Automated collection and analysis of phonological data

James Myers
National Chung Cheng University
Lngmyers@ccu.edu.tw

International Conference on Linguistic Evidence 2008
Tübingen, Germany Jan 31 - Feb 2

Acknowledgments

• **Research assistants** Tsung-Ying Chen, Chen-Tsung Yang, Yu-Guang Ko
• **Suggestions** from James S. Adelman, Harald Baayen, Wayne Cowart, John C. Pezzullo
• **Grants:** National Science Council (Taiwan) grants NSC 94-2411-H-194-018, NSC 95-2411-H-194-005, NSC 96-2411-H-194-002
• MiniJudgeJS, MiniJudgeJava, MiniCorpJS are co-copyrighted by National Chung Cheng University

Overview

• Automating traditional linguistic methods
• An application to Mandarin phonotactics
• How the tools work
• Plans for the future

Reform, not revolution

• Traditional linguistic methods have limits
  – Phonological patterns in lexicons need not be synchronically active (Ohala, 1986)
  – Informal acceptability judgments may be unreliable (Schütze, 1996; Cowart, 1997)
  – Without quantification, inferences are weak
  – Differences in methodological traditions hinder interdisciplinary collaboration
• Yet traditional methods should be built on; they don’t need to be fully replaced

In praise of tradition

• Two major methods in testing grammars
  – Corpus analysis (particularly in phonology)
  – Psycholinguistic experimentation (judgments)
• These methods deserve respect
  – Similar to common psycholinguistic methods
  – Data are often stable and replicable
  – Implicitly quantitative (as we’ll see)
  – Appropriate for testing long-term knowledge
• Easier than “full-fledged” psycholinguistics

Making reform easy

• Education
  – This conference, Cowart (1997), etc
• Automation
  – Experiments (e.g., WebExp: [www.webexp.info](http://www.webexp.info))
  – Corpus data (e.g., Praat: [www.fon.hum.uva.nl/praat](http://www.fon.hum.uva.nl/praat))
• Software tools that implement and extend traditional methods
MiniCorp and MiniJudge

- www.ccunix.ccu.edu.tw/~lngproc/MiniGram.htm
- **MiniCorp**
  - Software for creating, exploring, and analyzing (lexical phonological) corpora
- **MiniJudge**
  - Software for designing, running, and analyzing linguistic judgment experiments
- Free, open-source, and cross-platform
  - JavaScript (or Java) & R (www.R-project.org)

Example: Mandarin phonotactics

- Mandarin disallows syllables with identical first and last vowels (e.g., Duanmu, 2007)
  1. uai⁴ “outside” uei⁴ “for”
     iau⁴ “want” iou⁴ “again”
  2. *uau *oue *iei *iai
- But some speakers have exceptions
  3. iai² “cliff” (also 哇、唯、啱)

An analysis

- Identical vowels are blocked by the Obligatory Contour Principle (OCP)
- Why are the exceptions permitted?
  - Performance: Lexicons reflect not just grammar, but also processing and accidents
  - Or competence…? (e.g., Pater, to appear)
  - Exception-specific Optimality-Theoretic (OT) faith constraints (cf. exception diacritics)

Empirical challenges

- **Quantitative** questions about corpus data
  - Do the exceptions undermine the OCP?
  - Yet are the exceptions too rare to support the exception-specific constraint?
  - Even if both constraints are reliable, is their claimed ranking supported by the data?
- Corpus data as evidence for a proposed grammar, not learning of a grammar

MiniCorpJS

- **Prepare corpus**
  - Load raw corpus
  - Tag corpus items
  - Save tagged corpus
- **Explore corpus** (optional)
  - Classify corpus items (under construction)
  - Learn OT grammar (under construction)
  - Find corpus neighbors
  - Compute transitional probabilities (under construction)
- **Test hypotheses**
  - Download and install R (if you haven’t already)
  - Define OT hypothesis
  - Generate R command code
  - Paste R command code into R
  - R will summarize the analysis in an easy-to-read format

Tagging the corpus

- Items are represented in terms of constraint violations (cf. Golston, 1996)
  - Scrolling and sorting
  - Regular expression matching (i.*i)|(u.*u)
Testing the OT hypothesis

• Generate R command code, run it in R

Constraint test:

<table>
<thead>
<tr>
<th>Constraint</th>
<th>Weight</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>FaithEx</td>
<td>-8.8253</td>
<td>*</td>
</tr>
<tr>
<td>OCP</td>
<td>-7.2087</td>
<td>*</td>
</tr>
</tbody>
</table>

(* significant constraint)

Both constraint weights are significantly negative (violated less than obeyed)

But the claimed ranking is not supported

(No significant rankings)

How MiniCorp works

• Poisson regression
  Count-based loglinear modeling; cf. Hayes & Wilson
  Independent variables: Constraint violations
  Dependent variables: Counts of items violating different constraint combinations

<table>
<thead>
<tr>
<th>Counts</th>
<th>FaithEx</th>
<th>OCP</th>
</tr>
</thead>
<tbody>
<tr>
<td>13603</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

(a kluge to ensure convergence)

• Weights are regression coefficients

Testing constraint rankings

• Compare regression equations that do vs. do not assume identical weights
  (1) Different: Counts ~ w1FaithEx + w2OCP
  (2) Identical: Counts ~ w(FaithEx + OCP)
  [Algebra shows that (1) is an additive extension of (2), permitting a likelihood ratio test]
  • This logic generalizes to the strict OT ranking of any number of constraints
    A>>B>>C ⇒ A>>(B, C) (no “ganging up”)

Theoretical implications

• OCP is reliable in the Mandarin lexicon
• No reason to reject the performance-based interpretation of its exceptions
• Exception constraints are possible...
  – 7 or more exceptions would be enough to make the ranking statistically reliable
• But is this lexical pattern still active synchronically?
  – What do native speaker judgments show...?

MiniJudgeJS

• Design experiment
  Choose experimental factors
  Choose set of prototype items
  Choose total number of item sets
  Select prototype set (optional)
  Choose replacement prototypes for additional item sets (optional)
  Approve and save master list of test items

• Run experiment
  Choose number of subjects
  Write instructions for subjects
  Save electronic survey information
  Copy survey forms to print or email

• Analyze experiment
  Download and install R (if you haven’t already)
  Enter raw results
  Generate datafile
  Add neighborhood densities (optional, for judgments of word-sized items)
  Save data file
  Generate R code
  Print R command code into R
  R will summarize your findings in an easy-to-read format

Designing the experiment

• Linguists understand factorial designs
  – They just call them minimal pairs or sets
• So first choose basic set of (nonword) items, defined by one or two factors:
  [+FirstU +LastU]: tuou² ㄉㄨㄡ
  [+FirstU –LastU]: tuei² ㄉㄨㄟ
  [-FirstU +LastU]: tiou² ㄉㄧㄡ
  [-FirstU –LastU]: tiei² ㄉㄧㄟ

  -MiniJudge then guides the user to create new item sets to improve generalizability
Surveys
- Items (here, 16) put into random order
- Surveys (here, 20) emailed or printed (here, printed)
- Responses are quick yes/no judgments
  - Guessing allowed, but must judge all items, in order
  - Binary judgments can detect gradience (Cowart, 1997)

Results

How MiniJudge works
- Generalized linear mixed effect modeling (GLMM) (Agresti, 2002; Baayen, to appear)
  - GLMM is like logistic regression, familiar from VARBRUL (Paolillo, 2002) – PLUS:
  - Random variables (e.g., speakers & items) are included inside the same model
- MiniJudge includes item order as covariate
  - May reduce order-related nuisance effects
  - Option to factor out interactions with order (change in judgment contrasts over time)

A stricter analysis
- Judgments of nonlexical items are affected by analogy with “neighboring” lexical items (e.g., Bailey & Hahn, 2001)
- MiniCorp can count neighbors
  - Items differing in one segment (Luce, 1986)
  - MiniJudge then provides the option to add neighborhood density as a covariate
  - This factors out the effects of superficial analogy on acceptability judgments

Effect of neighbors
- Original analysis

|          | Estimate | Std. Error | z value | Pr(>|z|) |
|----------|----------|------------|---------|----------|
| Intercept| -1.35027 | 0.47769    | -2.827  | 0.00470  |
| Factor1  | -0.36059 | 0.14289    | -2.523  | 0.01162  |
| Factor2  | -0.12841 | 0.14170    | -0.885  | 0.38114  |
| Order    | 0.02018  | 0.03080    | 0.988   | 0.32723  |
| Factor1:Factor2 | -0.04917 | 0.14360 | -0.337 | 0.73624 |

- Analysis including neighborhood density

|          | Estimate | Std. Error | z value | Pr(>|z|) |
|----------|----------|------------|---------|----------|
| Intercept| -1.962791 | 0.573796   | -3.421  | 0.000623 |
| Factor1  | -0.199942 | 0.163569   | -1.222  | 0.221568 |
| Factor2  | -0.317748 | 0.170956   | -1.865  | 0.062152 |
| Order    | 0.039377  | 0.031495   | 1.257   | 0.208866 |
| Neighbors| 0.012890  | 0.006371   | 2.038   | 0.041561 |
| Factor1:Factor2 | -0.111937 | 0.220739 | -0.496 | 0.620656 |

Theoretical implications
- A lexical pattern need not be encoded synchronically in terms of “grammar”
- Other lexical patterns in Mandarin syllables fail to affect judgments at all
  - Not even via neighbors (Myers, 2008)
- Caveat: Small studies aren’t conclusive
  - Grammatical constraints can affect judgments even when neighborhood density is factored out (e.g., Frisch & Zawaydeh, 2001)
Other applications of the tools

• Quickly resolving judgment ambiguities in morphology & syntax (e.g., Myers 2007)
• Studying the interaction between grammar and processing (e.g., Ko, 2007)
• Quick piloting for large-scale experiments (e.g., Lawrence, 2007)
• Surveying an entire linguistic system
  – Quick, small-scale studies of each pattern
  – Studies can be run in parallel by assistants without much prior training

Plans for the near future

• New options
  – Rule ordering tests (cf. Sankoff & Rousseau, 1989)
  – Corpus exploration (cf. Uffmann, 2006)
  – Non-binary judgments (cf. Featherston, 2005)
  – Tools to help generate nonword items
• Improved statistics
  – Built-in analyses (though keep R code writer)
  – Exact statistics to avoid kluges (Myers et al., 2007)
• Improved interface
  – All Java, native language help, etc

Conclusions

• Traditional methods are a good start
• To build on them, linguists need help
• Automation is one way to do this
• MiniCorp & MiniJudge are already usable
• They have helped test theoretically interesting claims of various sorts
• Yet they are in need of improvement
• Collaborators and competitors are both most welcome!

Automated collection and analysis of phonological data

James Myers
National Chung Cheng University
Lngmyers@ccu.edu.tw

International Conference on Linguistic Evidence 2008
Tübingen, Germany Jan 31 - Feb 2

References (1/4)


References (2/4)

References (3/4)


References (4/4)


