LEARNING PHONOLOGY AS A WAY TO LEARN HOW THEORIES ARE IMPROVED

Andrew Carstairs-McCarthy: *Department of Linguistics, University of Canterbury, Private Bag 4800, Christchurch, New Zealand.*
<andrew.carstairs-mccarthy@canterbury.ac.nz>

1. Recent history of phonology

In some disciplines, theoretical innovations come thick and fast, even at the level of undergraduate study; in others, background assumptions remain pretty much unchanged for decades. Also, in some disciplines, a particular analytical framework can be identified, reasonably uncontroversially, as the 'mainstream' framework at any one time; in others this is not so. Many of the difficulties and the opportunities presented by teaching any discipline are affected by its settings for these two parameters (so to speak). Phonology, broadly speaking, combines rapid theoretical innovation with an identifiable mainstream.

In the early 1970s, the mainstream framework was that of Chomsky and Halle's *The Sound Pattern of English* (SPE) (1968). The subtitle of Anderson's *Phonology in the Twentieth Century* (1985) draws attention to the contrast between 'rules' and 'representations'. In the SPE framework, rules were in the ascendant. Phonological representations were relatively uninteresting. They were basically linear strings of segments and boundaries, the segments consisting of unordered matrices of binary distinctive features, each with its value (plus or minus). But, from the late 1970s on, analytical emphasis shifted in the direction of more sophisticated structuring of representations, both syntagmatic and paradigmatic.

On the syntagmatic dimension, the syllable reasserted itself, in recognition of the inadequacies of the binary feature [±syllabic]. The syllable came to be established as a unit in a hierarchy whose higher-level units were the foot and the (phonological) word. Independently, the foot and the phonological word were deriving support from new approaches to prominence and metrical structure, inspired in part by inadequacies of the binary feature [±stress]. Meanwhile, on the paradigmatic dimension, 'feature geometry' revived insights of Firthian phonology (Palmer 1970), allowing one-to-many and many-to-one links between feature values on distinct autosegmental tiers, and between these tiers and the syntagmatic units of phonology, particularly the individual segment. At the same time, the distinction between 'feature' and 'feature value' became blurred, with a new readiness to recognise nonbinary features: thus, instead of a binary feature [±coronal] many phonologists now recognise a single-valued feature [coronal] in a mutually exclusive relationship with [dorsal] and [labial] and perhaps [guttural], all linked to one node for '(consonantal) place' in feature geometry's three-dimensional 'bottle-brush' structure — not hard to visualise but difficult to represent on the printed page.

The next big innovation involved the relationship between inputs and outputs, or between underlying and surface representations. The terms 'input' and 'output' conjure up an image of a factory production line, with raw materials entering at one end, being subjected to a succession of manufacturing operations ('ordered rules'), and emerging at the other end as a finished product (a pronounceable 'surface representation'). When Optimality Theory came on the scene in the mid-1990s, the production line metaphor gave place to a metaphor of a job interview panel: the output representation is not necessarily ideal, but it is 'optimal' in terms of the priorities attached to the various criteria that the applicants are asked to fulfil (or in terms of the 'ranking' of the 'constraints' governing the 'evaluation' of the 'candidates').

The popularity of Optimality Theory does not mean that students no longer need to know about earlier approaches. In particular, the concerns that gave rise to the recognition of autosegmental tiers are as valid as they ever were. The candidates in an OT tableau are typically represented not in feature-geometric diagrams but in traditional symbols, perhaps with indication of syllable boundaries and stress — but that is largely because of the typographic difficulty of fitting anything more elaborate into the space available. And SPE-style linear rule notation has not disappeared, because it is such a convenient device for representing phonological processes informally.

Alongside these changes, phonological theory has remained in one respect

stable over the last thirty years. Is the allomorphy displayed by the stems of *divine* and *divinity*, or of *leave* and *left*, something that phonology should try to account for, or should it be regarded as a purely morphological phenomenon? Notoriously, *The Sound Pattern of English* treated as the business of phonology all morphophonology except gross suppletion. Most phonologists have retreated from that position now, but not in an orderly fashion. There is still an expectation that phonology should take care of as much morphophonology as possible, but determining the point at which phonology stops remains in the 'too hard' basket. A well-known alternation such as that between [t], [d] and [əd] in regular English past tense forms is typically regarded as the business of phonology, because the allomorphs are phonologically similar, the choice of between them is phonologically determined, and (assuming underlying /d/) the distribution [t] and [əd] can be seen as phonotactically motivated. But what about the [t] of *kept*, *bent* and *bought*? Is it the same as the [t] (spelled -*ed*) of *seeped* and *chased*? If it is, what should we say about the differences in the past tense forms of *bend* (*bent*) and *mend* (*mended*)? If not, must we classify -*t* and the [t] allomorph of -*ed* as purely accidental homonyms? And what about the 'zero past' of some verbs that end in coronals, such as *spread*? Should these be treated as underlyingly suffixed, e.g. /spred + d/ or /spred + t/? Theorists and textbook writers alike have generally been silent on such questions in recent years.

There is thus much that students need to know about. The morphophonological issues that I mentioned last are the most difficult to handle satisfactorily at the undergraduate level. One way out for the phonology instructor is to relegate them to a separate course on morphology. For the purposes of this article, I will assume that that is what is done. So far as the other issues are concerned, the challenge of covering the ground can be turned into an opportunity to teach important general lessons about scientific method, as I will illustrate.

2. Characteristics of good theories and the nature of theoretical improvement

In any discipline, students learn about both content and method. In history, for example, students learn about what happened when, and why; they also learn about how to interpret the available evidence (documents of all kinds, oral traditions, archaeological remains, and so on). This is equally true in linguistics. An important difference between linguistics and history, however,

is that the methodology of many areas of linguistics is more akin to that of so-called 'bench sciences' such as physics and chemistry. That is, it is possible in linguistics to conduct replicable experiments. An analysis of a problematic grammatical or phonological phenomenon will often entail predictions about grammatical or phonotactic acceptability that can be tested against the reactions of native speakers. A good analysis, as in physics or chemistry, will be one in which these predictions are generally confirmed, and a particularly good analysis will be one that leads to correct predictions that go beyond the original problem.

A good theory, however, is not just one that leads to valid predictions. A theory may be so accommodating that it is compatible with not merely everything that is observed but with everything that might conceivably be observed. Karl Popper long ago pointed out that such 'theories' are not really theories at all, because they are empirically empty. A good theory is not merely compatible with as much as possible of what is observed; it is also incompatible with as much as possible of what is not observed. Theoretical improvement is often not a matter of accommodating problematic observations about things that do happen, but rather of supplying new principled reasons for why many things don't happen. One does not have to be a 'naive falsificationist' to agree with Popper on this. Even if the relationship between linguistic theory and the observed phenomena of speech is as indirect as is claimed by Chomskyan linguists, they too agree that a good theory should exclude as much as possible of what does not happen.

In phonology, there are excellent opportunities to illustrate precisely this. I will present three case studies. All of them are based on experience in the phonology section of Canterbury University's 300-level *Theory of Linguistics II* course, which I have taught regularly since 1982.

3. Case study 1: Linear rules, ranked constraints, and English /z ~ s ~ əz/

The English pattern of allomorphy exhibited by the regular noun plural and verbal 3rd singular suffix -*s* can easily be described by invoking two SPE-style linear rules: a rule of schwa-epenthesis between sibilants, and a voicing assimilation rule applying to clusters of obstruents. This is how the situation is described by Spencer in his textbook *Phonology* (1996: 49–55). Both of these rules seem satisfactorily 'natural' (whatever precisely that means). Yet in order

to achieve the right outcome, we need to make sure that schwa-epenthesis applies before voicing assimilation. In terms of *SPE*-style phonology, is there a deep reason why these rules are so ordered, or is it just an accidental fact about English?

It is hard to think of any deep reason in SPE terms. Theories of rule-ordering developed in the 1970s distinguished feeding and bleeding orders; but that distinction does not help here. It was suggested then that rules should tend to apply in an order that would maximize feeding and minimize bleeding. The underlying idea seems to have been that speakers (or rather the brains of speakers) will favour a grammar in which every rule is exploited as much as possible. Yet in this English instance, to get the proper plural form for a noun such as *place* ([pleisəz], not *[pleisəs]), schwa-epenthesis must precede and thus bleed voicing assimilation. Could it be, then, that it is just an accident that no dialect of English exists in which the two rules apply in counter-bleeding order, so that the plural of *place* is [pleisəs]?

Native English speakers in my phonology classes tend to agree that a hypothetical English dialect with plural [pleisəs], and in which the plural of *piece* rhymes with *thesis*, seems unlikely. They readily agree that a version of phonological theory that would exclude such a dialect in principle would be an improvement. I therefore conclude my discussion of SPE-style phonological rules by assuring them that such a theory does indeed exist, and that they will encounter it in a few weeks' time, when I introduce Optimality Theory (OT).

In discussing how theoretical preoccupations change, I do not gloss over the influence of mere fashion. I point out that it is probably not an accident that Optimality Theory began to take off at just the time when word-processing applications acquired the ability to manipulate tables more efficiently, with cells that can be shaded and within which text can be centred or left-justified. Constructing OT tableaux in the 1990s was a novel and enjoyable typographic challenge. This usually elicits a wry laugh. But twenty-first-century students too find OT tableaux fun to use.

One of my first illustrations of OT involves the /z ~ s ~ əz/ and /d ~ t ~ əd/ alternations. Let us suppose that in English a version of the Obligatory Contour Principle ranks high, enforcing 'schwa-insertion' between coda obstruents that are identical in their supralarygeal feature values. Another high-ranked (indeed, never violated) constraint imposes voicing agreement on tautosyllabic obstruent clusters. To be precise, both these constraints outrank two 'Faithfulness' constraints: DEP-IO and IDENT-IO[voice]. Students find it easy to see that, if the input for the suffixes is /z/ and /d/ respectively, no

manipulation of the ranking of these four constraints can ever yield [pleisəs] as optimal for the plural of *place*. Likewise, no manipulation can yield [weitət] as optimal for the past tense of *wait*. The question left hanging by the SPE-style analysis ('Could there ever be a dialect of English in which the plural of *piece* rhymes with *thesis*?') now receives a clear answer: 'No.'

It may seem that a question is begged by assuming /z/ and /d/ as inputs. Surface allomorphs [əs] and [ət] would indeed be possible in a dialect where the inputs for the suffixes were /s/ and /t/ rather /z/ and /d/. But such an analysis would render it hard to account for the choice of the allomorphs /z/ and /d/ after a sonorant other than the 'inserted' schwa, as in *boys*, *bells, played* and *called*. After all, the realization of an underlying /s/ and /t/ by [s] and [t], so as to yield e.g. [bɔis] as the plural of *boy* or [pleit] as the past tense of *play*, would be favoured not merely by Faithfulness but also by the unmarked status of voiceless obstruents generally, captured in OT by a constraint such as 'No Voiced Obstruents' or 'OBSTR[-voice]'.

The claim that OBSTR[-voice] applies in English may well provoke a protest from some students. It is good that it should do so, because answering that protest supplies another pedagogical opportunity. 'Surely English just doesn't have any 'No Voiced Obstruents' constraint', the protest runs, 'because English has voiced obstruents all over the place, in codas as well as in onsets. It's silly to suggest that this constraint just happens to be ranked low in English!' A more sophisticated complainer may take a Popperian line: 'This suggestion is not just silly but cheating, because the possibility of low ranking renders it impossible in principle to show that any constraint is absent from the grammar of any language!'

The instructor may perhaps sympathize with this complaint. But the instructor is also, for the time being, a spokesperson for John J. McCarthy on behalf of OT. So would McCarthy regard this complaint as fatal? No. He would argue that even low-ranked constraints can turn out to play an active part in the phonology of a language through the phenomenon known as 'the emergence of the unmarked' (McCarthy and Prince 2004). And English supplies neat illustrations of this with respect to OBSTR[-voice]. In most varieties of English, nasal-plosive clusters in coda include [mp, ŋk], where the plosive is voiceless, but not *[mb, ŋg]. Moreover, monomorphemic items can contain voiceless obstruent clusters in coda such as [pt, kt, ps, ks] (*apt, act, lapse, box*), but not voiced ones (*[bd, gd, bz, gz]). In clusters, at least, voiced obstruents are disfavoured.

The determined objector may go on to say: 'But what about [nd] as a coda

cluster, as in *land, mend, sound* and many other examples? And what about the voiced obstruent clusters at the end of *robbed, dragged, seethed, grubs, dogs, wreaths, loaves* and so on?' But again, the instructor can turn these examples to good account.

The availability of [nd] as a coda reflects the relatively unmarked status of coronal consonants vis-à-vis those at other places of articulation. This shows up also in two further facts that native English-speaking students can easily discover for themselves, given one or two hints. Heavy rhymes consisting of a diphthong or long vowel followed by an intamorphemic consonant cluster are possible in English, but only if the consonants in the cluster are coronal. Thus we have *wild, field, pint, paint, pound, pounce, strange* and many other such words, but no similar examples with dorsal or labial consonants. Also, the diphthongs [au] and [ɔi] as nuclei tolerate coronals in an accompanying coda, as in *loud, lout, louse, coin, quoit, owl, oil*, but not labials or dorsals (except in phonologically aberrant proper names such as *Doig*). In OT terms, this suggests that the NoCODA constraint made famous by Prince and Smolensky's pioneering analysis of Tagalog (2004: 40-45, originally published in 1993) must be treated as a family of constraints, with the unmarked status of coronals reflected in a universal ranking of NoCODA[labial] and NoCODA[dorsal] above NoCODA[coronal].

The voiced obstruent clusters in *robbed, grubs* and so on contain a morpheme boundary. They can thus be used as an opportunity to discuss how OT treats the tug-of-war in morphology and phonology between Faithfulness and Markedness. In these regularly inflected forms, Faithfulness outranks Markedness, thus yielding optimal forms that contain phonologically awkward combinations. However, in irregular inflection, phonological Markedness constraints are generally not violated: contrast *pens*, the plural of *pen*, with *pence*, the irregular plural of *penny*, and contrast the past tense forms of *ooze* and *lose*: *oozed* has both a 'long' nucleus and a voiced obstruent coda, a combination that is impossible in a syllable with no morpheme boundary, whereas *lost* has the phonological shape of a monomorphemic items such as *cost*.

There are admittedly loose ends here. Precisely how are the relevant constraints to be formulated and ranked so as to take care of all these observations? And are we to assume that Faithfulness and Markedness constraints are ranked differently for different sections of the vocabulary, thus sacrificing the analytic discipline imposed by the doctrine that any one language variety observes a single uniform constraint ranking? However,

these are encouraging rather than discouraging loose ends. They merely confirm what sensible students already know: that not all problems are yet solved. A few students may indeed be motivated to go on and try to solve them later in their careers. In the light of all this, students are happy to agree that, so far as those extremely familiar English suffixal alternations are concerned, Optimality Theory supplies not merely a new way of analyzing the phenomena but a way that is an improvement on earlier analyses.

4. Case study 2: Syllable margins, extrametricality and stress in Tibetan, Attic Greek and Latin

A common error among beginners in phonology is the assumption that the form in which a stem appears when it is on its own must be a better guide to its underlying representation than the forms in which it appears when it is affixed or part of a compound. Every instructor in phonology encounters this assumption. Standard examples used to dislodge it include German lexemes such as *Bund* 'bundle' and *bunt* 'many-coloured', both realizable as [bʊnt] but shown to be phonologically distinct by their contrasting plural forms [bʊndə] and [bʊntə].

   Halle and Clements' *Problem Book in Phonology* (1983:105) provides a more exotic illustration that yields a satisfying 'Aha!' experience as soon as one cottons on to what is happening. In Tibetan, numerals from eleven to nineteen (or, at least, such of them as are presented by Halle and Clements) take the form 'ten-one', 'ten-two' and so on, while multiples of ten take the form 'two-ten', 'three-ten' and so on. Forming these higher numerals in Tibetan thus ought to be a simple matter of compounding the numerals corresponding to 'one' through 'ten'. Students are puzzled to find, however, that some of these compounds contain unexpected medial consonants not found in either of the two elements of the compound when they are on their own. Where do these extra consonants come from? Formulating rules to insert them turns out to be troublesome, and the rules themselves look contrived and unnatural. But what if these consonants are present in underlying representations — specifically, in the underlying representation of the second element of the compounds where they appear? The problem now becomes one of explaining not why they get inserted but why they sometimes get deleted.

   The answer to this problem is simple: the first consonant in a word-initial cluster gets deleted. In SPE-style linear notation, this is easy to represent:

(1) C → Ø / #__C

In terms of syllable structure, the phenomenon is equally easy to describe: Tibetan does not tolerate complex onsets. The reason why two-consonant clusters are permitted medially is that the first of the two can be slotted into the coda of the preceding syllable. But a question now arises that is somewhat similar to our earlier question about a hypothetical variety of English in which the plural of *piece* rhymes with *thesis*. There is in principle another way in which an initial consonant cluster could be simplified, namely by deletion of the second consonant. In SPE notation, this is equally easy to express:

(2) C → Ø / #C__

So could there be a dialect of Tibetan in which rule (2) applies rather than rule (1)?

Again, invocation of post-SPE developments suggests that the answer is no. In the notation of (1) and (2), 'C' is an abbreviation for '[-syllabic]'. It therefore has no part in the newer style of representation that exploits the syllable as a structural unit. In this newer framework, let us make the usual assumption that syllable nuclei are sonority peaks, and that non-peak segments are assigned to syllables on the basis of two default assumptions: (a) onsets are maximized; (b) onset maximization is subject to sonority sequencing in accordance with the sonority hierarchy (vowels, glides, liquids, nasals, fricatives, plosives). (I call these 'default assumptions' rather than 'universals' for reasons that I will explain later.) On this basis, syllabification is a matter of attaching material to sonority peaks, working outwards from syllable nuclei towards syllable margins. Any segment that is unsyllabifiable (that is, that cannot be attached to a syllable without violating the language's syllabification constraints) will be a segment that is a remote from rather than close to a sonority peak. It follows that, if one of two consonants in word-initial position is left unsyllabified, it will be the first, not the second. Actual Tibetan thus behaves as expected, and the hypothetical Tibetan with rule (2) is revealed as impossible (or at least, extremely unlikely).

There are likely to be some students who question the validity of sonority sequencing as a default assumption. If there aren't, instructors themselves should take the initiative in questioning it! Sonority sequencing is violated by forms such as the Russian monosyllables *mgla* 'gloom' and *rta* 'mouth (genitive singular)', and by numerous examples in cluster-friendly languages

such as Polish and Georgian. It is also violated by an English monosyllable such as *sprints,* where the initial and final [s] constitute extra sonority peaks. But the instructor can again turn these complaints to good account. In English it may not be an accident that the extra consonants that violate sonority sequencing are always coronal, and that they are often inflectional affixes too. Besides, need it be the case that all unsyllabified consonants are 'deleted'? In Tibetan, it seems, if a consonant is not in either an onset or a coda, it cannot be pronounced. However, Spencer's (1996) analysis of the final [ts] of *convicts* suggests that not all languages are so ruthless. For English, Spencer favours a kind of half-way house: a consonant can belong to a syllable as an 'appendix' at the end without being part of the coda. This raises the possibility of having two layers of syllable structure: the core syllable (consisting of onset and rhyme) and the extended syllable (consisting of onset, rhyme and appendices). Are there any other pedagogically useful phenomena that support the notion of the extended syllable — that is, phenomena which show that is more than just a device to protect the sonority sequencing generalization? The answer is yes. Kenstowicz (1994: 264-9) cites an intriguing analysis by Steriade (1990) of certain morphological-cum-phonological data in classical Attic Greek.

In Attic Greek verbs, one can distinguish a root and a 'perfect' stem, which is derived from it in various ways. What 'perfect' means in syntactic terms does not matter for present purposes. What matters is that, from roots beginning in a consonant, some perfect stems are formed with a prefixed *e-* (the so-called 'augment') and some with reduplication. At (3) are examples of both:

(3) *Reduplicated perfects*   *'Augmented' perfects*

luː	le-luk-a	'untie'	sper	e-spar-mai	'sow'
sɛːman	se-sɛːmɛn-a	'signify'	zdeug	e-zdeug-mai	'yoke'
klepʰ	ke-klopʰ-a	'steal'	kten	e-kton-a	'kill'
tlaː	te-tla-men	'endure'	psau	e-psauk-a	'touch'
knai	ke-knɛːs-mai	'scrape'	smukʰ	e-smug-mai	'smoulder'
pneu	pe-pneuk-a	'breathe'			
grapʰ	ge-grapʰ-a	'write'			

What determines whether a verb takes reduplication or the augment? Steriade argues that the crucial factor is sonority distance. In the examples with reduplication, the root begins either with a single consonant or with a cluster in which there is a relatively large sonority gap between the first and second consonants. On the other hand, in the examples with the augment, the root

begins with a consonant cluster within which the sonority gap is small (e.g. [ps], [sm]) or nonexistent ([kt]), or in which sonority sequencing is violated ([sp]).

This is where the distinction between core syllabification and extended syllabification comes in. Let us assume that, for Greek, core syllabification requires in onset clusters a sonority gap at least as big as that between plosives and nasals. On that basis, reduplication occurs when core syllabification is able to assign the whole of any root-initial consonant cluster to a syllable onset. But if one consonant of a root-initial cluster is not syllabifiable in this way, then the perfect stem is formed with augment rather than with reduplication.

Let us consider the root *kten*- 'kill'. (In the perfect stem, the vowel is ablauted to -*o*-, which need not concern us.) Core syllabification leaves [k] out in the cold. With the addition of the augment *e*-, [k] becomes syllabifiable as a coda consonant: [ek.to.na]. What happens, however, when the stem of this verb is not preceded by an affix of any kind? This does indeed occur, for example in the present tense. No coda slot is available for the [k] here. So does Greek, like Tibetan, ruthlessly refuse to let the initial [k] be pronounced? It turns out that Greek is more tolerant than Tibetan. In forms such as [kteinɔː] 'I kill', [k] is permitted to remain — just as, in English, segments can be pronounced even though they lie outside the core syllable (in Spencer's analysis). We may reasonably regard [ktei] here as an extended syllable, with [k] belonging to the syllable but not to its onset. Students can thus be led to see that the idea that lies behind syllable 'appendices' has an application outside English.

I mentioned earlier that the syllable as a syntagmatic phonological unit came to be seen from the late 1970s as just one unit within a prosodic hierarchy, below the foot and the phonological word. So do we find at higher levels phenomena comparable to the syllable 'appendix'? Our theory leads us to expect that peripheral units at other levels too should be able to behave as if they are somehow outside the core structure. Therefore it is satisfying to be able to present to students data that support this.

Within metrical theory, there developed at an early stage a doctrine of 'extrametricality', according to which syllables that are endmost in the phonological word may be deemed extrametrical from the point of view of foot formation. A nice illustration of this is supplied by the Latin stress rule: 'stress the penultimate syllable if it is heavy, otherwise the antepenultimate'. Here is an SPE-style formulation:

(4) $V \rightarrow$ [+stress] / _____ C_0 ([V, –long] C^1) V C_0 #

But could there be a pseudo-Latin with the following stress rule instead?

(5) $V \rightarrow [+\text{stress}] / \underline{\quad} C_0 ([V, +\text{long}] C^1) V C_0 \#$

In this pseudo-Latin, what the parentheses require us to skip is precisely a certain kind of heavy syllable, not a light one. That seems highly improbable. Yet in SPE terms the two processes are equally easy to state. So students are easily persuaded that the SPE-style formulation is unsatisfactory.

   Happily, a better way of describing things emerges once we recognize the foot as a metrical unit — provided that we permit a word-final syllable to be 'extrametrical', i.e. outside any foot, just as a syllable-final consonant can be outside the syllable's coda. If we treat the final syllable in Latin words as extrametrical, then Latin word stress exactly fits of the following metrical parameter settings (Spencer 1996: 250-251):

   (i)   words are right-headed
   (ii) foot-formation starts at the right edge of the word
   (iii) foot-formation is quantity-sensitive (i.e. every heavy syllable is the head of its foot)
   (iv) feet are left-headed
   (v)   feet are maximally binary

Latin stress behaviour thus turns out to be not one of an almost infinite range of conceivable patterns, as the SPE-style rule implies. Instead, it is constrained along a number of dimensions. What's more, if the freedom implied by the SPE formulation were genuine, it would be hugely surprising if some language unrelated to Latin were to observe precisely the same rule. On the other hand, if it is a matter of choosing settings for a small list of parameters, such a coincidence would be less remarkable. And it turns out that Bedouin Hejazi Arabic behaves just the same as Latin (or at least, that fragment of Bedouin Hejazi Arabic that is presented by Spencer as an exercise [1996: 264]).

   This chain of examples has led from a problem about determining underlying representations in Tibetan to the issue of how to describe stress in Latin. The route was determined by the search for theoretical improvements, at a level readily graspable by undergraduates. The search has been broadly successful. Admittedly, there are loose ends. For example, is it really the case that settings for parameters (i)–(v) are independent? But these are (once again) loose ends of an encouraging kind. They are consistent with a view of

phonological theory as illustrating how analyses that merely 'work' can be replaced by analyses that, at least in some degree, explain.

5. Case study 3: The distribution of final /e/ on Spanish nouns

A substantial minority of Spanish nouns end in *-e* or a consonant in the singular (e.g., *nube* 'cloud', *mes* 'month'). All of these end in *-es* in the plural (*meses, nubes*). Let us call these Type A and Type B respectively. An obvious question arises: is membership of these patterns arbitrary (that is, lexically determined) or is it predictable? That is, are Types A and B really variants of a single pattern? This case study shows how students can be led to see that the distribution of nouns between the two patterns is predictable (with very few exceptions), but that stating the relevant generalizations poses frustrating problems in a rule-based framework — problems that disappear once one looks at it in terms of ranked constraints.

Here are some relevant data:

(6) *Type A*

Singular	Plural		*Type B* Singular	Plural	
['xefe]	['xefes]	'chief'	[reð]	['reðes]	'net'
['sjeɾpe]	['sjeɾpes]	'snake'	[mjel]	['mjeles]	'honey'
['nuβe]	['nuβes]	'cloud'	[maɾ]	['maɾes]	'sea'
['ombɾe]	['ombɾes]	'men'	['oɾðen]	['oɾðenes]	'order'
['pljeɣe]	['pljeɣes]	'fold'	[mes]	['meses]	'month'
['buke]	['bukes]	'boat'	[beθ]	['beθes]	'time'
[pa'leŋke]	[pa'leŋkes]	'palisade'			
[al'kalde]	[al'kaldes]	'mayor'			
['paɾte]	['paɾtes]	'part'			
['paðɾe]	['paðɾes]	'father'			

The search for a pattern can be conducted as a brain-storming exercise in class. Before long, the students will collectively conclude that the Type B nouns all end in the singular in a single coronal consonant, whereas in the Type A nouns what precedes the final [e] or [es] is anything but a single coronal consonant: a labial or dorsal consonant, or else a consonant cluster. Honesty impels me at this point to admit to students that this pattern is not absolutely exceptionless: for example, the noun *reloj* [re'lox] 'clock' belongs to Type B despite the fact

that it ends in a noncoronal consonant. But students are generally willing to accept that lexical exceptionality, or else failure so far to discover interfering factors of general relevance, do not vitiate the usefulness of exploring patterns such as that involving coronal articulation here.

Now the question arises of what this pattern implies for underlying representations. Is it that the final [e] is underlyingly present in both Type A and Type B (hence [ˈxefe] and [reð] reflect /xefe/ and /reðe/), so that it gets deleted in the singular of Type B? Alternatively, is this [e] absent underlyingly, so that it gets inserted everywhere that it occurs?

I enourage students to explore first the insertion hypothesis. To take care of the plurals of both types, it is easy enough to devise an SPE-style rule that splits up a word-final cluster /Cs/, especially as Spanish lacks such clusters except in a very few learned words such as *tórax* 'thorax'. But what of the singulars of Type A? The environment for insertion is a disjunction of conditions: after a consonant cluster, or a single consonant that is labial or dorsal. In SPE notation, such an environment is expressible with curly brackets. But I put to my students the widely-held view that curly brackets represent an admission of defeat: a failure to find an appropriate generalization, due to a weakness in either the particular analysis or the theory itself. They are accordingly not satisfied with this analysis. Even if one's theory treats '[-coronal]' as designating a natural class of sounds — something rejected in Spencer's (1996) framework, where CORONAL, LABIAL and DORSAL are treated as single-valued features, or as three values for a feature PLACE — , '[-coronal]' will not suffice here, because of examples such as [alˈkalde], [ˈpaɾte] and [ˈpaðɾe], where /e/ remains after a coronal cluster.

What of the deletion hypothesis, then? It is easy enough to formulate a rule that deletes word-final /e/ after a single word-final coronal consonant. But students who have become used to recognizing the syllable as a phonological unit are rightly unsatisfied with this. After the deletion, the coronal consonant in question is necessarily syllabified in a coda. The distribution of this final [e] seems therefore to be tied closely to what is permissible in Spanish codas. The data suggest a preference in Spanish for codas to consist of a single coronal consonant. (Prima facie counterexamples such as [ˈom.bɾe] and [pa.ˈleŋ. ke], assuming that they are syllabified in this way, are easily accounted for in terms of assimilation by nasals to the place of articulation of a following plosive.) But there is no way to formulate as an SPE-style rule the instruction 'Delete word-final /e/ if the result would be a well-formed coda'. Even if input strings are syllabified before /e/-deletion takes place, we are no better

off: input strings such as /xefe/ and /reðe/ will naturally be syllabified as [xe.fe] and [re.ðe], so that Types A and B are not differentiated, with both [f] and [ð] assigned to onsets.

What is the way forward, then? By this stage students need little encouragement to explore an Optimality-Theoretic solution. Let us assume that Spanish assigns a high rank to the already familiar constraint CODA[coronal], a member of the family of coda constraints whose patriarch, so to speak, is NOCODA. Let us assume in particular that this outranks the anti-insertion constraint DEP-IO, and that inputs for both Type A and Type B lack the final [e], as in the insertion hypothesis. It is easy to see now that CODA[coronal] and DEP-IO, thus ranked, will favour [xe.fe] and [reð] over *[xef] and *[re.ðe] as outputs for /xef/ and /reð/ respectively. To ensure that [paɾ.te] and [al.kal.de] beat *[paɾt] and *[al.kald] as the optimal outputs for /paɾte/ and /alkald/, something more is necessary; plausibly, perhaps, a high-ranked constraint NOCOMPLEXCODA (another member of the NOCODA family). Anyway, there is plenty of scope for students to work out as an exercise what constraints are needed, and with what ranking, to exclude various wrong but not implausible candidates for selected nouns belonging to Types A and B.

Yet again, there are loose ends. For example, why in particular should [e] be the vowel that gets inserted in Spanish, rather than some other vowel? And what about the rare noncoronal codas that do exist: not only in *reloj*, *tórax*, *ombre* [ˈom.bɾe] and *palenque* [pa.ˈleŋ.ke] but also in words such as *pacto*, *apto*, *elección* (Harris 1983: 18)? Again, however, these are loose ends that should inspire rather than discourage those students who are keen enough to explore further on their own.

6. Conclusion

Very few of the students that I have taught since 1982 have gone on to become phonologists by profession. But many of them, with a little encouragement, have become quite adept at hypothesis testing and at understanding the difference between a good theory and a less good theory. What happens in phonology, and why, seems to me less mysterious than what happens in syntax (a field that I know less about), and much less mysterious than what happens in morphology (a field that I know more about but still do not understand well). That is a subjective reaction. However, even instructors whose subjective reaction is different from mine may agree that there is unusual

scope in phonology to look at the same body of data from the point of view of a range of theoretical approaches, and thus build students' understanding of how a science progresses.

References

Anderson, Stephen R. 1985. *Phonology in the Twentieth Century: Theories of Rules and Theories of Representations*. Chicago: University of Chicago Press.

Chomsky, Noam and Morris Halle. 1968. The Sound Pattern of English. New York: Harper & Row.

Halle, Morris and G.N. Clements. 1983. *Problem Book in Phonology*. Cambridge, MA: MIT Press.

Harris, James. 1983. *Syllable Structure and Stress in Spanish*. Cambridge, MA: MIT Press.

Kenstowicz, Michael. 1994. *Phonology in Generative Grammar*. Oxford: Plackwell.

McCarthy, John J. and Alan Prince. 2004. 'The Emergence of the Unmarked.' In John J. McCarthy (ed.) *Optimality Theory in Phonology: A Reader*. Oxford: Blackwell. 483–494.

Palmer, F. R. (ed.). 1970. Prosodic Analysis. Oxford: Oxford University Press.

Prince, Alan and Paul Smolensky. 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Oxford: Blackwell. (Originally distributed in 1993 by Rutgers University Center for Cognitive Science as RuCCS TR-2.)

Spencer, Andrew. 1996. *Phonology: Theory and Description*. Oxford: Blackwell.

Steriade, Donca. 1990. *Greek Prosodies and the Nature of Syllabification*. New York: Garland.