Patterns, pervasive patterns and feature specification

K. David Harrison and Abigail Kaun [1]

1. Introduction

In this paper, we explore how surface apparent patterns can lead speakers to posit underspecification in lexical entries. Further, we show that current interpretations of Lexicon Optimization (Prince and Smolensky 1993), in particular that of Archiphonemic Underspecification (Inkelas 1995), incorrectly predict the distribution of underspecification in lexical entries. We present new data and advance two theoretical claims bearing on the above issues. First, we claim that predictably alternating feature values are not necessarily underspecified. In support of this claim we introduce patterns of reduplication in Hungarian word games. Second, we claim that segments that never alternate may still be underspecified. In support of this claim we introduce vowel harmony patterns from Tuvan (or Tyvan), a Turkic language of Siberia. Both claims demonstrate that the incidence of underspecification is neither that which is predicted by Archiphonemic Underspecification, nor that predicted by earlier Auto-segmental models. We point the way toward a revised model that would diagnose environments in which underspecification occurs. The model is designed to allow for a principled distinction between pervasive patterns that give rise to underspecification, and other patterns that do not.

Transparency to spreading (or assimilation), susceptibility to spreading, failure to initiate spreading and various other types of behavior prompted analysts working within a derivational model of phonology to hypothesize that certain featural specifications are absent for the relevant segments throughout at least a portion of the phonological derivation. Contrastive Underspecification (Steriade 1987) and Radical Underspecification (Archangeli 1984) were the dominant formal models designed to predict in a principled manner the incidence of underspecification. Contrastive Underspecification theory posited non-contrastiveness as the criterion for potential underspecification. A feature value might be missing from underlying representations if it failed to serve a contrastive function for the segment class in question. Under Radical Underspecification, the commitment to a redundancy-free lexicon had the consequence of eliminating all predictable feature values from underlying representations.

Within Optimality Theory (OT), Richness of the Base specifically rules out the systematic exclusion of featural specifications from input representations (Prince and Smolensky 1993, Smolensky 1995). The input space is assumed to be infinite, thus unrestricted. The lexicon, by contrast, is assumed to be finite. A learner's construction of lexical representations is guided by Lexicon Optimization (Prince and Smolensky 1993), which heavily favors fully specified inputs. It is assumed that a speaker will choose the most harmonic (i.e., the fullest) input-to-output mapping. Outputs can be mapped to fully specified inputs without the accrual of gratuitous faithfulness violations. Given two competing input forms, one fully specified and one partially specified, the fully specified alternative will be preferred, all else being equal. The OT model nevertheless leaves room for the possibility that partially underspecified lexical entries will on occasion be posited.

Archiphonemic Underspecification (Inkelas 1995) seeks to predict when underspecified inputs will in fact be deployed. It demonstrates that the principles of Lexicon Optimization dictate that predictable feature values will be underspecified only when they enter into surface alternations. Underspecification will arise only in instances where there are fully regular and predictable alternations, as the following schematic shows:

(1)

	Alternating	*Non-alternating*
Predictable	**underspecified**	**specified**
Unpredictable	**specified**	**specified**

In a backness harmony language like Turkish, for instance, only those vowels that are involved in allophonic alternations will be underspecified for backness in lexical entries. In the words given in (2), all post-initial vowels agree in backness with the initial vowel.

(2) Two words of Turkish

a. kilim-im 'rug-1'

b. kɨlɨc-ɨm 'sword-1'[2]

However, it is only the suffix vowels that both alternate and have a predictable value for backness. Archiphonemic Underspecification predicts that only they will be underspecified for backness, as indicated in (3):

(3) Assumed Lexical Representations

```
a. kilim-im          b. kɨlɨc-ɨm
     | |                   | |
    -B -B                 +B+B
```

This model essentially claims that harmony targets only suffix vowels in Turkish, because it is only for suffix vowels that a backness value lacking in the input representation is introduced in the corresponding output. Root vowels, whether harmonic or not, will be fully specified (and presumably identical) in input and output representations. Root vowels, on this view, even though they may appear trivially to obey harmony, cannot be thought of as undergoing harmony. Clements and Sezer (1982, p. 226), motivated by entirely different theoretical considerations, take a similar position, stating that "…the burden of proof is on the linguist who wishes to demonstrate that roots [in Turkish] are governed by vowel harmony at all."

In section 3.1, we introduce vowel harmony patterns from Tuvan. These demonstrate that speakers can underspecify not only alternating vowels in affixes but also non-alternating vowels within roots. The Tuvan data are corroborated by speakers' distinct treatment of harmonic and disharmonic roots in two other vowel harmony languages: Finnish and Turkish. The patterns that we discuss for all three languages indicate that harmony is indeed active within roots. We thus challenge the position advanced by Clements and Sezer (1982) and the predictions of Archiphonemic Underspecification.

Before turning to the harmony data, we first discuss Hungarian word game data in which full specification allows for the only satisfactory account of the observed patterns. The Hungarian data are problematic for Archiphonemic Underspecification because while the game forms require across-the-board full specification, only a subset of the relevant segments are predicted by the model to be fully specified.

2. Predictable and alternating segments are not necessarily underspecified

2.1. Hungarian word games

Hungarian is relevant to the present discussion because Hungarian vowels exhibit a number of predictable feature values, some of which participate in alternations and others which do not. Those that enter into surface alternations are predicted by Archiphonemic Underspecification to be underspecified in lexical entries, while those that maintain a fixed, non-alternating value are predicted to be fully specified.

The vowel inventory of Hungarian is given in (4). Vowel length is contrastive, as shown. Crucial to the argument developed below is the fact that the long and short variants of the vowels represented by orthographic <e> and <a> are qualitatively quite distinct. Long mid [eː] alternates with short low vowel [æ], and long unrounded [ɑː] alternates with the short rounded vowel [ɔ]. The predictably alternating pairs are enclosed in boxes below.

(4) Hungarian Vowel Inventory

	Front		*Back*
High	i, iː	ü, üː	u, uː
Mid	eː	ö, öː	o, oː
			ɔ
Low	æ		ɑː

Within the frameworks of both Contrastive Underspecification and Radical Underspecification, these quality differences would be probable candidates for underspecification due to their predictability from length. Quantity is demonstrably contrastive for the vowel system of Hungarian as a whole. In Optimality Theory, however, the principles of Lexicon Optimization lead us to expect the accompanying quality differences to be explicitly recorded in lexical entries when they do not enter into surface alternations. Only in cases where surface alternations do occur does the model predict underspecification.

In certain nouns, the vowel of the final syllable is at times short, and at other times long, as we shall show. For the majority of vowels, this length difference does not substantively affect vowel quality, however for the vowels indicated in boxes in the chart in (4) above, the length difference is accompanied by the expected quality differences. Thus, vowels in the first syllable of so-called 'shortening bases' (Whitney 1949) are long when the noun is unaffixed, but short when certain affixes are added, e.g., *uːr* 'master,' and *urɔk* 'master-PL,' *u̲rɔ* 'master-3,' and *e̲ːr* 'vein,' but *æ̲ræk* 'vein-PL' and *æ̲ræi* 'vein-3.' In another class of nouns, the vowel of the final syllable is short when unaffixed, but long when certain affixes are added, e.g., *kæfæ̲* 'brush,' but *kæfe̲ːk* 'brush-PL' and *fɔ̲* 'tree,' but *fa̲ːk* 'tree-PL.' Note, however, that it is not the case that all affixes trigger length alternations of these types of stems. It is instead particular morphemes that have this property.

There are also stem vowels that resist length alternations entirely. These can be either consistently long or consistently short. For example we finds roots such as *keːp* 'picture' and *haːz* 'house,' which retain their long vowel under plural affixation, yielding *keːpæk* 'picture-PL' and *haːzɔk* 'house-PL' (not **kæpæk* or **hɔzɔk)*. Similarly, we find roots like *kært* 'garden,' which retain the short vowel under suffixation, yielding *kærtæk* 'garden-PL' (not **keːrtæk*, for example).

For the two alternating noun-classes, Archiphonemic Underspecification predicts underspecification of the predictable vocalic features. The model predicts full specification for the non-alternating classes and for all vowels occurring in pre-final syllables.

Evidence bearing on this prediction comes from a Hungarian reduplicative word game similar to the English 'Ubbi Dubbi' game.[3] The Hungarian game, known as *Veve*, works as follows.[4] For a given word, a sequence /-Vv-/ is inserted before the rhyme of each syllable. The overall quality (feature values for height, backness and rounding) of the reduplicant V is identical to that of the following vowel. We report here on two versions of the game (Veve I and Veve II). In both versions, the reduplicated vowel is always short: this is a 'rule' of the game. In Veve I, the length (quantity) of base vowels is retained, as shown in (5). In Veve II, however, all vowels, both reduplicant and base, surface as short. Some examples of this are shown in (6).

(5) Veve I examples

	base	base+reduplicant	
a.	itt	iv-itt	‘here’
b.	tiːz	t-iv-iːz	‘ten’
c.	sæm	s-æv-æm	‘eye’
d.	eːr	ev-eːr, *æv-eːr	‘vein’
e.	bɔb	b-ɔv-ɔb	‘bean’
f.	ɑːr	ɑv-ɑːr, *ɔv-ɑːr	‘price’
g.	neːvmɑːʃ	n-ev-eːvm-ɑv-ɑːʃ, *n-æv-eːvm-ɔv-ɑːʃ	‘pronoun’

(6) Veve II examples

	base	base+reduplicant	
a.	itt	iv-itt	‘here’
b.	tiːz	t-iv-iz	‘ten’
c.	sæm	s-æv-æm	‘eye’
d.	eːr	ev-er, *æv-ær	‘vein’
e.	bɔb	b-ɔv-ɔb	‘bean’
f.	ɑːr	ɑv-ɑr, *ɔv-ɔr	‘price’
g.	neːvmɑːʃ	n-ev-evm-ɑv-ɑʃ, *n-æv-ævm-ɔv-ɔʃ	‘pronoun’

Note that by doing away with all length distinctions, as is done in Veve II, it becomes unclear which portion of the string constitutes the reduplicant, and which constitutes the base. Thus, the reduplicant could be construed as either -Vv- or as -vV-. For concreteness, we will assume that, as in Veve I, speakers of Veve II insert the sequence -Vv-.

Of particular importance are examples d, f, and g, shown in (5) and (6). In each of these forms the reduplicants contain vowels that never surface outside the context of Veve. The examples in (5d) and (6d) contain short [e] while those in (5f) and (6f) contain short [ɑ].

The examples in (5g) and (6g) contain instances of both of these otherwise non-surfacing vowels.

2.2. A formal account of Hungarian

A Correspondence-based analysis of the pattern (McCarthy and Prince 1995), which we shall adopt herein, might run as follows. First, we represent the absence of [æː], [e], [ɔː], [ɑ] in the general inventory of Hungarian as a constraint on inventory structure, given in (7).[5] (cf. Tableau A, third column)

(7) Inventory Structure (*æː,*e,*ɔː,*ɑ) **IS**

Following the analysis of correspondence developed in McCarthy and Prince (1995), we invoke two general faithfulness constraints. The first requires identity between input and output forms with regard to quantity (length).

(8) Input-Output Identity for length **IDENT-I/O[LG]**

Any output vowels that do not match their corresponding input vowel in length will run afoul of this constraint (Tableau A, second column). A second constraint requires identity between input and output forms with regard to quality (i.e. all vowel features other than length).

(9) Input-Output Identity for quality **IDENT-I/O[QUAL]**

Any vowel in either base or reduplicant that does not match the input vowel in all features will violate this (Tableau A, fourth column).

We begin with an account of the basic, non-reduplicated forms. Given that Richness of the Base requires that no restrictions be imposed on input structures, it must be the case that inputs containing vowel qualities that are not part of the Hungarian vowel inventory must be mapped to outputs containing only legitimate vowels. This means that INVENTORY STRUCTURE must outrank at least one of the two faithfulness constraints.

A general principle of optimality theory is that a given constraint hierarchy must be able to map any input onto some well-formed output. So, for example, an input such as /er/ (which is not an actual or even a possible word of Hungarian, because it contains [e]) will be mapped to the well-formed surface form [ær] if IDENT-I/O[QUAL] is low-ranked (tableau A). Note that while the output [ær] is well-formed and thus a possible word of Hungarian, it is not an actual word.

(10) Tableau A

/er/	**IDENT-I/O[LG]**	**IS**	**IDENT-I/O[QUAL]**
a. eːr	*!		
☞ b. ær			*
c. er		*!	
d. æːr	*!	*	

Alternatively, if IDENT-I/O[LG] is low-ranked, then the surface form [eːr] (corresponding to an actual Hungarian word meaning 'vein') will be selected.

(11) Tableau B

/er/	**IDENT-I/O[QUAL]**	**IS**	**IDENT-I/O[LG]**
☞ a. eːr			*
b. ær	*!		
c. er		*!	
d. æːr	*!	*	*

We have shown that there is more than one constraint ranking that will give rise to a licit vowel on the surface when the input contains an vowel that violates the inventory structure constraint. The difference between Veve I and Veve II demonstrates that at least two distinct rankings are posited by speakers of what would appear to be the same dialect of Hungarian (i.e., educated Budapest). Speakers of Veve I invoke the general constraint ranking in tableau A while speakers of Veve II invoke that in tableau B. The full rankings for Veve I and Veve II are given in tableaux C and D respectively.

Two additional correspondence-theoretic (McCarthy and Prince 1995) constraints will be posited to characterize the patterns of Veve. These require the base and the reduplicant to be identical in quantity and quality, respectively. Note that 'base' refers to the *surface* form, which the reduplicant must resemble, and not to the input (underlying) form. Identity between base and reduplicant is thus independent of the degree to which either element is faithful to the input form. In this analysis, we do not invoke a separate constraint on input-reduplicant identity.

(12) Base-Reduplicant Identity for length **IDENT-B/R[LG]**

(13) Base-Reduplicant Identity for quality **IDENT-B/R[QUAL]**

In order to ensure a quality-match between a vowel of the base and its corresponding vowel in the reduplicant, the Veve forms allow otherwise unattested vowels (specifically [e] and [ɑ]) to surface. Base-Reduplicant Identity must therefore outrank INVENTORY STRUCTURE. For simplicity, we will only consider Veve candidates in which the reduplicant vowel is short. Both documented versions of the game require short reduplicant vowels and we assume that this restriction is undominated. Thus for input /eːr/, there is no possible Veve output [eːveːr].

The constraint rankings are as follows: for Veve I, the faithfulness constraints IDENT-B/R[QUAL] and IDENT-I/O[LG] outrank INVENTORY STRUCTURE, as shown in tableau C for /eːr/ 'vein.' A single violation of IDENT-B/R[QUAL] is assigned for each vowel in the reduplicant (underlined) that fails to match its corresponding base vowel in one or more features. A violation of IDENT-I/O[LG] is assigned for every vowel in the base (but not in the reduplicant) that fails to match its corresponding input vowel in length. No violations are assigned for a length mismatch between input and reduplicant, as I-R identity is superfluous to the present analysis.

(14) Tableau C: Veve I with full specification

/Vv, eːr/	IDENT-I/O [LG]	IDENT-B/R [QUAL]	IS	IDENT-I/O [QUAL]	IDENT-B/R [LG]
☞ a. eveːr			*		*
b. æveːr		*!			*
c. ever	*!		**		
d. æver	*!	*	*		
e. evæːr		*!	**	*	*
f. ævæːr			*	*!	*
g. evær	*!	*	*	*	
h. ævær	*!			*	

For Veve II, both base-reduplicant faithfulness constraints outrank IS, as shown. Note that the Input-Output constraint IDENT-I/O[QUAL] dominates IS, giving rise in the optimal candidate (c) to the usually ill-formed surface quality [e].

(15) Tableau D: Veve II with full specification

/Vv, eːr/	IDENT-B/R [LG]	IDENT-B/R [QUAL]	IDENT-I/O [QUAL]	IS	IDENT-I/O [LG]
a. eveːr	*!			*	
b. æveːr	*!	*			
☞ c. ever				**	*
d. æver		*!		*	*
e. evæːr	*!	*	*	**	
f. ævæːr	*!		*	*	
g. evær		*!	*	*	*
h. ævær			*!		*

The word *eːr* belongs to the so-called shortening class (its plural form is *ær-æk*). The root vowel alternates predictably, surfacing as either [eː] or [æ], depending on its length. Archiphonemic Underspecification thus predicts the vowel should be underspecified for some feature(s). Since length is contrastive for the Hungarian vowel system as a whole, we assume that Archiphonemic Underspecification would predict underlying suppression of the quality features (most significantly, the height difference between [eː] and [æ]). If the input for alternating words like *eːr* is underspecified, however, the Veve forms cannot be accounted for. We repeat the above tableaux, this time allowing for the predicted underspecification of vowel quality, indicated in the input below by an archiphone /E/.

(16) Tableau E: Veve I with underspecification

/Vv, Eːr/	IDENT-I/O [LG]	IDENT-B/R [QUAL]	IS	IDENT-I/O [QUAL]	IDENT-B/R [LG]
☞ a. eveːr			*		*
b. æveːr		*!			*
c. ever	*!		**		
d. æver	*!	*	*		
e. evæːr		*!	**		*
☞ f. ævæːr			*		*
g. evær	*!	*	*		
h. ævær	*!				

(17) Tableau F: Veve II with underspecification

/Vv, Eːr/	IDENT-B/R [LG]	IDENT-B/R [QUAL]	IDENT-I/O [QUAL]	IS	IDENT-I/O [LG]
a. eveːr	*!			*	
b. æveːr	*!	*			
Λ c. ever				**!	*
d. æver		*!		*	*
e. evæːr	*!	*		**	
f. ævæːr	*!			*	
g. evær		*!		*	*
M h. ævær					*

Under this scenario, the Veve I grammar shown in tableau E has no basis for choosing between candidates (a) *eveːr* and (f) æ*vææːr*, due to the fact that IDENT-I/O[QUAL] does not assign any violations. Similarly, the Veve II grammar in tableau F selects (h) æ*vær,* shown with a 'bomb,' as a better candidate than the attested (a) *ever*, shown with a 'sad face.' The adoption of underspecified representations thus fails in this case to select the attested output.

Finally, Hungarian children reportedly play a word game in which they sing a ditty (18) while replacing all the vowels with a single vowel. The replacement vowel may be any vowel from the Hungarian inventory. The real song goes like this:

(18) seːp ɔz itsi pitsi nöːj tsipöː
'beautiful' 'that' 'teeny tiny' 'woman's' 'shoe'
'Teeny tiny women's shoes are beautiful.'

Given a replacement vowel [ɔ], the song is sung as follows:

(19) sɔːp ɔz ɔtsɔ pɔtsɔ nɔːj tsɔpɔː

Given replacement vowel [æ], it sounds like this:

(20) sæːp æz ætsæ pætsæ næːj tsæpæː

In 'itsi pitsi,' vowel quality may be changed, but the vowel quantity of the input is always preserved. This scenario potentially yields all four of the INVENTORY STRUCTURE violations listed in (7). In (19) and (20), IS violations are presumably assigned for long [ɔː] and long [æː]. If the replacement vowel were instead [a] or [e], then IS violations would be incurred for each occurrence of short [a] and [e].

The 'itsi pitsi' facts raise the same problem for Archiphonemic Underspecification we exposed in Veve. Namely, no surface difference between putatively specified and underspecified vowels (i.e., predictably alternating and non-alternating) can be discerned. To the contrary: Across-the-board full specification is required. One might attempt an explanation of such word games based on the claim that a reduplicative game, as a morphological process, requires that the input contain a well-formed (i.e., fully specified) word. On this view, underspecified inputs such as those in tableaux E and F would never be submitted to the grammar for analysis.

This explanation cannot be correct, however, as we will show in the remainder of this paper. In the harmony cases to be discussed below, we show that speakers' performance in novel reduplicative tasks indicates that underspecified inputs are indeed utilized. Moreover, the particular distribution of underspecified features is explicitly not that which the principles of Lexicon Optimization and Archiphonemic Underspecification would lead us to expect.

3. Predictable but non-alternating segments may be underspecified

We will present empirical evidence from three vowel harmony languages—Tuvan, Finnish, and Turkish—which poses a further problem for the predictions of Archiphonemic Underspecification. In these languages, we argue, predictable segments must be underspecified even though they are non-alternating. If non-alternating segments are underspecified then alternation is not an adequate diagnostic of underspecification. Our claim that speakers underspecify non-alternating segments rests on the different patterning of harmonic vs. disharmonic vowels in the novel context of reduplication and re-harmonization.

3.1. Tuvan reduplication

Tuvan (Anderson and Harrison 1999) has an eight-vowel inventory, plus contrastive length.

(21) Tuvan vowel inventory

	Front		*Back*	
High	i	ü	ɨ	u
Non-high	e	ö	a	o

Like most Turkic languages, Tuvan enforces strict backness harmony: only front vowels [i ü e ö] or back vowels [u ɨ a o] may co-occur within a word. Backness harmony is fully productive within roots and affixes. Nonetheless, the language tolerates some disharmony in loanwords, native compound forms, and in one exceptional non-alternating morpheme. Tuvan also has rounding harmony, simply characterized by two basic principles: (i) the vowels [ö o] may not occur in post-initial syllables, and (ii) a high vowel must be rounded [ü u] when it follows any rounded vowel [ü u ö o], otherwise it must be unrounded. As we will show, harmonic and disharmonic vowels pattern differently under reduplication. Our claims about underspecification and harmony rest on this difference in patterning. In this paper, we limit our attention to backness harmony. The interaction of rounding harmony with reduplication is considerably more complex (Harrison 1999, Kaun 2000), but consistent with our general analysis.

Tuvan has a morphological process of full reduplication (Harrison 1999, 2000). For example *nom* 'book' when reduplicated becomes *nom-nam* 'books and the like'. Its use is restricted, however, to a subset of speakers, to a special register, and to certain dialects. Thus, not all native speakers have been exposed to this type of reduplication. Nonetheless, the process is sufficiently transparent that we were able to teach it to both adults and young children in a matter of seconds. Speakers who had just learned the rule were able to produce novel reduplicants, and their output matched that of speakers who use reduplication regularly. The potentially *novel* character of reduplication for some speakers will be central to our argument.

Reduplication takes the entire base and repeats it, while replacing the vowel of the initial syllable. The replacement vowel is a pre-specified [a], except when the base vowel happens to be [a], in which case the replacement vowel is [u] (22i, j).

(22) Full reduplication of monosyllabic bases

	base	base + reduplicant	
a.	nom	nom-nam	'book'
b.	er	er-ar	'male'
c.	seːk	seːk-saːk	'mosquito'
d.	is	is-as	'footprint'
e.	ög	ög-ag	'yurt'
f.	süt	süt-sat	'milk'
g.	qɨs	qɨs-qas	'girl'
h.	xol	xol-xal	'hand'
i.	at	at-ut	'name'
j.	aːr	aːr-uːr	'heavy'

There is clearly a faithfulness relation between the base and reduplicant: except for the replacement vowel, the two are always identical.

3.2. Tuvan re-harmonization

However, reduplicated polysyllabic roots (along with any suffixal morphology) may surface as considerably less faithful to their bases. In polysyllabic forms, post-initial vowels generally agree in backness with the replacement vowel. Note that since the replacement vowel is either [a] or [u], it is always [+back]. To achieve this agreement, speakers subject post-initial vowels to re-harmonization. To call attention to potential re-harmonization effects, we underline all post-initial vowels in reduplicants.

(23) Full reduplication of polysyllabic bases with re-harmonization

a.	idik	idik-adɨk (*adik)	'boot'
b.	fiːdik	fiːdik-faːdɨk (*faadik)	'video cassette'[6]
c.	teve	teve-tava (*tave)	'camel'
d.	tevelerim	tevelerim-tavalarɨm (*tavelerim)	'camel'-PL-1

Disharmonic segments, whether native or borrowed, fail to undergo re-harmonization, and remain disharmonic.

(24) Full reduplication of polysyllabic bases with no re-harmonization

a.	maʃina	maʃina-muʃina (*muʃɨna, *muʃuna)	'car'
b.	ajbek	ajbek-ujbek (*ujbak)	'Aibek' (name)
c.	ʒiguli	ʒiguli-ʒaguli (*ʒagulɨ,*ʒagulu)	'Zhiguli' (car)
d.	aːl=ʒe	aːl=ʒe-uːl=ʒe (*uːl=ʒa)	'yurt'=ALL[7]

3.3. A formal analysis of Tuvan

Adopting an Optimality Theoretic framework, we model Tuvan harmony and reduplication with the following constraints.

Morphological vowel replacement rule	**REPLACE.V1**
Backness Harmony	**ALIGN[BK]**
Input-Reduplicant Identity	**IDENT-I/R**
Base-Reduplicant Identity	**IDENT-B/R**

The first of these, REPLACE.V1, is a Tuvan-specific morphological constraint that simply requires the speaker to replace the vowel when doing reduplication. The actual choice of replacement vowel, as we have argued elsewhere, falls out from general markedness constraints (Harrison 1999, 2000). The alignment constraint is adopted from Smolensky (1993). Though it employs the metaphor of 'aligning' underlying features with the edge of the word domain, it could be equally well construed as a garden-variety constraint on surface well-formedness. We will thus assign an alignment violation in the tableaux below to vowel segments which fail to agree in backness with a vowel in an adjacent syllable. The identity constraints, adopted from McCarthy and Prince (1995), require the surface form of the reduplicant to be identical with the *input* form and the *base* (surface form), respectively.

To begin, let's consider an input form such as *idik* 'boot.' In accord with the predictions of Lexicon Optimization, we assume a fully specified representation for this harmonic form, even though the backness value of the second vowel is predictable on the basis of the backness of the initial vowel. For *idik*, as for all harmonic words, the following constraint ranking allows us to capture the pattern of re-harmonization:

(25) REPLACE.V1 >> ALIGN[BK] >> IDENT-I/R, IDENT-B/R

Later, we will have cause to modify this ranking (28, 29). But first, to show how the constraints interact, consider Tableau G, where we have included all constraints except for REPLACE.V1 (the Tuvan-specific morphological constraint).

(26) Tableau G: Harmonic form with fully specified input.

/idik, RED/ | / -B	ALIGN[BK]	IDENT-I/R	IDENT-B/R
a. idik-adik	*!* [8]	*	*
☞ b. idik-adɨk		**	**

The attested output form can only emerge if the harmony constraint ALIGN[BK] outranks both faithfulness constraints. This is due to the fact that the unattested candidate (a) is more faithful to the input, both in terms of IDENT-I/R and IDENT-B/R. Re-harmonization under reduplication provides ample evidence for the undominated status of the harmony constraint.

But for a disharmonic form with a fully specified input, the ranking shown in (25) selects the wrong candidate, as shown in tableau H. Note that for disharmonic forms, we assess a violation of ALIGN[BK] for each vowel that fails to align its underlying backness specification within its harmony domain (here, the base or the reduplicant, taken separately, cf. endnote 8).

(27) Tableau H: Disharmonic form with fully specified input.

/ajbek, RED/ | | +B -B	ALIGN[BK]	IDENT-I/R	IDENT-B/R
Λ a. ajbek-ujbek	***!*	*	*
M b. ajbek-ujbak	**		**

The attested disharmonic candidate (a) should win, but does not (Λ). Instead, the unattested candidate (b) wins (M). For candidate (a) to win, the harmony constraint must rank below at least one of the faithfulness constraints. We thus propose an alternative ranking in which ALIGN[BK] ranks below input-reduplicant faithfulness. Again, we assume a fully specified input.

(28) Tableau I: Disharmonic form with fully specified input (new ranking)

/ajbek, RED/ | | +B -B	IDENT-I/R	ALIGN[BK]	IDENT-B/R
☞ a. ajbek-ujbek	*	****	*
b. ajbek-ujbak	**!	**	**

This ranking correctly selects the attested disharmonic form. We are thus faced with a ranking paradox by which harmonic and disharmonic sequences seem to require separate constraint rankings (i.e., separate grammars). This apparent paradox may be resolved if we allow harmonic words to be represented by underspecified inputs (contra the predictions of Lexicon Optimization and Archiphonemic Underspecification).

(29) Tableau J: Harmonic form with underspecified input

/idik, RED/ | -B	IDENT-I/R	ALIGN[BK]	IDENT-B/R
a. idik-adik	*	*!*	*
☞ b. idik-adɨk	*		**

Since disharmonic forms do not undergo re-harmonization (tableau H), while harmonic ones do (tableau G), we have assumed the former are fully specified for the harmonic feature and the latter are partially underspecified. Underspecification thus has a desirable result in that it obviates the need to posit separate constraint rankings for various subsets of the lexicon. Partial underspecification has the following consequence, illustrated in tableau J. A faithfulness violation is incurred only for the vowel of the initial syllable in each candidate. The underspecified, non-initial vowels undergo 'cost-free' re-harmonization (candidate b). In tableau I, on the other hand, both vowels are fully specified such that an output which obeys harmony (candidate b), does so at a cost, namely the violation of input-reduplicant faithfulness. The same disparity in the treatment of harmonic and disharmonic vowels is found in Finnish (section §3.4) and Turkish (section §3.5).

3.4. A Finnish word game

Facts similar to those of Tuvan have been documented in a Finnish reduplicative word game known as *kontti kieli* ‘knapsack language’ (Campbell 1986, Vago 1988). The game adds the word *kontti* ‘knapsack’ after a word, then preposes the initial (C)V sequences of the two words. Speakers then re-harmonize the remaining vowels according to front/back harmony, with the exception of the neutral [e] and [i]. As in Tuvan, speakers re-harmonize harmonic segments (30a, b), but consistently fail to do so with disharmonic ones (30c, d). Potential re-harmonization targets are underlined in the data below.

(30) Finnish *kontti kieli*

a.	mitä kontti ‘what’ ‘knapsack’	◊	ko-ta mi-ntti(*ko-tä)
b.	sikiö kontti ‘embryo’ ‘knapsack’	◊	ko-kio si-ntti (*ko-kiö)
c.	konglööri kontti ‘juggler’ ‘knapsack’	◊	ko-nglööri jo-ntti (*ko-ngloori)
d.	manööveri kontti ‘maneuver’ ‘knapsack’	◊	ko-nööveri ma-ntti (*ko-nooveri)

Note that root vowels of Finnish never alternate except within the special context of *kontti kieli*. These results show that in the novel context of a language-external word game, speakers treat harmonic and disharmonic segments differently. Campbell (1986) used this to argue for the psychological reality of a rule of vowel harmony within Finnish, even in the face of numerous surface counter-examples to harmony. From our perspective, the Finnish data are comparable with Tuvan and may be analyzed in the same way, with the same implications.

We have claimed that speakers of vowel harmony languages such as Tuvan and Finnish underspecify predictable but non-alternating segments within roots. This claim entails that vowel harmony must be an active process even in cases where it appears to be little more than a static, phonotactic pattern. In Tuvan and Finnish, we have presented evidence that the quality of root vowels is established by means of alignment (Kaun 2000), except in the cases of disharmonic roots. We turn now to Turkish, where we extend our argument that harmony (rendered formally as alignment) actively determines feature values of all harmonic segments, even those that never enter into surface alternations. We argue that Turkish vowel co-occurrence patterns within roots are not merely static, phonotactic patterns but active harmonic processes that have psychological reality and accessibility for speakers (contra Clements and Sezer 1982).

3.5. Turkish root harmony

Turkish has both backness harmony and a vowel inventory like that found in Tuvan, but it lacks a Tuvan or Finnish-style reduplication process that would subject roots to novel alternations. Turkish root vowels thus never alternate in any context. As a pilot study, we taught a Tuvan-style reduplication rule to two speakers of Turkish. The speakers were instructed to replace the initial vowel of a set of 40 Turkish words (real lexemes of Turkish, both harmonic and disharmonic) with [a] or [u]. After making the replacement, they could then make any other changes—or no changes at all—to the reduplicant. The resulting form (the reduplicant) was to be a nonsense word that ‘sounds like a Turkish word.’ The speakers produced multiple reduplicants for some bases; they then selected the best-sounding ones. Preliminary results of our pilot study show effects quite similar to those of Tuvan. The Turkish speakers showed a clear preference for re-harmonizing harmonic words according to the pervasive pattern of backness harmony (31a, b). But they generally failed to re-harmonize disharmonic words (31c, d). Starred forms given below show speakers’ failure to apply backness harmony to reduplicants. (We note that in some cases speakers also failed to apply rounding harmony, e.g., *muli̱*. The facts of rounding harmony are more complex and left for further study.)

(31) Turkish novel reduplication

a.	kibrit	◊	kabrɨ̱t (*kabri̱t)	‘match’
b.	bütün	◊	batɨ̱n (*batü̱n, *bati̱n)	‘whole’
c.	mali	◊	muli̱ (*mulɨ̱ *mulü̱)	‘Mali’
d.	butik	◊	bati̱k (*batɨ̱k)	‘boutique’

Our Turkish speakers, like Tuvan or Finnish speakers, apparently manipulated underlyingly unspecified segments by re-harmonizing them in this novel, reduplicative context. A fuller study of Turkish speakers’ harmony preferences under reduplication will, we predict, provide additional empirical evidence that speakers of harmony languages underspecify predictable but non-alternating segments. They do so, we suspect, in response to an observed harmonic pattern attested in most (but by no means all) lexemes. We hypothesize that the presence of a pervasive pattern of vowel co-occurrence in roots, in combination with regular alternations in suffixes is sufficient to drive speakers to posit a general system of vowel harmony that obtains across both roots and affixes.

4. Systematicity vs. idiosyncrasy

We have shown that feature values, though they may be both predictable and alternating, are not necessarily underspecified. This was the case in Hungarian. Moreover, we have shown that feature values that are predictable but non-alternating may be underspecified. This was the case in the harmony patterns of Tuvan, Finnish, and Turkish. These findings indicate that Inkelas' schematic chart (1) depicting the distribution of underspecification must be revised as follows:

(32)

	Alternating	*Non-alternating*
Predictable	**underspecified *or* specified**	**specified *or* underspecified**
Unpredictable	**specified**	**specified**

Given the indeterminacy of this revised model in predicting the exact distribution of underspecification, we propose an alternative and more flexible model represented in (33). This schematic differs from Inkelas' in several respects. First, for 'alternating' and 'non-alternating,' we substitute 'systematic' and 'idiosyncratic.' Secondly, rather than construing these as discrete categories, we recognize that they constitute a continuum. We do the same for the categories 'predictable' and 'unpredictable'.

(33)

	Systematic ←	→ *Idiosyncratic*
Predictable ↕	**underspecified**	**specified**
Unpredictable	**specified**	

Between the poles of systematic and idiosyncratic, and similarly between the poles of predictable and unpredictable, there is, we predict, an identifiable boundary (shown as a thick, shaded line above). outside this boundary, speakers will cease to deploy underspecification.

It is not yet clear how far along the Systematic-Idiosyncratic continuum a particular pattern may progress before speakers are no longer tempted to view the pattern as predictable. Harmony languages, we have observed, tolerate a considerable amount of disharmony (the unpredictability factor introduced by disharmonic lexemes or morphemes) while still retaining an overall systematicity (the harmony system as a whole). We thus pose the question, "what counts as a *pervasive* pattern?" The real empirical challenge lies in mapping out the two continua and in demonstrating that a given (pervasive) pattern fits at a certain point along them. We have proposed the location of several such patterns herein.

The vowel quality-length connection in Hungarian is indeed predictable. However, the incidence of alternation is idiosyncratically conditioned by word-class. Thus, our model predicts full specification in Hungarian. In Tuvan, Turkish, Finnish, and other harmony languages, predictability is manifested in both alternating and non-alternating segments. Nonetheless, a pervasive pattern permeates the language as a whole. This sort of systematicity correlates with underspecification, we claim. Idiosyncrasy in these harmony languages is limited to an exceptional class of disharmonic foreign borrowings and a very restricted class of non-alternating morphemes. For idiosyncratic/unpredictable vowels in these classes, we of course predict full specification.

In addition to the cases we have discussed in this chapter, our model allows languages to be located at different points along the continua. For example, we would expect a language having strict vowel co-occurrence patterns like those of Turkish and Tuvan, but lacking any affixal morphology at all, to exhibit underspecification of predictable feature values. Over time, we might expect a language to shift its position along the continua. This is, we hypothesize, what has taken place in Uzbek, a Turkic language that appears to be in the process of losing its vowel harmony system. As a greater proportion of disharmonic words enters the language, the vowel harmony pattern in Uzbek becomes less systematic and more idiosyncratic. As this occurs, speakers may at some point cease to consider the pattern predictable and will no longer posit underspecification.

Our goal is to construct a model that will allow us to characterize precisely the circumstances under which speakers will posit abstract lexical entries. This research program, which we have referred to as *pattern-responsive lexicon optimization*, is outlined here and in Harrison and Kaun (2000). An adequate model, we propose, should anticipate speakers' propensity to underspecify in response to pervasive, surface-true patterns (the prime example being vowel harmony), and their propensity to fully specify in response to more limited or idiosyncratic patterns.

[1] We would like to thank the participants at the Conference on Distinctive Feature Theory (October 1999) at ZAS Berlin for their insightful comments on an earlier version of this paper. We also wish to acknowledge the contributions of our many Tuvan, Hungarian and Turkish language consultants. Funding for fieldwork in Tuva in 1998 was provided by the International Research and Exchanges Board (IREX).

[2] The tag '-1' denotes first person (singular). All tags herein follow the EUROTYPE Guidelines of the Committee on Computation and Standardization, European Science Foundation (1993).

[3] 'Ubbi Dubbi' is an English speech disguise game that was propagated by the 1970's children's television show 'Zoom,' which aired in the United States. The game works by inserting [ʌb] before the rhyme of each syllable within a word.

[4] We would like to thank our Hungarian consultants Ms. B. Ürögdi and Mr. B. Varady for their patience and generosity in helping us document Hungarian word games.

[5] We find violations of INVENTORY STRUCTURE in the context of Veve as we have just seen, where short [e] and [ɑ] are allowed to surface. Such violations are observed elsewhere in Hungarian: for some speakers foreign diphthongs are

realized as long monophthongs, e.g. *europa* ◊ [æː]ropɔ, and *automata* ◊ [ɔː]tomɔtɔ. Here we find long vowels [æː] and [ɔː] surfacing despite their violation of IS.

[6] Note that accidentally harmonic loanwords e.g., *fiːdik* (23b) also undergo re-harmonization, even though they may violate other phonotactic constraints of the language. For instance, Tuvan has no native phoneme [f], however the loanword *fiːdik* 'video cassette' is fully harmonic and treated like any native, harmonic word with respect to re-harmonization. This shows that an analysis invoking two grammars, one for native vocabulary and one for borrowed or foreign vocabulary, is untenable.

[7] The Tuvan Allative marker /=tʃe/ is an enclitic.

[8] We assess violations of ALIGN[BK] within the domain of the base or reduplicant, taken separately. For example *idik-adik* accrues just two violations, one for each vowel in the reduplicant (underlined). If a disharmonic base and reduplicant e.g., *ajbek-ujbek* (27a), were counted together as a single domain, there would be more violations, but the same candidates would win out.

5. References

Anderson, Gregory D. S. and K. David Harrison
1999 *Tyvan*. [Tuvan] Languages of the World/Materials, vol. 257. München: LINCOM Europa.

Archangeli, Diana
1984 Underspecification in Yawelmani phonology and morphology. Doctoral dissertation, MIT, Cambridge, Mass.

Campbell, Lyle
1986 Testing phonology in the field. In *Experimental phonology*, ed. John J. Ohala and Jeri J. Jaeger, 163-186. Orlando, Florida: Academic Press.

Clements, George N. and Engin Sezer
1982 Vowel and consonant disharmony in Turkish. In *The structure of phonological representations Vol. II*, ed. Harry van der Hulst and Norval Smith, 213-255. Dordrecht: Foris.

Harrison, K. David
1999 Tuvan reduplication and harmony. In Proceedings of the Berkeley Linguistics Society 25: Parasession on Turkic, Dravidian and Caucasian Linguistics.

Harrison, K. David
2000 *Topics in the phonology and morphology of Tuvan*. Doctoral dissertation, Yale University, New Haven, Conn.

Harrison, K. David and Abigail Kaun
2000 Pattern-Responsive Lexicon Optimization. In Proceedings of NELS 30.

Inkelas, Sharon
1995 The consequences of optimization for underspecification. In *Proceedings of the Twenty-fifth Northeastern Linguistics Society*. ed. Eugene Buckley and Sabine Iatridou. Amherst, Mass.: GLSA.

Kaun, Abigail
2000 Rounding harmony. In *The phonetic basis of phonology,* ed. Bruce Hayes, Robert Kirchner and Donca Steriade. Cambridge: Cambridge University Press.

McCarthy, John and Alan Prince
1995 Faithfulness and reduplicative identity. In *University of Massachusetts Occasional Papers in Linguistics 18: Papers in Optimality Theory*, ed. Jill Beckman, Suzanne Urbanczyk, and Laura Walsh, 249-384. Amherst, Mass.: GLSA.

Prince, Alan S. and Paul Smolensky
1993 Optimality Theory: Constraint Interaction in Generative Grammar. Ms., Rutgers University, Rutgers, N.J. and University of Colorado at Boulder, Boulder, Col.

Smolensky, Paul
1993 Harmony, markedness and phonological activity. Paper presented at the Rutgers Optimality Workshop-1, Rutgers University.

Smolensky, Paul
1995 The initial state and "richness of the base" in Optimality Theory. Technical report, Cognitive Science Department, Johns Hopkins University, Baltimore, Md.

Steriade, Donca
1987 Redundant values. Papers from the twenty-third regional meeting, Chicago Linguistic Society, vol 2, 339-62. Chicago: Chicago Linguistics Society, University of Chicago.

Vago, Robert
1988 Vowel harmony in Finnish word games. In *Features, segmental structure and harmony processes part II,* ed. Harry van der Hulst and Norval Smith, 185-205. Dordrecht: Foris.

Whitney, Arthur H.
1949 Colloquial Hungarian. London: Routledge.