Exploring the Link between Grammar and Speech Processing

Jennifer Cole
University of Illinois

The traditional view

- Phonology:
  - Where discrete categories of contrasting speech units are defined;
  - Where constraints are expressed on the contexts in which a speech unit may appear;
  - The input to phonetics.
The traditional view

Phonetics:
- Where articulatory and/or acoustic features are associated with phonological structures;
- Where the duration and extent of a speech unit are defined in terms of continuous parameters;
- The input to the processing domains (speech production and perception).

Lexicon:
- A catalogue of the phonological, morphological, syntactic and semantic features of each word,
- In terms of categorical, contrastive features.
Unresolved Issues

(Lindblom 2000)

- The psychological reality of the phonological encoding;
- Lack of phonetic invariance.

---

Unresolved Issues

(Pierrehumbert 1999)

- Grammatical knowledge is based on statistical generalizations
  - over speech tokens, expressing fine phonetic detail;
  - over the lexicon, expressed in terms of phonetic detail.
Emergent Phonology
(a.k.a. Exemplar-based phonology)

- Speech forms are encoded in terms of directly observable properties of the input signal.
- The encoding preserves all contextual information.
- Categories emerge from accumulated phonetic experience.

Resolves...

- Phonological encoding:
  - tied to perceived properties of stimulus.
- Invariance “problem”:
  - Variation is embraced.
- Stochastic grammars:
  - input data not reduced, statistical properties available.
- Lexical influence:
  - phonological categories emerge from exemplar-based lexicon.
Link between grammar and speech perception

- Phonological categories emerge directly from perceived features of the stimulus.

- Thus: perceptual salience is a criterion for the formation of a phonological category.

Link between grammar and speech production

- Articulatory reduction in unscripted, relaxed speech determines the acoustic variants of a speech unit. (Kohler 2000)

- Thus: articulatory effort is a criterion for the formation of a phonological category.
Challenge for Emergent Models

- Do all perceivable and statistically robust patterns of variation provide an equal basis for grammatical generalization?

Prediction of Emergent Models

- All sound patterns have their origin in variation at the articulatory or acoustic level.

- Other sources?
  - Other cognitive mechanisms involved in language processing.
Bases for sound patterns

- Auditory
  - time and frequency resolution.
- Vocal Tract/ Respiratory
  - movement and coordination of speech articulators; aerodynamic conditions
- Cognitive-Linguistic
  - category formation, speech planning, lexical access,...

Research direction

Seek correspondences between sound patterns and speech processing.

- Linking sound patterns to speech processing;
- Linking properties of speech production & perception to sound patterns.
Sound patterns based in perception

Example: distinctively nasalized vowels can result from contextual nasalization.

Sanskrit  
Hindi

\[ ñ \text{NC} \rightarrow ñ \text{NC} \rightarrow ñ \text{C} \]

Perceptually-driven Neutralization

Rhee 1998 UIUC diss.

Stop manner contrast in \( C_C \)

Korean Hindi
3-way contrast suspended 4-way contrast maintained
tense/aspirated/lax voiced/voiceless; asp./unasp.
Quantity-sensitive stress

Ahn 2000 UIUC diss.

- QS systems
  - ‘heavy’ syllables attract stress.
- Heavy syllables
  - have long vowel (CVV)
  - or coda consonant (CVC)
- CVC syllables attract stress
  - only in positions where CV syllables undergo lengthening.

---

Quantity-sensitive stress

- Latin
  - L L L
  - L H L
Stress repulsion in QS stress

Claims:
- CVC is not an inherent stress attractor.
- Contextual lengthening of CV is cumulative.
- Vowel duration is contrastive: CV vs. CVV
- CV repels stress when subject to lengthening, to preserve salient durational distinction.
- CVC does not repel stress, is not subject to lengthening.

Sound patterns based in articulation

Example: light vs. dark /l/ in American English.  
light, ball
- Dorso-velar constriction present in both.
- Gestures phased differently in onset vs. coda positions: /l/ and nasals.
  - Asynchronous gestures in coda.
  - Synchronous gestures in onset.
(Browman & Goldstein 1998)
A production account of assimilation (Lee 1999 UIUC diss.)

- Consonant assimilation reflects patterns of CV coarticulation.
- Increasing degrees of coarticulation results in shift in C place of articulation.
  - ti $\rightarrow$ ti $\rightarrow$ ti $\rightarrow$ tsi
- Assimilation of C place:
  - common in CV
  - rare in VC

Example of CV assimilation

- CV Palatalization:
  - ty, ti $\rightarrow$ [tsi]  dy, di $\rightarrow$ [dzi]
  - Optional in English: last year, Goodyear
  - Obligatory in Japanese:
    - ita “board”  utsi “house”
- A rare case of VC Palatalization:
  - Basque: [Hualde 1991]
    - amai-tu $\rightarrow$ amaitšu “to finish (perf.)”
Coarticulation is greater in CV than in VC

- Acoustic variation in C_i V C_i (English speakers).
- More variation in CV sequences than in VC sequences:
  - Consonant release burst
  - F2 transition onset/offset
- Patterns of phonetic variation parallel patterns of phonological assimilation.

A production account of coalescence (Iskarous 1998)

- Vowel coalescence observations
  - Common: a+i → e
  - Very rare: i+a → e
- Why? Dynamic patterning
  - (schematic from ultrasound tongue imaging)
How tightly linked are speech production and linguistic sound patterns?

Cole, Hualde & Iskarous 1999:
- Consonant lenition in VCV in Spanish.
  - *pide* vs. *digo*
- A production account:
  - the consonant is “pulled open” by the vowel gesture.
- What factors condition variation in degree of lenition?

Expectations based on production hypothesis

- Greatest degree of lenition when adjacent vowels are “open” /a/.
  - Example: *amagar*
- Lesser lenition in the context of adjacent “close” vowels /i,u/.
  - Example: *sagus*
Continuum of constriction degree

<table>
<thead>
<tr>
<th>fully open</th>
<th>fully constricted</th>
</tr>
</thead>
<tbody>
<tr>
<td>utterance</td>
<td>initial</td>
</tr>
<tr>
<td>-----------</td>
<td>------------------</td>
</tr>
</tbody>
</table>

Generalization: less constriction in intervocalic position

Vowel Assimilation

- A restriction on the vowel sequences that may occur in adjacent syllables in a word.
- A vowel must agree in one or more quality with the preceding or following vowel: **Vowel Harmony**
- Why?
### Vowel Qualities

<table>
<thead>
<tr>
<th></th>
<th>Front</th>
<th>Back</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>i</td>
<td>u</td>
</tr>
<tr>
<td>Non-High</td>
<td>e</td>
<td>o</td>
</tr>
<tr>
<td></td>
<td></td>
<td>a</td>
</tr>
</tbody>
</table>

#### Vowel Harmony Languages

- **Back/Front:**
  - Turkic, Mongolian, Finnish, Hungarian, Germanic (umlaut), ...

- **Turkish example:**
  - *ayak-tan*  “from the foot”
  - *inek-ten*  “from the cow”

- **High/Non-High:**
  - Algonquian, Australian aboriginal (Warlpiri), Romance (dialects of Italian & Spanish), Bantu...
Articulator stability with vowel harmony? No.

- Vowel assimilation across an intervening consonant.
  - *e.g.*, *ayak-tan*

- Tongue body position:
  - *a - t - a*
  - Lo   Hi   Lo
  - Bk   Fr   Bk

Speech planning

- Speech production study
  - [Khasanova, Cole & Dell, *in progress*]
- Hypothesis:
  - Sequences of vowels that share at least one feature will be easier to articulate.
  - Shared segment and syllable features are known to facilitate fast speech production.
    - -- Sevald & Dell 1994; Sevald, Dell & Cole 1995
A perceptual advantage to vowel assimilation?

Hypothesis:
Maintaining constant vowel quality in at least one feature on both sides of a vowel may facilitate consonant identification.

[b/d/g] identification

Cued by:
- Formant transitions preceding and following the stop closure in VCV. [Delattre, et al. 1955; Öhman 1966]
- Spectral properties of the stop burst. [Stevens & Blumstein 1978, Stevens 1985]
- Both cues heavily conditioned by the flanking vowel quality.
**Experiment 1: C-identification**

- Perception of C-place in VCV
  - front/back “harmonic” vs. “disharmonic” vowel sequences.
- Target stimuli:
  - /b,d,g/
- Adjacent vowels:
  - **Back** (u,o)
  - **Front** (i,e)

---

**Perception in Noise**

- Noise reduces the perceptual salience of C-Place features.
  --- Miller & Niceley 1955; Wang & Bilger 1973

- Pilot study: main effect of noise level on C-Place identification.
Stimuli

- Nonsense words: \textit{omedu, omogi, omubo}
- 3 consonants: b/d/g
- 16 vowel contexts for VCV: (i/e/u/o)
- Uniform noise signal at two amplitude levels:
  - .25 and .75 times the amplitude of clear speech
- 288 distinct stimuli (3 \times 16 \times 2 \times 3)

Participants

- **Speaker:** 3 male speakers of American English

- **Subjects:** 17 American English speakers, UIUC undergrads, monolingual, from Chicago area.
Measurements

- Responses converted to sensitivity measurement (I).
  \[ I = 1 - P(\text{fa}) + P(\text{hit}) / 2 \]

“\( I \)” increases with identification accuracy.

Results

Main Finding:

Vowel harmony has no significant effect on Consonant place identification for intervening stop consonants.
Perceptual biases for /b,d,g/: 

- Identification is worst in Fr_Fr.
- A potential context for the suspension of a Consonant place contrast??
  - No evidence from existing languages...

- Implication: language can tolerate weakly-cued contrasts.

---

Perceptual biases for /b,d,g/: 

- No consistent facilitating context.

- Bk_Bk is best for /b,d/.

- _Bk reduces /g/ identification.
  - /g/ is lost in this position in some Spanish words: *agua* → *a(g)wa*
A perceptual advantage to vowel assimilation?

Hypothesis:
Maintaining at least one common vowel feature across a V-V sequence in adjacent syllables may facilitate vowel identification.

Vowel place identification

- Vowel quality (place of articulation) is cued by first three formants.
  ---Chiba & Kajiyama 1941

- Vowel formants vary due to V-V coarticulation, in both directions, even across a consonant. ---Öhman 1966
Experiment 2: V-Identification

- Perception of V1 in V1-C-V2
  - front/back “harmonic” vs. “disharmonic” vowel sequences.
- Target stimuli
  - /I,E,O,U/ lax vowels
- Following vowel contexts
  - Back (“ow”)
  - Front (“ay”)

Stimuli

- Nonsense words
  - *bootay, bikow*
  - Stress on final vowel, target vowel unstressed.
  - 16 nonsense words
    - 4 target vowels /I,E,U,O/
    - 2 intervening consonants /t,k/
    - 2 following vowel contexts /“ay”, “ow”/
  - 2 noise levels
  - 2 speakers (American English)
  - 16 x 2 x 2 = 64 distinct stimuli
Results

Main Finding:

Vowel harmony has no significant effect on vowel place identification for the preceding and unstressed vowel in the harmony sequence.

Combined Results: Experiments 1 and 2

- Front/Back vowel harmony offers no advantage for the perception of place of articulation of stop-consonants or vowels in a harmony domain.
A Cognitive-Linguistic constraint

- Vowel harmony introduces redundancy.
  - The harmony feature only counts once per harmony domain (typically the morpheme or word).
- Vowel harmony introduces shared features across syllables.
  - Shared features facilitates speech production.

Conclusions

- (Old) Some constraints on speech production and perception are reflected in linguistic sound patterns.
- But: The linguistic sound pattern may generalize the environment in which the constraint holds. (Spanish /g/-Lenition)
Conclusion

- (New) Some constraints on perception seem unrelated to linguistic sound patterns.
  - /b,d,g/ identification: worst in Fr_Fr

- Language can tolerate some weakly-cued contrasts.

Conclusion

- (New) Some sound patterns lack a clear basis in articulation or perception.

- Motivation may be found in other cognitive mechanisms related to speech processing (Back/Front vowel harmony)
  - An advantage for lexical redundancy?
  - Shared features facilitate speech planning?