side of the disjunction. To see the effect of *Complex-Syllable and *$C^{unsyll}$ without interference from Foot-Binarity(syllable), we need a competition where both the winner and loser contain at least two syllables. An example is given in (56). Since Foot-Binarity(syllable) assesses *blank* in this tableau, we can safely conclude that *Complex-Syllable and *$C^{unsyll}$ really do dominate *V#.

(56) Tableau for [haj.wis.kʔa] 'don't laugh!' (Newman 1944: 118)

| /hajwis-kʔa/ | *Comp-Syll | *$C^{unsyll}$ | *V# | Max | Ft-Bin(syll) |
|---|---|---|---|---|---|
| a. → haj.('wis.kʔa)$_{foot}$ | | | * | | |
| b. ('haj.wiskʔ)$_{foot}$ | *W | | L | *W | |
| c. ('haj.wis.kʔ)$_{foot}$ | | *W | L | *W | |

Another way to pull apart this ranking disjunction is to show that Foot-Binarity(syllable) must be ranked too low to dominate *V#. In fact, it has to be ranked below Max, and since we already know that Max is ranked below *V#, it follows that Foot-Binarity(syllable) cannot possibly dominate *V#. The argument for ranking Foot-Binarity(syllable) below Max is based on the observation that Yawelmani has some monosyllabic content words, such as [tiʔ] 'house (subjective)' (Newman 1944: 240). The word [tiʔ] violates Foot-Binarity(syllable) and obeys Max. To make the ranking argument, we require a losing candidate that obeys Foot-Binarity(syllable) and violates Max. The idea, which goes back to the beginning of OT (Prince and Smolensky 1993/2004: 57), is that Foot-Binarity(syllable) can be vacuously satisfied by deleting all of the segmental material in the input. With no segments to parse, there are no syllables, no feet, and no phonological word – hence, vacuous satisfaction of Foot-Binarity(syllable). The ranking argument is given in (57).

(57) Ranking argument: Max >> Foot-Binarity(syllable)

| /tiʔ/ | Max | Ft-Bin(syll) |
|---|---|---|
| a. → ('tiʔ)$_{foot}$ | | * |
| b. Ø | ***W | L |



In summary, constraints that favor losers need to be dealt with in the analysis because they threaten the winner, and constraints that favor winners need to be dealt with because they threaten the previously established rankings. The threat from loser-favoring constraints needs to be eliminated by ranking them below winner-favoring constraints, and any disjunctions that come from winner-favoring constraints need to be resolved whenever possible. Constraints that favor neither winners nor losers can be ignored at no peril to the analysis. In many cases, these constraints might simply be irrelevant to the phenomenon under discussion, and including them in the analysis would be a distraction.

EXERCISES

25 In exercise 21, you were not asked to consider the potential effect of $Max_{stem\text{-}final}$ on the analysis of Diola Fogny. Now you are. Can it be omitted from discussion, according to Prince's criteria? Explain your answer. (You should assume that all of the assimilating nasal consonants in the data are stem-final, so this constraint is at least potentially relevant.)

26 Assume that there is a markedness constraint *ŋ that is violated by velar nasals. Can this constraint be omitted from discussion of Diola Fogny, according to Prince's criteria? Explain your answer.

## 2.8 Inputs in Ranking Arguments

Which inputs need to be dealt with in an analysis? The answer might seem obvious: the analysis needs to deal with the inputs for all of the data that are being analyzed. It is of course correct that the analysis needs to deal with all of those inputs, but that isn't enough. In phonology especially, the data sets that are the focus of an analysis are paradigms with alternations, such as (15) and (16). The inputs for data sets like these are sometimes insufficient for constructing a solid OT analysis. There are two reasons for this.

First, because the data set was probably constructed with a focus on forms that alternate, it may be biased toward inputs that map to unfaithful output forms. In an OT analysis, however, inputs that map to *faithful* output forms are also relevant to the analysis, since they tell us which markedness constraints are crucially dominated. For example, Yawelmani /ʔilk-al/ → [ʔil.kal] shows that the language has codas, so No-Coda must be dominated by Max and Dep, to rule out

unfaithful codaless candidates like *[ʔi.ka] and *[ʔi.li.ka.li]. Similarly, the faithful mapping /la:n-al/ → [la:.nal] shows that long vowels are permitted in open syllables, so the markedness constraint against long vowels must be ranked below IDENT(long). Faithful input–output mappings like these are relevant to the analysis because of the assumption that all constraints are universally present in the grammars of all languages (§1.3). It's important to discover which markedness constraints a language violates, since those constraints must be ranked below faithfulness. The only way to discover these constraints is to look at faithful input–output mappings. Like Sherlock Holmes's "curious incident of the dog in the night-time,"[12] the inputs where nothing happens can be crucial.

Second, the inputs that are involved in alternations almost never include all of the logical possibilities that the grammar must treat unfaithfully. One of OT's basic hypotheses is that constraint ranking is the *only* systematic difference between languages (§1.7). If this is true, then languages cannot differ systematically in their lexicons. This has important implications for what inputs need to be considered, what the grammar must do with them, and what ranking arguments we therefore need to construct.

The idea that languages cannot differ systematically in their lexicons is called *richness of the base* (Prince and Smolensky 1993/2004: 205, 225). This phrase is a little obscure, so I will first explain its origin. The word "base" refers to the input to the grammar, since in early syntactic theory the base was the phrase structure component that produced inputs to the transformational component.[13] The word "richness" is used here in the sense of "profusion." In OT, the base (= lexicon, as input to the grammar) contains a profusion of diverse forms because it isn't subject to any language-particular restrictions.

If richness of the base is assumed, then OT cannot rely on certain analytic techniques that are common in other theories. In phonology, for example, non-OT analyses frequently employ devices like lexical redundancy rules, morpheme-structure constraints, or lexical underspecification – e.g., /bn/ is prohibited morpheme-initially in English, or voicing is lexically unspecified in labials in Arabic (which has [b] but not [p]). Similar ideas are common in contemporary syntactic theory as well – e.g., the claim that languages differ systematically in whether their *wh*-words carry a feature that requires movement to [Spec, CP]. Because of richness of the base, these methods of analysis aren't available in OT. Instead, all aspects of well-formedness are under the control of EVAL and the constraint hierarchy, and all



systematic differences between languages can and must be obtained only from differences in constraint ranking.

There are two main arguments in support of richness of the base. One is parsimony: since ranking can differ from language to language, the strongest hypothesis is that ranking is the *only* possible difference between languages. The other argument goes back to the study of phonological conspiracies in the 1970s (see §1.1). Researchers at that time noticed that restrictions on the lexicon often had the same effect as the phonological rules. For example, Yawelmani has a morpheme structure constraint that prohibits initial consonant clusters, and it also has a phonological rule of epenthesis that eliminates unsylabifiable clusters (see (58)). This kind of conspiracy was referred to as the *Duplication Problem*, since the restrictions on the lexicon duplicate the effects of the rules (Clayton 1976, Kenstowicz and Kisseberth 1977). Some researchers proposed to solve the Duplication Problem by eliminating restrictions on the lexicon and using just rules or output constraints to account for all generalizations. Richness of the base is OT's instantiation of this idea. See McCarthy (2002: 68–91) for further explanation of the Duplication Problem.

As a practical matter, richness of the base means that the grammar has to deal with a much wider range of inputs than the analyst might normally consider. Even though English has no words that alternate in a way that would require the underlying form /bnæg/, the grammar of English still has to handle the input /bnæg/. "Handle," in this context, means "account for the unpronounceability of the faithful candidate *[bnæg]." The grammar must be designed so that it selects something other than unpronounceable *[bnæg] as the most harmonic member of /bnæg/'s candidate set. In this way, it's the grammar alone, rather than the grammar aided by restrictions on the lexicon, that accounts for the set of possible words or grammatical sentences of English or any other language.

In general, the grammar of every language has to map every possible input to some well-formed output. (We'll see an interesting twist on this idea in §6.5.) To check whether the grammar really does this, the analyst needs to ask questions that are sometimes not very obvious from the data being studied. In phonology, it's necessary to ask about inputs containing various configurations that are unpronounceable in the language, such as initial clusters in Yawelmani. Nothing in the Yawelmani data would suggest that inputs with initial clusters are important, since there is no reason to set up underlying representations with initial clusters. In syntax, similar questions have to be

asked about inputs where case is assigned inappropriately or required verbal arguments are missing or there are unwanted instances of dummy elements like English unstressed *do*. (See Smolensky, Legendre, and Tesar (2006: 529) for an example of richness of the base in syntax.) In general, it isn't enough that the analysis works when the inputs are well-behaved; the analysis has to work over all possible inputs.

In actual practice, the analyst's job isn't nearly so daunting. The important thing is to make sure that the analysis does not rely on convenient regularities in the inputs. Because these regularities are so often unstated, even in theories that allow language-particular restrictions on inputs, it takes some effort to realize that they're there. But no analysis in OT is complete until the inputs have been checked for regularities. If input regularities are found, the analysis needs to be fixed so that it can handle, in the sense used above, inputs that do not conform to these regularities.

Let's look at the Yawelmani example once again. The active markedness constraints – *Complex-Syllable, *$C^{unsyll}$, and *V# – limit the shapes of syllables and words. We therefore need to make sure that the inputs do not exhibit any convenient regularities of syllable or word shape that might be helping the analysis along. Since the analysis has only been checked with underlying representations that were obtained from the data sets (15) and (16), we shouldn't be surprised to find that there are significant regularities in the inputs. These regularities disclose holes in the analysis that need to be filled.

So far, the Yawelmani analysis has focused on roots with the shapes CVC, CV:C, and CVCV:. The gaps are fairly obvious. For example, what about roots like CVCC or VC? Often, such questions can be answered simply by looking at more data. For instance, further data like (58) show that Yawelmani has underlying CVCC roots and that these roots undergo vowel epenthesis before consonant-initial suffixes. This tells us something about the ranking of Dep.

(58) Yawelmani epenthesis (Kenstowicz and Kisseberth 1979: 85)

| | *Underlying* | *Surface* | |
|---|---|---|---|
| a. | /ʔilk-hin/ | [ʔi.lik.hin] | 'sing (nonfuture)' |
| | /lihm-hin/ | [li.him.hin] | 'run (nonfuture)' |
| b. | /ʔilk-al/ | [ʔil.kal] | 'sing (dubitative)' |
| | /lihm-al/ | [lih.mal] | 'run (dubitative)' |

The question about VC roots is harder to answer. Yawelmani has no vowel-initial syllables or words, so an input VC root must not be mapped to a faithful output form – i.e., hypothetical /ap-hin/ cannot become *[aphin], since *[aphin] isn't pronounceable in this language (Newman 1944: 27). But the only way for *[aphin] to lose is for some other candidate to win, and there is no evidence from alternations to tell us what that other candidate is.[14] We might conjecture that it's [ʔaphin], but this really is nothing but a conjecture. On how to deal with questions like this, see §2.10.4.



Richness of the base is particularly important when analyzing systems of contrast and neutralization. Take Yoruba, for example. Vowels contrast in nasalization except after a nasal consonant, where all vowels neutralize to nasal (Pulleyblank 1988: 258). Thus, [a], [ã], [ba], [bã], and [mã] are allowed, but not *[ma]. Traditional analyses would rule out *[ma] by ruling out /ma/ from the input, using something like a morpheme structure constraint or lexical underspecification. Richness of the base requires that the grammar do all of the work of explaining the ill-formedness of *[ma]. In particular, the grammar of Yoruba must treat /ma/ unfaithfully, mapping it to something well-formed like [mã].

A grammar that does exactly that is given in (60) and (61). This grammar is based on the three constraints in (59). The markedness constraint *$V_{[+nasal]}$ is a general, context-free constraint against nasal vowels. The other markedness constraint, *$NV_{[-nasal]}$, prohibits oral vowels in a specific context, when they are preceded by a nasal consonant. In other words, *$V_{[+nasal]}$ is a general force in opposition to nasal vowels, whereas *$NV_{[-nasal]}$ exerts a pressure in their favor in the right environment. The third constraint in (59) requires faithfulness to input nasality.

(59) Nasality constraints (McCarthy and Prince 1995)

a. *$V_{[+nasal]}$
Assign one violation mark for every nasalized vowel.

b. *$NV_{[-nasal]}$
Assign one violation mark for every sequence of a nasal consonant followed by a nonnasalized vowel.

c. Ident([nasal])
Assign one violation mark for every segment that changes its value for the feature [nasal] between input and output.

Since the grammar maps /ma/ to [mã], the pro-nasal constraint *$NV_{[-nasal]}$ must dominate the anti-nasal constraint *$V_{[+nasal]}$ as well as the faithfulness constraint. This ranking result is shown in (60).

(60) Yoruba: *$NV_{[-nasal]}$ >> IDENT([nasal]), *$V_{[+nasal]}$

| /ma/ | *$NV_{[-nasal]}$ | IDENT([nasal]) | *$V_{[+nasal]}$ |
|---|---|---|---|
| a. → mã | | * | * |
| b. ma | *W | L | L |

Tableau (60) presents an incomplete picture, however, because it doesn't yet explain why there is a nasality contrast in [a] vs. [ã] and [ba] vs. [bã]. That requires the further ranking in (61): faithfulness to nasality in the input overrides the anti-nasal force of *$V_{[+nasal]}$.

(61) Yoruba: (*$NV_{[-nasal]}$ >>) IDENT([nasal]) >> *$V_{[+nasal]}$

| /bã/ | *$NV_{[-nasal]}$ | IDENT([nasal]) | *$V_{[+nasal]}$ |
|---|---|---|---|
| a. → bã | | | * |
| b. ba | | *W | L |

This analysis treats the distribution of nasalized vowels in Yoruba as a fact about surface forms: there is no nasalization contrast after nasal consonants because the markedness constraint *$NV_{[-nasal]}$ dominates the faithfulness constraint IDENT([nasal]), but there is a nasalization contrast elsewhere because IDENT([nasal]) dominates *$V_{[+nasal]}$. In a traditional analysis, neutralization of a contrast is dealt with by a restriction on the inputs to the grammar: vowels preceded by nasal consonants must be nasal or must be underspecified for nasality. In OT, contrast or the lack of it is determined by the grammar, so contrast and distribution are facts about surface structure alone.

This sort of reasoning and analysis isn't limited to phonology; similar things need to be said to account for the distribution of unstressed *do* in English (§2.9). In general, if some linguistic item has a restricted distribution, then faithfulness to that item is ranked below some markedness constraint or constraints that control the distribution. In Yoruba, *$NV_{[-nasal]}$ controls the distribution of nasality in vowels but *$V_{[+nasal]}$ does not. That is why Yoruba neutralizes the contrast in one environment (after a nasal consonant) and preserves it elsewhere.

The same method of analysis is applicable when there is no contrast at all. Madurese has [a], [ba], and [mã], but not *[ã], *[bã], or *[ma]



(Stevens 1968).[15] In Madurese, then, there is perfect complementary distribution, so there is no environment where nasality is contrastive in vowels. This means that IDENT([nasal]) is ranked below both of the markedness constraints, as shown in (62) and (63).

(62) Madurese: *$NV_{[-nasal]}$ >> *$V_{[+nasal]}$, IDENT([nasal])

| /ma/ | *$NV_{[-nasal]}$ | *$V_{[+nasal]}$ | IDENT([nasal]) |
|---|---|---|---|
| a. → mã | | * | * |
| b. ma | *W | L | L |

(63) Madurese: (*$NV_{[-nasal]}$ >>) IDENT([nasal]) >> *$V_{[+nasal]}$

| /bã/ | *$NV_{[-nasal]}$ | *$V_{[+nasal]}$ | IDENT([nasal]) |
|---|---|---|---|
| a. → ba | | | * |
| b. bã | | *W | L |

Richness of the base is the source of more confusion and misunderstanding than any other aspect of OT. One misunderstanding is the belief that richness of the base requires all languages to have identical lexicons. Actually, richness of the base says that there are no *systematic* differences in lexicons. In other words, linguistic patterns or generalizations cannot be attributed to lexical differences. Richness of the base does not exclude the possibility of unsystematic differences in lexicons, of which there are many. Languages unsystematically differ in the meanings that they associate with specific segmental sequences; the meaning *felis catus* is associated with the segment sequence [kæt] in English but not other languages. Lexicons are full of accidental properties like this, and richness of the base says nothing about them.

Another misunderstanding is the idea that richness of the base requires absurd underlying representations. Suppose the grammar of English maps /ŋkæt/ to [kæt], because the markedness constraint violated by *[ŋkæt]'s initial cluster dominates MAX. This doesn't mean that the underlying representation for the actual word [kæt] is /ŋkæt/. The underlying representation for [kæt] is /kæt/, of course. When children acquire English, they have no reason to set up any other underlying representation for [kæt].[16]

This misunderstanding is the result of failing to distinguish between inputs and underlying representations. The set of inputs is a construct of the theory: it's simply the result of freely combining all of the representational primitives, such as features, in all possible ways. But underlying representations are a construct of learners: they are inferences about the shared properties of a group of related words. The actual underlying representations that some learner has acquired are a finite subset of the infinite set of phonological inputs. In general, richness of the base says nothing about how to analyze specific surface forms; it's about the general structure of the language rather than individual words or sentences.

Another source of confusion is the incorrect assumption that the inputs of the rich base must be transformed into actual words of the language. On this view, English phonology couldn't map /bnæg/ to, say, [blæg] because there is no such word in the language. This assumption is wrong because it misconceives the goal of phonological analysis. We aren't concerned with the phonology of the *actual* words of the language so much as the *possible* words. Clearly, [blæg] is a phonologically possible word of English,[17] and our analysis should say as much. In this respect, the goal of phonology is much like the goal of syntax. The goal of a syntactic analysis of English is to construct a grammar of all possible sentences and not, say, just the sentences that have been uttered since the beginning of modern English around 1550.

A final source of confusion about richness of the base is the incorrect assumption that it somehow overrides the theory of representations. An anonymous reviewer for the same widely-admired journal mentioned in §2.2 faulted an author for not considering inputs with underspecification, since this was supposedly required by richness of the base. This complaint would only make sense if the author had assumed the possibility of underspecified representations, which he did not. You are free to assume *universal* restrictions on inputs, such as full specification, without running afoul of richness of the base.

Richness of the base presents a special problem of indeterminacy in some analyses. See §2.10.4 for an explanation of this problem and some suggestions about how to deal with it.

EXERCISES

27 The data in (58) tell us something about the ranking of DEP in Yawelmani. What do they tell us? In exercise 17, you saw other evidence for how DEP is ranked in this language. Considering both sources of evidence and everything you know about Yawelmani's constraint hierarchy from §2.3, is there a problem? What is it? How could you solve it?



28 Hawai'ian has no consonant clusters whatsoever. From this fact alone, what (if anything) can we conclude about constraint ranking in Hawai'ian?

29 In Nancowry (Radhakrishnan 1981), nasal and oral vowels contrast in all environments, so all of the following are well-formed: [a], [ã], [ba], [bã], [ma] and [mã]. Analyze Nancowry using the constraints in (59). Be sure to present ranking arguments like (60)–(63).

30 In Spanish, the voiced stops [b, d, g] are in complementary distribution with their fricative counterparts [β, ð, ɣ]. Using the following data (from a problem set in Halle and Clements 1983), formulate a descriptive generalization and construct an OT analysis that is consistent with richness of the base.

| | | | |
|---|---|---|---|
| [aɣrio] | 'sour' | [komuniðað] | 'community' |
| [gustar] | 'to please' | [deðo] | 'finger/toe' |
| [xweɣo] | 'game' | [droɣas] | 'drugs' |
| [alβondiɣas] | 'meatballs' | [seða] | 'silk' |
| [gastos] | 'expenses' | [ganaðo] | 'cattle' |
| [gonsales] | a surname | [usteð] | 'you (sg. polite)' |
| [jaɣa] | 'sore, boil' | [bastante] | 'plenty' |
| [uβa] | 'grape' | [brinkar] | 'to jump' |
| [futbol] | 'soccer' | [suβo] | 'I climb' |
| [alɣo] | 'something' | [uβo] | 'there was' |
| [sombra] | 'shade' | [kluβ] | 'club' |
| [saβino] | 'cypress' | [karβon] | 'coal' |
| [kaβe] | 'it fits' | [berðe] | 'green' |

## 2.9 Working through an Analysis in Syntax

The methods of analysis described here are equally applicable to syntax. Since I am not a syntactician, I will be using a published analysis as an example, Grimshaw's (1997) account of *do* support in English. What I say here shouldn't be taken as an accurate summary of Grimshaw's work; rather, it's a pedagogically oriented partial restatement that focuses on illustrating the various analytic techniques that I've been presenting here.

The data that will be analyzed are given in (64). (See Grimshaw's article for additional relevant data, such as *Who ate apples?*.)