Categories and processes in phonological acquisition and adult speech perception

Sharon Peperkamp

Laboratoire de Sciences Cognitives et Psycholinguistique
(EHESS, DEC-ENS, CNRS)
A simple model of speech processing

PERCEPTION:  
DECODING  

underlying form  

phonological decoding:  
compensation for phonological processes  

phonological surface form  

phonetic decoding:  
discretizing continuous representation in terms of native categories  

phonetic surface form  

PRODUCTION:  
ENCODING  

underlying form

PHONOLOGY

PHONETICS

[Diagram showing the process of speech processing with phonological and phonetic decoding.]
Part 1: phonetic decoding

PERCEPTION:
DECODING

underlying form

phonological surface form

phonetic surface form

phonic decoding:
discretizing continuous representation in terms of native categories

PRODUCTION:
ENCODING

underlying form

phonological surface form

phonetic surface form

PHONOLOGY

PHONETICS
Perceptual assimilation

• Non-native speech is distorted and perceived in terms of native categories (Sapir 1921; Polivanov 1931; Trubetzkoy 1939)

• Experimental evidence
  – segments: Japanese listeners, English /l/-/r/ contrast (Goto 1971)
  – suprasegments: English listeners, Mandarin tone contrasts (Kirilof, 1969)
  – syllable structure: Japanese listeners, French consonant clusters (Dupoux et al., 1999)
Acquisition of categories

• L1 phonological categories are acquired early
  – vowels: 6 months (Kuhl et al., 1992; Polka & Werker, 1994)
  – consonants: 10 months (Werker & Tees, 1984)
  – tones: 9 months (Mattock & Burnham, 2006)
  – syllables: 14 months (Dupoux et al., in press)

• L2 phonological categories are hard to learn
  – vowels Pallier et al., 1997; Flege et al., 1999
  – consonants Takagi & Mann, 1995; Lively et al., 1993
  – tones Stagray & Downs, 1993; Wang et al., 1999
  – syllables Dupoux et al., in press
Case study: stress

Stress

– French: non-contrastive
– Spanish: contrastive
  • /'bebe/ ‘s/he drinks’
  • /beˈbe/ ‘baby’

Sequence repetition task

– AABAB → reply 11212
– sequences with length 2 to 6
– conditions:
  • stress: mípa vs. mipá
  • phoneme: tuku vs. tupu

Dupoux, Peperkamp & Sebastián-Gallés (2001)
Questions

French listeners exhibit stress ‘deafness’

→ When is fixed stress acquired?

→ To what extent can stress ‘deafness’ be overcome?
Stress perception in 9-month old French and Spanish infants

- Head-turn preferential listening

- Two familiarization groups:
  - initial stress (/datu, sapi, kiba, nuki, latu, buki, luma, tiku/)
  - final stress (/da'tu, sa'pi, ki'ba, nu'ki, la'tu, bu'ki, lu'ma, ti'ku/)

- Eight test sequences with novel non-words:
  - four stress-initial (/lapi, naku, nila, tuli/)
  - four stress-final (/ki'bu, lu'ta, pi'ma, pu'ki/)
Results

Skoruppa, Pons, Christophe, Bosch, Dupoux, Sebastián-Gallés, Limissuri & Peperkamp (2009)
Infants’ sensitivity to the acoustic correlates of stress

• Are French infants incapable of perceiving acoustic correlates of stress?

• Familiarization: multiple tokens of a single non-word
  – initial stress (/pima/)
  – final stress (/pi'ma/)

• Test: eight trials
  – four stress-initial (/pima/)
  – four stress-final (/pi'ma/)
Results

Skoruppa, Pons, Christophe, Bosch, Dupoux, Sebastián-Gallés, Limissuri & Peperkamp (2009)
Discussion

• The role of stress (predictable or contrastive) is acquired before 9 months
• At this age, not much lexical information is available
  – no minimal pairs
  – no reliable word segmentation
• A possible acquisition mechanism: examination of stress at utterance edges (variable in Spanish vs. final in French)
Stress perception in French-Spanish bilinguals

• 39 late learners
  – native speakers of French
  – learned Spanish after age 11
  – 6 in a Spanish dept. in Parisian university, 33 in Barcelona (> six months)
  – three groups according to degree of practice (questionnaire): beginners, intermediate, advanced

• 23 native (simultaneous) bilinguals
  – one French-speaking and one Spanish-speaking parent
  – roughly half of them lived and were tested in Paris, the other half in Barcelona
Prelexical processing

• Sequence recall task: AABA → reply 1121

• Conditions: *phoneme* [fiku] - [fitu]
  *stress* [númi] - [numí]

Lexical processing

Lexical decision

- stress: melón – *mélon ‘melon’
- (phoneme: blanco – *blanto ‘white’)


Summary of results

• late bilinguals:
  – not different from French monolinguals
  – no differences according to degree of practice

• native bilinguals
  – in between French monolinguals/late bilinguals and Spanish monolinguals
  $\rightarrow$ monomodal or bimodal distribution of individual scores?
Density

0.0 0.2 0.4 0.6 0.8 1.0

-2 0 2 4 6 8

Native bilinguals

Normalized individual composite stress 'deafness' index

French late bilinguals

Spanish monolinguals
Language dominance measures

**Binary measures**: language spoken in country of residence

**Gradient measures**: estimated amount of exposure/use (except ‘prenatal’: language of the mother)

Correlation coefficients between Z-scores and lang. dominance measures

- **biographic**
- **subjective**

---

- prenatal
- 0-2 yrs
- 2-4 yrs
- 4-10 yrs
- 10-18 yrs
- current
- fluency
- importance
Discussion

• Stress perception is highly non-plastic in adults
  – late French-Spanish bilinguals have as much difficulty perceiving stress as monolingual French speakers
  – about half of native French-Spanish bilinguals have difficulty perceiving stress
    (see also Grosjean 1989; Cutler et al., 1989, 1992; Sebastián-Gallés et al., 2005)

• The stress parameter is fixed early in life
  – the performance of native bilinguals is correlated with language exposure during their first 4 years of life
    (see also Sebastián-Gallés et al., 2005)
How to account for perceptual assimilation?

• **Two-step model** (Church, 1997; Berent et al., 2007)
  – step 1: segmental repairs
    • computed by acoustic or articulatory distance metric (Kuhl 1993, Best 1994)
  – step 2: syllable structure repairs
    • computed by the phonological grammar (Berent et al., 2007)
How to account for perceptual assimilation?

- **One-step model** (Peperkamp & Dupoux, 2003; Peperkamp, Vendelin & Nakamura, 2008)
  - simultaneous processing of segments and syllables

- **Hidden Markov Model** (Dupoux et al., submitted):
  
  maximize the probability of sequence $s_1...s_N$ given speech signal $x_1...x_N$:
  
  $p(s_1...s_N \mid x_1...x_N) \sim p(s_1...s_N) \cdot p(x_1...x_N \mid s_1...s_N) = p(s_1...s_N) \cdot \prod_{i=1}^{N} p(x_i \mid s_i)$
Empirical test:
Perception of illegal consonant clusters

Three languages with restrictions on coda consonants:
- Japanese
- Brazilian Portuguese
- European Portuguese

Phonology:
- Japanese: /wu/- and /i/-epenthesys in morphologically complex words

Phonetics:
- Japanese: [wu] is the shortest vowel (Han, 1962)
- Brazilian Portuguese: [i] is the shortest vowel (Escudero et al., 2009)
- European Portuguese: fast speech unstressed vowel deletion
Predictions

• Two-step model
  – perceptual epenthesis only in Japanese
    (or in all three languages if ‘hidden rankings’, Berent et al., 2009)
  – epenthetic vowel can be either /u/ or /i/
  – the choice of the epenthetic vowel is independent of segmental environment

• One-step model
  – perceptual epenthesis in Japanese and Brazilian Portuguese
  – default epenthetic vowel is the phonetically minimal one: /u/ in Japanese, /i/ in Brazilian Portuguese
  – choice of the epenthetic vowel influenced by segmental environment
Experiment 1: vowel classification

• Stimuli
  – 13 VCCV nonwords
  – three stimuli per item
    • VCCV
    • VC(u)CV
    • VC(i)CV

• Task
  – forced choice:
  – [VC?CV] → [a], [e], [i], [o], [u], or nothing
Global results

Dupoux, Parlato, Frota, Hirose & Peperkamp (submitted)
Effect of coarticulation

![Graph showing the effect of coarticulation between [i] and [u] sounds in Brazilian Portuguese and Japanese. The graph displays the i-u Labelization Difference Score for [i]-coarticulated, natural, and [u]-coarticulated sounds, with error bars indicating variability. The data shows that there are more [i] responses in Brazilian Portuguese and more [u] responses in Japanese.]

Dupoux, Parlato, Frota, Hirose & Peperkamp (submitted)
Experiment 2: ABX discrimination

• Conditions
  
  \[\text{ebizo} - \text{ebzo} \quad \text{ebuzo} - \text{ebzo}\]
  
  \[\text{ebizo} - \text{eb(i)zo} \quad \text{ebuzo} - \text{eb(i)zo}\]
  
  \[\text{ebizo} - \text{eb(u)zo} \quad \text{ebuzo} - \text{eb(u)zo}\]

• Stimuli:

  same 13 items as Exp. 1
  
  • A and B: female speakers
  • X: male speaker
Global results

Dupoux, Parlato, Frota, Hirose & Peperkamp (submitted)
Effect of coarticulation

/i/ vs. cluster more difficult
/u/ vs. cluster more difficult

Dupoux, Parlato, Frota, Hirose & Peperkamp (submitted)
Discussion

• Perceptual assimilation of illegal segments and of illegal phonotactics takes place in a single step

• Epenthetic vowels are not inserted by the phonological grammar, but by a low-level mechanism such as a HMM
Part 2: phonological decoding

PERCEPTION: DECODING

underlying form

phonological decoding:
compensation for phonological processes

phonological surface form

PHONOLOGY

PHONETICS

phonetic surface form

PRODUCTION: ENCODING

underlying form

phonological surface form

phonetic surface form
Assimilation

French:
regressive **voice** assimilation in obstruent clusters

\[mec\ doux \rightarrow me[g]\ doux\] ‘soft guy’

English:
regressive **place** assimilation of /t,d,n/ before labials and velars

\[sweet\ girl \rightarrow swee[k]\ girl\]
Compensation for assimilation

Word detection task

\textit{ex., bottes:}

Elle amis ses bottes rouges aujourd’hui.

‘She’s put on her red boots today.’

<table>
<thead>
<tr>
<th></th>
<th>Native: voice</th>
<th></th>
<th>Non-native: place</th>
</tr>
</thead>
<tbody>
<tr>
<td>control</td>
<td>…bottes rouges…</td>
<td>‘red boots’</td>
<td>…bottes rouges…</td>
</tr>
<tr>
<td>viable assimilation</td>
<td>…bo[d]es grises…</td>
<td>‘grey boots’</td>
<td>…bo[d]es grises…</td>
</tr>
<tr>
<td>unviable assimilation</td>
<td>…bo[d]es noires…</td>
<td>‘black boods’</td>
<td>…bo[k]es noires…</td>
</tr>
</tbody>
</table>
Results

Darcy, Ramus, Christophe, Kinzler & Dupoux (2009)
Questions

Adult listeners compensate more for native than for non-native assimilation rules
(see also Marslen-Wilson et al., 1995; Gaskell & Marslen-Wilson 1996, 1998, 2001; Coenen et al., 2001; Gow 2001, 2002; Mitterer & Blomert 2003; Mitterer et al., 2006a; but cf. Gow & Im 2004, Mitterer et al., 2006b)

→ When are native assimilations acquired?

→ To what extent can adults learn assimilations in L2?
Compensation for assimilation in French toddlers

familiar object: bus

unfamiliar object: bu[z]

control_familiar
Montre le bu[s]
Show the bus

control_unfamiliar
Montre le bu[z]
Show the bu[z]

viable assimilation
Montre le bu[z] de Paul
Show Paul’s bus/[z]

unviable assimilation
Montre le bu[z] là-bas
Show the bu[z] over there
Results (N=27, 29–36 months)

Skoruppa & Peperkamp (submitted)
Compensation for assimilation in French toddlers

familiar object: clown  
unfamiliar object: clow[m]

<table>
<thead>
<tr>
<th>control_familiar</th>
<th>Montre le clown</th>
<th>Show the clown</th>
</tr>
</thead>
<tbody>
<tr>
<td>control_unfamiliar</td>
<td>Montre le clow[m]</td>
<td>Show the clow[m]</td>
</tr>
<tr>
<td>‘viable’ assimilation</td>
<td>Montre le clow[m] par ici</td>
<td>Show the clow[m] over here</td>
</tr>
<tr>
<td>unavoidable assimilation</td>
<td>Montre le clow[m] là-bas</td>
<td>Show the clow[m] over there</td>
</tr>
</tbody>
</table>
Results (N=27, 29–36 months)

Skoruppa & Peperkamp (submitted)
33-month-old French toddlers compensate for native voice assimilation but not for non-native place assimilation
How about L2 learners?

- Two groups of English learners of French
  - low proficiency
  - high proficiency (Paris residents)

Darcy, Peperkamp & Dupoux (2007)

![Bar chart showing % detection for Low proficiency with categories Voice and Place](chart.png)
How about L2 learners?

- Two groups of English learners of French
  - low proficiency
  - high proficiency (Paris residents)

Darcy, Peperkamp & Dupoux (2007)
Compensation index =
\[
\frac{\text{detection(viable)} - \text{detection(unviable)}}{\text{detection(control)} - \text{detection(unviable)}}
\]
Discussion

- Adults can learn L2 phonological processes
- Ties in with previous findings:
  - Adults can learn to compensate for dialectal processes
    (Scott & Cutler, 1984; Sumner & Samuel, 2009)
  - Adults can learn phonological processes in artificial language-learning experiments
    (Pycha et al., 2003; Peperkamp et al., 2006; Peperkamp & Dupoux, 2007; Wilson, 2006; Moreton, 2008; Finley & Badecker, 2009)
Investigating learning mechanisms

• How do adults learn phonological processes?

• Artificial language-learning experiments:
  – distributional learning
    (Pycha et al., 2003; Wilson, 2006; Moreton, 2008; Finley & Badecker, 2009)
  – feature-based learning
    (Pycha et al., 2003; Finley & Badecker, 2009)

• Aim: investigate mechanisms of adult phonological learning using an *implicit* learning paradigm
Accented French

Alternation: changes in rounding of front vowels

- round + round

**Harmonic French**

- *liqueur* → *liquère* ‘liqueur’
- *adultère* → *adulteur* ‘adultery’

**Disharmonic French**

- *pudeur* → *pudère* ‘decency’
- *misère* → *miseur* ‘misery’
Design

Exposure (40 min.):
- Short stories spoken in accented French
  (cf. Maye, Aslin & Tanenhaus, 2008)
- Natural recordings

**Standard French:**  *Sans pudeur, il boit de la liqueur.*
  ‘He drinks liqueur shamelessly.’

**Harmonic French:**  *Sans pudeur, il boit de la liqueur.*

**Disharmonic French:**  *Sans pudère, il boit de la liqueur.*
Test phase

• Forced choice accent identification: *liquère* - *pudère*

• 60 harmony-disharmony pairs
  – matched for
    • frequency
    • grammatical category
    • number of phonemes
    • number of morphemes
    • number of syllables
  – no items yielding real words
    (hideux ‘hideous’ → idée ‘idea’)
  – no items yielding identical vowels
    (tissu ‘fabric’ → tissi)
  – 20 exposure pairs (both words used 2-6 times in stories),
    40 novel pairs
Mixing the alternations
(cf. Pycha et al., 2003)

**Mixed French:**

- **Harmony for mid vowels**
  - *liqueur* → *liquère*
  - *adultère* → *adulteur*

- **Disharmony for high vowels**
  - *désir* → *désure*
  - *laitue* → *laitie*
Possible learning mechanisms

• Feature-based learning
  – Harmonic French:
    • $V_{[+\text{front}, \alpha \text{ round}]}$ occurs after $V_{[+\text{front}, \alpha \text{ round}]}$
  – Disharmonic French:
    • $V_{[+\text{front}, \alpha \text{ round}]}$ occurs after $V_{[+\text{front}, \beta \text{ round}]}$
  – Mixed French:
    • $V_{[-\text{high}, +\text{front}, \alpha \text{ round}]}$ occurs after $V_{[+\text{front}, \alpha \text{ round}]}$
    • $V_{[+\text{high}, +\text{front}, \alpha \text{ round}]}$ occurs after $V_{[+\text{front}, \beta \text{ round}]}$

• Prediction
  – mixed French is *more* difficult than either Harmonic or Disharmonic French
Possible learning mechanisms

- Feature-less learning
  - Harmonic French:
    - [i,e,ε] occur after [i,e,ε]
    - [y,φ,œ] occur after [y,φ,œ]
  - Disharmonic French:
    - [i,e,ε] occur after [y,φ,œ]
    - [y,φ,œ] occur after [i,e,ε]
  - Mixed French:
    - [y,e,ε] occur after [i,e,ε]
    - [i,φ,œ] occur after [y,φ,œ]

- Prediction:
  - mixed French is *equally* difficult as Harmonic and Disharmonic French
trained items novel items

Harmony Disharmony Mixed

mean percentage of correct response

exposure items novel items

Skoruppa & Peperkamp (submitted)
Discussion

• Same results as Pycha et al. (2003)
  – no difference between harmony and disharmony
  – worse performance on mixture of both

• Hence, vowels are represented with internal structure, not as atomic elements

• Note: no evidence for abstract, phonological features!
Conclusion

- Phonetic decoding
  - is acquired during the first year of life
  - is responsible for perceptual assimilation of illegal phonological **categories** (segments, suprasegments and syllables)
  - treats illegal segments and illegal syllables (as well as illegal suprasegments?) in a single processing step
  - is highly non-plastic
Conclusion

• Phonological decoding
  – is acquired before three years of age
  – is responsible for undoing the effects of
  – is learned - at least in adults - with reference to
    the internal structure of segments
  – is to a large extent plastic
Conclusion

• Different cortical structures? (Marian, Spivey, and Hirsch 2003)
  – Superior Temporal Gyrus
    • same centers of activation for L1 and L2
    • activated during phonetic processing
  – Inferior Frontal Gyrus
    • different centers of activation for L1 and L2
    • activated during lexical processing
PERCEPTION:
DECODING

underlying form

phonological decoding:
Inferior Frontal Gyrus

phonological surface form

PHONOLOGY

PHONETICS

phonetic decoding:
Superior Temporal Gyrus

phonetic surface form

underlying form

PRODUCTION:
ENCODING

underlying form

phonological surface form

phonetic surface form

PHONOLOGY

PHONETICS
Acknowledgments

- Emmanuel Dupoux, LSCP
- Núria Sebastián-Gallés, Pompeu Fabra University
- Laura Bosch, University of Barcelona
- Katrin Skoruppa, University College London
- Isabelle Darcy, Indiana University
- Erika Parlato, Universidade Federal de Minas Gerais