Adult learners process phonology from first exposure

S. E. Carroll, S. Jackson & D. MacDonald
University of Calgary
carrolls@ucalgary.ca

Mind-Context Divide, University of Iowa April 30th-May 2nd, 2009

Research questions of "Naming" Project

- If words are form-meaning mappings, how early in L2 acquisition can adult learners
  - acquire representations of forms?
  - map forms to meanings (directly or via syntactic mediation)?
  - map meanings to forms (directly or via syntactic mediation)?
- What processes are involved in doing these things?
- Does L1 knowledge play a role in doing these things?

Research questions of Studies 1 & 2 of project

- How early in L2 acquisition can adult learners segment word forms from the signal?
  - Answer: Right from first exposure
- Are cognate words easier to segment? (Rast 2008)
  - Yes
- Can learners simultaneously map forms to meanings?
  - Answer: If “meaning” = reference: “yes”.
  - This fact, however, says nothing about the emergence of lexical semantics

Where we are headed …

- If segmentation presupposes the ability to compute transitional probabilities over syllables (a unit of phonology), then *ab initio* adult learners can compute phonological units from first exposure.
- If the ability to produce speech sounds on the basis of perception requires the mediation of phonological representations, then *ab initio* adult learners can compute phonological units from first exposure.
- If the activation of L1 "cognate" words from L2 stimuli requires the activation of Lexical Access Units (LAUs) that are phonological representations, then *ab initio* adult learners can compute phonological units from first exposure.

Words as interfaces across language modules

- Tri-partitite modular functional architecture of language à la Jackendoff (1989, 1997)
  - Prosodic Representations (PR)
  - Morpho-syntactic Representations (M-SR)
  - Conceptual Representations (CR)
    - that get input from PR and/or M-SR and link back to PR and/or M-SR
- These representations are linked by Correspondence Rules
  - Some are defaults, e.g. referential expressions ⇔ X^{max}; Prosodic Word ⇔ X^{0}
  - Some are acquired, e.g. Focus Tier […] ⇔ Focal Accent
- Words may have representations of each sort, but
  - one type of representation can be acquired independently of the others
  - which means that they need not be acquired at the same time

Grammatical assumptions

- Prosodic Representations include
  - Hierarchical metrical structures (syllables, feet, prosodic words, intonated phrases, …), cf. Nespor & Vogel (1986)
- Morpho-syntactic Representations are
  - Monostratal (no UTAH, no movements), cf. Culicover & Jackendoff (2005)
- Conceptual Representations
  - include Function-Argument Structures
    - in particular, TOKENi, TOKENj (referential use of noun phrases in context)
  - 2 tiers to represent semantic roles (Thematic Tier, Action Tier)
  - Focus Tier
  - Link to Spatial Structures that interface with sensory-motoric systems (including vision)
- Lexical units can be any size from
  - Morpheme, to
  - Sentence (e.g. X *isn't what* pro*'s cracked up to be*; *There's many a slip twixt cup and lip; as you know*, …)

Acquisitional assumptions

- *Concrete Minimalism* (Culicover 1999): the language learner posits the minimum structure necessary to relate sound and meaning
- Learners can make use of domain-general computational abilities (statistical learning)
  - to segment sound forms (Saffran et al. 1996; Peña et al. 2002; Bonatti et al. 2005, inter alia)
    - that are constrained in the language faculty to operate on critical units (syllables)

Studies 1 & 2

Within-subjects design

- Study 1: Task in English (baseline) and in German (cognate words)
- Study 2: Both tasks in German: cognate words vs. non-cognate words
- Stimuli presented via computer using E-Prime

Ab initio learners

- In each experiment, Anglophones with no previous exposure to German
  - listened to 20 New Information (deictic) sentences that introduce a referent
  - while looking at 20 line drawings of people.
- They were asked to "learn the names" of the people.

Example Stimuli

- Training phase

Das ist Sonja.

+

Verification Measure

- At end of all 20 items of the Training Trial
- Same picture
- Verification phase

Ist das NP_1 oder NP_2?

Ist das Sandra oder

Sonja?

If all 20 Qs answered correctly → Test phase

If any error → Re-do Training Trial (up to 10)

Test Phase

- Different picture
- Same Q as before
- (1/2 items in medial position; ½ items in final position)

Re-test Measure

- 2 weeks later with original picture
And different Q
Ist das Sonja oder Sandra?
(reversing order of NPs)

To recapitulate

- Training items:
  - 20 statements, names are sentence-final
  - Sentences fully randomised
  - 4 German/3 English presentational contexts
    - *Hier ist Johann; Das ist Silke; Hier sehen Sie Frank; Da steht Ingelore*
    - *This is Peter; Here we have Angela; Now we see Carol*

Procedures I

Session 1

- Questionnaire
- Digital recall test
- Exp1 (either English or German)
- Exp2 (either German or English)

Session 2 (2 weeks later)

- Re-test1 (either German or English)
- Re-test2
- Re-test1 (either English or German)
- Re-test2

Measures

- Accuracy on first training trial
- Number of training trials to criterion
- Accuracy on Test
- Accuracy on Re-test1 (order of items reversed)
- Accuracy on Re-test2 after feedback

We also collected RTs and pronunciations from Test phase onwards

Participants: Study 1

- 24 UofC students on English experiment, all learned to criterion.
  - 22 returned for Re-test session
- 25 began German experiment, 23 learned to criterion
  - 22 returned for Re-test session

Results: Study1

	Trial1	# Trials	Test	Retest 1	Retest2
English	92.6%	E<G: 2.38* G<E: 1.63	97.12%	82.5%	92.95%
German	85.9%	E<G: 3.83 G<E: 2.54	96.2%	87.9%	97.6%

Results: Study1

	Trial1	# Trials	Test	Retest1	Retest2
English	92.6%	E<G: 2.38* G<E: 1.63 = 2	97.12% * ***	82.5%	92.95% ***
German	85.9%	3.18	96.2% *** ***	87.9%	97.6% ***

Results: Study1

	Trial1	# Trials	Test	Retest1	Retest2
English	92.6% **	E<G: 2.38 G<E: 1.63 **	97.12%	82.5%	92.95%
German	85.9%	E<G: 3.83 G<E: 2.54	96.2%	87.9% *	97.6%

Study 2

- Comparison of cognate names from Study 1 with non-cognate names, e.g. *Benno, Ingelore, Lutz, Regina*
- Stimuli prepared same way
- Same procedures
- Same measures

Participants (Study 2)

- 33 tested,
  - 24 learned to criterion on cognate test/23 on non-cognate test
  - 21 returned for Re-test session

Results: Study2

	Trial1	# Trials	Test	Retest1	Retest2
Cognates <Data from S1>	82.7% <85.9%>	C<NC: 4.08* NC<C: 2 = 3.04 <3.18>	95% <96.2%>	85.9% <87.9%>	93.4% <97.6%>
Non-cognates	70.6%	5.56	95%	79.04%	90%

Results: Study2

	Trial1	# Trials	Test	Retest1	Retest2
Cognates	82.7%	C<NC: 4.08* NC<C: 2 = 3.04	95% *** ***	85.9%	93.4% ***
Non-cognates	70.6%	5.56	95% *** ***	79.04%	90% ***

Results: Study2

	Trial1	# Trials	Test	Retest1	Retest2
Cognates	82.7% ***	C<NC: 4.08 NC<C: 2 = 3.04	95%	85.9% **	93.4%
Non-cognates	70.6%	C<NC = 4.9 NC<C = 6.16 =5.56 *** for NC<C group	95%	79.04%	90%

Summary

- Task is easy (subjects well above chance performance on Trial1, Test, and Re-tests on all item types)
  - easier in English than in German but once mappings are in place, not much difference in ability to recall mappings (Test, Re-test2)
  - Task is easier for German cognate names than for non-cognates but once mappings are in place, not much difference in ability to recall mappings (Test, Re-test2)

Why is this task so easy ?

- task is ecologically valid (we do map names to pictures a lot)
- since subjects understand the instructions, they understand the task (inference)
- in L1
  - because of knowledge of L1 phonology, grammar & probably vocabulary, and
- in the L2
  - because learners can segment the names from signal, and map to a TOKENi representation created from visual inputs

“Bottom-up” stuff

- Acoustic-to-prosodic mappings
  - Learners are able to locate syllable boundaries
  - and compute *transitional probabilities* to segment a sound form

What is a Transitional Probability (TP)?

- Given σ1, TP is likelihood σ2 will occur
- A form of statistical learning
  - Ability to compute TPs on syllables is robust in both adults & young children (Saffran et al. 1996)
- Statistical learning can occur on visual stimuli and tones
  - Ability is domain-general (Fiser & Aslin 2001, 2002; Creel et al. 2004)
  - Not specific to humans (Hauser et al. 2001)

Transitional Probabilities (TPs)

- Das#ist#Benno

  1 .1 1

- Da#steht Johanna

  1 .1 1 .3333

But notice

- statistical learning alone will not explain the differences between cognate and non-cognate items
  - because the TPs are the same between the "frame" and the 1st syllable of both type of name

Explanation

- Performance is * better with German cognates because
  - this is L1-word activation
  - L1-based lexical activation is automatic
  - L1-based word recognition is extremely robust (German speech signal = “noisy English input”)

⇨ Effects of L1 knowledge manifest at initial stage of L2 learning

⇨ This L1-based word recognition also requires computation of phonological representations (LAUs)

But …

- Not everyone thinks LAUs are phonological
- Minimally normalised acoustic templates (Klatt 1979)?
  - Not likely
    - because the German sounds are quite different from the English and the processing is too fast (subjects do not need multiple exposures)
- Phonological representations?
  - Underspecified segmental/featural representations computed left-to-right? (Marslen-Wilson & Lahiri 1991)
  - Or consonantal tiers only? (Bonatti, Peña, Nespor, & Mehler, 2005)

Trial1 (both studies)

- It is hard to argue for a given model
- We looked at several variables
- Explaining the "Josef" effect (it's easy), "Albert" is hard?

No. of syllables	No
Stress placement	No
Matching stress	No
Complex onsets	No
Complex codas	No
<r> in onsets	No
<r> in codas	Yes

Phonetics matters (<r> effect)

⇒ Subjects have enriched phonological representations on the forced-choice task.

- We think the task also matters

[fʁaŋk] … [fʁants] [anə] … [anita]

and are properties of entire PW utilised?

Top-down stuff

- Creation of visual image of “person” leads to construction of TOKENi in conceptual representations
- These might be used to activate Nouns in lexicon (speeding up the activation of English proper names)
- So top-down mapping might “win out” over a bottom-up computation in the case of the cognates

Conclusions

- If segmentation relies on identifying syllables and computing transitional probabilities over syllables,

⇨ subjects are computing phonological representations from first exposure

- If our cognate results show lexical activation of L1 names from L2 input and lexical access codes are phonological, then

⇨ subjects are computing L1 phonological representations from first exposure

⇨ Subjects are also computing meanings from first exposure

Implications for interface theories

- Our subjects can use domain-independent processes to extract prosodic words from the signal virtually instantaneously (phonetic-phonology mappings)
- They can map across the Phonological-Conceptual modules virtually instantaneously (encoding TOKENi)
- But no one should assume that predicate-argument structures or semantic roles (focus/topic, quantification …) are instantly available from the context
  - because reference does not determine "lexical semantics"
  - and proper names prove it (they have virtually no "semantic" features, cf. Katz 1977)

Acknowledgements

Canada Research Chairs Program (grant no. 950-202408)

Canada Research Chairs

Chaires de recherche du Canada

Canada

Chaires de recherche du Canada

Canada Research Chairs

Canada

UNIVERSITY OF CALGARY

References I

Bonatti, L., Peña, M., Nespor, M., & Mehler, J. 2005. Linguistic constraints on statistical computation: The role of consonants and vowels in continuous speech processing. *Psychological Science 16*: 451-59.

Creel, S.C., Newport, E.L., & Aslin, R.N. 2004. Distant melodies: Statistical learning of non- adjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory & Cognition 30*: 1119-30.

Culicover, P. 1999. *Syntactic nuts: Hard cases in syntactic theory*. Oxford: Oxford University Press.

Culicover, P., & Jackendoff, R. 2005. *Simpler syntax*. Oxford: Oxford University Press.

Fiser, J., & Aslin, R.N. 2001. Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science 12*: 499-504.

Hauser, M.D., Newport, E.L., & Aslin, R.N. 2001. Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition 78*: B53-B64.

Jackendoff, R. 1987. *Consciousness and the computational mind*. Cambridge, Ma: MIT Press.

Jackendoff, R. 1997. *The architecture of the language faculty*. Cambridge, Ma: MIT Press.

References II

Katz, J. 1977. A proper theory of names. *Philosophical Studies 31*: 1-80.

Klatt, D.H. 1979. Speech perception: A model of acoustic phonetic analysis and lexical access. *Journal of Phonetics 7*: 279-342.

Lahiri, A., & Marslen-Wilson, W. 1991. The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition 38*: 245-94.

Nespor, M., & Vogel, I. 1986. *Prosodic Phonology*. Dordrecht: Foris.

Peña, M., Bonatti, L., Nespor, M., & Mehler, J. 2002. Signal-driven computations in speech processing. *Science 298*(5593, Oct. 18): 604-7.

Rast, R. 2008. *Foreign language input: Initial processing*. Clevedon, U.K: Multilingual Matters.

Saffran, J.R., Newport, E.L., & Aslin, R.N. 1996. Word segmentation: The role of distributional cues. *Journal of Memory & Language 35*: 606-21.

Words without <r>: Trial 1

100% - √ match - Claudia:klawdiə/klɑdiV/ kld
Dipthong vs. simple vowel
97.3% - √ match - Otto: ɔtoː/ɑto/t
94.6% - √ match - Agnes: agnɛs/ægnVs/gns
85.3% - √ match - Anita: anitə/ænitV/nt
98.6% - x match - Edmund: ɛdmunt/ɛdmVnd/dmnd
differences are at end of word, so no problem
82.6% - x match - Sonja: zonja/sonja/snj
85.3% - x match - Eva: efa/iːv/v
98.6% - x match - Josef: joːsɛf/dʒosVf/dʒsf
94.6% - x match - Johanna:johaːnə/dʒoænV/dʒn

<r> effect

100% - Margarete: mrgrt – m_rgr$t
97.3% - Karin: krn – k$r_n
94.6% - Andreas: ndrs – _ndr_s
93.3% - Carolina: krln – krl$n
93.3% - Frank: frŋk – fr_ŋk
90.6% - Bruno: brn – br$n
88% - Harald: hrlt – h$r_lt
- Laura: lr - l$r
92% - Georg: grk – g$_rk
72% - Hermann: hrmn – h_r$m_n
61.3% - Albert: lbrt – _l$b_rt

No obvious lexical neighbourhood effects

- *Agnes, Anita, Bruno, Eve, Herman, Karin, Otto, Sonja,*

occupy a unique slot in lexical organisation

- crucially *Margareta, Margaret, Margo, Margie (Andreas, Andre, Andrew, Claude, Claudia, Claudine, Frank, Frankie, Francis, Francine, Franciska, Harald, Harry … Laura, Laurie, Lorna,* Joseph, Joe, Joanne, …)

do not