LINGUISTIC TYPOLOGY AND CROSS-LINGUISTIC PSYCHOLINGUISTICS

James Myers
Graduate Institute of Linguistics
National Chung Cheng University
Taiwan

Tsung-ying Chen
Dept. of Foreign Languages & Literature
National Tsinghua University
Taiwan

50th Annual Meeting of the Societas Linguistica Europaea, Zürich
12 Sept 2017
Welcome to our workshop!
OVERVIEW

- Why do typological psycholinguistics?
  - To see how language experience affects learning and processing, and thus...
  - ... to seek out and explain linguistic universals

- Two methodological approaches
  - Empiricist: Induce universals from large samples
  - Rationalist: Deduce universals from learnability

- Expanding the typology
TYPOLOGY AND PSYCHOLINGUISTICS

- Typologists are doing more experiments
  + E.g. Bickel et al. (2015)
- Psycholinguists are doing more typology
  + E.g., Norcliffe et al. (2015)
- Both could learn more from each other
  + Experiments to test processing-based explanations for language universals
  + Experiments on a wider variety of languages to test hypothesized processing universals
  + These goals cannot be achieved in isolation!
UNIVERSAL PROCESSING MODELS?

O'Seaghdha et al. (2010), modifying Levelt et al. (1999)
OR UNIVERSAL LEARNING MODELS?

Universal learning algorithm

Chinese experience

English experience
TESTING UNIVERSALS

- Rationalist/deductive approach
  - Chomsky (1965): Learners need universals
  - So assume all is universal unless shown otherwise

- Empiricist/inductive approach
  - Greenberg (1963): Languages vary in limited ways
  - So study the typology of cross-linguistic variation

- What about psycholinguists?
  - Empiricists when studying individual languages
  - But rationalists when making universal claims
A typical psycholinguistic study

A behavioral database for masked form priming

James S. Adelman · Rebecca L. Johnson · Samantha F. McCormick · Meredith McKeague · Sachiko Kinoshita · Jeffrey S. Bowers · Jason R. Perry · Stephen J. Lupker · Kenneth I. Forster · Michael J. Cortese · Michele Scaltritti · Andrew J. Aschenbrenner · Jennifer H. Coane · Laurence White · Melvin J. Yap · Chris Davis · Jeesun Kim · Colin J. Davis

Published online: 1 February 2014
© Psychonomic Society, Inc. 2014

Abstract Reading involves a process of matching an orthographic input with stored representations in lexical memory. The masked priming paradigm has become a standard tool for investigating this process. Use of existing results from this paradigm can be limited by the precision of the data and the need for cross-experiment comparisons that lack normal experimental controls. Here, we present a single, large, high-precision, multicondition experiment to address these problems. Over 1,000 participants from 14 sites responded to 840 trials involving 28 different types of orthographically related primes (e.g., castfe–CASTLE) in a lexical decision task, as well as completing measures of spelling and vocabulary. The data were indeed highly sensitive to differences between conditions: After correction for multiple comparisons, prime type condition differences of 2.90 ms and above reached significance at the 5% level. These analyses will form a basis for the use of these data in quantitative model fitting and evaluation and for future exploration of these data that will inform and motivate new experiments.

Keywords Visual word recognition · Lexical decision · Orthographic priming · Megastudies

Introduction

The everyday activity of reading involves correctly selecting from one’s vocabulary the viewed word from among a variety of candidate words with some or many of the same features.
Piepenbrock, & Gulikers, 1995). All word targets had a frequency above zero in CELEX, HAL (Burgess, 1998), and SUBTLEX (Brysbaert & New, 2009) and had lexical decision accuracy of at least 80% in the English Lexicon Project (Balota et al., 2007). No target was listed in CELEX as having
Rowed to Recovery: The Use of Phonological and Orthographic Information in Reading Chinese and English

Gary Feng and Kevin Miller
University of Illinois at Urbana–Champaign

Hua Shu and Houcan Zhang
Beijing Normal University

To examine how readers of Chinese and English take advantage of orthographic and phonological features in reading, the authors investigated the effects of spelling errors on reading text in Chinese and English using the error disruption paradigm of M. Daneman and E. Reingold (1993). Skilled readers in China and the United States read passages in their native language that contained occasional spelling errors. Results showed that under some circumstances very early phonological activation can be identified in English, but no evidence for early phonology was found in Chinese. In both languages, homophone errors showed a benefit in measures of later processing, suggesting that phonology helps readers recover from the disruptive effects of errors. These results suggest that skilled readers take advantage of the special features of particular orthographies but that these orthographic effects may be most pronounced in the early stages of lexical access.
Two-language studies imply a factorial design

But cross-linguistic variables are confounded
  - English vs. Chinese orthography (ABC vs. 字)
  - English vs. Chinese homophones (few vs. many)

Some variables are also gradient
  - Number of homophones: few, more, many, ...

Additional variables keep getting discovered
  - English vs. Chinese relative clauses: N-RC vs. RC-N
  - ... but also many language-specific factors (Lin, 2017)
THE ROLE OF EMPIRICISM/INDUCTION

- Lexical researchers know these problems well
- Participants come into lab knowing their lexicon
  + Words cannot be experimentally manipulated
- Lexical variables are also gradient & confounded
  + Gradience: Frequency, length, ...
  + Confounding: High-frequency words are shorter
- Designing factorial lexical experiments is thus a “confounded nuisance” (Cutler, 1981)
- A recently popular solution: megastudies
  + Regression analyses on large, diverse data sets
THE MEGASTUDY MOVEMENT

- The English Lexicon Project (Balota et al. 2007)
  + 1000s of visual lexical decisions and naming responses for 1000s of written English words and non-words:
    [http://elexicon.wustl.edu/](http://elexicon.wustl.edu/)

- Similar projects in Dutch, French, Chinese, Malay (Balota et al. 2012; Keuleers & Balota 2015)
  + Also auditory lexical decision, priming, picture naming, ...
SO HOW ABOUT META-MEGASTUDIES...?

- Myers (2016)
  - Generalizing megastudy logic cross-linguistically
SAMPLING LANGUAGES FOR GRAMMAR

(e.g., Cysouw, 2005)

“Natural”

Universal principle

Type A

Type B

Borrowing

Descent

Learning from speech community hides natural patterns
SAMPLING LANGUAGES FOR PROCESSES

Language experience

“Natural”

Universal learning algorithm

Process A
Lang  Lang  Lang

Process B
Lang

Borrowing & descent can only bias up here

Processing ignores (?) community conventions
For psycholinguistics, language samples only need to deconfound processing-related variables + Borrowing and descent unlikely to affect processing independently of the typological features themselves

“Small” samples OK (e.g., Vittinghoff & McCulloch, 2007) + Only need around 10 x number of variables

Partial confounds OK (e.g., O’Brien, 2007) + Even with $r^2 = .9$ or higher
LANGUAGE SAMPLES IN TODAY’S TALKS

- Dressler et al.
- Chen & Myers
- White et al.
- Namboodiripad
- Culbertson & Braquet
- Dimitriadis et al.
- Sauppe et al.
- Blumenthal-Dramé & Kortmann
THE ROLE OF RATIONALISM/DEDUCTION

- If learners need universals, then why not study universals by looking at learners?

- Natural language learning
  - Children are biased (e.g., Lidz & Gagliardi, 2015)
  - Adults treat “natural” and “unnatural” patterns within their own language differently (e.g., Hayes et al., 2009)

- Artificial grammar learning
  - Simulates second language learning, allowing for controlled tests of learning biases (Culbertson, 2012)
LEARNING IN TODAY’S TALKS

- Natural first language acquisition by children
  + Dressler et al.

- Artificial language learning by adults
  + Culbertson & Braquet
  + Dimitriadis et al.
  + White et al.

- Presumed effects of first language on adults
  + Blumenthal-Dramé & Kortmann
  + Chen & Myers
  + Namboodiripad
  + Sauppe et al.
ARTIFICIAL GRAMMAR LEARNING

Language A experience

Universal learning algorithm

Language B experience

Learning succeeds

Learning fails
ARTIFICIAL LEARNING IN CONTEXT

English experience

Universal learning algorithm

Chinese experience

English-modulated learning algorithm

Lang. A

A>B

Lang. B

Chinese-modulated learning algorithm

A>B
HOW TO EXPAND THE TYPOLOGY?

- Make lots of friends...
  - Bates et al. (2003): 7 languages, 22 authors
- ... and/or emulate the typological grammarians
  - Thousands of independent scholars collect data on specific languages for their own purposes...
  - ... and then a relatively small number of typologists compile databases (e.g., Haspelmath et al., 2005)
  - Could psycholinguistics be done this way too...?
<table>
<thead>
<tr>
<th>Time</th>
<th>Authors</th>
<th>Domain</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:30-9:55</td>
<td>Dimitriadis &amp; al.</td>
<td>Morphology</td>
<td>Artificial grammar</td>
</tr>
<tr>
<td>11:30-11:55</td>
<td>Culbertson &amp; Braquet</td>
<td>Syntax</td>
<td>Artificial grammar</td>
</tr>
<tr>
<td>12:00-12:25</td>
<td>White &amp; al.</td>
<td>Phonology</td>
<td>Artificial grammar</td>
</tr>
<tr>
<td>12:30-13:00</td>
<td>Namboodiripad</td>
<td>Syntax</td>
<td>Acceptability</td>
</tr>
<tr>
<td>16:00-16:25</td>
<td>Chen &amp; Myers</td>
<td>Phonology</td>
<td>Acceptability</td>
</tr>
<tr>
<td>16:30-16:55</td>
<td>Dressler &amp; al.</td>
<td>Morphology</td>
<td>Corpus analysis</td>
</tr>
<tr>
<td>17:30-17:55</td>
<td>Blumenthal-Dramé &amp; Kortmann</td>
<td>Syntax</td>
<td>Comprehension</td>
</tr>
<tr>
<td>18:00-18:25</td>
<td>Sauppe &amp; al.</td>
<td>Syntax</td>
<td>Production</td>
</tr>
<tr>
<td>18:30-18:55</td>
<td>Discussion</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
REFERENCES (1/4)


Cutler (1981). Making up materials is a confounded nuisance, or: Will we be able to run any psycholinguistic experiments at all in 1990? Cognition, 10.


REFERENCES (3/4)


REFERENCES (4/4)


O’Seaghdha et al. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition, 115*.