Phonology and the lexicon: a tutorial

Ricardo Bermúdez-Otero & Donca Steriade
University of Manchester & MIT

Phonological Theory Agora 3
Tours
14 October 2016

PROGRAMME

10.00 - 10.30 Part I: A survey of some current issues
Ricardo Bermúdez-Otero

Part II: Focus on paradigmatic dependencies and the phonological lexicon

10.30 - 11.15 Cyclicity and its extensions
Donca Steriade

11.15 - 11.45 Paradigmatic dependencies without cyclic containment as UR acquisition
Ricardo Bermúdez-Otero

11.45 - 12.15 Open discussion

[NB: Timings are approximate.]

§1 How much phonological information is stored in the lexicon?
Part I of the tutorial surveys two prominent topics of debate bearing on this question:
• The effects of usage factors: phonetic detail in the lexicon?
• Morphological decomposition: the size of lexically stored exponents
Throughout, conceptual considerations and arguments based on internal evidence are related to external data, specially from historical change and from psycholinguistic experiments.

§2 The status of underlying representations is discussed in Part II of the tutorial with particular reference to the analysis of paradigmatic dependencies.

THE EFFECTS OF USAGE FACTORS: PHONETIC DETAIL IN THE LEXICON?

The phonological lexicon in the classical modular feedforward architecture of grammar

§3 The modular feedforward architecture:

\[
\begin{align*}
\text{Underlying representation} & \quad \text{phonological rules} \\
\text{Surface representation} & \quad \text{phonetic rules} \\
\text{Auditory and articulatory representations} & \quad \text{(continuous)}
\end{align*}
\]

• The phonological representations stored in the lexicon consist of discrete categories.
• The lexicon contains no ‘fine’ (gradient, subcategorical) phonetic information.
§4 Argument 1: the double articulation of language
Architectures like §3 capture the intuition that phonology is a discrete combinatorial system:
• an arbitrarily large number of signifiants (Saussure 1916) or formes vocales (Martinet 1960) is set up through the recombination of a small number of discrete meaningless units;
• meaningful expressions do not have holistic phonetic properties (e.g. whole-word duration targets).
In a system that relies on holistic signals, in contrast, parsing error imposes a tight upper bound on the number of possible signals: see Nowak et al. (1999) for a mathematical demonstration.

§5 Argument 2: neogrammarian change
Architectures like §3 explain the existence of neogrammarian change, i.e. phonetically gradient but lexically regular change (Bermúdez-Otero 2015: 379-82 and references therein):

neogrammarian change is change in the implementation rules assigning phonetic targets to discrete categories in surface representations.

SR

phonetic implementation rule

[F] [F]

target at time \( t_1 \)

target at time \( t_2 \)

Paul  Prinzipien (Paul 1886[1880]: 62) appeals to discrete combinatorics as the explanation of neogrammarian change; cf. Auer’s (2015) Paul as a usage-based linguist avant la lettre.

Bloomfield  Dissociation of lexical from phonetic knowledge: ‘two layers of habit’, one linking words to phonemes, the other linking phonemes to phonetic parameters (Bloomfield 1933: §20.11, pp. 364-5).

§6 The modular architecture in §3 also underpins hugely influential psycholinguistic models such as that of Levelt (1989).

§7 The empirical challenge: phonetic effects of usage factors
In recent decades, phonologists working in usage-based frameworks (e.g. Bybee 2001) and psycholinguists have identified an empirical challenge to the modular architecture:

gradient usage-related lexical properties

<table>
<thead>
<tr>
<th>neighbourhood density</th>
<th>contextural predictability</th>
</tr>
</thead>
<tbody>
<tr>
<td>token frequency</td>
<td>have an effect on</td>
</tr>
</tbody>
</table>

...  ...

gradient phonetic properties

<table>
<thead>
<tr>
<th>duration</th>
<th>gesture amplitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>coarticulation (gestural overlap)</td>
<td></td>
</tr>
</tbody>
</table>

...

§8 The case of lexical token frequency
High-frequency words are relatively hypoarticulated:
• shorter duration (Whalen 1991; Gahl 2008 on time vs thyme)
• more vowel centralization (Wright 2003)
• more coarticulatory vowel nasalization (Zellou & Tamminga 2014) etc.

§9 The case of neighbourhood density
Neighbourhood density is defined as the number of phonologically similar neighbours

\[ \text{e.g., for } /k\text{x}t/, \quad \{/k\text{x}t/, /k\text{x}t/, /k\text{x}t/\ldots \} \]

weighted by frequency.

Words in high-density neighbourhoods are relatively hyperarticulated:
• less vowel centralization (Wright 2003)
• longer VOT in fortis plosives (Baese-Berk & Goldrick 2009) etc.


§10 The problem (cf. §3):
• If the underlying phonological representations stored in lexical entries consist of discrete categories, they cannot encode word-specific phonetic detail.
• Even if other attributes of lexical entries have continuous values, this gradient information cannot reach the phonetic module via surface phonological representations if the latter consist solely of discrete phonological categories.
A moderate response: gradient activation of discrete categories

§11 Key ideas
• The modular feedforward architecture provides a correct account of the facts at Marr’s (1982: 25) computational level of description.
• The phonetic effects of usage factors arise at Marr’s algorithmic (processing) level.
• Lexical phonological representations consist solely of discrete categories: there is no direct encoding of fine phonetic detail in the lexicon.
• But discrete symbolic representations can be gradiently activated.
• Gradient activation cascades higher to lower levels of representation before processing at the higher levels is complete.

Synthesis of ideas from classical symbolic (e.g. Pylyshyn 1984, Marcus 2001) and connectionist (Rumelhart et al. 1986) approaches to cognition.


§12 Application to neighbourhood density effects (Baese-Berk & Goldrick 2009)
Activation cascades from lemmas through phonological forms to phonetic parameters even before lemma selection is complete.

![Diagram of lexical hierarchy with labels: CAT, CAB, CAD, long VOT, high F2, and line thickness representing activation strength.]

§13 Problems
• Neighbourhood effects and frequency effects pattern differently, and so must involve different mechanisms (Munson 2007; Goldrick et al. 2011: 69).
• To my knowledge, there is no fully developed account of frequency effects from this perspective.

A radical response: Exemplar Theory

§14 Key ideas
• The lexicon contains exemplar clouds, i.e. collections of episodic memory traces containing fine phonetic detail.
• Usage causes lexical representations to be constantly updated as old exemplars decay and new exemplars are added to the cloud.


§15 Problems for ‘pure’ Exemplar Theory raised in the phonetic and psycholinguistic literature
(1) Parsing challenge in matching auditory input to holistic fine-grained targets: see §4 above (German et al. 2013: 230)
(2) Difficulty in accounting for the existence of neogrammarmian change: see §5 above (Pierrehumbert 2002, Bermúdez-Otero 2007)
(3) Failure to explain the instantaneous generalization of a newly learnt phonetic pattern to the whole lexicon (McQueen et al. 2006, Peperkamp & Dupoux 2007, Curier et al. 2010, Nielsen 2011, Curier 2012: §21.1).
(4) Failure to explain “deafness” to postlexical properties (e.g. French speakers’ stress deafness) (Rahmani et al. 2015)
(5) Primacy of unreduced canonical forms in word recognition (Curier 2012: 416–7, Ernestus 2014)
(6) Phonotactic learning driven by type frequency, not token frequency (Richtsmeier 2011, Pierrehumbert 2016: §2)
(7) Absence of expected lexically-specific effects of word’s phonetic environment (Cohen-Goldberg 2015)
(8) Little evidence that episodic detail primes word recognition under naturalistic conditions (McLennan 2007: 68, Hanique et al. 2013)
• etc.

§16 Hybrid models

but, so far, such models fail to specify the division of labour between their two components.
A test case: the effect of frequency on diachronic change

§17 The existence of neogrammarian change is problematic for Exemplar Theory because exemplar clouds encode word-specific phonetic properties; cf. §5. Indeed, Bybee (1998, 2002) and Pierrehumbert (2001, 2002) assert that no change is truly neogrammarian:

**Key empirical prediction of Exemplar Theory**

In diachronic changes involving phonetic reduction (lenition, coarticulation), high-frequency words are ahead of low-frequency words and change faster.

§18 Postulated mechanism

- High-frequency words undergo greater exposure to reductive phonetic biases during use.
- The gradient effect of these biases is registered separately for each word in its own cloud.

§19 Predicted diachronic trajectory (but cf. Sóskuthy 2014)

- Only one study (Hay & Foulkes 2016) reports high-frequency words changing faster (§19), but the observation is unreliable:
  (i) mixes two corpora collected 50 years apart and separated by a 36-year gap in apparent time;
  (ii) the old bad-quality corpus shows no frequency effect at all;
  (iii) the new good-quality corpus shows a constant rate effect;
  (iv) the time:frequency interaction is obtained by interpolating across the two corpora.


§20 In contrast, approaches to usage effects that rely on gradient activation (§11-§12 above) predict truly neogramarian change involving constant rate effects (Kroch 1989):

- High-frequency words are ahead synchronically but change at the same rate diachronically.
- This is because the processing mechanisms that cause usage effects are
  (a) time-invariant (as long as the usage factors themselves do not change)
  (b) orthogonal to innovation in the phonetic implementation rules (no word-specific loops).

This prediction is explicitly stated in Bermúdez-Otero et al. (2015) and Kiparsky (2016: 482).

§21 The empirical record so far


My own view

§22 Debate continues to rage, but my assessment of the current situation is as follows:

- 'Pure' Exemplar Theory (§14) is untenable for the reasons listed in §15.
- Currently, hybrid models combining classical symbolic grammars and exemplar memory (§16) are poorly specified and so have little empirical content; each mechanism (symbolic computation or exemplar memory) is invoked in a purely post hoc manner.
- The most interesting line of research is gradient symbolic activation (§11), which preserves intact the empirical content of the classical modular architecture (§3).
MORPHOLOGICAL DECOMPOSITION: THE SIZE OF LEXICALLY STORED EXPONENTS

Morphological decomposition and the balance between computation and storage

§23  The question

Which morphosyntactically complex items are recognized via whole-form lexical entries? entries for their constituents?

The question has direct implications for the balance between computation and storage in phonology: alternants that are not stored must be derived.

§24  Example: English keep

syntax-semantics

<table>
<thead>
<tr>
<th>stored exponents</th>
<th>phonetic signal</th>
<th>computations required</th>
</tr>
</thead>
<tbody>
<tr>
<td>/kɛpt/</td>
<td>[kɛpt] or [kɛpt]</td>
<td>allomorph selection (2x)</td>
</tr>
<tr>
<td>/kɛp/</td>
<td>/t/</td>
<td>allomorph selection</td>
</tr>
<tr>
<td>/t/</td>
<td></td>
<td>closed syllable shortening</td>
</tr>
<tr>
<td></td>
<td></td>
<td>short vowel shift</td>
</tr>
</tbody>
</table>

§25  Theoretical choices

Higher levels of decomposition are typically accompanied by
• minimal storage (e.g. SPE’s evaluation measure)1
• relaxations of modularity (e.g. readjustment rules in DM,2 indexed constraints in OT3)
• relaxations of locality (e.g. morphophonological rules free from morphological locality)4
• high amounts of opacity (e.g. SPE)5

Lower levels of decomposition are typically accompanied by
• redundancy between storage and computation (e.g. Jackendovian redundancy rules)6
• duality of symbolic rules and connectionist association (e.g. Pinker’s dual-route theory)8
• denial of synchronic reality for some patterns7

E.g.


§26  Effects of frequency on recognition speed

• Two measures of frequency:
  • surface frequency = frequency of taking
  • base frequency = frequency of TAKE = sum of the frequencies of take, takes, took, taken, and taking

• General observation:
  higher frequency ⇄ higher recognition speed1 (e.g. Forster & Chambers 1973)

§27  Priming

• Priming: exposure to form a speeds up the recognition of form b

• Full priming:
  e.g. German Waggon-s ‘train_carriage-PL’ primes Waggon ‘train_carriage’[SG] as much as Waggon primes itself (Clahsen et al. 2003)

• Full priming ⇒ evidence for decomposition
  Reduced priming ⇒ evidence for own entry in the lexicon

§28  Affix shift errors

E.g. let go-ing for lett-ing go, tell us-ing for tell-ing us5 ⇒ evidence for decomposition of letting, telling (Stemberger & MacWhinney 1986: 23)

§29  Convergence of internal and psycholinguistic evidence?

Needless to say, the psycholinguistic evidence by itself does not suffice to settle the debate, as witnessed by broad disagreements among psycholinguists themselves. But hypotheses supported by convergent arguments from internal evidence and psycholinguistic data arguably have a particularly strong claim on our attention.

2. I refer to ‘higher recognition speeds’ rather than ‘lower reaction times’ or ‘lower recognition latencies’ so as to make the relationship with frequency direct rather than inverse.
Convergent psycholinguistic and internal evidence for stem storage

§30  Spanish nominal classes (Bermúdez-Otero 2013):

<table>
<thead>
<tr>
<th>Class</th>
<th>Theme</th>
<th>Singular</th>
<th>Plural</th>
<th>Gloss</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>o-stem</td>
<td>/-o-/</td>
<td>[li-o]</td>
<td>[li-o-s]</td>
<td>‘muddle’</td>
<td>M</td>
</tr>
<tr>
<td>a-stem</td>
<td>/-a-/</td>
<td>[kán-a]</td>
<td>[kán-a-s]</td>
<td>‘grey hair’</td>
<td>F</td>
</tr>
</tbody>
</table>

(Also e-stem and athematic stems.)

Hypothesis: the lexicon does not store bare roots, but complete stems with their theme vowels.

§31  Psycholinguistic evidence

E.g. the adjective CIEGO ‘blind’

<table>
<thead>
<tr>
<th>lexieme</th>
<th>/θje-g-o/</th>
<th>/θje-g-a/</th>
</tr>
</thead>
<tbody>
<tr>
<td>wordforms</td>
<td>/θjéγ-o/</td>
<td>/θjéγ-a/</td>
</tr>
<tr>
<td>M.SG</td>
<td>F.PL</td>
<td></td>
</tr>
</tbody>
</table>

The box highlights the items whose frequency is predicted to govern recognition speeds (see §26).


(i) CIEGO ‘blind’ vs VIUDO ‘widowed’

- CIEGO is masculine-dominant: frequency of cieg-o(-s) > frequency of cieg-a(-s)
- VIUDO is feminine-dominant: frequency of viud-o(-s) < frequency of viud-a(-s)

- recognition speed for cieg-o(-s) > recognition speed for cieg-a(-s)
- recognition speed for viud-o(-s) < recognition speed for viud-a(-s)

(ii) cult-o ‘cultivated M’ vs bell-o ‘beautiful M’

- frequency of cult-o(-s) = frequency of bell-a(-s)
- recognition speed for cult-o(-s) = recognition speed for bell-a(-s)

even though

- frequency of CULTO < frequency of BELLO
- frequency of cult-a(-s) < frequency of bell-a(-s)

(iii) rat-o-s ‘while.PL’ vs bot-a-s ‘boot.PL’

- frequency of wordform rat-o-s = frequency of wordform bot-a-s
- recognition speed for wordform rat-o-s > recognition speed for wordform bot-a-s
- frequency of stem rat-o(-s) > frequency of stem bot-a(-s)
- frequency of wordform rat-o (SG) > frequency of wordform bot-a (SG)

§33  Internal evidence

• Storing stem allomorphs predicts the right local domains for allomorph selection:
  e.g. kwént-a kont-a-dó/FL027Eh-∅ Alternation between /wé/ and /o/ count-TH count-TH-er-TH governed by stress

  a. cyclic domain structure
     \[1^c \text{koNt-}\] \[1^c \text{kweNt-a}\] \[\text{do/FL027Eh-}\] \[\text{e}\]
  b. first cycle
     \[\text{kon.ta}\]
  c. second cycle
     \[\text{kwen.ta}\]

• Storing root allomorphs predicts the wrong local domains for allomorph selection:

  a. cyclic domain structure
     \[1^c \text{koNt}\] \[1^c \text{kweNt}\] \[\text{a}\] \[\text{do/FL027Eh-}\] \[\text{e}\]
  b. first cycle
     \[\text{kwen.ta}\]
  c. second cycle
     \[\text{kwen.ta.dó}\]

Halle et al. (1991) avoid this problem by treating diphthongization as a regular phonological rule instead of allomorph selection, but the cost is

(i) four abstract vowels
(ii) extrinsing ordering of diphthongization before stress assignment at the word level.

See §25 above on the theoretical baggage associated with full decomposition.
Convergent psycholinguistic and internal evidence for analytic lexical entries

§34  The hypothesis: Some morphologically complex forms have their own lexical entry, but their lexical phonological representation is decomposed into pieces

E.g. vocabulary item  FEATHERY  \(\leftrightarrow /\text{fəθər}/\)  
or in psycholinguistic terms  lemma  FEATHERY  \(\rightarrow /\text{fəθər}/\)  
form  /\text{fəθər}/  /-\text{er}/  (Tafli 2004: 747)

§35  Psycholinguistic evidence

German inflection and derivation (Clahsen et al. 2003)

<table>
<thead>
<tr>
<th>Type of item</th>
<th>Full priming?</th>
<th>Surface frequency effect?</th>
</tr>
</thead>
<tbody>
<tr>
<td>regular -s plural: e.g. Wagen-s</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>diminutive: e.g. kind-schen</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>irregular -er plural: e.g. kind-er</td>
<td>no</td>
<td>yes</td>
</tr>
</tbody>
</table>

Recall that full priming ⇒ evidence for decomposition (§27)  
surface frequency effect ⇒ evidence for own entry in lexicon (§26)  
Solution: a decomposed (analytic) entry KINDCHEN  \(\leftrightarrow /\text{kndl}/+\text{cən}/\)

§36  Internal evidence

Analytic entries are independently needed to explain the behaviour of items that are semantically noncompositional but phonologically complex.

Salient case: complex place names (Kohnlein 2015, Mascaro 2016)

E.g.  
Dutch Wagen-ing-[s]n  morphological complexity revealed by  
violation of trisyllabic stress window  
schwa after stressless syllable  
but the meaning of Wagen-ing-en is not compositionally derived from Wagen-ing- and -en.

Solution: ‘town in Gelderland’  \(\leftrightarrow\) WAGENINGEN  \(\leftrightarrow /\text{wægəning}/+\text{-en}/\)

References


Cutler, Anne, Jason Eisner, James M. McQueen & Dennis Norris. 2010. How abstract phonemic categories are necessary for coping with speaker-related variation. In Cécile Fougeron, Barbara Kühnert, Mariapaola


