

Lexicon Optimization in Languages without Alternations

Moira Yip

University of California, Irvine

1 Introduction

The development of phonological theory has been largely driven by languages with alternations, where considerations of lexical economy make the postulation of abstract underlying forms and productive rules which transform these into surface forms very natural. Languages with few or no alternations, however, have never fitted smoothly into such theories. To derive rich surface inventories from more parsimonious underlying inventories, it was necessary to postulate abstract underlying forms even for morphemes which only ever surfaced with one particular allophone. Even if lexical economy was demoted as a paramount consideration, the occurrence of alternations in one small corner of the grammar, such as in loanwords, still forced the linguist back to the abstract and rule-based analysis. This was so because the alternative, a set of phonotactic statements about the surface distribution of allophones, could not alone produce alternations: only rules could do that, and once the grammar included rules, they could be made use of for other purposes, including the non-alternating forms.

Output-based theories, such as Declarative Phonology, or Optimality Theory, are tailor-made for languages of this type. Surface-true generalizations can be trivially dealt with. When alternations are encountered, they can be understood as the direct result of the pressure to observe these surface constraints, and no special rules are needed. Notions of lexicon optimization drive one to the view that the learner will naturally internalize the forms closest to the surface, absent paradigm pressure to do otherwise.

It is argued that abstract underlying representations and rules that produce surface forms are highly inefficient for non-alternating systems, in that they frequently require both rules that derive A from B, and rules that derive B from A, in the same contexts. It is proposed that language is learnt on the basis of core data, and that non-core data - language games, poetry, speech errors, onomatopoeia, loanwords - can be used as a probe to investigate the nature of the underlying representations. This paper finds inconclusive evidence for abstract underlying representations, and concludes that the balance of the evidence suggests that learners acquire something rather close to what they hear, unless information from

alternations or paradigms forces them to do otherwise.

1.1 Theoretical Assumptions

Much of this paper argues against a rule-based model with a commitment to lexical economy, essentially the model of SPE (Chomsky and Halle 1968), and in favor of an output-based approach in which it is possible to remain non-committal about the nature of the underlying representation because *any* starting point will lead to the output that best satisfies these constraints. Theories with these properties include Declarative Phonology (Bird 1990, Scobbie 1991, Coleman 1992), Unification-based Phonology (Broe 1993), and Optimality Theory (OT) (Prince and Smolensky (1993), McCarthy and Prince (1993), since developed in a large body of work by a list of authors too numerous to mention here). Many of the arguments in this paper support any output-based approach, although I will present the analyses in OT terms, for concreteness. In section 4, on Mandarin palatalization, and section 6, on Chaoyang nasalization, I present cases that provide arguments specific to OT.

Optimality Theory (OT), proposes that the grammar consists of a set of ranked, violable, universal constraints. A set of outputs is generated for each input, and inspected by the ranked constraint set. The winning candidate is that which best satisfies the constraints, where satisfaction is assessed as follows. Each candidate is checked against the highest ranked constraint. Those that fail are eliminated. Those that pass are then assessed by the next ranked constraint, and so on until a single candidate remains. This is the optimal candidate. In the event of a tie at any stage in the process, the decision is passed on to the next ranked constraint. Although these constraints may be surface true, as they are in many of the cases in this paper, they need not be. A constraint will be violated on the surface if necessary in order to satisfy a higher-ranked constraint. Surface-true constraints are thus undominated constraints. Constraints may be positive or negative, either insisting on some configuration or banning it. Examples of the former would be Generalized Alignment statements, that require the alignment of morphological and/or phonological structure (McCarthy and Prince 1994). Examples of the latter would be feature co-occurrence constraints (segment structure constraints), which block combinations of features.

The constraints include a set that attempt to enforce FAITHFULNESS, ensuring that the output is identical to the input. Deviations from the input (violations of FAITHFULNESS) happen only in order to satisfy more highly ranked constraints, such as syllable-structure requirements.

Particularly important here is that the presence of FAITHFULNESS in the system has consequences for the structure of the lexicon. Prince and Smolensky (1993:191) develop a notion of Lexicon Optimization which states that for a given phonetic form, the chosen UR will be the one that maps onto the surface form with the fewest violations of high-ranked constraints. Since FAITHFULNESS is a set of constraints, the

chosen UR will, *ceteris paribus*, be the one with the fewest FAITHFULNESS violations, the one most similar to the surface form. For non-alternating forms, this means that output constraints will be reflected in the lexicon.

This approach naturally results in the postulation of a larger inventory of more concrete UR's, where by concrete I mean close to the phonetic surface. I take this "concrete" level to be the output of the phonology and the input to the phonetic realization rules, as in Pierrehumbert and Beckman (1988).¹ A smaller more abstract set of UR's enjoys no particular advantage under this view. As a consequence, if it is then deemed necessary to hold this proliferation of UR's in check, a separate notion of lexical economy will be required. This will be discussed in section 2.1.4.

It is important to note that even if non-alternating forms have concrete UR's, any alternations that may occur in sub-areas of the vocabulary, such as loans, speech errors, or language games, can still be accounted for if we assume that such data arise from novel inputs which still have to satisfy the same set of output constraints. This point is developed in sections 2 and 4.

An SPE type-system takes a different path. The presence of alternations in one corner of grammar inevitably implies the existence of rules. Once these rules exist, they can now be used *free*, to produce economies elsewhere. This results in the postulation of a smaller inventory of more abstract UR's, where by abstract I mean distant from the phonetic surface. Although this extension of the rules to the non-alternating forms, and the postulation of unique abstract underlying forms, is not absolutely required by the grammar, Broe (1993:160-161) points out that a commitment to a unique underlying form is an axiom of generative approaches. There have of course been proposals that aim to reduce or eliminate the abstractness of the underlying forms, but they are separate mechanisms superimposed on the basic architecture of the grammar, which, left to itself, encourages abstractness. In contrast, OT does not need to make the same commitment to a unique UR that rule-based systems do, so although FAITHFULNESS leads us to expect concrete UR's, in fact any reasonable UR will produce the right output if FAITHFULNESS is ranked low. See Itô, Mester and Padgett 1995 on this point.

I will first argue against rule-based analyses on two grounds. First, redundancy: in several cases discussed here it is necessary to posit rules that go in two directions to converge on the same result. Second, rules are sometimes needed only for sub-areas of the grammar such as loanwords. This suspicious innovation is unnecessary in an output-based system. These points are illustrated in three different Chinese dialects in sections 2-4: section 2 looks at the ATR/Length connection in Cantonese vowels, section 3 looks at Fuzhou high vowels, and section 4 focusses on Mandarin palatal consonants. These three sections all support an output-based approach in which any reasonable underlying form would produce the right output, and in which high-ranked FAITHFULNESS would result in a lexicon which stayed close to the surface forms. The second half of the paper looks more closely at UR's: is there clear evidence that tells us whether UR's are abstract or concrete. Sections 6 and 7

use non-core phenomena to probe the nature of underlying representations; section 6 discusses Chaoyang nasalization and section 7 discusses Mandarin poetic rhyme.

2 The General Problem of Chinese vowel systems

The typical Chinese language has a large number of allophones in complementary distribution, which can therefore be derived from a smaller number of phonemes. The following chart illustrates this situation with data from Cantonese. The surface vowels are given down the left-hand side, and the rhyme contexts go across the rows.

(1) Cantonese:

a:	a:	a:i	a:u	a:m	a:n	a:ŋ	a:p	a:t	a:k
ʌ		ʌi	ʌu	ʌm	ʌn	ʌŋ	ʌp	ʌt	ʌk
ɛ:	ɛ:					ɛ:ŋ			ɛ:k
e		ei							
œ: œ:						œ:ŋ			œ:k
ö		öü				ön			öt
ɔ:	ɔ:	ɔ:i				ɔ:n	ɔ:ŋ	ɔ:t	ɔ:k
o			ou						
i:	i:		i:u	i:m	i:n		i:p	i:t	
ɪ						ɪŋ			ɪk
u:	u:	u:i			u:n			u:t	
ʊ						ʊŋ			ʊk
ü:	ü:				ü:n			ü:t	

Usually, no alternations offer the learner evidence for any given underlying representation (UR). Hashimoto (1972:153) summarizes 9 different phonemic analyses of the 13 surface vowels; the number of phonemes ranges from 11 to 5. Surface [ön] is variously derived from /ön, œn, on, yon/ and /wen/ by different researchers. For Mandarin, the 12 surface vowels are frequently analyzed as derived from /i,u,a, ə/ (eg Cheng 1973).

In rule-based systems, there is an intimate connection between the form of rules, and the choice of UR. If there is a tense/lax alternation, and the grammar contains a laxing rule but not a tensing rule, then the UR's must have tense vowels. In OT, by contrast, any UR will give the right output, if the phonotactic constraints outrank FAITHFULNESS. (Itô, Mester and Padgett 1995). The form of the phonotactic constraints is driven by the observed surface forms.

Sinologists have always realized the important role played by phonotactics, but certain phenomena forced the recognition of a rule component as well. In corners of the language, alternations sometimes appear. Consider loanwords (see Silverman

1992, Yip 1993b, LaCharité and Paradis 1995). Suppose that when a word enters Cantonese from English, vowels change so as to observe the phonotactics. Then rules are needed to perform these changes, but they are rules whose action is not visible in the core language, and thus rules that would seem to be unlearnable prior to loanword acquisition. In OT, though, these kinds of alternations are straightforward: they are the result of submitting to the system a new kind of input, one it has not encountered before, and holding these new inputs up to the same set of output constraints as the native vocabulary. No new rules are needed. The first defect of rule-based systems, then, is that they are forced to postulate otherwise unmotivated rules to explain this peripheral data. The second defect, as we shall see shortly, is that they are sometimes forced to postulate pairs of rules that go in opposite directions and conspire to produce the right output. This highly suspicious state of affairs does not of course arise in OT. The final conclusion of this section will be that there is no evidence that the child learns abstract underlying vowel phonemes; the question becomes almost irrelevant from an OT perspective, because any reasonable input will give the right output, and the abstract input enjoys no special advantage.

2.1 Cantonese ATR and Length

The surface vowel inventory we saw earlier can be organized as follows. There is a mutual dependence of length and ATR of a rather unusual kind.

(2) Surface Vowel Inventory

Long:	α hi, α ATR	+hi, +ATR	i:	ü:	u:			
		-hi, -ATR			ε :	æ :	ɔ :	
							a:	
Short:	α hi, $-\alpha$ ATR	+hi, -ATR	ɪ		ʊ			
		-hi, +ATR			e	ö	o	
							ʌ	

Note that the long vowels match Archangeli and Pulleyblank's (1994) unmarked inventory, /i, u, ε , ɔ , a/. In the non-low vowels, where length or ATR is not contrastive, long vowels are usual; the short variants are found only before certain codas:

(3)	ei	ɪŋ/k	ön/t
	ou	ʊŋ/k	öü

Length or ATR is only contrastive on low vowels in syllables closed by a glide or consonant, [a:] vs. [ʌ], in (4a). All open syllables are long [a:], (4b).

- (4) a. sʌi⁵³ 'west' sai⁵³ 'to waste'
 sʌn⁵³ 'new' sa:n⁵³ 'mountain'
 tʌp² 'to hammer' ta:p² 'pile'
 b. ta:³⁵ 'hit' fa:⁵⁵ 'flower'

Our discussion will focus primarily on the low vowels. In a rule-based system with a commitment to lexical economy, one must ask the question "Which is underlying: length or ATR?". In the next two sections I explore this question, and show that one set of facts suggests ATR is underlying and length derived, while a different set of facts shows the reverse. A rule-based account thus needs rules deriving length from ATR, and rules deriving ATR from length, a clearly redundant system.

2.1.1 ATR determines Length

In the core language, there are two arguments that ATR is underlying and length is merely its surface counterpart. (i) Stop-final syllables cannot carry underlying contour tones. This suggests they are one TBU, or one μ . This is true for both short [tʌp] and long [ta:p], suggesting that the phonological contrast is one of ATR, not length, and that length is secondary. (ii) The short allophones of the non-low vowels are conditioned by the presence of certain features (velar codas, high glides, etc), and not by syllable structure (yi:n vs. yIŋ). It is much more plausible that a featural environment should condition a featural change than a length change. For example, velars may well be [-ATR], and cause the preceding high vowel to lax. If this is right, the featural, ATR, change must be basic and the length change secondary. (iii) This is confirmed by a productive example of this process from outside the core language, from syllable contraction. In various high frequency expressions, two syllables may contract to one (Cheung 1986): e.g. mei ts^hʌŋ → mɛ:ŋ 'not yet'. If this contraction adds a velar coda to a previously short tense vowel, as it does in this example, the vowel laxes. It also lengthens, so again we see that the ATR change is basic and the length change secondary and automatic.

In a rule-based account, these facts will require rules that predict length from ATR values. In the non-alternating cases, at least, they could be feature-filling redundancy rules, but rules nonetheless. Two rules will be needed. The first rule makes vowels short, and is needed to explain why a high vowel laxated by a velar coda surfaces as short. The second rule makes vowels long, and explains why a mid vowel laxated by a velar coda surfaces as long. I rough out a pair of rules below:

- (5) [α hi, -α ATR] → μ(tʌp, yIŋ)
 [α hi, α ATR] → μμ (ta:p, mɛ:ŋ)

2.1.2 Length determines ATR:

Another set of arguments appears to show the reverse: that length determines ATR.

(i) In the core language, all open syllables are long. This is of course common cross-linguistically; in Cantonese, the minimum word is the syllable so this is presumably due to a minimum word size of $\mu\mu$. For all vowels, even the low vowels where normally ATR/length is contrastive, the vowel quality goes along with this: -ATR in non-high vowels, +ATR in high vowels. Not only is it implausible that a rule assigns ATR on the basis of syllable structure, but it is also not possible to have a simple rule that does this, since the particular value of ATR depends on the vowel height. If the vowel is first lengthened, and then ATR is secondary, these problems do not arise. Two more arguments come from outside the core language. (ii) In syllable contraction, there are two sources for a long vowel. First, the same open syllable lengthening causes an ATR switch: $y\lambda u k^h ei \rightarrow y\varepsilon$: 'especially' shows the mid front /e/ lengthening and laxing to $[\varepsilon:]$. Second, Cheung (1986) shows that in contraction at least one sonorant TBU mora must be preserved from each syllable; if the syllable is stop final, a vocalic mora must be preserved (since the stop is not a TBU) resulting in a long vowel in the contracted syllable. The ATR value of this vowel is adjusted accordingly: $y\lambda u t\lambda k \rightarrow ya:k$ 'available'. (iii) The third argument comes from loanwords. Firstly, English /ə, ʌ/ in open syllables become $[a:]$ in Cantonese; in closed syllables, they remain $[\Lambda]$.

(6)	bus	pa:si:
	cover	k ^h ʌpfa:
	number one	ʌmpa:wʌn
	porter	p ^h ɔ:t ^h a:
	major	mɛ:tʂa:

Conversely, when short stressed syllables are closed by a consonant (a regular process in the loanword phonology) English /a/ becomes $[\Lambda]$: 'copy' k^hʌpp^hi:. Finally, open syllable lengthening consistently causes ATR adjustments whenever necessary:

(7) Other ATR/Length adjustments caused by open syllable length:

film	[ɪ]	fi:lʌm	[i:]
major	[e(y)]	mɛ:tʂa:	[ɛ:]
soda	[o(w)]	sɔ:ta:	[ɔ:]

In a rule based account, these facts force the formulation of rules in which ATR values are derived from length. I formulate two such rules below. Note again that both rules are needed, as indicated by the examples in parentheses after each rule:

- (8) $\mu \rightarrow [\alpha \text{ hi, } -\alpha \text{ ATR}]$ (copy)
 $\mu\mu \rightarrow [\alpha \text{ hi, } \alpha \text{ ATR}]$ (film)

A rule-based grammar thus needs four rules, (5) and (8), which conspire to produce forms which observe the ATR/Length constraints. Importantly, they carry

out adjustments in both directions, adjusting length to fit ATR and ATR to fit length. What is being missed here is of course the sense of a well-formed output, which is precisely what OT and other constraint-based phonologies are well-suited to explain. Of course, a rule-based framework can be adapted to deal with these sorts of mutual dependencies by the addition of redundancy rules, most notably by Stanley (1967), see also Anderson (1985) and for a particularly interesting non-rule based recent treatment of mutual feature dependency, see Broe (1993). Lexical Phonology (Kiparsky 1982), so successful in extending our understanding of redundancy and underspecification in a wide range of cases, still encounters problems with mutual dependency, since a commitment must be made as to which feature is specified underlyingly, and which one supplied by rule. I conclude that it remains true that data like Cantonese require the addition of an extra mechanism to a rule-based grammar, whereas they fall out naturally in an output-based system.

2.1.3 Output-based grammar: an OT account

The ATR/Length dependencies can be stated as the following pair of constraints:

(9) ATR-Length Connection:

- | | |
|--|--|
| <p>a. *a *μ</p> <p style="text-align: center;"> </p> <p style="text-align: center;">[αhi, αATR]</p> | <p>b. *Λ: *μ μ</p> <p style="text-align: center;">\ /</p> <p style="text-align: center;">[αhi, -αATR]</p> |
|--|--|

The second of these is a grounded condition in the sense of Archangeli and Pulleyblank (1994), who argue that high vowels are preferably [+ATR] and non-high vowels are preferably [-ATR], for reasons "grounded" in the articulation. It is thus a strong candidate for membership in the set of UG constraints. The first constraint is less obviously of cross-linguistic validity, so a final determination must await a greater understanding of ATR-length connections cross-linguistically. These two constraints replace the four rules of the rule-based system, eliminating the redundancy of rules going in two directions.

In an output-based system, asking whether the contrast between [a:] and [Λ] is underlyingly a distinction between /a:/ and /a/, or between /Λ:/ and /Λ/, or between /a/ and /Λ/ , is pointless. Their distribution follows from *any* reasonable choice of underlying contrasts, the ATR-Length constraints, and a bi-moraic minimum syllable size,² all ranked above FAITHFULNESS. Depending on the input, FAITHFULNESS violations will vary. The following tableau compares different outputs without commitment to what the input might be. [a:] will be optimal, because it alone passes all three constraints.

	* σ_{μ}	*a	* Λ :
ta	*!	*	
t Λ	*!		
ta:			
t Λ :			*!

[Table I. *Open Syllables*]

If we now consider the possible inputs that might lead to [a:], clearly underlying /a:/ will proceed to surface [a:] with no FAITHFULNESS violations. Any other input will violate FAITHFULNESS with respect to either ATR (/ Λ :/), or length (/a/), or both (/ Λ /). Prince and Smolensky 1993 propose that Lexicon Optimization is the result of the tension between a desire to minimize lexical specification, called *SPEC, and a desire to minimize FAITHFULNESS violations. The higher-ranked of the two will decide matters. Although /a:/ minimizes FAITHFULNESS violations, it does badly with respect to *SPEC, because it includes both ATR and length information, whereas the main alternative /a/ includes only one of these. What about even more distant UR's? Suppose we posit /tæ/ as the UR for surface [ta:]. Such a UR arguably ties with /ta/ in terms of lexical economy, i.e. *SPEC, but it incurs FAITHFULNESS violations for both length and [backness, whereas /ta/ incurs only a length violation. I therefore conclude that the learner will have no reason to posit even more distant UR's, unless they are notably more economical when assessed by *SPEC. The set of UR's that need to be considered is thus blessedly small, but it is still greater than one: the grammar underdetermines the choice of UR. In the next section, I will argue that the unmarked ranking ranks FAITHFULNESS above lexical economy, (such as *SPEC), and thus that in the unmarked case UR's are close to the surface form.

2.1.4 Lexical Economy

There is a fundamental difference between a rule-based system and an output-based system. In a rule-based system if [a:] comes from /a:/, then there are no rules in the grammar to derive [a:] from any other source. Thus if we suddenly encounter an instance of [a:] that clearly has another source, we are forced to write rules especially for the occasion. It is this situation that gives rise to analyses which postulate things like "rules of the loanword phonology", or "rules of the secret language". The alternative reaction to the discovery of such new data is to back up and reconsider the idea that [a:] comes from /a:/. Perhaps, after all, it comes from a more abstract source, the one that is visible for the first time in the new data. In an

output-based system, no such difficulty arises. Even if [a:] is normally the reflex of /a:/, other inputs subjected to the same set of constraints may have [a:] as their optimal output. There is no reason to change the grammar in any way; [a:] may still come from /a:/, and the same set of constraints is operative in all areas of the grammar.

It follows that it is very hard to find empirical arguments in favor of any particular underlying representation in an output based system. In sections 6 and 7 I will investigate some arguments from Chaoyang nasalization and Mandarin vowels that remain, unfortunately, inconclusive.

Absent such arguments, the language learner must surely proceed on the basis of the observed data. Hearing [a:], the child learns /a:/. For discussion on this point, see Venneman 1973, Inkelas 1994, Golston 1995. What would then drive the child to change her mind and switch to some other UR? The usual answer is "lexical economy". Let us look more closely at this notion

I will divide lexical economy into four types, which do not necessarily coincide. A generalized notion of economy thus requires trade-offs.

- (11) a. economy of individual lexical entries
- b. economy of phoneme inventory
- c. economy of phonotactic combinations
- d. economy of paradigms

a. Economy of individual UR's: This is the notion of economy captured by *SPEC (Prince and Smolensky 1993). This notion of economy does not necessarily lead to the choice of a more abstract underlying representation. For example, in Fuzhou (see section 3), a traditional abstract analysis would posit /ei/ as the UR for surface [i], and yet /ei/ is in no sense "simpler" than the more concrete /i/.

b. Economy of phoneme inventory: This and subsequent judgements of economy must be made across the entire lexicon, not on the basis of individual lexical entries. OT in its current form includes no mechanism for doing this. For example, Mandarin palatals are in complementary distribution with retroflexes, and could be derived from them, thus eliminating palatals from the phoneme inventory. Type (a) economy has nothing to say about such cases, since retroflexes are not obviously "simpler" than palatals.

c. Economy of phonotactic combinations: Traditionally Morpheme Structure Constraints (MSC's) limit the possible combinations of phonemes. For example, if Fuzhou [i] is derived from underlying /ei/, no economy of phoneme inventory is achieved (since /i/ is still a phoneme), but there is an economy of combinations: solitary /i/ may not be an underlying rhyme, but /i/ in combination with /e/ in the rhyme /ei/ is fine. Note that economy of this type does not necessarily coincide with economy of type (a): in the Fuzhou case, positing /ei/ satisfies type (c) economy but violates type (a) economy.

d. Economy of paradigms: Morphemes which participate in paradigms are usually assumed to have a fixed UR. For any given form, this UR may not be the most economical by any of measures (a-c), but looking across the paradigm it will be the form which best explains the paradigm as a whole.

PARSIMONY will be used to refer to a principle, not stated here, which enforces lexical economy of all four types, and incorporates trade-offs. Given the obvious difficulties in formulating such a principle precisely, one might start to doubt its role in lexical acquisition. The usual rationale for assuming that PARSIMONY exists is that lexical storage is expensive, and must be minimized. Recent work has thrown doubt on this notion: see particularly Steriade 1994. In cases like those that are the focus of this paper, this rationale is certainly unconvincing, because the numbers are so small. Consider Cantonese. If each surface vowel is stored, then 13 vowels must be learnt; if these are reduced to some smaller number of phonemes, the maximum reduction seems to be down to 5 phonemes. The 8 phoneme difference is hardly a major memory burden. An alternative way of counting would be to count combinations, or rhymes. Each vowel can occur in nine rhyme environments (see (1)). If 13 vowels are assumed, then we have 117 possible rhymes, of which 48 occur. If we assume 5 phonemes, then we have 45 possible rhymes, plus some extra as a result of allowing pre-nuclear glides. (see Hashimoto 1972:154 for details). The saving is then about 72 possible rhymes. But these do not actually have to be learnt: all that has to be learnt are the actually occurring rhymes, which will be 48 under any account.

I conclude that the null hypothesis is that learners of languages like Cantonese, with few or no alternations, learn something very close to what they hear, and that PARSIMONY is overridden by FAITHFULNESS.³

3 Fuzhou Vowel Alternations

Fuzhou (see Yip 1980, Wright 1983, Chan 1985, and Jiang-king 1995) shows vowel alternations in metrically weak positions. These alternations correlate with the loss of complex LHL or HLH tones. The first syllables in the following examples show a sub-set of the alternations:⁴

- | | | | | |
|-------------|-----------------------|-------------|-------------------------------------|------------------|
| (12) ei > i | a. lei ^{LHL} | 'sharp' | li ^{HL} lei ^{LHL} | 'very sharp' |
| ou > u | b. sou ^{LHL} | 'colorless' | su ^{HL} sou ^{LHL} | 'very colorless' |
| öy > y | c. töy ^{HLH} | 'resentful' | ty ^H töy ^{HLH} | 'very resentful' |

Underived monosyllabic morphemes show the same correlations; note that words with high vowel nuclei in citation forms never undergo any alternations:

- | | | | | |
|------|----------------------|------------|-----------|-----------------------------|
| (13) | HLH, LHL | | HL, H, ML | |
| ei | peiŋ ^{HLH} | 'combine' | i | piŋ ^H 'guest' |
| ou | tsouŋ ^{HLH} | 'handsome' | u | tsuŋ ^{ML} 'permit' |
| öy | töyŋ ^{LHL} | 'middle' | y | tyŋ ^{HL} 'repeat' |

The standard generative analysis (Wang 1968, Yip 1980) argues that there are no underlying /i, u, y/ rhymes. All surface [i,u,y] are derived by rule from /ei, ou, öy/ in a μ environment, since this rule is needed anyway for the reduplicative data. This analysis runs into two problems. First, as pointed out by Jiang-King (1995), this gives an unusual set of underlying rhymes, with no plain /i,u,y/. Second, as with any abstract analysis, learnability issues arise. Positing /peij/ for [piŋ] in the non-alternating cases like (13) can only be explained if lexical economy plays a major role.

An output-based grammar encounters no such problems. Alternations will result from satisfying a set of undominated constraints, at the expense of FAITHFULNESS. I retain Wright's insight that the rhymes differ in moraic structure. The following sketch draws heavily on Jiang-king (1995), which offers a detailed OT analysis of the full set of Fuzhou rhymes.

I posit two constraints. The first penalizes short diphthongs, and the second requires that the reduplicant be a light syllable (so that the entire reduplicated form is iambic).

(14) a. $*\mu$ No short diphthongs.
 VV

b. RED= μ Reduplicant is a light syllable

Both of these dominate MAX, which enforces maximal reduplication.

The following tableau shows the reduplicated form of /lei/.

$\mu \mu$ RED-/l e i/	RED= μ	$*\mu$ VV	MAX
$\mu \mu \quad \mu \mu$ l e i l e i	*!		
$\mu \quad \mu \mu$ l e i l e i		*!	
$\mu \quad \mu \mu$ li l e i			*

[Table II. *Reduplicated Adjectives*]

Non-alternating mono-morphemic [piŋ] would surface correctly from either, trivially, mono-moraic /piŋ/, or monomoraic /peij/, shown below, provided FAITHFULNESS is ranked low.

μ /peij/	* μ VV	FAITHFULNESS
μ peij	*!	
μ ☞ p<e>ij		*

[Table III. *Non-alternating high vowel nuclei*]

(See Jiang-king (1995) for a discussion of why /i/ rather than /e/ surfaces in Tables II and III.) A tableau of tableaux shows that the output (a) based on underlying /pij/ has a violation-free path through the constraints, but the output (b) based on underlying /peij/ does not:

	* μ VV	FAITHFULNESS
a. μ ☞ pij		
b. μ p<e>ij		*!

[Table IV. *Tableau of Tableaux comparing [pij, μ] output from either /pij, μ / or /peij, μ / input*]

The UR /pij/ wins on several counts here: it has a more faithful path to the output, and a simpler UR (i.e. economy of type (a)). Proponents of the more abstract /peij/ would have to appeal to economy of type (c), combinatorial economy. In particular, phoneme inventory economy will not distinguish between /i/ and /ei/, since /i/ must still be a phoneme to distinguish /a/ from /ai/.

Under the output-based account, we see a familiar picture: the alternations found in reduplication are unproblematically derived, but they in no way force us to conclude that *all* [i,u,y] are derived from diphthongs.

4 Mandarin Alveo-Palatal Consonants

I now shift to a well-known case of consonants, this time in Mandarin. The palatals

tʃ, tʃ^h, and ʃ are found only before [i, ü] and the corresponding glides. They are in complementary distribution with three other series, velars, retroflexes, and dentals, none of which occur before [-back, +high].

(15)	Velars	kai	ku	*ki	*kü
	Dentals	tsai	tsu	*tsi	*tsü
	Retroflexes	<u>t</u> sai	<u>t</u> su	* <u>t</u> si	* <u>t</u> sü

	Palatals	*tʃai	*tʃu	tʃi	tʃü

In an abstract analysis, economy of phoneme inventory supplies pressure to eliminate the palatals as phonemes, and derive them from one of the other series, but as Chao (1934) points out there is no unique solution to this problem: any of the three series is a possible source.

It is often claimed that a velar source is supported by onomatopoeia (Chao 1934, Lin 1989, Chiang 1992, Wu 1994), since the onomatopoetic vowel /i/ causes palatalization of an underlying velar. The (a) data shows the general pattern, and the (b) data shows the palatalization. The underlying form shows up in the third syllable, and the palatalizing environment /i/ is in the first syllable:

(16)	CV → Ci li CV IV	
a.	p ^h i li p ^h a la	noise of fire crackers
	ti li ta la	sound of rain drops
	t ^h i li t ^h u lu	slurping
b.	tʃi li k ^(w) a	lachatting noise
	tʃ ^h i li k ^{h(w)} a la	noise of falling objects
	ʃi li xu lu	eating fast

There are no inputs with dentals or retroflexes. A second argument comes from the May-ka language game (Chao 1931), which supplies a velar /k/, which is then palatalized before an input vowel /i/, as shown below:

(17)	ma	→	mai-ka	
	liŋ	→	liai tʃiŋ	*kiŋ (later, liai > liɛ)

These data show that velars palatalize before /i/, but they do not show that this is the *only* (or even the main) source for palatals.

An argument from a different source comes from speech errors. I assume that speech errors take as input a phonological, as opposed to phonetic, representation, and that the output of speech errors conforms, at least in part, to the usual phonology of the language. Since Fromkin (1973) it has been known that velar nasals can be

decomposed into their underlying /ng/ parts in speech errors (e.g. *Spri[ŋ]time for Hitler* -> *Spri[g]time for Hi[n]tler*) and also that outputs undergo rules such as voicing assimilation (e.g. *cow track[s]* -> *track cow[z]*). Shattuck-Hufnagel (1986:142) points out that these data show that when speech errors take place, segments "have not yet taken on their surface phonetic shapes" and that they then "adjust... to fit their new environment". She also concludes (p.118): "To the extent that we find positive speech error evidence to support a proposed phonological structure, we can infer that the structure is reflected in the speaker's processing representations." With this as background, let us look at Chinese. Shen (1992) gives an interesting example

involving an exchange of nuclei, or perhaps the feature [-back], which shows both retroflexes palatalizing to palatals in σ_2 , and palatals de-palatalizing to retroflexes in σ_1 .

- (18) Target: Palatal ... Retroflex tʃ^hü s_u (Shen: C7)
 Error: Retroflex.. Palatal ts^hu çü

A rule-based account of these data will be forced to postulate three rules, two of which will have to be "speech error" rules, because there is no evidence for them in the core phonology. In this respect the data are quite different from English: voicing assimilation is found both in the core language, and in speech errors:

- (19) (i) k → tʃ/_ [-back] (onomatopoeia, language games)
 (ii) ts → tʃ/_ [-back] (speech errors)
 (iii) tʃ → ts/_ [+back] (speech errors)

Rules (i-ii) could be collapsed, but rule (iii) is separate. This is strongly reminiscent of the Cantonese situation: rules going in opposite directions to achieve the same result.

As in Cantonese an output based system encounters no particular problems, under the assumption that speech error outputs are subject to the same OT grammar as the core language. There will be a constraint that requires palatalization, by spreading [-back] onto the onset:

- (20) ALIGN-L[-]: [-back, σ] : [-back] must be aligned with the left -edge of the syllable

The failure of labials, and coronal stops, to undergo palatalization can be handled by a segment structure constraint outlawing [Coronal, -cont, -back] and [Labial, -back]; I assume that affricates have a [+cont] component, following Sagey (1986), Lombardi (1990). For a different view, see Rubach (1994). Lastly, the non-existence

of palatals before [+back] vowels can be explained if we assume that [-back] (and probably [+back] as well) requires a vowel as its primary licenser, in the sense of Goldsmith 1990:

(21) LICENSEBACK: [-back] is only licensed by [-cons] segments.

Of course, once [-back] is attached to a vowel it may also attach to a consonant, but it may not attach to a consonant alone. Given these constraints, outputs like *[t̥s̥ü] will be rejected in favor of [t̥çü] because only the latter satisfies (20), the alignment requirement. Conversely, outputs like *[t̥çu] will be rejected in favor of [t̥su] because the former violates (21), the licensing requirement. Finally, outputs like [ti] will be preferred to *[t̥^yi] because the latter violates the segment structure requirement banning [Coronal, -cont, -back]; this will be so even though the optimal output, [ti], violates (20), the alignment requirement, showing that *[Coronal, -cont, -back] >> ALIGN-L[-back, σ]. Looked at in this way, palatalization provides another case of the violability of constraints typical of OT. Although ALIGN-L(-back,σ) is often violated on the surface, it plays an important role in the system.⁵

As in our other cases, in OT any reasonable UR leads to the right output if the palatalization constraints outrank FAITHFULNESS. A commitment to a single, consistent UR is not required. So it is not problematic that the source must be a retroflex in the speech error data, a velar in onomatopoeia and language games, but indeterminate elsewhere. Further, the fact that palatals depalatalize to retroflexes does not force us to the conclusion that they are underlying retroflexes. They could be underlying palatals, the concrete analysis, but still be forced to change when their environment is changed such that they violate a constraint. Finally, the details of the palatalization analysis show that the constraint responsible for palatalization, ALIGN-L(-back, σ), is *not* surface-true, because it is over-ruled by the more highly-ranked segment-structure constraint that blocks *t^y. This provides a further illustration of the violability of constraints expected within OT.

5 Non-core Phenomena as Probes of Underlying Representations

To summarize the discussion so far, constraint based theories are very good at dealing with languages with no alternations or ones where the alternations are limited to a small corner of the grammar, because (i) they do not encourage one to posit very abstract UR's for the vast majority of non-alternating morphemes and (ii) they do not lead one to posit multiple rules that converge on a single target, often by both deriving A from B and B from A. Although I have argued that OT does not encourage the postulation of abstract underlying representations, it remains possible that learners value PARSIMONY so highly that they eschew the more faithful route in favor of lexical economy. Finding empirical (as opposed to theory-internal)

arguments that bear on whether UR's are abstract or concrete has proved very hard. Indeed, we do not even know whether all learners are consistent in their own grammar, or with each other. The last two sections of this paper look for evidence that bears directly on discovering the underlying representation. First, I set out some background assumptions.

It is reasonable to suppose that language acquisition is based on the "core" language, where I intend this term to exclude data from atypical, specialized domains like language games, onomatopoeia, speech errors, loanwords, and poetic rhyme. If this is right, these specialized sub-areas can be used as possible probes into the grammar formed without them as input. (Some of these are obviously more likely to be available to the child than others - poetic rhyme and onomatopoeia being the two most obviously present in the young child's environment. If all this data is included in the acquisition input data, then they have no privileged status that allows us to use them as a probe at all, and we can conclude nothing special from their behavior. However, given the difficulty of accessing UR's in these languages, this line of attack seems worth exploring.) There are two types of situation. Firstly, there can be new inputs to the core grammar, as in loanwords. I assume, with Singh 1987, Silverman 1992, and Paradis and Prunet 1995, that when a loanword first enters the vocabulary it becomes subject to the grammar of the host language. If the word becomes fully assimilated, it conforms entirely to the host phonology. This is, one might propose, what full assimilation means.

	Lexicon: UR's	Grammar
Core Language	Core	Core
Loanwords	New	

[Table VI. *Loanwords: New inputs to an existing grammar.*]

The core grammar is presented with an entirely new type of input, and has to respond. The form of its response can be informative. Suppose that excess consonants are handled by epenthesis, and yet the language has shown no sign of any epenthesis at its core. Assuming that the grammar has not changed, only the inputs have, there cannot be any epenthesis rule; however, the constraints that restrict consonant clusters in the core are still in effect, and epenthesis as a response to them is entirely understandable. Such data can thus be seen as support for a constraint-based grammar rather than a rule-based grammar. This argument is developed in Yip 1993, in connection with Cantonese loanwords.

In other cases non-core phenomena can probe the nature of UR's. Consider poetic rhyme. Here the grammar is enlarged by the addition of some rule/constraint/principle of poetic rhyme, but the inputs are unchanged lexical entries

formed on the basis of the core.

	Lexicon: UR's	Grammar
Core Language	Core	Core
Poetic Rhyme		New

[Table VII. *Poetic Rhyme: New grammar and existing inputs.*]

If the rhyme needs access to a particular set of UR's, then it constitutes evidence for those UR's. Onomatopoeic reduplication can be viewed in the same way: existing UR's are subjected to a new grammatical addition, the onomatopoeic reduplication process. The outcome can thus provide insight into the nature of the UR. The next section uses this strategy to search for evidence that bears on the nature of the underlying representations in non-alternating systems; it turns out that the little evidence I have discovered so far is inconclusive.

6 Evidence for Abstract UR's: Chaoyang Non-localized Features

Onomatopoeic reduplication, and some other types of reduplication, are accompanied by loss of certain input features, and retention of others. I will argue that all and only the segment-level features Place, voice, and continuant are lost, and all and only the morpheme-level features tone, constricted glottis, and nasal are retained, and that this distinction must therefore be present in the lexical entries. Since even morpheme-level features surface on individual segments, UR's in which they are not yet localized on those segments are relatively abstract, suggesting that PARSIMONY is more important than FAITHFULNESS in Chaoyang. Surface placement of morpheme-level features is controlled by a set of constraints, two of which are shown to be violated on the surface. Again, then, we see an instance of this central OT property: constraints may be violated under the pressure of dominant and competing constraints.

Chaoyang has nasal and oral vowels, and these vowels exhibit a restricted distribution with respect to syllable structure, and the distribution of nasal and oral consonants. Although there are no alternations of the kind found in familiar nasal harmony languages, I will show that Chaoyang nasalization is represented only by a single morpheme-level specification, whose surface localization results from a set of constraints. I begin by setting out an OT analysis of [nasal] as a morpheme-level feature, and then use the reduplication data to show that this analysis is correct.

6.1 Background

Chaoyang is a S. Min dialect spoken in Guangdong province, near Chaozhou. The data comes from several papers by Zhang (1979a,b,c, 1980, 1981, 1982). Chaoyang has five vowel phonemes /i,u,e,o,a/, plus nasalized counterparts, and the following consonants /p, p^h, b, m, t, t^h, l, n, ts, ts^h, s, z, k, k^h, g, ŋ, h/. /l/ plays the role in the system of a voiced coronal stop. Final consonants may be [m,ŋ,p,k] and glottalization. There are eight tones. Open syllables are of the form (C) (G) V (G) (?), with oral or nasal vowels. Closed syllables are (C) (G) V C, and vowels are oral only.⁶ I follow Duanmu (1990) in assuming that CV syllables are actually CVV, as in other dialects, to satisfy a minimum word requirement. Some sample syllables are given below:

(22)	pou ¹¹	'chew'	p ^h aŋ ³³	'fragrant'
	me: ⁵³	'fast'	tsi:ʔ ⁵⁵	'stone'
	ŋiam ⁵⁵	'surname'	bi:ʔ ¹¹	'hide'
	lok ¹¹	'shake'	siap ¹¹	'forty'
	oiʔ ⁵⁵	'narrow'	lauʔ ¹¹	'lick'
	ĩãũ ^{ʔ11}	'fold'		

6.2 Glottalization:

In Yip (1994) I argue in detail that glottalization is not a segment [ʔ], but a feature of the entire morpheme. There are a number of arguments for this; similar arguments for the related dialect Taiwanese can be found in Roberts and Li 1963, Li 1989, and Chung 1995a,b. The simplest argument is based on the observation that syllables cannot end in both a glide and a consonant *[auk], *[aim], and yet [auʔ] is fine.

Syllables with glottalization pattern with stop-final syllables in their inability to take contour tones. Since final stops are glottalized, I argue that they always carry the feature [constricted glottis], henceforth [c.g.], and that [c.g.] moras, whether stops or glottalized sonorants, may not bear tone. Nasal-final syllables, by contrast, may have contour tones and may never be glottalized. Final oral and nasal stops may then be seen as allophones conditioned by the presence or absence of the privative feature [c.g.], as shown below. On the privative character of laryngeal features, see Lombardi (1991, 1995).

(23)	unmarked	[c.g.]
	μμ	μμ
	V: a:	a:ʔ
	VG au	auʔ

VC am ap?

Under this view, codas have only Place features in UR, no [nasal] or [voice].⁷ The voicing and nasality of "nasal" codas are not phonologically present; they are the surface realization of a [+consonantal] closure in the absence of [c.g.]. This will play an important role in what follows.

6.3 Nasalization

The following table shows the distribution of nasality in the syllable; *p* stands for all voiceless unaspirated stops, *m* for all nasals, and so on.

	A	B	C	D
	a: au a:ʔ auʔ ã: ãũ ã:ʔ ãũʔ aŋ akʔ	aŋ akʔ	ãŋ ãkʔ	
p,p ^h ,∅	+	+	+	-
b	+	-	+	-
m	-	+	+	-

[Table VIII. *Nasality on Onsets and Rhymes*]

This distribution holds true both as an MSC, and in derived forms; the second example below from Chiang (1992: 209) shows that the /l/ introduced as the onset of the second syllable in reduplication surfaces as [n] before nasal vowels:

- (24) kuaʔ luaʔ tŋ 'cut off'
 kĩã nĩã lai 'walk-come'

The restricted distribution of nasality suggests that nasality is marked only for the morpheme (and hence the syllable), not the segment, and that there is a maximum of one specification of the privative feature [nasal] per morpheme.⁸ (On the privative nature of [nasal], see Steriade 1993, Trigo 1993.) Since a purely output-based system does not include a morpheme-structure-constraint component, any syllable may have a [nasal] specification in UR. It is thus important to explain why nasality surfaces in a variety of different positions in the syllable, and why in certain syllables, such as those with consonantal codas and voiceless onsets, like [tak] or [tam], nasality never surfaces at all. In other words, if /tak, [nasal]/ is a possible UR, why is *[tãk] an impossible surface form?

The surface occurrence of nasality is controlled by a set of ranked, violable constraints. The first constraint is motivated by the observation that in purely vocalic rhymes, where all segments are voiced, nasalization surfaces on the entire rhyme, as in [pãũ].

- (25) **"RHYME" HARMONY:** All moraic segments must share any nasal specification.

The second constraint is based on a similar observation about syllables composed entirely of voiced segments: they are entirely nasal, as in [mãũ], or entirely oral, as in [bau].

- (26) **SYLLABLE HARMONY:** All segments in the syllable must share any nasal specification.

The two harmony constraints could of course be formulated as Align constraints (see Kirchner 1993, Chung 1995b). A third constraint explains why voiceless segments are exempt from nasalization. This is quite clear in the case of voiceless onsets, since they may precede nasal rhymes; I shall also argue that it is true for codas. The NAS-VOI constraint in (27) should be read as "the presence of [nasal] implies/requires the presence of [voice]". Itô, Mester and Padgett (1995) propose this to achieve linking of [voice] to [nasal] segments; here it is used to block linking [nasal] to a voiceless segment, either an onset or a coda.

- (27) **NAS-VOI:** [nasal] → [voice]

Finally, we have a member of the PARSE family:

- (29) **PARSE [nasal]:** [nasal] must be parsed

Here "parse" means associated to at least one segment, and thus realized. These constraints, appropriately ranked, will correctly position [nasal] in the syllable. Below I show where [nasal] actually surfaces on four different syllable types. The syllables have voiced or voiceless onsets, and are open or closed. The final /k/ stands for *any* consonantal closure, "oral" or "nasal" (recall that *all* codas are unspecified for voice and nasal). Crucially, I assume that [nasal] will inevitably occur underlyingly on all syllable types, including /tak/, and thus its failure to surface on such a syllable (in other words, the absence of *[tãk]), demands an explanation.

UR		PR	Position of [nasal] in PR
tau	[nasal]	tãũ	entire rhyme
lau	[nasal]	nãũ	entire syllable
tak	[nasal]	tak	does not surface
lak	[nasal]	nak	onset only

[Table IX. *Positioning of [nasal] in different syllable types*]

The ranking arguments are summarized here. First, PARSE-NASAL >> SYLL-HARMONY, because nasal surfaces even if syllable harmony is impossible because a voiceless segment is present. For example, /tau, [nasal]/ > [tãũ], and /lak, [nasal]/ > [nak]. Second, RHYME-HARMONY >> PARSE-NASAL, because if a voiceless segment (i.e. a coda) is present in the rhyme, the nucleus may not be nasalized. For example /tak, [nasal]/ > [tak] (or [taŋ] if it is not [c.g.]), not *[tãk]. Third, NAS-VOI >> PARSE-NASAL, because if both onset and coda are voiceless, nasality cannot surface: for example, /tak, [nasal]/ > [tak] (or [taŋ] if it is not [c.g.]), not *[tãk] or *[tã̃]. The complete ranking is given below:

(30) NAS-VOI, RHYME-HARMONY >> PARSE NAS >> SYLL-HARMONY

It is important to note that this analysis of Chaoyang makes crucial use of the OT claim that constraints are violable. Neither PARSE NAS nor SYLL-HARMONY is surface true, and yet they play a central role in the analysis. Analyses which rely solely on surface-true output constraints are unable to explain why [tãk] and [tãŋ], [nãk] and [nãŋ] are impossible. Here then is an instance where OT offers a more straightforward account than other output-based theories.⁹

I have shown that there is a viable OT analysis of Chaoyang nasalization in which UR's have a relatively abstract morpheme-level nasal specification. The final section produces independent evidence in favor of the morpheme-level nature of [nasal].

6.4 Confirmation from pIpaLa Reduplications

Onomatopoeic reduplication in Chaoyang changes the onset of the third syllable to /l/ and the vowel of the first syllable to /i/. Some other forms of reduplication in the language show the same onset and/or vowel replacement, although other details vary.

- (31) L, I Insertion by Syllable Type:
- | | | | |
|----|------|--|-------------|
| 1. | CV | k ^h i k ^h a la kio | ML ML ML ML |
| 2. | CGVG | tsi tsiau liau kio | H H H ML |
| | | ĩ uãĩ nuãĩ kio | H H H ML |
| 3. | CVN | piŋ paŋ laŋ kio | L L L ML |
| | | or pi paŋ laŋ kio | |
| | | k ^h iŋ k ^h om kio | L L ML |
| | | hi hom lom kio | L L L ML |
| 4. | CVO | kik kiak liak kio | L L L ML |
| | | or ki? kiak liak kio | |

The changes are in fact more widespread than just insertion of /i/ and /l/. In the first syllable, codas are also neutralized to velar or glottal, and all vocalic segments (nuclear

vowel and glides) become /i/. As a result the set of rhymes after /i/ "insertion" is [i, iŋ, ik, iʔ, ĩ, ĩʔ]. Notice that all Place distinctions are lost, but nasality, glottalization on the vowel, and tone are retained, as shown below. Examples are drawn from a variety of reduplications, not limited to onomatopoeia.

(32) Replacement of nuclei by /i/:

a. Loss of Place

...ti ta... ...hi he... ... li lo...

b. Retention of [c.g.], tone, nasality

... nĩʔ nĩʔ?... L L ...tsi tsiau... H H ...ĩ ũĩ... H H

Crucially, nasalization on the coda is not retained; if the coda is lost, no residual nasalization surfaces on the remaining vowel:

(33) Loss/Replacement of Codas:

a. Loss of Place: neutralization to velar or glottal

...piŋ paŋk^hiŋ k^homtiʔ top... ...kiʔ kiak....

b. Retention of [c.g.]

...tiʔ top.....kiʔ kiak....

c. No nasal residue from deletion: Codas are not phonologically [nasal].

... hi hom... ... li lom...

The data in (33) are significant because in many languages, such as French, loss of a [nasal] coda leaves [nasal] behind, and here in Chaoyang we see that loss of a [c.g.] coda leaves [c.g.] behind. The failure of a deleted /m/ to leave [nasal] behind is thus most easily understood if it is not phonologically [nasal] at all, as claimed here.

The /l/ similarly eliminates nearly all onset distinctions: the only one that remains is nasality, so that if the original onset was nasal, /n/ is used instead of /l/.

(34) Replacement of onsets by /l/:

a. Loss of Place, Manner, Voice, and Aspiration:

...kiak liak... ...k^ha la... ...zue lue... ...sau lau... ...p^hoi loi...

b. Retention of Nasality

...ŋaŋ naŋ... ...mã ã...

In sum, the features that are retained are (i) tone (ii) [c.g.] and (iii) [nasal] on onsets or vocalic segments, but not on codas. These are exactly the features that have been argued to be morpheme-level features in this paper.¹⁰ I conclude that reduplication is diagnostic of localized vs. morpheme level features in that only the former delete. [nasal] must be morpheme-level, although this is the more abstract UR.

The Chaoyang case is particularly interesting because the realization of nasal is dependent not just on the presence of another feature, [voice], but also on syllable structure. It is not obvious how Broe's approach to mutual dependency, which can handle cases like Mandarin in section 4, can be extended to cases of this kind.

7 Evidence both For and Against Abstract UR's: Mandarin

Poetic Rhyme

Mandarin poetry has been argued to calculate its rhyme at the underlying level. If this is right, it offers insight into the UR's of the language. The evidence turns out to be contradictory: certain rhymes appear to require an abstract UR, others appear to require a more concrete one.

7.1 For Abstract UR's:

It is usual to analyze Mandarin as having an underlying four vowel system of /i,u,a,ə/, and to derive the mid vowels [e,ɛ,o,ɔ] by spreading frontness or rounding from adjacent segments (Chao 1934). High nuclei are also derived from glide plus schwa sequences. The arguments are distributional, not backed up by alternations, but PARSIMONY (applied to phoneme inventories and combinations) requires this approach.

Chen (1984) shows that rhyme is calculated before these spreading rules apply, and looks at the nucleus and any following material only. Crucially, his description holds of modern vernacular poetry and song, not just of traditional literary forms, so I follow Chen in taking it to provide insight into the synchronic phonological system. According to Chen, poetic rhyme rhymes VX. The rhyme scheme is overt in rhymes with the nucleus [a], from /a/; the following all rhyme: [pən, t'ian, k'uan, tsüan], whereas [lau, lou] or [kaŋ, kəŋ] do not. This schema is not surface true in the next set of data. The following all rhyme:

- (35) PR [fəŋ, tiŋ, nuŋ, ɕiuŋ]
UR /fəŋ, tiəŋ, nuəŋ, ɕüəŋ/

As Chen points out, if rhyme is calculated on the underlying representation, then the same schema, VX, explains why all these words rhyme - they all have [əŋ]. This constitutes an argument in favor of the abstract UR's, otherwise the rhyme is impossible to understand in any principled way.

7.2 Against Abstract UR's:

Most authors (e.g. Chao 1934, Cheng 1973, Wu 1994) derive all mid vowels from a single central unrounded schwa-like vowel. This proposal receives some support from speech errors (Shen 1992), where $\text{ə} \rightarrow \text{e}/\text{---i}$, in *nai hə* \rightarrow *nai hei*.

If [ɛ] and [ɔ] are both /ə/, then according to Chen's algorithm they should rhyme in [wɔ] and [yɛ], since these come from /wə/ and /yə/ respectively. Chen does not discuss these rhymes, but in fact they do not rhyme, suggesting that these two

vowels may be underlyingly identical to their surface forms rather than descended from an abstract /ə/, and that PARSIMONY may not be so important here.

We are thus left with a puzzle: why is there a difference between surface high and surface mid vowels? If the rhyming arguments go through, we must conclude that Mandarin speakers learn abstract UR's for the high vowels, but concrete UR's for the mid vowels. This is particularly perplexing given that the high vowels are phonemes under any analysis, and the only saving made by deriving some from sequences of high glide plus schwa is in the combinatorial arena. For the mid-vowels, deriving them from schwa actually simplifies the phoneme inventory, so one might think the pressure to select an abstract UR would be stronger.

8 Conclusions

I have argued that rule-based systems are inappropriate for languages without alternations, and that output-based systems avoid the duplication problem encountered by rule-based accounts. Output based grammars naturally result in a less economical and indeterminate lexicon, but this is not a problem. In fact, some authors have argued that OT removes the need for *any* UR: all the work is done by constraints. (Hammond 1995, Russell 1995, Golston 1995, Yip 1995). Alternations in corners of the grammar result from entering novel inputs (loans, speech errors, language games) or making additions to the grammar (a poetic rhyme requirement). They can operate as probes into the core grammar or core lexicon. In practice, however, the evidence is unclear. There is a residue of evidence for abstract UR's, but the abstraction in one case (nasalization) involves an absence of localization in UR, not an absence of melody, and the other case, poetic rhyme, is inconsistent. One suspects that finding conclusive arguments for any particular UR may prove elusive. This raises the possibility that it is also a difficult acquisition task, and that the easiest strategy may be to stick close to the phonetic surface form absent clear guidance to the contrary.

NOTES

* I would like to thank audiences at Royaumont and at UC Santa Cruz, for comments on an earlier version of this paper. I am especially grateful to thoughtful reviews by John Coleman and Janet Pierrehumbert. All errors are of course my own. This work was made possible in part by a generous grant from the Chiang Ching Kuo Foundation.

REFERENCES

- Anderson, S. (1985). *Phonology in the twentieth century: Theories of rules and theories of representations*. Chicago, University of Chicago Press.
- Archangeli, D. and D. Pulleyblank (1994). *Grounded phonology* MIT Press.
- Bird, S. (1990). *Constraint-based phonology*. Ph.D. Dissertation, Univ. of Edinburgh.
- Broe, M. (1993). *Specification theory: the treatment of redundancy in phonology*. Ph.D. Dissertation, Univ. of Edinburgh.
- Chan, M. (1985). Fuzhou phonology: a non-linear analysis of tone and stress. Doctoral dissertation, U. of Washington.
- Chao, Y.R. (1931). Fan-qie yu ba zhong [Eight varieties of secret language based on the principle of fanqie]. *Bulletin of the Institute of History and Philology Academia Sinica* II 320-354
- Chao, Y-R. (1934). The non-uniqueness of phonemic solutions of phonetic systems. *Bulletin of the Institute of History and Philology, Academia Sinica* 4, 363-397
- Chen, M. (1984). Abstract symmetry in Chinese verse. *Linguistic Inquiry* 15.1:167-170
- Cheng, C-C. (1973). *A synchronic phonology of Mandarin Chinese* Mouton, The Hague.
- Cheung, Kwan-Hin (1986). The phonology of present-day Cantonese. Doctoral dissertation, University College, London.
- Chiang, W-Y. (1992). *Prosodic phonology and morphology of affixation in Chinese languages* PhD Dissertation, U. of Delaware.
- Chomsky, N. and M. Halle (1968). *Sound Pattern of English*. New York, Harper and Row.
- Chung, R.F. (1995a) Aspects of Southern Min phonology. Ms, MIT

- Chung, R.F. (1995b) Southern Min nasality in optimal domains. Ms, MIT
- Coleman, J. 1992 The phonetic interpretation of headed phonological structures containing overlapping constituents. *Phonology* 9.1: 1-44
- Duanmu, S. (1990). *A formal study of syllable, tone, stress and domain in Chinese languages* MIT Ph. D.Dissertation.
- Fromkin, V.A. (1973). *Speech errors as linguistic evidence*. Mouton.
- Goldsmith, J. (1990). *Autosegmental and Metrical Phonology* Oxford: Blackwell.
- Golston, C. (1995) Direct optimality theory: Representation as constraint violation. Ms., Heinrich-Heine Universitat Dusseldorf.
- Hammond, M. (1995). There is no lexicon. Ms., University of Arizona, Tucson. Pp. 16. ROA-43.
- Hashimoto, A. O-K.Y. (1972). *Studies in Yue dialects 1: phonology of Cantonese* Cambridge: Cambridge University Press.
- Inkelas, S. (1994). The consequences of optimization for underspecification. Ms., University of California, Berkeley. Pp. 28. ROA-40.
- Itô, J., A. Mester, & J. Padgett. (1995). Licensing and underspecification in Optimality Theory. *Linguistic Inquiry*. 26.4:571-614
- Jiang-King, P. (1995). Fuzhou tone-vowel interaction. Ms., U. of British Columbia
- Kiparsky, P. (1982). Lexical phonology and morphology. In I-S Yang, ed., *Linguistics in the Morning Calm* Seoul: Hanshin. 3-91
- Kirchner, R. 1993. Turkish vowel harmony and disharmony: an Optimality-Theoretic account. Ms., UCLA. Pp. 20. ROA-4.
- Li, P. J.K. (1985). A secret language in Taiwan. *Journal of Chinese Linguistics* 91-121
- Li, P.J.K. (1989). Minnanyu hosaiyinwei xingzhi de jiantao [The nature of Southern Min Glottal Stop]. *Bulletin of the Institute of History and Philology* 60.3: 487-492

- Li, P.J.K. (1992). Minnanyu de biyin wenti [The problems of Southern Min nasality]. *Proceedings of the First International Symposium on Chinese Languages and Linguistics*. 423-435 Taipei: Academia Sinica.
- Lin, Y-H. (1989). *Autosegmental treatment of segmental processes in Chinese phonology* PHD Dissertation, University of Texas, Austin
- Lombardi, L. (1990). The non-linear organization of the affricate. *Natural Language and Linguistic Theory* 8:374-425
- Lombardi, L. (1991). *Laryngeal features and laryngeal neutralization*. Ph. D. Dissertation, U. of Massachusetts, Amherst
- Lombardi, L. (1995) Laryngeal features and privativity. *Linguistic Review*.
- McCarthy, J. & A. Prince (1994). *Generalized alignment*. Ms, U. of Massachusetts, Amherst, and Rutgers University.
- McCarthy, J. & A. Prince. (1993). *Prosodic morphology I: constraint interaction and satisfaction*. To appear, MIT Press. Technical Report #3, Rutgers University Center for Cognitive Science. Pp. 184.
- Noyer, R. (1994). Palatalization and vowel place in San Mateo Huave: the competition of syntagmatic and paradigmatic well-formedness. Presented at Winter LSA. Ms., Princeton University. Pp. 16.
- Paradis, C. & D. LaCharité (1995). Preservation and minimality in loanword adaptation. Ms, U. Laval.
- Prince, A. and P. Smolensky (1993). *Optimality Theory: constraint interaction in generative grammar* Ms., Rutgers University and U. of Colorado, Boulder
- Roberts, T. & Y-C Li (1963). Problems in the phonology of the Southern Min dialect of Taiwan. *Journal of Tunghai University* 5, 95-108
- Rowicka, G. (1994). Palatal assimilation in prefixed words in Polish. To appear in Reineke Bok-Bennema & Crit Cremers (eds.) *Linguistics in the Netherlands 1994*. Amsterdam: John Benjamins. Pp. 12.
- Rubach, J. (1994). Affricates as strident stops in Polish. *Linguistic Inquiry* 25.1:119-144

Russell, K. (1995). Morphemes and candidates in Optimality Theory. Ms., University of Manitoba. Pp. 46. ROA-44.

Sagey, E.. (1986) *The representation of features and relations in non-linear phonology*. MIT PH D Dissertation.

Scobbie, J. (1991). *Attribute-value phonology*. Ph.D. Dissertation, Univ. of Edinburgh.

Shattuck-Hufnagel, S. (1986). The representation of phonological information during speech production planning: evidence from vowel errors in spontaneous speech. *Phonology* 3: 117-150

Shen, J-X. (1992). Types of slips of the tongue. *Zhongguo Yuwen* 4:306-316)

Silverman, D. (1992). Multiple scansions in loanword phonology: evidence from Cantonese. *Phonology* 9.2

Singh, R. (1987). Well-formedness conditions and phonological theory. *Phonologica* 1984. W. Dressler, ed., pp 273-286. Cambridge University Press, Cambridge.

Stanley, R. (1967). Redundancy rules in phonology. *Language* 43:393-436.

Steriade, D. (1993). Closure, release, and nasal contours. In M. Huffman and R. Krakow, eds., *Nasality* San Diego: Academic Press

Steriade, D. (1994). Underspecification and markedness. In J. Goldsmith, ed., *Handbook of phonology*. Basil Blackwell, Oxford.

Trigo, L. (1993). The inherent structure of nasal segments. In M. Huffman and R. Krakow, eds., *Nasality* San Diego: Academic Press.

Venneman, T. (1973). Phonological concreteness in natural generative grammar. In R. Shuy and C. Bailey, ed.s, *Towards tomorrow's linguistics*. Washington DC, Georgetown University Press.

Wang, S-Y (1968). The many uses of F₀. *Project on Linguistic Analysis Reports*, 2nd series 8:w1-w35. Reprinted in Valdman, A. (ed) 1972 *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre* Mouton, The Hague.

Wang, S. (1992). Nasality as an autosegment in Taiwanese. Paper presented at the First International Conference on Languages in Taiwan. National Normal University, Taipei.

Wright, M. (1983). *A metrical approach to tone sandhi in Chinese dialects*. Doctoral dissertation, U. Mass Amherst

Wu, Y-W. (1994). *Mandarin Segmental Phonology* U. of Toronto Ph D Dissertation.

Yip, M. (1980). *The tonal phonology of Chinese*. Doctoral Dissertation, MIT.

Yip, M. (1993). Cantonese loan word phonology and optimality theory. *Journal of East Asian Linguistics* 2, 261-291.

Yip, M. (1994). Morpheme-level features: Chaoyang syllable structure and nasalization. To appear in *Proceedings of the Sixth North American Conference on Chinese Linguistics (NACCL VI)*. University of Southern California.

Yip, M. (1995). Identity avoidance in phonology and morphology. To appear in the *Proceedings of the Conference on Morphology and its Relation to Syntax and Phonology*, UC Davis.

Zhang, S. Y. (1979a). Chaoyang fangyan de wenbai yidu. [Colloquial and literary strata in the Chaoyang dialect]. *Fangyan* 1979, 4.241-267.

Zhang, S. Y. (1979b). Chaoyang fangyan de chongdieshi. [Reduplication in the Chaoyang dialect]. *Zhongguo Yuwen* 1979, 2.106-114.

Zhang, S.Y. (1979c). Chaoyang fangyan de liandu biandiao. [Tone sandhi in the Chaoyang dialect]. *Fangyan* 1979, 2.93-121.

Zhang, S. Y. (1980). Chaoyang fangyan de liandu biandiao. [Tone Sandhi in the Chaoyang dialect (II)]. *Fangyan* 1980, 2.123-136.

Zhang, S.Y. (1981). Chaoyang fangyan de yuyin xitong. [An outline of Chaoyang phonology]. *Fangyan* 1981, 1.27-39.

Zhang, S.Y. (1982). Chaoyang fangyan de xiangshengzi chongdie shi. [The reduplicated onomatopoeic particles in the Chaoyang dialect]. *Fangyan* 1982, 3.181-182.

Zhu, D.X. (1982). Chaoyang hua he Beijing hua chongdie shi xiangshengci de

gouzhao. [A comparison between onomatopoeic words in reduplicated forms in the Chaoyang and Beijing dialects]. *Fangyan* 1982, 3.174-180.

[1] For example, a Mandarin word like [yɛ:²¹⁴] 'also', with the falling-rising "full third" tone and the resulting extra long vowel has been argued to have an abstract UR with a schwa, /yə²¹⁴/. The more 'concrete' UR I have in mind would not be identical to the phonetic form, but rather one degree removed, lacking the extra-long vowel necessary to realize the complex tone. It would thus be /yɛ²¹⁴/, or perhaps /yɛ:²¹⁴/, incorporating the open-syllable lengthening that satisfies a bi-moraic Min Wd requirement. In this paper, references to 'concrete' UR's, or "the speaker learns what he hears" should be taken to refer to this level of representation, which I assume to be the same for both perception and production (although this is not a logical necessity).

[2] If stop-final syllables are monomoraic, the *σ_μ constraint must be dominated by some other constraint applying to such syllables.

[3] It is quite possible that different learners may make different choices for the same vowel, perhaps even for individual words. I will not explore this possibility here.

[4] The data is a sub-set of a more complex set of vowel changes, all conditioned by the same environment. Wright (1983), Jiang-King (1995) show that the environment is μ vs. μμ. The language has no context in which underlying μ become μμ, so there are no alternations clearly showing derived ei, ou, öy diphthongs.

[5] For another view of palatalization in OT, see Noyer 1994, Rowicka 1994.

[6] Zhang (1982) gives three reduplicated forms with rhymes containing nasal vowels followed by [p]. In his 1981 paper, however, he is quite explicit that such rhymes are not possible in the core language. I have nothing to say about these three unexpected rhymes, and the analysis in this paper will account only for the core language.

[7] Chung (1995b) draws my attention to one case of optional nasalization caused by a "nasal" coda. Zhang (1981) states that there are two pronunciations for 'music': [im gau?] or [im ŋau?]. Chung concludes that codas must have a nasal specification, contra the claims of Yip 1994, and this paper. However, in this example, both syllables carry full tones, showing that they are separate phonological words. In the related dialect Taiwanese, according to Chung, there is obligatory nasal assimilation within a single phonological word, but no nasal assimilation across two phonological words. I suspect that the Chaoyang assimilation differs from Taiwanese in being a late optional phonetic process insensitive to prosodic word boundaries; since the codas are indubitably phonetically nasal, phonetic nasalization tells us nothing about their phonological specification.

[8] Somewhat similar arguments, in very different frameworks, have been made for Taiwanese by Li (1985, 1992), Wang (1992) and Chung (1995a,b).

[9] Devices such as logical disjunctions or defaults may be able to handle this data in other theories; the difference, I think, is that OT takes as fundamental the violable nature of constraints, so such cases involve no extra mechanisms or costs whatsoever.

[10] In Yip (1994) I address the issue of whether these features can be identified as morpheme level as opposed to some other possibility such as syllable-level, rhyme-level etc. On the assumption that lexical entries do not include syllable structure, features must be underlyingly properties of either segments, or morphemes. Affiliations with syllable structure must be created by the phonology, as assumed here.