Chapter 1

Introduction

> [T]he principle of linear ordering must in fact be qualified, since in certain cases it is untrue that a linear ordering can be imposed on phonological rules.
>
> *Chomsky (1967: 118)*

BASIC GENERALIZATIONS in generative phonology are typically expressed in the form of MAPPINGS from input representations to output representations. To the extent that a given phonological theory individualizes these mappings, the otherwise expected application of a given mapping is sometimes observed to be prevented or BLOCKED under systematic circumstances attributable to other, independently motivated generalizations.

I take it as a given that an adequate phonological theory must accurately define the ways in which mappings interact such that one can block another and the conditions under which blocking of a mapping is expected. (These are of course not the only conditions on an adequate phonological theory; they are simply the ones on which I choose to focus attention here.) The relative adequacy of competing theories can be measured in part by the extent to which these manners of interaction and conditions of blocking follow from more basic assumptions.

A blocked mapping is a generalization about the phonology of a language that sometimes does not hold due to its systematic INTERACTION with some other generalization. Blocking shares this definition with certain types of OPACITY (Kiparsky 1971, 1973a), about which much ink has been spilled over the past decade and a half or so. Blocking and opacity are both examples of OBSCURED GENERALIZATIONS (Baković 2007, 2011), which have been the lynchpins of arguments between proponents of Optimality Theory (henceforth OT; Prince and Smolensky 1993) and proponents of Chomsky and Halle's (1968) original theory of generative phonology (henceforth *SPE*, after the title of that landmark text).

In the broadest possible terms, the argument in favor of OT comes from blocking and the argument in favor of *SPE* comes from opacity. (Though there are significant wrinkles in both arguments; see Chapter 7 for specific discussion.) The latest and perhaps most comprehensive rehearsal of the argument from opacity

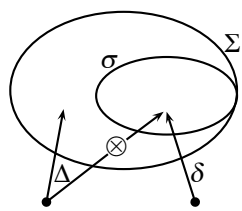
can be found in this series (McCarthy 2007b); my overall aim in this book is to provide a comparably comprehensive treatment of blocking.

1.1 Disjunctive application

I focus on a particular type of blocking that I refer to in this book as DISJUNCTIVE APPLICATION, for three reasons. First, because disjunctive application was the first type of blocking identified in generative phonology and it has thus received recurring special attention in the literature. Second, because key phenomena in other domains of linguistics (morphology and morphosyntax in particular) have been argued to be best described in terms that are virtually identical to those that define disjunctive application; phonologists and nonphonologists alike may thus benefit from the focused attention that this topic receives here. And third, because some serious work advocating *SPE* over OT has somewhat mysteriously invoked disjunctive application as a key argument (Halle 1995; Halle and Idsardi 1997, 1998); my hope is that this book will put such arguments to rest, regardless of the ultimate verdict on which of *SPE* and OT is the superior theory.

Briefly, schematically, and mercilessly glossing over the many and varied details that are investigated in depth in later chapters, disjunctive application refers to the following type of situation. Suppose there are two mappings, δ and Δ, such that their individually predicted outputs are *incompatible* in some specifiable set of contexts σ. Suppose further that σ is the only set of contexts in which δ is applicable, while Δ is applicable in a larger set of contexts Σ that *properly includes* σ. Under these conditions, only δ actually applies in σ; Δ is blocked from applying in σ, and thus only applies in the complementary set of contexts σ' in Σ. This is illustrated in (1.1) below, where arrows indicate mapping applicability and '$\otimes$' indicates blocking of an otherwise applicable mapping.

(1.1) The complementarity of disjunctive application



COMPLEMENTARITY is thus at the heart of disjunctive application. The same is of course true of the one kind of phonological analysis that is all but guaranteed to be covered in an introductory linguistics textbook, the COMPLEMENTARY DISTRIBUTION of allophones in basic phonemic analysis. Complementary distribution problems invite the same descriptive language — one allophone occurs in one specific context, another occurs elsewhere — but, curiously, this is as far as the connection with disjunctive application usually goes. Complementarity is widely held to be an irreducible part of the statement of a principle that is responsible for the imposition of disjunctive application as a special mode of interaction

between mappings while at the same time being equally widely held to be *derived* from analytical devices particular to complementary distribution.

A central argument of this book is that the complementarity observed in cases of disjunctive application is an epiphenomenon of the formal relationship that happens to hold between the basic analytical elements responsible for the interacting mappings defined as δ and Δ are above; complementarity is thus just as much a derived property of disjunctive application as it is of complementary distribution. An important consequence of this argument is that disjunctive application does not deserve the special attention that it has received in the literature, though I of course hope that the current book is the exception that proves this rule. Disjunctive application is nothing more than a subspecies of a more inclusive type of blocking, following directly from the most basic assumptions of OT and highlighting the inherent limitations of the most basic assumptions of *SPE*.

1.2 Basic assumptions

So what *are* these 'most basic assumptions' of OT and *SPE*? In OT, they are CONSTRAINT RANKING and CANDIDATE COMPARISON. When two candidates being compared bring two constraints into conflict, the higher-ranked constraint $\mathbb{C}_1$ trumps the lower-ranked constraint $\mathbb{C}_2$ such that the winner is the candidate that fares better on $\mathbb{C}_1$. If this winner wins all candidate comparisons, and if $\mathbb{C}_2$ defines the structural description of an otherwise motivated mapping, then that mapping is effectively blocked in this scenario. Ranking thus settles the conflict between the constraints that define the inputs to the mappings δ and Δ in the abstract situation described above: if $[\![\mathbb{C}_\delta \gg \mathbb{C}_\Delta]\!]$, then δ will prevail in σ and Δ will carry the day in σ' (that is, elsewhere in Σ); if on the other hand $[\![\mathbb{C}_\Delta \gg \mathbb{C}_\delta]\!]$, then Δ will hold sway everywhere in Σ, as if δ didn't even exist.

In the case of *SPE*, the most basic assumptions are RULE ORDERING and SERIAL DERIVATION. Whether a rule applies is determined strictly by whether its structural description is met by the representation delivered by the immediately preceding rule in the ordering. This method is referred to as CONJUNCTIVE APPLICATION, a term of art that appropriately highlights the contrast with disjunctive application. It is not that Chomsky and Halle (1968) did not countenance disjunctive application; indeed, it is this work and Chomsky (1967) that first brought the significance of disjunctive application to the attention of phonologists and other linguists. However, the proposals made for determining disjunctive application in these works involved special interpretive conventions on rule abbreviation notations that were specifically designed to *circumvent* what would otherwise be conjunctive application of rules (see §2.2 for details).

1.3 The Elsewhere Condition

Interpretive conventions on rule notations were eventually abandoned in favor of a more general principle referred to as the ELSEWHERE CONDITION (henceforth the EC) by Kiparsky (1973b), adopted in some form in much subsequent

work in phonology as well as in other domains of linguistics. Individual definitions of the EC vary in their details, but the core of the principle lies in a required formal relationship between two rules, stated in two parts. Two rules that are formally related by these twin criteria apply disjunctively, with the properly included (= 'more specific') rule blocking the properly including (= 'more general') rule.

(1.2) The EC's core criteria for disjunctive application

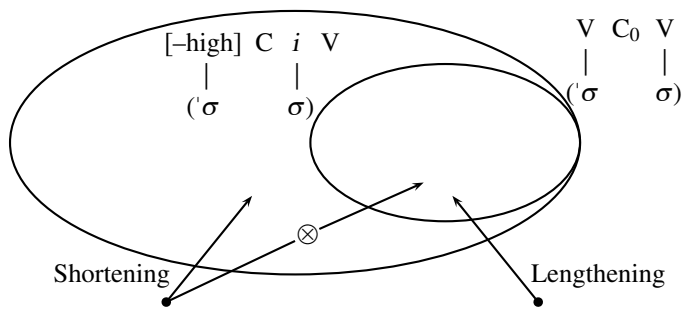
a. **Proper inclusion of contexts.** The contexts of applicability of one rule are properly included in those of the other.

b. **Incompatibility of changes.** The structural changes of the rules are incompatible with each other.

These core criteria do not necessarily originate with Kiparsky (1973b), but for a variety of reasons this one article seems to have had more influence than any other work — prior, contemporaneous, or since — on subsequent research pertaining to the topic of disjunctive application. Indeed, Kiparsky (1973b) is regarded by some as being "among the most important contributions to phonology" (Halle and Idsardi 1997: 344), and this significance appears also to be appreciated by some of our colleagues in other subfields of linguistic theory. But for all its promise of greater empirical coverage than the notational conventions of Chomsky and Halle (1968), the EC does not rise above the theoretical exigency of those conventions. The EC exists purely because the normal mode of interaction between rules in *SPE*, conjunctive application, must be circumvented in certain cases.

It is useful at this point to consider an example of the kind of case about which the EC has something to say. In English, two rules are applicable to the head vowels of branching main stress feet (Kenstowicz 1994b; Halle 1995; see §3.3 for more extensive discussion): one rule aims to lengthen the head vowel of a branching foot if it is [–high] and if the nonhead of the foot is an *i* followed immediately by another vowel, and another rule aims to shorten the head vowel of all branching feet. The structural changes of these rules are clearly incompatible, and the contexts of applicability of the lengthening rule are properly included in those of the shortening rule. The more general shortening rule is thus blocked by the EC from applying just to those heads of branching feet meeting the more specific conditions of the lengthening rule, as illustrated in (1.3).

(1.3) An example of disjunctive application



But note that the same rules applying conjunctively can achieve the same result: if the more general shortening rule applies before the more specific lengthening rule, then those vowels that are 'incorrectly' shortened in contexts of lengthening are simply (re-)lengthened later (Chomsky and Halle 1968; Kiparsky 1982; Prince 1997a).[1] Proponents of the EC point to the fact that this kind of conjunctive analysis results in so-called DUKE OF YORK DERIVATIONS (Pullum 1976), whereby e.g. an underlying long vowel is shortened only to be re-lengthened, and that these types of derivations are in principle undesirable (Halle and Idsardi 1998). Other work has demonstrated that Duke of York derivations are undesirable in fact (McCarthy 1999b, 2003c; Norton 2003); see §3.3.5 for discussion.

There is undoubtedly something right about the spirit of the EC, but as I explain in Chapter 5, the letter of this principle is riddled with stipulations; disjunctive application is not so much predicted as it is depicted by the EC. This is due to the very fact that the EC must be stated in the first place: its statement is necessary because no part of it follows from the most basic assumptions of *SPE*. The EC is a completely separate, normative principle grafted on to the architectural framework of *SPE*, dictating how certain very particular rule pairs must interact: disjunctively, rather than conjunctively. The EC can as such be refined and redefined, as it has in fact been several times in the literature, to accommodate this or that particular empirical example of (what looks like) an 'elsewhere effect'.

Even the core criteria in (1.2) have been tweaked from definition to definition, as we'll see, although there appears to be an implicit understanding that these are the least negotiable of the EC's terms — but again the very fact that these terms have to be stated at all belies this understanding. For example, why shouldn't incompatible rules with merely *overlapping* contexts of applicability apply disjunctively in those shared contexts? After all, conjunctive application of some such rules can and does also lead to undesirable Duke of York derivations (McCarthy 1999b, 2003c, 2007a, b); see §5.3. The apparent key to proper inclusion is that it is *asymmetrical*, allowing a unique determination of which rule blocks the other. The symmetry of overlap makes this determination impossible on these grounds alone, and so countenancing blocking in such cases requires that the EC be complicated even more than it already is.

1.4 OT logic

As I explain in Chapters 4 and 6, the foundational logic of OT directly predicts the disjunctive application of mappings in an elsewhere relationship without the need for a separate principle like the EC and its attendant problems of formulation. The various details of an empirically adequate formulation of the EC, each of which must be stipulated within *SPE*, fall out as necessary consequences of the same OT logic responsible for all mappings and their interactions, including other types of blocking. We can lead up to the more detailed discussion of this OT logic in later chapters with the following. Paraphrasing Grimshaw's (1997) compact statement, the relative harmony of any two competing candidates is measured by the

1. There are other cases for which a conjunctive analysis is not so readily available; see §2.4.

highest-ranked constraint on which those candidates conflict. Now consider two constraints $\mathbb{C}_1$ and $\mathbb{C}_2$ that conflict over the choice between an optimal `winner` candidate and a suboptimal `loser` candidate from some `input`, and suppose that $\mathbb{C}_1$ is the highest-ranked constraint on which `winner` and `loser` conflict. $\mathbb{C}_1$ must prefer `winner` and $\mathbb{C}_2$ must prefer `loser` in order to correctly resolve the conflict between the constraints' assessments of the candidates.

(1.4) Conflict resolution in OT

`input`	$\mathbb{C}_1$	$\mathbb{C}_2$
☞ `winner`		*
`loser`	* !	

All mappings and their interactions are defined by this one mechanism in OT, the particular differences among them being due to different logical relationships that arise naturally from different definitions of $\mathbb{C}_1$ and $\mathbb{C}_2$ and their expected interactions with other constraints. All that is peculiar about disjunctive application is that the constraints $\mathbb{C}_\delta$ and $\mathbb{C}_\Delta$ conflict only over a proper subset of the comparisons in which $\mathbb{C}_\Delta$ prefers `loser`; in other words — and simplifying somewhat — $\mathbb{C}_\Delta$ is violated whenever $\mathbb{C}_\delta$ is satisfied, but not vice versa. Such situations satisfy the twin criteria in (1.2). Given this logical relationship between the constraints, the independent coexistence of two mappings defined by $\mathbb{C}_\delta$ and $\mathbb{C}_\Delta$ in the same grammar requires the ranking $[\![\mathbb{C}_\delta \gg \mathbb{C}_\Delta]\!]$, and the disjunctivity of their application is entailed; the opposite ranking $[\![\mathbb{C}_\Delta \gg \mathbb{C}_\delta]\!]$, on the other hand, renders the mapping defined by $\mathbb{C}_\delta$ completely inoperative.

This last point follows from the result that Prince and Smolensky (1993) call PĀṆINI'S THEOREM ON CONSTRAINT-RANKING (henceforth PTC), which is often mischaracterized in the literature as the OT counterpart of the EC. There is a very important difference between the EC and PTC, however. PTC is a *theorem*, a provable proposition following from the most basic assumptions of OT. Unlike the EC, PTC is *not* a separate, normative principle grafted on to the independent architectural framework of OT, dictating how particular constraint pairs must interact; PTC instead descriptively identifies certain conditions under which the more specific of two constraints is expected to be INACTIVE — unable to participate in the selection of an optimal candidate from some input.[2]

But Prince and Smolensky's (1993) proof of PTC is incomplete, both in a way that Prince and Smolensky themselves note and in a way that emerges particularly clearly from the present study. The incomplete proof of PTC is nevertheless used in Chapter 6 as a necessary jumping off point for a more complete proof of another theorem that identifies the relevant set of conditions guaranteeing the inactivity of *the more specific of two constraint rankings that correspond to mappings* and from which elsewhere interactions between mappings meeting the criteria for disjunctive application in (1.2) follow. To distinguish it from PTC, I call this theorem the ELSEWHERE THEOREM ON CONSTRAINT-RANKING (ETC).

2. This point, made clearly enough by Prince and Smolensky (1993), has been lost on many linguists, perhaps due to a general unfamiliarity with the distinction between theorems and principles.

Much of the argument in this book thus amplifies and refines the following statement by Prince and Smolensky (1993), made in their brief but significant discussion of the "obvious affinities" between the EC and PTC.

> There is an important difference: PTC is merely a point of logic, but the [EC] is thought of as a principle specific to UG, responsible for empirical results which could very well be otherwise. ... [The EC] folds together a point of logic (PTC) with additional claims about what linguistic phenomena are incompatible. With the incompatibility claims properly factored out into substantive constraints of various types, what's left is PTC; that is to say, nothing. (Prince and Smolensky 1993: 119–120)

The difference is that the EC *must be stated* in order to guarantee blocking of a more general rule by a more specific incompatible rule in *SPE*; blocking of a more general constraint by a more specific conflicting constraint is *already guaranteed* by the logic of OT, and so PTC and the ETC needn't be stated at all.

1.5 Organization

The remainder of the book is organized as follows.

Chapter 2: A brief history of blocking, where I lay the groundwork for understanding — and for justifying — the provenance and privileged place of disjunctive application and other types of blocking in phonological theory.

Chapter 3: Elsewhere in *SPE*, where I distinguish two types of elsewhere interactions, named UNBOUNDED COMPLEMENTARY DISTRIBUTION (UCD) and BOUNDED COMPLEMENTARY DISTRIBUTION (BCD). I illustrate the ways in which both types can be analyzed in *SPE*, and I demonstrate that the need for disjunctive application analysis in *SPE* is restricted to a particular class of cases of BCD, which I refer to thereafter as DISJUNCTIVE BCD.

Chapter 4: Elsewhere in OT, where I present the OT analysis of UCD and disjunctive BCD and demonstrate how their basic sameness is captured in OT by the use of the same fundamental analytical elements.

Chapter 5: The Elsewhere Condition, where I discuss in detail various stipulations that have been and must be made in the statement of the EC in order to properly describe purported examples of disjunctive application.

Chapter 6: The Elsewhere guarantee, where I take on the task of demonstrating how elsewhere interactions follow from the most basic assumptions of OT, building on Prince and Smolensky's (1993) incomplete proof of PTC.

Chapter 7: Conclusion, where I conclude with summary remarks as well as discussions of other uses of disjunctive application in linguistic theory, other instances of apparently elsewhere-like interactions, and the weaknesses of the opacity argument for *SPE* and of the blocking argument for OT.