Phonologization of redundant contrasts and the Contrastivist Hypothesis

Pavel Iosad
The University of Edinburgh
pavel.iosad@ed.ac.uk

Eighth North American Phonology Conference Concordia University 9th May 2014

Outline

• Underlying vs. surface contrast in an emergent-features framework
• Smuggling redundant features into the input
• Two case studies:
  – Split with no underlying contrast: Ulster English FACE
  – Split with no surface contrast: Welsh mid vowels
• Refocusing the Contrastivist Hypothesis

1 Emergent features and underlying contrast

1.1 The framework

Basics

• Basically, in this talk I try to reconcile some facts with a framework that I like to think makes sense
  – Phonological representation is not trivial and based on phonological activity
  – Features (and other categories) are not universal but arise through learning (‘emergent’)
  – Only positive evidence acceptable to construct representations: methodological minimalism (for some definition thereof)
Emergent features

- Basic assumption: hearing a sound doesn’t immediately let the learner know what its phonological representation is


- Some insights here might be compatible in broad outline with an ‘unlearning’ system à la Hale, Kissock & Reiss (2007), Hale & Reiss (2008), if it is coupled with a contrastivist approach

- Minimalism: no requirement for a substantial, universal statement of substance-to-phonology mapping, categories only arise when needed

Contrastivism

- The Contrastivist Hypothesis (D. C. Hall 2007, p. 20): ‘The phonological component of a language \( L \) operates only on those features which are necessary to distinguish the phonemes of \( L \) from one another’


- Minimalism: lexical contrast provides independent evidence for the existence of the relevant categories, no need to draw on other sources

1.2 Quasi-phonemic contrast and its discontents

What is ‘phonemic contrast’?

- The Contrastivist Hypothesis as defined above makes a strong prediction: if some distinction is involved in phonological computation, it must be used for ‘phonemic’ contrast

- ‘Phonemic’ contrast used to mean surface contrast

- Under this interpretation, a simple version of Contrastivist Hypothesis has been argued to be untenable since at least Halle (1959) with the Russian \([d\ddag] \) and \([d\#] \)

- One possible response: define ‘phonemic’ contrast in terms of contrast in underlying representations

Cf. the reanalysis of Russian by Dresher (2009), where \([d\ddag] \) and \([d\#] \) can get non-redundant \([\pm\text{voice}] \) specification
Secondary splits and quasi-phonemes

• It has long been recognized that the classic approach to secondary split by Twaddell (1938) makes the mysterious assumption that the loss of the conditioning environment does not lead to a reversion of allophony (e.g. Kiparsky 1995, Janda 2003, Bermúdez-Otero 2007)

• If Germanic pre-umlaut [mīsiz] is /mūsiz/ and /i/ is lost, why does it become /mīs/ and not /mūs/?

• Solution: phonologization before phonemicization (Hyman 1976, Bermúdez-Otero 2007, Kiparsky 2014)

• The fact of surface contrast is irrelevant if there is clear evidence of phonologization

Derived contrasts

• This set of issues is closely related to the problem of ‘derived contrasts’ (e.g. Harris 1990, K. C. Hall 2013)

• Scottish Vowel Length Rule / Aitken’s Law: complementary distribution in underived forms, contrast (?) in derived forms

• tide [ˈtʌɪd] ≠ tied [ˈtæd], but only because of tie [tæ]

• Kiparsky (2014) calls such predictably distributed segments ‘quasi-phonemes’, seemingly rejects the last remainos of contrastivist thinking by disavowing Structure Preservation (which hinges on the notion of ‘allophony’)

The problem

• Can these facts can be reconciled with the framework I’m trying to adopt?

• If surface contrast cannot tell us anything useful about phonological status, what use is the Contrastivist Hypothesis?

• What is ‘contrast’ if not ‘surface contrast’?

1.3 Proposal

Computation: part of the answer

• All solutions to the ‘absent contrast problem’ agree in recognizing the importance of computation

• It is what forces the underlyingly distinct segments into predictable distributions
A phenomenon is ‘phonological’ if it is involved in proprietary phonological computation, i.e. the mapping from sets of proprietary phonological symbols to other sets of proprietary phonological symbols; modularity (Reiss 2007, Hale & Reiss 2008, Scheer 2010, Bermúdez-Otero 2012, Hale, Kissock & Reiss 2014; also Foley 1977, Keating 1984, 1988, Morén 2007, Samuels 2011)

• Predictability of distribution does not come into this definition of being phonological

Proposed solution

• The motivation for contrastivism here is methodological minimalism: minimize the use of entities for which there is no independent evidence

• This independent evidence is provided by the entities being stored in the lexicon

• The presence of something (e.g. a feature) in lexical entries is sufficient to achieve our goals

• Stating ‘contrast’ in terms of predictability of distribution is an unnecessary additional stipulation

Case studies today

• How do features get into underlying representation?

• Bottom-up learning promotes the lexicalization of predictably distributed categories

2 Case studies

2.1 Ulster English FACE vowel: no underlying contrast

No contrast in underlying representations

• Classic case of ‘derived contrast’

• If learning is bottom-up, the setting-up of the categories precedes the learning of relevant rules

• I suggest that the OT principle of Richness of the Base ‘smuggles’ the surface categories into underlying representations before they can be derived by rule

• Even after the rule is learned, there is no incentive to realign the underlying representations

• Distribution in URs technically predictable, but that’s OK
The facts

• Allophony of the Mid Ulster English FACE vowel (Wells 1982, Harris 1985, 1990; Warren Maguire p. c.)

• Two allophones:
  
  – Lower, monophthongal, written as [ɛː] but not necessarily identical to DRESS in quality (may be [ɛ:]))
  
  – Higher, usually diphthongal [ɛə]/[ɛ̝ə]/[ɪə], but may be [i], e.g. in some Co. Londonderry varieties (McCafferty 1999)

• Basic distribution: lower allophone when final in a stress domain at the stem level
  
  – Lower
    
    * Morpheme-final, underived: day [ˈdɛː], lay [ˈlɛː]
    
    * Foot-final, underived: Mayhew [ˈmeː.hjuː], latex [ˈlɛː.tɛks]
    
    * Non-final, derived: days [ˈdɛːz], laid [ˈlɛːd]
  
  – Higher: elsewhere
    
    * Non-final, underived: daze [ˈdɪəz], table [ˈtæbl̩]

Surface vs. underlying contrast

• Under a classic phonemic account, there is clearly a contrast between days [ˈdɛːz] and daze [ˈdɪəz]

• Boundary symbols complicate things a bit: /de-z/ vs. /dez/

• Problem for Lexical Phonology: Lowering is allophonic (non-neutralizing) but stem-level (to account for [ˈdɛːz] [SE ˈdɛːz])

The allophonic issue is a bit blurred: recall that in some varieties we do get [i] for the higher allophone, presumably neutralizing with KIT

• Issue for classic Structure Preservation, although accounted for in Stratal OT, which, like most post–Halle (1959) approaches, refuses to give theoretical status to ‘phonemes’: see the discussion of Chung’s Generalization by Bermúdez-Otero (2012)

• Crucial assumption: both day and daze actually have the same vowel underlyingly: /de/ vs. /dez/
Redundant contrasts and the Contrastivist Hypothesis

Why economical URs?

- Basic assumption in early generative phonology: storage is expensive, computation is cheap
- If you can derive something, you should
- Fewer ‘phonemes’ (= possible segments in URs) is more economical and thus preferable

Argument here

- Bottom-up learning promotes redundant URs
- No reason to think Ulster English [dɪəz] is anything but /dɪəz/
- Both ‘allophones’ of FACE are possible in URs ⇒ phonology making reference to them is consistent with the Contrastivist Hypothesis

Bottom-up learning

- A learner faced with Ulster English pairs like [dɛːz]/[dɪəz] is likely to set up [ɪə] and [ɛː] as distinct categories
  - Phonetically robust distinction
  - Relatively little variability
  - Distinction arguably unavoidable if days and daze are to be distinguishable at all (cf. Hale & Reiss 2008)
- To ‘undo’ the distinction, the learner has to learn additional conditions
  - Finality in stress domain (opaque generalization!)
  - Stem-level affiliation
- Segmental learning comes first (Bye 2011, Uffmann 2014), so it has to be unlearning

Lexicon Optimization

- If we assume an OT-based grammar, what are the consequences of the learner parsing daze as [ˈdɪəz] and days as [ˈdɛːz]?
- Before the additional stipulations come in, and given the lack of alternations, they will be stored faithfully
(1) 

<table>
<thead>
<tr>
<th></th>
<th>FAITH-[ɪə]</th>
<th>FAITH-[ɛː]</th>
<th>*[ɪə]</th>
<th>*[ɛː]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[dɪəz] a. /dɪəz/</td>
<td>/dɪəz/</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[dɛːz] b. /dɛːz/</td>
<td>*!</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[dɛːz] c. /dɛːz/</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[dɪəz] d. /dɪəz/</td>
<td>*!</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[dɛːz] e. /dɛːz/</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| [dɪəz] f. /dɪəz/ | *! | * | *

Refined learning

- The next step is promoting the horribly descriptive constraint *ɪə]_Ft_
- It doesn’t really influence the URs

(2) 

<table>
<thead>
<tr>
<th></th>
<th>*ɪə]<em>Ft</em></th>
<th>FAITH-[ɪə]</th>
<th>FAITH-[ɛː]</th>
<th>*[ɪə]</th>
<th>*[ɛː]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[dɪəz] a. /dɪəz/</td>
<td>/dɪəz/</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[dɛːz] b. /dɛːz/</td>
<td>*!</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[dɛːz] c. /dɛːz/</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[dɪəz] d. /dɪəz/</td>
<td>*!</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[dɛːz] e. /dɛːz/</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| [dɪəz] f. /dɪəz/ | *! | * | *

What does the constraint actually do?

- The job of *ɪə]_Ft_ is to rule out a candidate contained in the rich base (McCarthy 2005, Jarosz 2006)

(3) 

<table>
<thead>
<tr>
<th></th>
<th>*ɪə]<em>Ft</em></th>
<th>FAITH-[ɪə]</th>
<th>*[ɪə]</th>
<th>*[ɛː]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [ˈdɪə]</td>
<td>*!</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. [ˈdɛː]</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>
Redundant contrasts and the Contrastivist Hypothesis

• This would otherwise be useful if we had [ɪə]~[ɛː] alternations, but we don’t, as far as I am aware

• This is the stem-level ranking: what about other levels?

Since the stem level never produces candidates with foot-final [ɪə], we don’t have to worry about it

Derived contrasts aren’t derived

• The fact that the outputs of stem-level computation only allow [ɪə] foot-finally emerges from the learning of the lexicon

• While the relevant rankings are required to exclude phonotactically illicit candidates supplied by the rich base, without alternations there is no evidence that actual lexical items undergo unfaithful mappings

• The ‘complementary’ distribution in [ˈdɪəz] vs. [ˈdɛː] is a fact of the lexicon, not of the grammar

Actually, it is not obvious how complementary it is: not all prosodic structure is stored, so in what sense are the distributions in /dɛː/ vs. /dɪəz/ ‘complementary’?

Summary

• As long as the learner sets up the categories, they will become available to enter underlying representations

• Later, the learner needs to set up the grammar that promotes complementary distributions at the stem level

This crucially presupposes that the same categories should be available to Richness-of-the-Base arguments, conceived of as possible elements of underlying representations

• There is no pressure anywhere in the learning mechanism to realign the URs of non-alternating morphemes to make the complementary distribution derived

• ‘Derived contrast’ does not stop distinctions from entering underlying representations

No problem for the Contrastivist Hypothesis
2.2 South-West Welsh mid vowels: no surface contrast

No contrast in surface representations

- Predictable distribution in surface representations: ‘categorical allophony’ (Ladd 2006, Kim 2013)
- Phonologized: sensitive to proprietary phonological attributes (feature specification)
- Markedness outranks faithfulness, but presence in underlying representations relies on non-alternating cases and Richness of the Base

Vowel length: North Welsh

- In most Welsh varieties, short and long vowels contrast at least in some positions: examples from Dyffryn Alyn (Thomas 1966)

  - Minimally, in stressed monosyllables before [n l r]

(4) a. [ˈtʰaːn] tân ‘fire’
   b. [ˈkʰan] can ‘hundred’

  – Also not strictly speaking a contrast (Wells 1979), but vowels can be short or long depending on following consonants

(5) a. [ˈmaːb] mab ‘son’
   b. [ˈkʰum] cwm ‘valley’

Vowel quality: North Welsh

- Six vowels enter the length contrast in North Welsh: /i u e o a/ (G. E. Jones 1984, Mayr & Davies 2011)
- The non-low vowels can be either ‘tense’ or ‘lax’: [i]/[i], [u]/[u] etc.
- Unstressed syllables: poorly understood variation, may be to some extent influenced by open/closed character of the syllable (e.g. Thomas 1966, G. E. Jones 1984), but certainly no contrast
- Stressed syllables: long vowels are ‘tense’ [i: u: e: o:], short vowels are ‘lax’ [i u e ɔ]

(6) a. [ˈtʰoːn] tôn ‘tune’
   b. [ˈtʰɔn] ton ‘wave’
Vowel length: South Welsh

- South Welsh has a generally similar system, but the length contrast is also found in stressed penultimate syllables (G. E. Jones 1971, Awbery 1984, 1986, Wmffre 2003)

(7) a. [ˈtʰoːn] tôn ‘tune’
   b. [ˈtʰoˑne] tonau ‘tunes’
   c. [ˈtʰɔn] ton ‘wave’
   d. [ˈtʰɔnˑe] tonnau ‘waves’

- In most dialects, the quality system is generally similar to that seen in North Welsh

Analysis so far

- Vowel length appears to be contrastive, or at least phonologized: witness its interaction with clearly contrastive things like [n l r] moraicity or other segmental features in consonants
- The status of the ‘tenseness’ distinction is ambiguous: it is fully predictable, but it does not seem to participate in phonological computation in interesting ways
- The neutralization in unstressed syllables — precisely where there is no contrast — is somewhat suggestive of a [±tense] feature only present in stressed syllables (cf. Dyck 1995, 1996, Dresher 2009)

Vowel quality: South-West Welsh


(8) a. [ˈkʰɔˑdɪ] codi ‘to rise’
   b. [ˈkʰɔˑdɔð] cododd ‘((s)he) rose’
   c. [ˈɡwɛˑdʊχ] dywedwch ‘say’
   d. [ˈɡweˑdɔð] dywedodd ‘((s)he) said’

- I suggest that this distinction is phonologized
  - Categorical (tbc)
  - Sensitive to phonological structure
Not just undershoot

- Representative examples (speaker from Goodwick, Pembs.)

(9) a. ['seːrɛn] seren 'star'
   b. ['tʰɛːbiɡ] tebyg 'similar'

Phonologization

- Plausible origin: trade-off in inherent length, cf. East Slavic (Crosswhite 2000), Munster Irish (Ó Sé 1984)

- But there does not appear to be a continuous relationship between the length of the post-tonic vowel and the F₁ value of the stressed vowel

- (Data analysis ongoing)

- It would seem that the key characteristic of the allophony is its sensitivity to the phonological feature [±high] of the following vowel: dissimilation within a foot?

- Precise analysis depends on what high vowels do: again, analysis ongoing
Underlying representations

- How would the learner deal with these cases, assuming they set up the separate categories [ɛː ɔː]? 
- We do get alternations, as in *codi* vs. *cododd* (ex. 8) 
- Under a classic Lexicon Optimization scenario along the lines of Inkelas (1994), we expect alternating cases to be unspecified for the relevant feature

- Crucially, we still expect non-alternating forms to be specified, using the same scenario is above
- Non-alternating instances exist, of course: *wedyn* ‘after’, *nefodd* ‘heaven’ (Owens 2013)

Computation is king

- If we assume that [ɛː ɔː] can be phonological categories, the grammar has to provide mechanisms to ensure the correct complementary distribution
- The precise analysis hinges on the representational approach chosen
  - The binary option: alternating mid vowels are underlyingly unspecified for ‘[tense]’, grammar adds the correct specification in stressed position, leaves unstressed vowels unspecified to derive variation
  - The privative option: one of the categories is distinguished by an extra feature, grammar ensures its presence/absence in correct contexts

Obscuring URs

- Either way, /ɛː ɔː/ are allowed in underlying representations, they just surface in predictable ways
- If the grammar is set up correctly, we do not, in fact, have to posit underspecified inputs

\[
\begin{align*}
\text{(10)} & \quad \text{a. } [t\text{ˈreː}] & \quad tref & \quad \text{‘town’} \\
& \quad \text{b. } [t\text{ˈreːvð}] & \quad trefi & \quad \text{‘towns’}
\end{align*}
\]

- Lexicon Optimization leads us to expect an underspecified /E/ here, but forms such as these do not really show the sort of variation associated with underspecification
- Might the learner converge on /tre(ː)v/ as the UR, with an unfaithful mapping in the plural?
- Little additional cost, since the grammar needs to deal with disharmonic rich-base inputs
The precise analysis here really depends on the featural solution, and to an extent on the interpretation of faithfulness constraints. See Krämer (2012, ch. 8) for discussion of the latter point.

(Happy to answer questions)

Summary

- In the case of exceptionless complementary distribution, non-alternating instances will make their way into the lexicon.
- The grammar still needs to coerce the categories into the correct distribution:
  - Rich-base forms
  - Possibly unfaithful mappings rather than underspecification in the case of alternations
- No contradiction with the Contrastivist Hypothesis, since the relevant categories are at least possible to entertain entering the lexicon, and in all probability found there due to the existence of non-alternating forms.

3 Discussion

3.1 Integrating marginal contrasts

Abandoning surface contrast

- There are obviously more types of ‘marginal’ contrasts than could be covered here.
- The combination of bottom-up learning, Lexicon Optimization, and Richness of the Base can lead to categories with predictable distributions appearing in the lexicon.
- If phonologization is involvement in phonological computation, we do expect phonologization and lexicalization to go hand in hand.
- In that sense, the Contrastivist Hypothesis is true, but has little to do with predictability of distribution.

Leaky patterns

- In at least one of the cases discussed here, the conditioning is fairly categorical: foot-final vs. non-foot-final.
- We do expect cases where the distributions emerge from fuzzier phenomena, giving the learner substantially more uncertainty in the categorization of the data.
Redundant contrasts and the Contrastivist Hypothesis

• Such patterns may ‘leak’, with learners assigning an ‘unexpected’ category (perhaps under pressure from hypercorrection à la Ohala 1981, eventually also morphological factors)

• Exactly the situation found by Uffmann (2014) for GOAT and GOOSE fronting in Southern England
  – Generally coda [l] inhibits fronting, with some cyclic transfer: r[ou]ll, r[ou]ller
  – But cyclic transfer may fail: p[ou]le but p[ɔv]lar
  – And plenty of variation in genuinely monomorphemic items, as in m[ou]lar

• This is, of course, even less of a problem for the Contrastivist Hypothesis
  – Exceptional (lack of) fronting in monomorphemic items must definitely be stored in the lexicon
  – Same goes for stem-level derivatives like polar under some stratal approaches (Bermúdez-Otero & McMahon 2006, Collie 2007, Bermúdez-Otero 2012)

| The Ulster English pattern might leak too: Warren Maguire (p. c.) reports [ˈrɪəˌdar], essentially a minimal pair with [ˈlɛːˌtɛks] |

Summary

• The marginality, or complete absence, of a surface contrast need not prevent its phonologization

• In the present framework, phonologization (as participation in phonological computation) entails lexicalization more or less automatically

• A major prediction here is that even fully predictable distinctions may be phonologized

| In a way, this isn’t much more than Halle’s Russian [dʒ] argument |

• But now we know these things can be much more ‘allophonic’, witness the case of post-coronal GOOSE (=/Tuw/) in Philadelphia (Labov, Rosenfelder & Fruehwald 2013)

3.2 Whither the Contrastivist Hypothesis?

Refocusing the CH

• If any of the preceding works, the Contrastivist Hypothesis appears to lose much of its power

• Surface, or even underlying, contrasts are of little importance to determining what counts as phonological

• That is actually a desirable result: the true criterion of phonological status is modularity
‘The phonological component of a language $L$ operates only on those features which are necessary to distinguish the phonemes of $L$ from one another’

... and not more than that

No redundant features

• The CH retains its importance when we focus on the ‘no redundancy’ aspect

• The reasoning is as follows:
  1. Establish the set of phonological elements using independent criteria
  2. Assign a set of minimally contrastive specification, e.g. using the Successive Division Algorithm (Dresher 2003, D. C. Hall 2007)
  3. The CH tells us that no other features should be available to the phonology

• This arguably preserves the bulk of the empirical results achieved so far through contrastivist enquiry

Conclusion

• It is possible to reconcile contrastivist thinking and bottom-up learning

• The price to pay is the rejection of traditional distributional criteria as diagnostic of ‘contrastive’, and consequently phonological, status

• This move is overdue in any case, even for avowed contrastivists

• A leaner Contrastivist Hypothesis still retains much of its attraction

References


Redundant contrasts and the Contrastivist Hypothesis


Kim, Yuni. 2013. Marginal contrast, categorical allophony and the Contrastivist Hypothesis. Presentation at GLOW 36, Lund University.


