Comprehending Your Neighbour’s English Errors: A Mixed Models Analysis

Research Methodology and Statistics 2014
Kristy James
Speaking in a lingua franca

› English as current choice
› Different norms from native speech
› CLI from L1 and limited proficiency can result in errors

Speaker and Listener effects:
Present the story accurately in comprehensible English (create stimulus properties), so listener can decipher meaning (dependent on listener factors)
Three language combinations

<table>
<thead>
<tr>
<th>Family</th>
<th>Listeners</th>
<th>Speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Romance</td>
<td>Portuguese</td>
<td>Spanish</td>
</tr>
<tr>
<td>Slavic</td>
<td>Slovenian</td>
<td>Croatian</td>
</tr>
<tr>
<td>Germanic</td>
<td>Danish</td>
<td>Swedish</td>
</tr>
</tbody>
</table>

vs. English produced by these speakers, eg Spanish-accented English
Contents

› Introduce Experiment and Motivation
› Options for statistical approaches
› Mixed Models – a little theory
› A Generated Data analysis
› Conclusion
Experimental Overview

- Model a storytelling situation – Speaker participants tell one story each in L2 English and L1.
- Listener comprehension tested with multiple-choice questions via an online platform.
- (Conclude whether comprehension is better in English or the related language.)
- **Explain which errors in English most negatively affect comprehension – pronunciation, vocabulary or grammar.**
Motivation: Comprehending a related language

› Trending theme – cross-border communication
› Investigations into lexical/phonological distance
› Effects of prior exposure, schooling etc, as well as individual factors
› Our choice of normal speech (not foreigner-oriented)

Listener factors: Speaker speaks easily in native language (SP), listener factors determine comprehension (LF).
Methodology

> Elicit retellings based on two silent short films from 20-30 participants per language combination
> Select core speakers that cover canonical topics – approx. 20 stimuli selected
> Segment recordings into audio fragments that are relevant to a particular topic (12 ‘questions’ generated)
> Assign participants (listeners) to a random speaker (approx. 5 per stimuli), expose to audio fragment once and reveal comprehension question
Eliciting Retellings
Measuring Comprehension

› 12 multiple-choice questions, 6 in English, 6 in related language
› Crossed-design – Film A and Film B
Variation at:

› Speaker Level:
  • S01
  • S06
  • S09
    • Among others no. pronunciation errors, no. grammar errors, no. vocabulary errors

› Listener Level:
  • English proficiency, exposure to English, exposure to a Spanish accent
Explaining the variance

Stimulus (eg L1 Spanish) → Comprehension score → MANY Listeners (Portuguese)

Grammar errors → Pronunciation errors → Comprehension score

Vocabulary errors
Nested Data

(Level 3: Speaker L1)

Level 2: Stimulus (Speaker) x 20

Speaker A
Speaker B
Speaker n

Level 1: Listener x approx. 5

Listener 1
Listener 2
Listener 3
Listener m

(Level 0: Question) x 6

Qns 1-6
Qns 1-6
Qns 1-6
Qns 1-6
# Possible statistical methods

<table>
<thead>
<tr>
<th></th>
<th>RM ANOVA</th>
<th>Linear/Multiple Regression</th>
<th>Logistic regression</th>
<th>Mixed Models</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>When used in</strong></td>
<td><strong>For comparing overall success of ELF vs ReLa</strong></td>
<td><strong>When one listener hears each speaker (numerical response)</strong></td>
<td><strong>When one listener hears each speaker (categorical response)</strong></td>
<td><strong>Multiple listeners hear each speaker, can handle either numerical/categorical response</strong></td>
</tr>
<tr>
<td><strong>this experimental</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>context</strong>:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Assumptions</strong>:</td>
<td><strong>Assumptions of independence; homogeneity of variance, balanced design</strong></td>
<td><strong>Assumes homogeneity of regression slopes (eg vocabulary errors may aid understanding); Error rating per stimulus</strong></td>
<td><strong>See left; ideally requires annotation of errors per question</strong></td>
<td><strong>No assumption of independence</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>Robust against missing data</strong></td>
</tr>
</tbody>
</table>
Linear Regression

\[ y_i = a + bx_i + e_i \]

- a - intercept
- b - slope
- e - error
- Numeric (or categorical) independent variables, numerical response variable
- Numerical response: eg how many correct answers
- Intercept and slope represent average over all data points
- This experiment: with repetition of stimuli (these vary randomly) and potentially unbalanced numbers, therefore require taking averages – throwing away data
Logistic Regression

\[ P(Y) = \frac{1}{1 + e^{-\left(b_0 + b_1 x_1\right)}} \]

- Numeric (or categorical) independent variable, one categorical response variable
- Eg for each question, P that answer is correct
Advantages of MM

Level 1: \( \text{speech.rate}_i = a_{j[i]} + b \times \text{context}_i + e_i \)

Level 2: \( a_j = \mu_{\text{subject}} + e_j \)

› Can deal with nested variables (stimuli are repeated)
› No need to average data – retain information
› Random slopes could account for account for fatigue/becoming accustomed to accent
› Represents variance as coefficients
› Has equations to handle both categorical and numerical response variables (allows both linear and logistic analyses of data)
What are random and fixed effects?

- Random effects:
  - Levels randomly sampled from a larger population
  - Varies over time/between samples
  - May expect a different slope/intercept for each instance
  - Useful for analysing items tested (avoiding language as a fixed effect fallacy, here difficulty of question), or capturing subject variation

- Fixed effects:
  - Fixed number of levels
  - Expect the variable to contribute equally regardless of context
  - Generally do not vary over time
## Random and Fixed Effects

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Fixed Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker</td>
<td>English exposure</td>
</tr>
<tr>
<td>Listener</td>
<td>Errors in pronunciation, grammar, vocabulary</td>
</tr>
<tr>
<td>Stimulus (movie vs question)</td>
<td>Speaker Age</td>
</tr>
</tbody>
</table>
> summary(lm.mod1)

Call:
  lm(formula = tot ~ lst.enexposure + lst.enlevel + spk.age + spk.enyears +
      spk.enfreq + spk.enpron + spk.english + spk.envocab + movie.id,
      data = d5)  # d5 is numerical response variable (qns correct)

Residuals:
     Min       1Q  Median       3Q      Max
-3.3488 -1.1858  0.1416  1.2433  3.4159

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)      0.008783   3.573124   0.002  0.99804
lst.enexposure  0.456863   0.189410   2.412  0.01789 *
lst.enlevel     0.293529   0.175791   1.670  0.09844 .
spk.age         0.009743   0.119449   0.082  0.93518
spk.enyears    -0.104554   0.092604  -1.129  0.26188
spk.enfreq     0.641164   0.248287   2.582  0.01143 *
spk.enpron     0.504496   0.222272   2.270  0.02561 *
spk.english    -0.366587   0.527467  -0.695  0.48885
spk.envocab   -0.156958   0.204041  -0.769  0.44376
movie.idm2    -1.279662   0.456806  -2.801  0.00623 **

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.82 on 90 degrees of freedom
Multiple R-squared:  0.2856,    Adjusted R-squared:  0.2142
F-statistic: 3.998 on 9 and 90 DF,  p-value: 0.0002525

>
> summary(lmer.mod1)
Linear mixed model fit by REML ['merModLmerTest']
Formula: tot ~ (1 | spk.id) + (1 | movie.id) + lst.enexposure + lst.enlevel + spk.enyears + spk.enfreq + spk.enpron + spk.engrammar + spk.envocab
   Data: d5
REML criterion at convergence: 375.7
Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.99752 -0.63924  0.03119  0.53328  2.13735
Random effects:
 Groups     Name        Variance Std.Dev.
    spk.id (Intercept) 2.5916   1.6098
movie.id (Intercept) 0.2684   0.5181
     Residual             1.6514   1.2850
Number of obs: 100, groups: spk.id, 20; movie.id, 2
Fixed effects:
                  Estimate Std. Error df t value Pr(>|t|)
(Intercept)       0.46277    7.35172 12.85000  0.063 0.950778
lst.enexposure   0.32004    0.13944  80.90000  2.295 0.024311 *
lst.enlevel      0.48521    0.13343  81.03000  3.636 0.000485 ***
spk.age          -0.01497    0.24949 12.12000 -0.060 0.953127
spk.enyears     -0.19302    0.17544 11.85000 -1.100 0.293081
spk.enfreq      0.63005    0.51935 11.95000  1.213 0.248506
spk.enpron      -0.42779    1.10097 12.00000 -0.389 0.704408
spk.engrammar   -0.42779    1.10097 12.00000 -0.389 0.704408
spk.envocab     -0.19506    0.42619 12.08000 -0.458 0.655304
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
     (Intr) lst.nx lst.nl spk.ag spk.ny spk.nf spk.np spk.ng
lst.enexpser  -0.012
 lst.enlevel   0.002  0.054
Linear mixed model fit by REML ['merModLmerTest']
Formula: tot ~ (1 | spk.id) + (1 | movie.id) + lst.enexposure + lst.enlevel + spk.age + spk.enyears + spk.enfreq + spk.enpron + spk.gram + spk.envocab

Data: d5
REML criterion at convergence: 375.6679
Random effects:
  Groups   Name       Std.Dev.
           spk.id    (Intercept) 1.6098
           movie.id (Intercept) 0.5181
           Residual             1.2850
Number of obs: 100, groups: spk.id, 20; movie.id, 2
Fixed Effects:
(Intercept) lst.enexposure lst.enlevel spk.age spk.enyears spk.enfreq
spk.enpron spk.gram spk.envocab
  0.46277 0.32004 0.48521 -0.01497 -0.19302 0.63005
  0.55820 -0.42779 -0.19506
> lmer.mod3
Linear mixed model fit by REML ['merModLmerTest']
Formula: \( \text{tot} - \text{mean}(\text{d5$tot}) \sim (1 \mid \text{spk.id}) + (1 \mid \text{movie.id}) + \text{lst.enexposure} + \text{lst.enlevel} + \text{spk.age} + \text{spk.enyears} + \text{spk.enfreq} + \text{spk.enpron} + \text{spk.grammar} + \text{spk.envocab} \)
Data: d5
REML criterion at convergence: 375.6679
Random effects:
<table>
<thead>
<tr>
<th>Groups</th>
<th>Name</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>spk.id</td>
<td>(Intercept)</td>
<td>1.6098</td>
</tr>
<tr>
<td>movie.id</td>
<td>(Intercept)</td>
<td>0.5181</td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>1.2850</td>
</tr>
</tbody>
</table>
Number of obs: 100, groups: spk.id, 20; movie.id, 2
Fixed Effects:
<table>
<thead>
<tr>
<th>(Intercept)</th>
<th>lst.enexposure</th>
<th>lst.enlevel</th>
<th>spk.age</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2.72724</td>
<td>0.32004</td>
<td>0.48521</td>
<td>-0.01497</td>
</tr>
<tr>
<td>0.19302</td>
<td>0.63005</td>
<td>0.55820</td>
<td>-0.42779</td>
</tr>
<tr>
<td>-0.19506</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
>
> anova(lmer.mod1, lmer.mod2)
refitting model(s) with ML (instead of REML)
Data: d5
Models:
object: tot ~ (1 | spk.id) + (1 | movie.id) + lst.enexposure +
lst.enlevel +
object: spk.age + spk.enyears + spk.enfreq + spk.enpron +
spk.engrammar +
object: spk.envocab
..1: tot ~ (1 | spk.id) + (1 | movie.id) + lst.enexposure *
spk.envocab +
..1: lst.enlevel + spk.age + spk.enyears + spk.enfreq +
spk.enpron *
..1: lst.crexposure + spk.engrammar

<table>
<thead>
<tr>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>object</td>
<td>12</td>
<td>391.87</td>
<td>423.13</td>
<td>-183.94</td>
<td>367.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>..1</td>
<td>13</td>
<td>393.80</td>
<td>427.67</td>
<td>-183.90</td>
<td>367.80</td>
<td>0.0726</td>
<td>1</td>
</tr>
</tbody>
</table>

> anova(lmer.mod1, lmer.mod3) #equal AIC
Explaining the variance

- Random effects – sd
- Fixed effects – visible from model
- Still to consider: centring data
Next steps

› Data collection: release and promote survey to participants in Portugal, Slovenia and Denmark.

› Possible future study: Include L1 English speakers as listeners and see if variance is different

› Model for predicting comprehension? Areas to target in language teaching?
Questions?

› Missing data – ecological validity or to nullify subject?
› Averaging ‘listeners’ to still use linear regression?
References

