6e Journées Internationales du Réseau
Français de Phonologie
June 3-4, 2004, Orl´eans

Underspecification, Markedness, and the Acquisition Process

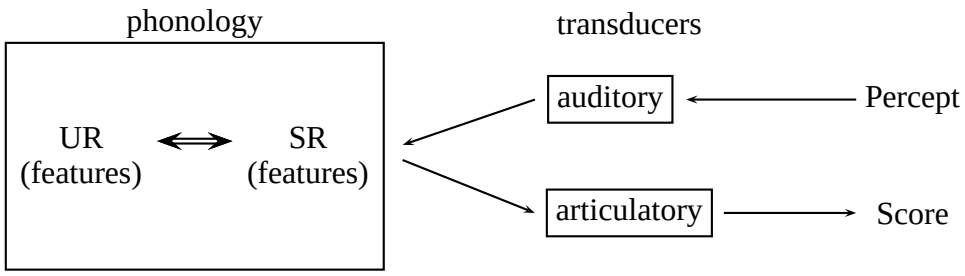
Mark Hale, Madelyn Kissock and Charles Reiss

Overview

1. Background Assumptions
2. Phonological Underspecification
   - Without Phonetic Underspecification
   - With Phonetic Underspecification
3. Constraint-based Approach to Perseverant (Phonetic) Underspecification
   - Small Peculiarities
   - Larger Peculiarities with Potentially More Serious Implications
4. Acquiring phonetic underspecification (with phonological underspecification) in an OT grammar
   - Initial Ranking — Markedness Constraints Above Faithfulness Constraints
   - Initial Ranking — Faithfulness Constraints Above Markedness Constraints
5. Further evidence that phonetic underspecification exists (and therefore must be acquirable)
6. Conclusions

1. Background Assumptions

- Phonology – a mapping of feature representations to feature representations.
- Features – symbolic, formal primitives drawn from a universally available, but fi nite, inventory.
- Transduction – the function which converts physical stimulus to symbolic representations (e.g., feature bundles) and, conversely, symbolic representations to physical events. Following Keating (1988): the transducers are not part of the phonology since they do not map features to features (one may contrast Keating's stance with that of Hammarberg 1976, who holds that the transducers involve cognition and are therefore also a kind of phonology).



2. Phonological Underspecification with and without Phonetic Underspecification

2.1. Phonological Underspecification with full phonetic specification – the case of Turkish plosives presented in Inkelas (1994)

	Nom	Acc	gloss
Non-alternating	sanat	sanat-ɨ	'art'
Non-alternating	etüd	etüd-ü	'etude'
Alternating	kanat	kanad-ɨ	'wing'

Inkelas argues that, in the case of predictable alternating forms such as the 'wing' forms above, the process of Lexicon Optimization will force an underspecified representation. The plosive will be underspecified for voice in the input and the voice value in the winning output candidate will be determined by the ranking of structure-filling constraints.

2.2. Underspecification in both Phonological and Phonetic Representations ('perseverant' underspecification)

General Characteristics

- no apparent articulatory target for one or more features
- articulation completely determined by context
- existence of alternating (but entirely predictable) articulations

The case of Russian [x]

Keating (1988) convincingly argues that Russian [x], in the case where no context rules apply, consists of a phonetic feature bundle with no specification for the feature [back]. (Note that this means that feature bundle will *never* have a feature specification for [back] because the transducer does not 'fill in' features.) Therefore, the following two instances of [x] are crucially different with respect to phonetic feature specification in the output of the phonology: the (a) case has a fully-specified feature bundle and the (b) case is underspecified, having no [back] feature.

(a) /axi/ → fully fronted fricative; context rule filled in [-back]

(b) /ixa/ → transient fricative, gradual transition throught its duration from the [back] values of the adjacent sounds ([-back] to [+back]); fricative remained underspecified

The case of Marshallese vowels

The Marshallese vowel system, discussed in Hale (2000) inter alia, is quite striking. The 'surface' vowels are given below, where the 'tie' symbol (as in i͡u) represents a smooth transition from one vowel to another, e.g. in this case, *i* to *u*.

i	ɯ	u	i͡ɯ	i͡u	ɯ͡i	ɯ͡u	u͡i	u͡ɯ
ɪ	ɤ	ʊ	ɪ͡ɤ	ɪ͡ʊ	ɤ͡ɪ	ɤ͡ʊ	ʊ͡ɪ	ʊ͡ɤ
e	ə	o	e͡ə	e͡o	ə͡e	ə͡o	o͡e	o͡ə
ɛ	a	ɔ	ɛ͡a	ɛ͡ɔ	a͡ɛ	a͡ɔ	ɔ͡ɛ	ɔ͡a

	Oral Stops			Nasals					
	Labial	Dental	Velar	Labial	Dental	Velar	liquids & glides		
'light' [-bk,-rnd]	p^j	t^j		m^j	n^j		l^j	r^j	j
'heavy' [+bk,-rnd]	$b^ɯ$	$t^ɯ$	k	$m^ɯ$	$n^ɯ$	ŋ	$l^ɯ$	$r^ɯ$	ɯ̯
'round' [+bk,+rnd]			k^w		n^w	$ŋ^w$	l^w	r^w	w

As Choi (1992) demonstrates, there is a steady transition between the articulatory position reflecting the back and round features of the preceding consonant to the articulatory position reflecting the back and round features of the following consonant.the back and round of the following consonant in every instance. These are phonetically distinct from diphthongs which have a relatively long-duration nucleus and a brief on- or off-glide. As Bender (1968) showed, the most coherent phonological analysis of the Marshallese vowel inventory is one in which the vowels themselves bear no features along the dimensions front-back and round-unround. That is, they differ from one another *only* along the height and ATR dimensions.

a. $C_{light}VC_{round}$: /n^j[+hi,+ATR]k^w n^j[+hi,+ATR]k^w/ > [n^j[+hi,+ATR]k^w n^j[+hi,+ATR]k^w] > ‡n^ji͡uk^w n^ji͡uk^w‡ 'clothing'

b. $C_{light}VC_{heavy}$: /n^j[-hi,+ATR]$t^ɯ$/ > [n^j[-hi,+ATR]$t^ɯ$] > ‡n^je͡ə$t^ɯ$‡ 'squid'

c. $C_{light}VC_{light}$: /t^j[-hi,-ATR]t^j/ > [t^j[-hi,-ATR]t^j] > ‡t^jɛt^j‡ '*Lutjanus Flavipes*'

Note that this gives Marshallese what appears superficially to be a large and rather unique vowel inventory, while, in fact, for all grammatical purposes, the Marshallese inventory is quite small.

The case of /t^j[-hi,+ATR]t^j/ is particularly interesting. This vowel will show apparent steady-state realization in the [ɛ] space, much like English [ɛ], from which it is, however, quite distinct, representationally. The existence of such cases is especially relevant to acquisition and will be discussed in more detail later.

3. Constraint-based approach to perseverant (phonetic) underspecification

At first glance, the Russian and Marshallese cases appear to be made for an Optimality Theoretic approach. (Note, however, that this is true only if one assumes the formal approach to phonology sketched above – one where features are simply symbolic representations manipulated by the grammar.)

/pʷV[+hi,+ATR]pʷ/	*V[±round]	*V[±back]	MAX-IO	DEP-IO	*[+ATR]	*[+hi]
a. [pʷupʷ]	*!	*		**	*	*
b. [pʷipʷ]	*!	*		**	*	*
c. [pʷapʷ]	*!	*		**		
d. [pʷV[+hi]pʷ]			*!			*
☞ e. [pʷV[+hi,+ATR]pʷ]					*	*

Constraints against the vowel features of [±round] and [±back] must dominate Faith constraints in order to rule out candidates which are fully specified along the back and round dimensions, and thus accurately capture the Marshallese inventory (which contains no vowels fully specified for backness and roundness). This ranking results in candidates (a)-(c) in the tableau above being eliminated. The even more underspecified, and thus seemingly less marked, candidate in (d) is unfaithful to the [+ATR] specification on the input vowel. Since this vowel is *too* underspecified for Marshallese, the Faithfulness constraint MAX-IO (which requires that the [+ATR] specification in the input be respected in the output) must outrank the Markedness constraint *[+ATR] (which would require the elimination of the underlying [+ATR] specification) in Marshallese. This leaves as the optimal candidate the V[+hi,+ATR] vowel of (e), underspecified in the output along the backness and roundness dimensions.

Small Peculiarities

- Insofar as the features of the vowel itself go, the optimal candidate violates neither Faithfulness constraints (MAX-IO and DEP-IO) nor the relevant Markedness constraints e.g., *V[±+round], *V[±-back].
- The 'missing' features of the input candidate's vowel are treated, for purposes of IO constraints, in the way that epenthetic vowels (or consonants) are treated. "Given the fact that an epenthetic segment has no input features to be faithful to, their feature content is delegated to markedness constraints." [Kager 1999:125] However, unlike in the epenthesis cases, there is no violation of DEP-IO by the winning candidate since *no features are present in the output candidate which were not present in the input representation.*
- Based solely on vowel features, there will be no metric for determining the relative ranking of F and M constraints, except with respect to Markedness constraints regarding height and ATR features for vowels (both of which must be ranked above the IO-Faith constraints).

Larger Peculiarities with potentially more serious implications

- Markedness in OT, while formally defined as 'having violation marks' (= a more marked form is one which has incurred more violation marks for its output structure than some other competing form) is, in practice, closely associated with typological markedness and notions such as cross-linguistic frequency of occurrence.

- By formal markedness definitions, Marshallese vowels will be highly unmarked, with a mid vowel which is less marked than a fully-specified ə. (Note that, as indicated earlier, Faith constraints are not relevant to this determination.)

- In terms of typology or frequency of occurrence, the inventory of vowels is highly marked

  a. at the phonological level. From a featural standpoint, the vowel inventory is extremely small, a total of 4 contrastive vowels. Due to the lack of features, there can be no real notion of distribution throughout the vowel space – arguably the same as having a 'poor' distribution.

  b. at the articulatory level. The majority of the articulations appear to fall into the 'not a natural language sound' category. From a more formal standpoint, it appears that the articulations of the vowels give physical outputs that virtually cover the acoustic space, such that many of the vowels would be minimally contrastive acoustically. [Note that this is actually the 'correct' result from the point of view of the lexicon, where there are only 4 contrastive vowels.]

  The input vowels are listed here using only their featural representations. The notation for surrounding consonants are: F=nonback, nonround consonants; B=back, nonround consonants, R=back, round consonants):

Input V-features		consonantal environment								
hi	ATR	F—F	B—B	R—R	F—B	F—R	B—F	B—R	R—F	R—B
+	+	i	ɯ	u	i͡ɯ	i͡u	ɯ͡i	ɯ͡u	u͡i	u͡ɯ
+	-	ɪ	ɤ	ʊ	ɪ͡ɤ	ɪ͡ʊ	ɤ͡ɪ	ɤ͡ʊ	ʊ͡ɪ	ʊ͡ɤ
-	+	e	ə	o	e͡ə	e͡o	ə͡e	ə͡o	o͡e	o͡ə
-	-	ɛ	a	ɔ	ɛ͡a	ɛ͡ɔ	a͡ɛ	a͡ɔ	ɔ͡ɛ	ɔ͡a

An alternation: [rʲi-] ∼ [rʲi͡ɯ-] ∼ [rʲi͡u-] 'agent noun prefix':
[rʲi-tʲerʲpʷalʷ] 'worker' (cf. [tʲerʲpʷalʷ] 'to work')
[rʲi͡ɯ-pʷɯ͡itʲpʷɯ͡itʲ] 'one who kicks' (cf. [pʷɯ͡itʲpʷɯ͡itʲ] 'to kick')
[rʲi͡u-ŋʷo͡erʲtʷak] 'snorer' (cf. [ŋʷo͡erʲtʷak] 'to snore')

Finally, a slightly bigger problem...

4. Acquiring phonetic underspecification (with phonological underspecification) in an OT grammar

Acquiring an adult-state constraint ranking in OT is done through the process of Constraint Demotion (Tesar and Smolensky, 1993; 1998 inter alia). The majority position is that, at the initial state, all Markedness constraints are ranked above all Faithfulness Constraints (but see Hale and Reiss (1997) for the alternative view). Certain Markedness constraints are then demoted based upon positive evidence – evidence which indicates that some marked candidate wins over some lesser-marked candidate.

As Tesar et al. (2003) note, the awkwardness lies in the fact that the acquirer must determine both lexical representations and a constraint ranking, each of which is dependent to a greater or lesser extent upon the other. They propose that the acquirer approaches the problem with a bias toward changing the ranking as a first step (with a further bias toward keeping Markedness constraints high whenever possible) and later, only if their ranking fails, modifying the lexical representation.

Recall that the adult-state grammar that needs to be acquired has *V[±round],*V[±back] >> MAX-IO,DEP-IO

The data that the acquirer gets includes:

- both alternating and non-alternating forms.
- forms that are indistinguishable (acoustically) from fully-specified forms such as [ɨɛtʲ] (from /tʲ[-hi,+ATR]tʲ/).
- forms which, as far as we know, cannot be represented by any combination or geometry of our current feature set – e.g., the steady transition from the [ɯ]-space to the [u]-space in [ɯ͡u] – and therefore cannot be outputs of the grammar itself (nor inputs for a human grammar).

Initial Ranking – Markedness constraints above Faithfulness constraints.

- First data to acquirer = [tʲɛtʲ]
- Acquirer's hypothesis = vowel is a fully-specified [ɛ];
  Action taken: change initial ranking such that MAX-IO and DEP-IO >> *V[±round] and *V[±back]
- Second piece of data to acquirer = [nʲe͡ətʷ]
- Acquirer's hypothesis = vowel has height and ATR features but is underspecified for [back] and [round];
  Action taken: keep derived ranking because the form [nʲe͡ətʷ] will still be the winning candidate since competitors which have back and round features specified will incur gratuitous DEP-IO violations.
- The order in which the acquirer receives the data will not change the ranking outcome.

- Lexicon Optimization will presumably lead to the acquirer positing underspecification for *only* those vowels which show alternations – a vowel which is invariably realized as [ɛ], for example, will maintain its fully-specified underlying representation and its fully-specified phonetic output form, since the grammar will produce an acceptable input-output mapping for such cases (i.e., it will not 'fail'). This grammar, once hypothesized by the learner, will also be supported by (1) the fact that the *V[±round] and *V[±back] constraints have *already been demoted* relative to MAX-IO and DEP-IO, and thus their 're-elevation' would be bad under OT learning-theoretic assumptions, and by the fact that Richness of the Base requires that no constraints on input forms (such as one requiring all inputs to be underspecified) be imposed.

✗ This is the *wrong* result. It holds that Marshallese has fully-specified vowels for the [ɛ] of [tʲɛtʲ] as well as underspecified vowels such as the [-hi,+ATR] vowel of [n̪ʲe͡ʌtʷ]. It fails to account for the complete absence of words of the shape [kʷɛkʷ] with a fully-specified [ɛ] between [+back] and [+round] consonants.

Initial Ranking – Faithfulness constraints above Markedness constraints.

- First data to acquirer = [tʲɛtʲ]

- Acquirer's hypothesis = vowel is fully-specified [ɛ];
  Action taken: Do nothing to constraint ranking – winning candidate will be the one that is *most faithful* to fully-specified [ɛ].

- Second piece of data to acquirer = [nʲe͡ətʷ]

- Acquirer's hypothesis: vowel has height and ATR features but is underspecified on the back and round dimensions;
  Action taken: None. As in the earlier scenario, these forms will be correctly handled by the current grammar (because Faith constraints outrank Markedness constraints, and maximally faithful outputs in this case will be underspecified).

- Lexicon Optimization will lead to acquirer positing underspecification for vowels which show alternations, which reveals to a learner *not hampered by the assumption of Richness of the Base* and one who has not *already (mistakenly) reranked Markedness constraints* that even seemingly fully-specified vowels (like the [i] of [n̪ʲitʲerʲpʷalʷ] 'worker') may be only apparently fully-specified, and can in fact be derived from an assumption of underspecified inputs. More agressive attempts at Lexicon Optimization reveal that this is true of *all* apparently fully-specified vowels (even the non-alternating ones), and a restriction on *inputs* (contra RotB) is posited. This process also reveals that *all* outputs can be treated as underspecified along the back and round dimensions (even those which are apparently [+back,+round]), leading to the reranking of *V[±back] and *V[±round] relative to the MAX-IO and DEP-IO constraints. Note that this is the first change in constraint ranking under the assumption of high-ranking initial Faith.

✔ This is the *correct* result. It holds that all Marshallese vowels are underspecifi ed phonetically for [back] and [round] and makes the correct prediction about loanwords.

5. Further evidence that underspecification really exists (and therefore needs to be acquirable): Evidence from loanwords into Marshallese.

Note: Marshallese has a palatal (traditionally, /y/, here /j/), a back non-round (traditionally, /h/) and a back round (traditionally, /w/) glide, each with the expected effect on adjacent vowels. The glides themselves are said to be 'weakly articulated' and noticable principally through their effects on adjacent vowels.

CVC-loanwords without glide epenthesis

-back,-round both	+back,-round both	+back,+round both
/jitʲ/ [jitʲ] 'yeast'	/lʷak/ [lʷak] 'lock'	/kʷikʷ/ [kʷukʷ] 'cook'

CVC-loanwords with glide epenthesis

- flanking back consonants

front V target	round vowel target
/kejek/ [kə͡ejə͡ek] 'cake'	/tʷiwilʷ/ [tʷu͡uwu͡ulʷ] 'tool'
/kajaŋ/ [ka͡ɛja͡ɛaŋ] 'gang'	/kewetʷ/ [kə͡owo͡ətʷ] 'goat'

- flanking front consonants

back non-round V target	round V target
	/tʲiwinʲ/ [tʲi͡uwu͡inʲ] 'June'
	/tʲiwitʲ/ [tʲi͡uwu͡itʲ] 'shoes'

- mixed flanking consonants

front V target	back non-round V target	round V target
/kʷijinʲ/ [kʷu͡ijinʲ] 'queen'	/tʲahatʷ/ [tʲɛ͡ahatʷ] 'shot'	/tʲawapʷ/ [tʲe͡ɔwɔ͡apʷ] 'soap'
/tʲajak/ [tʲɛjɛ͡ak] 'check'		/tʲawak/ [tʲɛ͡ɔwɔak] 'chalk'
/pʲajak/ [pʲɛjɛ͡ak] 'back, bag'		
/pʷehetʲ/ [pʷəhə͡etʲ] 'base'		

6. Conclusions

- Formal notions of markedness and those based on typology, or cross-linguistic frequency, do not converge on the same set of 'Markedness constraints' —the latter should therefore not be used in arguing in support of the former (and vice-versa).

- The acquisition path is successful only if, at the initial state, Faithfulness constraints are ranked above Markedness constraints.
- The acquisition path is hampered by the assumption of Richness of the Base.
- An abstract, feature-based representation system requires that some aspects of articulation not be specified (e.g., the transitions from [g] to [ɛ] in 'get'), because the grammar provides no appropriate representational apparatus. More interestingly, Marshallese teaches us that the grammar may similarly leave some aspects of articulation not specified (e.g., the backness of the tongue and roundedness of the lips during the vocalic segments of a string) for which the representational apparatus is, in principle, available. It appears that the phonetic implementation of the underspecification in these two types of case is very similar if not identical.

References

Choi, J. 1992. 'Phonetic Underspecification and Target Interpolation: An Acoustic Study of Marshallese Vowel Allophony.' *UCLA Working Papers in Phonetics* 82.

Bender, B. 1968. 'Marshallese Phonology.' *Oceanic Linguistics* 7:16-35.

Hale, M. 2000. 'Marshallese Phonology, the Phonetics-Phonology Interface and Historical Linguistics.' *The Linguistic Review*. 17:241-257.

Hale, M., M. Kissock and C. Reis. 2003. 'Microvariation, Variation, and the Features of Universal Grammar.' ms.

Hale, M., M. Kissock and C. Reis. 2002. 'Phonological Accounts of Microvariation in Pronunciation,' paper presented at the 25th GLOW Colloquium, Amsterdam, April 9-11.

Hale, M. and C. Reiss. 2000 'Phonology as Cognition.' in Burton-Roberts, N., P. Carr, and G. Docherty (eds.) *Phonological Knowledge: Conceptual and Empirical Issues*. Oxford: Oxford University Press.

Inkelas, S. 1994. 'The Consequences of Optimization for Underspecification,' in Beckman, J. (ed.) *Proceedings of the North East Linguisitc Society 25* pp. 287-302.

Inkelas, S., O. Orgun, and C. Zoll. 1997. 'The Implicaitons of Lexical Exceptions for the Nature of Grammar,' in Roca, I. (ed.) *Derivations and Constraints in Phonology*. New York: Oxford University Press. 393-418.

Ito, J., A. Mester, and J. Padgett. 1995. 'Licensing and Underspecification in Optimality Theory,' *Linguistic Inquiry* 26, 4: 571-613.

Kager, R. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.

Kager, R., J. Pater, W. Zonnenveld (eds.). 2004. *Constraints in Phonological Acquisition*. Cambridge: Cambridge University Press.

Keating, P. A. 1988. 'Underspecification in Phonetics.' *Phonology* 5:275-292.

Kingston, J. and R. Diehl. 1994. 'Phonetic Knowledge.' *Language* 70:419-454.

Liljencrants, J. and B. Lindblom. 1972. 'Numerical Simulations of Vowel Quality Systems: The Role of Perceptual Contrast.' *Language* 48:839-862.

McCarthy, J. 2002. *A Thematic Guide to Optimality Theory*. Cambridge: Cambridge University Press.

Pylyshyn. S. 1984. *Computation and Cognition: Toward a Foundation for Cognitive Science*. Cambridge, MA: MIT Press.

Steriade, D. 1995. 'Underspecifi cation and Markedness,' in Goldsmith, J. (ed.) *Handbook of Phonological Theory*, Oxford: Blackwell, 114-174.

Stevens, K.N. 1998. *Acoustic Phonetics.* Cambridge, MA: MIT Press.

_____. 1972. 'The Quantal Nature of Speech: Evidence from Articulatory-Acoustic Data.' in Denes, P. and E. David Jr. (eds.) *Human Communication: A Unified View.* New York: McGraw Hill.

Tesar, B. and P. Smolensky. 2000. *Learnability in Optimality Theory*. Cambridge, MA: The MIT Press.

Tesar, B., J. Alderete, G. Horwood, N. Merchant, K. Nishitani, & A. Prince. 2003. 'Surgery in Language Learning,' in Garding, G. and M. Tsujimura (eds.) *Proceedings of the 22nd West Coast Conference on Formal Linguistics*, 477-490.

Mark Hale
Concordia University, Montr´eal
hale1@alcor.concordia.ca

Madelyn Kissock
Oakland University
kissock@oakland.edu

Charles Reiss
Concordia University, Montr´eal
reiss@alcor.concordia.ca