data can also be analysed on the basis of another development in the theory of constraints: constraint conjunction. The rather specific *FINALOBSTRUENT/VOICE constraint can be replaced by a *conjunction* of two more general constraints, which both need to be satisfied at the same time: Ito & Mester (2003) argue that *OBSTRUENT/VOICE & NOCODA, both of which other analyses require independently, account for Final Obstruent Devoicing when conjoined.
 - In fact, things may be more complicated than what is shown here: Iverson & Salmons (eg, 2007) argue persuasively that the situation in German is, in fact, not a case of final devoicing at all, but rather a case of 'final fortition' (ie, the addition of a [spread glottis] specification to the obstruents involved). This would require a change to the analysis shown above, of course, but the grammar given above would still work for the devoicing pattern that exists in languages like Polish, Dutch and Catalan.
 - These references are to:
 - Ito, Junko & Mester, Armin (2003) 'On the sources of opacity in OT: Coda processes in German'. In Caroline Féry and Ruben van de Vijver (eds.) *The Syllable in Optimality Theory*. Cambridge: Cambridge University Press, 271-303.
 - Iverson, G. & Salmons, J. (2007) 'Domains and Directionality in the Evolution of German Final Fortition.' *Phonology* 24.1-25.
 - Lombardi, Linda (1999) 'Positional Faithfulness and the Phonology of Voicing in Optimality Theory'. *Natural Language and Linguistic Theory* 17, 267-302.
- Although OT has met with less acceptance in syntax than in phonology, the model is more like previous work in generative syntax than it is previous work in phonology, as McCarthy (2002) explains. In the 'Government and Binding' and similar models of syntax, a number of possible surface forms are derived by movements from underlying forms, yet most derivations 'crash' because they are ruled out by 'filters' or principles, which are, effectively inviolable constraints. OT analyses can often be linked to pre- or non-OT work by taking what such work assumed to be a universal principle and implementing it as a violable constraint. In this way, the constraint SUBJECT implements what standard generative syntax has called the Extended Projection Principle, which requires subjects in all clauses, except for verbs like *rain* in pro-drop languages.

To appear in: Chapman, S. & Routledge, P. (eds) *Key Ideas in Linguistics and the Philosophy of Language*. Edinburgh: Edinburgh University Press.

- An example of OT in syntax: the analysis of expletive *it* in English, developed in a number of places, including Speas (1997). This uses MAX, interpreted in the same way as in phonology, and two markedness constraints: SUBJECT requires that all clauses have a subject, and FULL INTERPRETATION (FULLINT) requires that there be no superfluous elements in an output. As in the analysis of German devoicing, the optimal candidate can violate constraints (and does violate FULLINT because *it* is semantically superfluous), but only if other candidates violate constraints which are more highly ranked. The tableau below shows that SUBJECT must be high ranked in English, so that expletive subjects can be made to occur in the output. SUBJECT will be ranked below FULLINT in pro-drop languages like Italian and Japanese, where expletive subjects are not possible.

'It rained today'	SUBJECT	MAX	FULLINT
☞ It rained today			*
rained today	!*		
It today		!*	*

- Constraint ranking replaces the notion of parameters found in much non-OT syntax (and in some phonological models, such as Government Phonology). Proponents of OT claim that constraint ranking is better than parameterisation, because it allows for *the emergence of the unmarked* ('TETU') effects to be recognised. Low-ranked constraints are never turned off, unlike parameters (which are either on/off: a language must either require onsets, or not require them). Even a low ranked constraint, like ONSET in languages which allow for onset-less syllables, can still play a role in aspects of their phonology, for example in infixation patterns, which can be driven by the need for the infixable syllables to have an onset (and hence be unmarked).
- There is now a vast OT literature, and, handily, much of it is available through the Rutgers Optimality Archive:
 - <http://roa.rutgers.edu/>
- Some other worthwhile things to read are:
 - Legendre, Géraldine, Grimshaw, Jane & Vikner, Sten (eds.) (2001). *Optimality-Theoretic Syntax*. Cambridge, MA: MIT Press.
 - Kager, René (1999). *Optimality Theory*. Cambridge: Cambridge University Press.
 - Ritter, Nancy (ed.) (2000). *A Review of Optimality Theory*. Special issue of *The Linguistic Review* 17.
 - Speas M. (1997) 'Optimality theory and syntax: null pronouns and control.' In Archangeli, D. & Langendoen, D.T. (eds) *Optimality Theory: An Overview*. Oxford: Blackwell.