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
  - www.ccunix.ccu.edu.tw/~lngproc/IWGE.htm
- Automation
  - Experiments (e.g., WebExp: www.webexp.info)
  - Corpus data (e.g., Praat: 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)

7

Example: Mandarin phonotactics

- Mandarin disallows syllables with identical first and last vowels (e.g., Duanmu, 2007)

(1) uai^4 "outside" uei^4 "for"
iau^4 "want" iou^4 "again"
(2) *uau *uou *iei *iai

- But some speakers have exceptions

(3) iai^2 "cliff" （崖） (also 睚、啀、[illegible])

8

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)

Faith$_{Exceptions}$ >> OCP

9

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
  - Cf. Tesar & Smolensky (1998), Boersma & Hayes (2001), Hayes & Wilson (to appear)

10

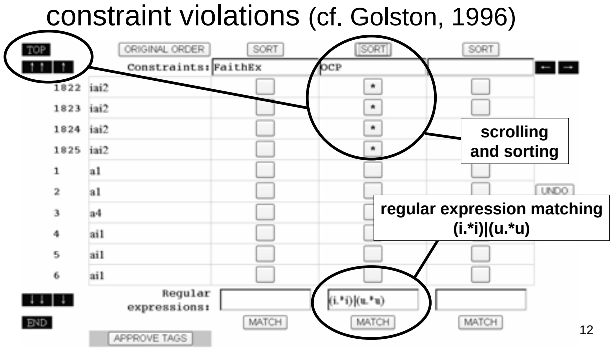
MiniCorpJS

- 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

11

Tagging the corpus

- Items are represented in terms of constraint violations (cf. Golston, 1996)



12

Testing the OT hypothesis

- Generate R command code, run it in R

```
Constraint test:

 Constraints    Weights p
     FaithEx    -8.8252 0        *
         OCP    -7.9087 0        *

 (* significant constraint)

Ranking test:

 Constraints    p
     FaithEx    0.2491

 (No significant rankings)
```

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

But the claimed ranking is not supported

13

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

Counts	$Faith_{Ex}$	OCP
13603	0	0
4	0	1
1	1	0
1	1	1

(a kluge to ensure convergence)

- Weights are regression coefficients

14

Testing constraint rankings

- Compare regression equations that do vs. do not assume identical weights
  (1) Different: Counts ~ w_1Faith$_{Ex}$ + w_2OCP
  (2) Identical: Counts ~ w(Faith$_{Ex}$ + 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 $\Rightarrow$ A>>{B, C} (no "ganging up")

15

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...?

16

MiniJudgeJS

- Design experiment
  - Choose experimental factors
  - Choose set of prototype items
  - Choose total number of item sets
  - Segment prototype set (optional)
  - Choose replacement segments for additional item sets (optional)
  - Approve and save master list of test items
- Run experiment
  - Choose number of speakers
  - Write instructions for speakers
  - Save schematic survey information
  - Copy survey forms to print or email
- Analyze experiment
  - Download and install R (if you haven't already)
  - Enter raw results
  - Generate data file
  - Add neighborhood densities (optional, for judgments of word-sized items)
  - Save data file
  - Generate R code
  - Paste R command code into R
  - R will summarize your findings in an easy-to-read format

17

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]: tuou[2] ㄉㄨㄡˊ
  [+FirstU –LastU]: tuei[2] ㄉㄨㄟˊ
  [-FirstU +LastU]: tiou[2] ㄉㄧㄡˊ
  [-FirstU -LastU]: tiei[2] ㄉㄧㄟˊ
- MiniJudge then guides the user to create new item sets to improve generalizability

18

Surveys

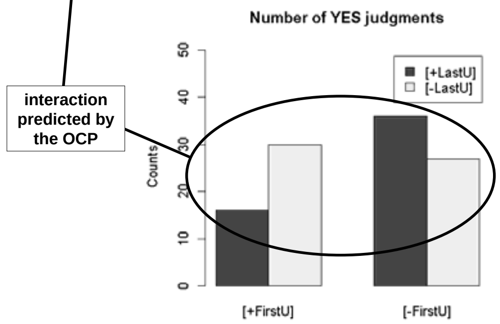
- Items (here, 16) put into random order
- Surveys (here, 20) emailed or printed (here, printed)
- Responses are quick yes/no judgments
  - Guessing allowed, but must judge all items, in order
  - Binary judgments can detect gradience (Cowart, 1997)

最後謝謝你的幫忙與合作!
##01

0	(1)	ㄅㄧㄞˊ	piai^2
1	(2)	ㄉㄨㄟˊ	tuei^2
0	(3)	ㄋㄨㄠˊ	nuau^2
1	(4)	ㄅㄧㄡˊ	piou^2
1×	(5)	ㄉㄨㄠˊ	tuau^2
1	(6)	ㄅㄧㄠˊ	piau^2
0×	(7)	ㄋㄨㄟˊ	nuei^2
1×	(8)	ㄉㄧㄞˊ	tiai^2
1	(9)	ㄉㄧㄠˊ	tiau^2
1×	(10)	ㄉㄧㄟˊ	tiei^2
1	(11)	ㄉㄨㄞˊ	tuai^2
0	(12)	ㄉㄨㄡˊ	tuou^2
1	(13)	ㄅㄧㄟˊ	piei^2
1	(14)	ㄋㄨㄞˊ	nuai^2
0	(15)	ㄋㄨㄡˊ	nuou^2
1	(16)	ㄉㄧㄡˊ	tiou^2

Results

```
The factor FirstU had a significant negative effect.
The interaction between FirstU and LastU had a significant negative effect.
There were no other significant effects.
```



20

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)

21

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

22

Effect of neighbors

- Original analysis

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-1.35027	0.47769	-2.827	0.00470	**
Factor1	-0.36054	0.14289	-2.523	0.01163	*
Factor2	-0.12541	0.14170	-0.885	0.37614	
Order	0.03018	0.03080	0.980	0.32723	
Factor1:Factor2	-0.46917	0.14360	-3.267	0.00109	**

- Analysis including neighborhood density

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-1.962791	0.573706	-3.421	0.000623	***
Factor1	-0.199942	0.163569	-1.222	0.221568	
Factor2	-0.317748	0.170356	-1.865	0.062153	.
Order	0.039577	0.031496	1.257	0.208906	
NeighDens	0.012984	0.006371	2.038	0.041561	*
Factor1:Factor2	-0.111937	0.225783	-0.496	0.620056	

23

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)

24

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

25

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

26

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!**

27

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

28

References (1/4)

Agresti, A. (2002). *Categorical data analysis*, 2nd ed. Wiley.

Baayen, R. H. (To appear). *Analyzing linguistic data: A practical introduction to statistics*. Cambridge University Press.

Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *JML, 44*, 569-591.

Boersma, P., & Hayes, B. (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry, 32*, 45-86.

Cowart, W. (1997). *Experimental syntax: Applying objective methods to sentence judgments*. London: Sage Publications.

Duanmu, S. (2007). *The phonology of Standard Chinese*, 2nd ed. Oxford University Press.

Featherston, S. (2005). Magnitude estimation and what it can do for your syntax: Some wh-constraints in German. *Lingua, 115* (11),1525-1550.

29

References (2/4)

Frisch, S. A., & Zawaydeh, B. A. (2001). The psychological reality of OCP-Place in Arabic. *Language, 77* (1), 91-106.

Golston, C. (1996). Direct Optimality Theory: Representation as pure markedness. *Language 72* (4), 713-748.

Hayes, B., & Wilson, C. (To appear). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*.

Ko, Yu-Guang. (2007). *Grammaticality and parsibility in Mandarin syntactic judgment experiments*. National Chung Cheng University MA thesis.

Lawrence, D. (2007, April). Using MiniJudge to test sentence acceptability. Talk presented at Thinking Matters Conference, University of Southern Maine.

Luce, P. A. (1986). *Neighborhoods of words in the mental lexicon*. Doctoral dissertation, Indian University, Bloomington, IN.

30

References (3/4)

Myers, J. (2007). MiniJudge: Software for small-scale experimental syntax. *International Journal of Computational Linguistics and Chinese Language Processing, 12* (2), 175-194.

Myers, J. (2008, July). From problem set to corpus analysis: MiniCorp and the Mandarin syllabary. Talk to be presented at the 18th International Congress of Linguists, Seoul, Korea.

Myers, J., Huang, S.-F., & Tsay, J. (2007). Exact conditional inference for two-way randomized Bernoulli experiments. *Journal of Statistical Software, 21*, Code Snippet 1, 2007-09-02.

Ohala, J. J. (1986). Consumer's guide to evidence in phonology. *Phonology Yearbook, 3*, 3-26.

Paolillo, J. C. (2002). *Analyzing linguistic variation: Statistical models and methods*. Stanford, CA: CSLI Publications.

References (4/4)

Pater, J. (to appear). The locus of exceptionality: Morpheme-specific phonology as constraint indexation. In S. Parker (Ed.), *Phonological argumentation*. Equinox.

Sankoff, D., & Rousseau, P. (1989). Statistical evidence for rule ordering. *Language Variation and Change, 1*, 1-18.

Schütze, C. T. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. University of Chicago Press.

Tesar, B., & Smolensky, P. (1998). The learnability of Optimality Theory. *Linguistic Inquiry, 29*, 229-268.

Uffmann, C. (2006). Epenthetic vowel quality in loanwords: Empirical and formal issues. *Lingua, 116*, 1079-1111.