Using linear mixed-effects models in psycholinguistic research

Shravan Vasishth, vasishth@acm.org

Universität Potsdam

28th March 2006, Groningen

Table of contents

The main goals today

- This is a non-technical and intuitive introduction to the use of linear mixed-effects models in psycholinguistic research.
- The focus is on hypothesis testing, not prediction.
- I will provide a real-life example from my own research to show how mixed-effects models can help us to build better and more informative statistical models.
- Towards the end I will also give some examples of how the R code relates to the content discussed in this document.

These slides can be downloaded from:
`http://www.ling.uni-potsdam.de/~vasishth/SFLS.html`.

The broad research question

- A current focus in sentence processing research is the issue of locality and anti-locality during online reading.
- The key claim in the literature is that argument-head distance is a (or perhaps the) major determinant of processing difficulty. This is the foundational idea behind several theories (such as Gibson's and Hawkins').
- But there are two issues:
  - There is significant evidence against locality (Konieczny 2000), (Vasishth 2003), (Vasishth & Lewis 2006), (Vasishth & Scheepers, in preparation)).
  - Current theories have no room for interference effects, except as an afterthought. The alternative we are exploring is that locality, anti-locality, and interference emerge from more general constraints on the human cognitive system.
- We'll explore one aspect of this issue today with some recent results (see Suckow et al. 2005, 2006).

Interference

The interference effect is a well-researched phenomenon in sentence processing. Gordon and colleagues, Van Dyke and Lewis, and others has shown that retrieving an element like an NP at a verb is harder when similar elements are available in working memory. Consider the following sentence:

(1) Der Anwalt, den der Zeuge/Säbel, den der Spion
The lawyer who the witness/sword that the spy
betrachtete, schnitt, überzeugte den Richter
looked-at cut convinced the judge

Interference predicts greater processing difficulty at the verb *betrachtete* when all three preceding NPs are human.

The missing VP effect

Previous research by Gibson and Thomas (1999) has shown for English that omitting the second verb in center embeddings results in *improved* grammaticality judgements.
However, only offline judgements or auto-paced reading have been brought to bear in the empirical issues. In (Suckow et al. 2006) we tried to replicate this missing VP effect using eyetracking.

(2) Der Anwalt, den der Zeuge/Säbel, den der Spion
The lawyer who the witness/sword that the spy
betrachtete, (schnitt,) überzeugte den Richter
looked-at cut convinced the judge

Experiment design and method

- There are two factors of interest: Similarity and Grammaticality.
- A 2 × 2 within-subjects (n=51) experiment was conducted, with 3 items per condition. Four counterbalanced lists were prepared and the 12 critical items in each list were interspersed with approximately 50 distractor sentences and the lists were pseudo-randomized. Subjects were randomly assigned to lists.
- The dependent measure of interest was total reading time at NP2, NP3, the first verb (V3), and the last (V1). Call these the critical regions.

```
[NP1+ [NP2+ [NP3+ VP3] (VP2)] VP1]
[NP1+ [NP2- [NP3+ VP3] (VP2)] VP1]
```


Hypothesis 1: Interference

Suckow, Vasishth, Lewis, and Smith (2006) hypothesized that interference could have a much more extensive effect.

```
[NP1+ [NP2+ [NP3+ VP3] VP2] VP1]
[NP1+ [NP2- [NP3+ VP3] VP2] VP1]
```

(a) ENCODING INTERFERENCE: Encoding an NP should be more difficult when a similar NP has recently been encoded, `NP2` and `NP3` should be harder to process when `NP2` is human (after factoring out frequency and length differences)

(b) RETRIEVAL INTERFERENCE: Interference predicts greater processing difficulty at V3 if all `NP`s previously seen are animate, since the verb is looking for a human argument and there are three candidates.

Hypothesis 2: The missing VP effect

- If VP2 is forgotten, then no disription (no increase in reading time) should occur at VP1 in the missing VP conditions (c, d) compared to the grammatical conditions.
- If VP2 is **not** forgotten, a disruption is expected at or just after the VP1 in the missing VP conditions (c, d).

a. `[NP1+ [NP2+ [NP3+ VP3] VP2] VP1]`

b. `[NP1+ [NP2- [NP3+ VP3] VP2] VP1]`

c. `[NP1+ [NP2+ [NP3+ VP3] - ] VP1]`

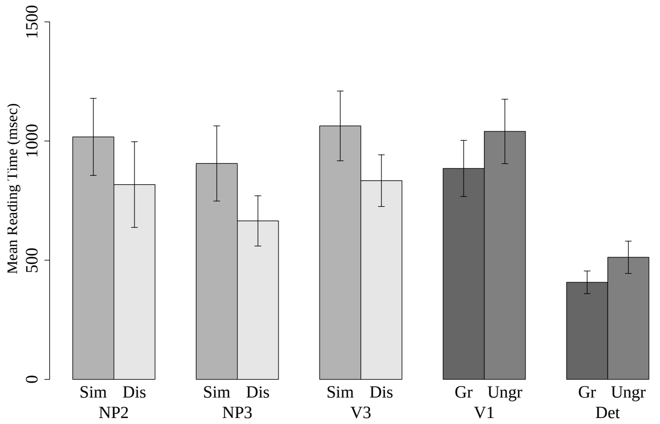
d. `[NP1+ [NP2- [NP3+ VP3] - ] VP1]`

Experiment results

In this talk the analysis results are almost beside the point, but I'll tell you anyway:

- Evidence for encoding and retrieval interference was found
- Surprisingly, German native speakers immediately detect the missing VP. This expresses itself as longer RTs at the final verb in the missing-VP conditions. This result goes against the simple "memory overload" explanation of Gibson and Thomas (1999), whereby the prediction for the second verb is forgotten–Germans don't forget.
- Interesting side note: in parallel English reading studies conducted at Ann Arbor, Michigan, we found that English speakers *do* seem to forget the middle verb's prediction. See Suckow et al. (CUNY 2006 poster) for details.

Critical regions



The data analysis problem

In this talk I will focus only on (a) the Similarity effect at NP2 and the region preceding NP3, and (b) the Grammaticality effect at V1. There are two complications in the data analysis:

- **Frequency** and **length** of NP2 differ in the manipulation: higher frequency NPs and shorter NPs will be processed faster and this could confound the results.
- Eyetracking data are sometimes unbalanced: sometimes subjects do not look at particular words or not look long enough, so we will probably not have exactly identical numbers of repeated measures for each subject.

Our basic statistical model (without interactions) will look something like this:

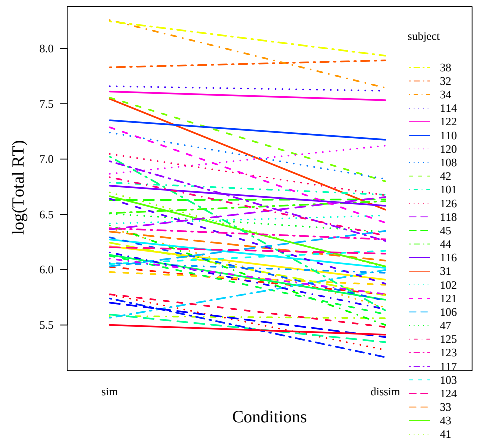
$$RT = \text{baseline RT} + \text{Sim} + \text{Gram} + \text{Freq} + \text{Len} + \text{residual} \quad (1)$$

Fixed and random effects in the model

- In general, effects can be “fixed” or “random”
- An example of a random effect is subjects: we are taking a random sample (well, in theory anyway) from a population.
- The factor(s) being manipulated (e.g. similarity) is a fixed factor, since we fixed it at +/- similar when we designed the experiment.
- However, fixed factors like similarity can also be treated as random factors. I will just explain what this amounts to.

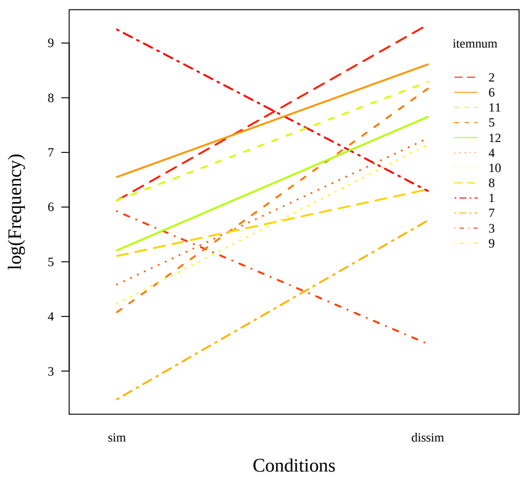
A mixed-effects model is one that has both fixed and random effects.

Subject-similarity interaction at NP2



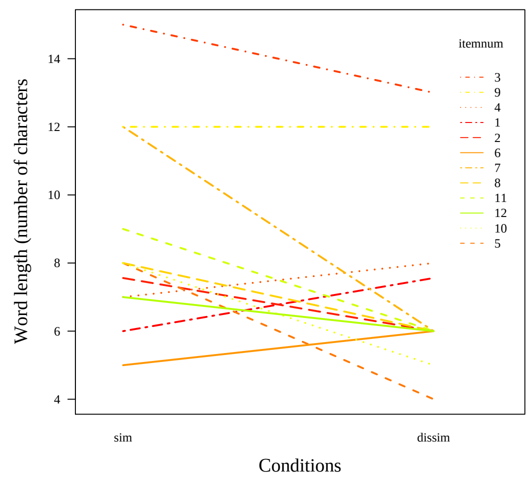
Most–but not all–subjects show a faster RT in the dissimilar condition.

Similarity-word frequency interaction at NP2



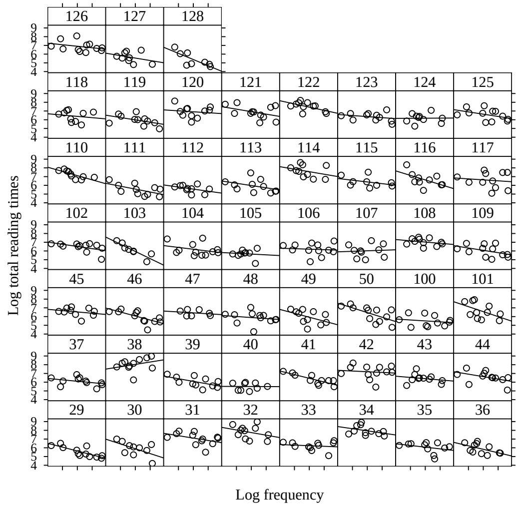
Dissimilar NPs are more frequent in all but two cases.

Similarity-word length interaction at NP2



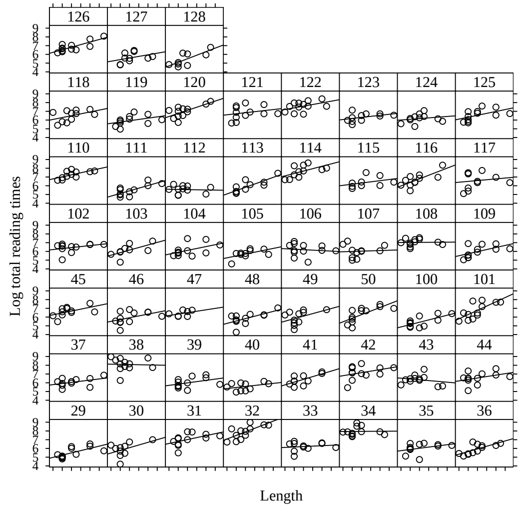
Dissimilar NPs happen to be sometimes shorter.

Log frequency-RT interaction at NP2



Unsurprisingly, high-frequency words have faster RT.

Length-RT interaction at NP2



Longer words have slower RT (also no surprise).

Starting small

The interaction plots confirm the importance of taking frequency and length of NP2 as explanatory covariates. Let's start by fitting a simple model:
Fixed factors:

- Similarity
- Word length
- Word frequency

Random factor:

- Subjects

$$y_{ij} = \mu + b_i + \beta_1 \mathsf{Sim}_{ij} + \beta_2 \mathsf{Len}_i + \beta_3 \mathsf{Freq}_i + \epsilon_{ij} \tag{2}$$

b_i is a separate coefficient (intercept) for each subject.

Model 1 estimates

Fixed effects coefficients (extracted from R output):

```
(Intercept)   Sim    Len    Freq
  6.09       -0.16   0.07  -0.03
```

Random effects (standard deviations):

```
Random effects:
 Formula: ~1 | subject
        (Intercept) Residual
StdDev:    0.640562 0.572838
```

$$y_{ij} = \mu + b_i + \beta_1 \mathsf{Sim}_{ij} + \beta_2 \mathsf{Len}_i + \beta_3 \mathsf{Freq}_i + \epsilon_{ij} \tag{3}$$

$$y_{ij} = 6.09 + b_i + (-0.16) \times \mathsf{Sim}_{ij} + 0.07 \times \mathsf{Len}_i + (-0.03) \times \mathsf{Freq}_i + \epsilon_{ij} \tag{4}$$

$$\mathsf{Var}(b_i) = 0.64^2,\ \mathsf{Var}(\epsilon_{ij}) = 0.57^2$$

Model 1 estimates

```
R> intervals(fm0.NP2)
Approximate 95% confidence intervals

 Fixed effects:
                      lower        est.        upper
(Intercept)       5.67040041  6.08701853  6.503636645
similaritydissim -0.26841254 -0.15744779 -0.046483036
len               0.04807326  0.07251362  0.096953981
lf               -0.07036548 -0.03263034  0.005104804
attr(,"label")
[1] "Fixed effects:"

 Random Effects:
  Level: subject
                    lower     est.     upper
sd((Intercept)) 0.5186496 0.640562 0.7911309

 Within-group standard error:
    lower      est.     upper
0.5382067 0.5728380 0.6096978
```


Model 1 estimates

- The three fixed effects each have a single (mean) coefficient across all subjects. A standard error is estimated for each coefficient as well.
- The mean coefficent and its SE together allow us to ask if there is a significant effect of each fixed factor. This is just asking, for each factor, whether we can reject the null hypothesis that β is zero ($\mathcal{H}_0 : \beta_k = 0$). We can just compute the t-statistic for each coefficient ($t = \frac{\beta_k - 0}{SE}$).

```
             Value  Std.Error  DF   t-value p-value
(Intercept) 6.088    0.212     494  28.706  0
sim        -0.157    0.056     494  -2.788  0.006
len         0.073    0.0124    494   5.829  0
lf         -0.033    0.0192    494  -1.699  0.09
```


Model 1 analysis of variance

We can also compute the ANOVA based on the fitted model:

	numDF	denDF	F-value	p-value
sim	1	494	55.52	4.165557e−13
len	1	494	76.50	0.000000e+00
lf	1	494	2.89	8.995200e−02

A more complex model

Fixed factors:

- Similarity
- Word length
- Word frequency

Random factors:

- Subjects
- Similarity is nested as a random factor inside Subject (separate term for each subject's exposure to similar and dissimilar conditions)

$$y_{ij} = \mu + b_i + \underline{b_{ij}} + \beta_1 \text{Sim}_{ij} + \beta_2 \text{Len}_i + \beta_3 \text{Freq}_i + \epsilon_{ij} \qquad (5)$$

Model 2 coefficients

The key change is in the random effects:

```
Random effects:
 Formula: ~1 | subject
        (Intercept)
StdDev:    0.640562

 Formula: ~1 | similarity %in% subject
         (Intercept) Residual
StdDev: 5.526901e-05 0.572838
```

$$y_{ij} = \mu + b_i + \underline{b_{ij}} + \beta_1 \mathsf{Sim}_{ij} + \beta_2 \mathsf{Len}_i + \beta_3 \mathsf{Freq}_i + \epsilon_{ij} \quad (6)$$

```
R> ranef(fm1.NP2)[2]#this is the nested random effect: b_ij
             (Intercept)
29/sim     -4.131275e-09
29/dissim  -1.938749e-09
30/sim      9.077410e-10
30/dissim  -3.705271e-09
```


Model 1 vs. 2 random effects

Model 1:

$$y_{ij} = \mu + b_i + \beta_1 \mathsf{Sim}_{ij} + \beta_2 \mathsf{Len}_i + \beta_3 \mathsf{Freq}_i + \epsilon_{ij} \quad (7)$$

```
Random effects:
 Formula: ~1 | subject
        (Intercept) Residual
StdDev:    0.640562 0.572838
```

Model 2:

$$y_{ij} = \mu + b_i + \underline{b_{ij}} + \beta_1 \mathsf{Sim}_{ij} + \beta_2 \mathsf{Len}_i + \beta_3 \mathsf{Freq}_i + \epsilon_{ij} \quad (8)$$

```
Random effects:
 Formula: ~1 | subject
        (Intercept)
StdDev:    0.640562

 Formula: ~1 | similarity %in% subject
         (Intercept) Residual
StdDev: 5.526901e-05 0.572838
```


Model 2 ANOVA

	numDF	denDF	F-value	p-value
similarity	1	50	55.52	1.183873e−09
lf	1	444	45.41	4.985123e−11
len	1	444	33.98	1.071035e−08

Table: NP2: similarity effect, logTRT, by subjects

Comparing Models 1 and 2

Which model is better? There are various ways to quantify this; the Akaike Information Criterion is one (it's based on log-likelihood and the number of parameters in the model). The lower the AIC the better the fit.

```
Model df      AIC       p-value
1     6     1108.965
2     7     1110.965    0.9996
```

There is not much motivation for fitting the nested random effect for similarity.

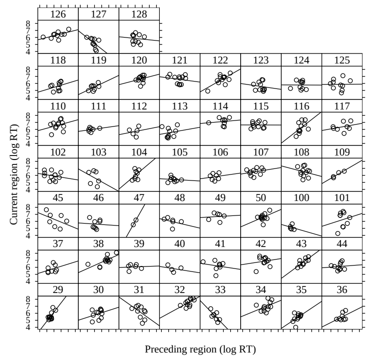
The region preceding the NP3

(3) Der Anwalt, den der Zeuge/Säbel, den **der** Spion
The lawyer who the witness/sword that the spy
betrachtete, schnitt, überzeugte den Richter
looked-at cut convinced the judge

Total reading time at the region preceding NP3 also includes parafoveal processing of NP3. We assumed that several factors would affect RT at *der*:

- Similarity, frequency, length of NP2 (as before)
- Spillover from the preceding region (*den*)

The motivation for spillover correction at *den*



Subjects show varying patterns of spillover from region $n-1$; there appears to be a subject-spillover interaction.

Model 1: Spillover as fixed effect

$$y_{ij} = \mu + b_i + \beta_1 \mathsf{Sim}_{ij} + \beta_2 \mathsf{Len}_i + \beta_3 \mathsf{Freq}_i + \beta_4 \mathsf{Spillover}_i + \epsilon_{ij} \quad (9)$$

```
Random effects:
 Formula: ~1 | subject
        (Intercept)  Residual
StdDev:   0.4059250 0.6029987

Fixed effects: logTRT ~ similarity + lf + len + logTRTn
                     Value Std.Error  DF   t-value p-value
(Intercept)       5.047819 0.3697377 394 13.652433  0.0000
similaritydissim -0.212096 0.0669203 394 -3.169379  0.0016
lf                0.019708 0.0231064 394  0.852926  0.3942
len              -0.021247 0.0144652 394 -1.468806  0.1427
logTRTn           0.193461 0.0444901 394  4.348397  0.0000
```


Model 1 ANOVA

	numDF	denDF	F-value	p-value
(Intercept)	1	394	9096.16	0.000000e+00
similarity	1	394	7.24	7.422626e−03
lf	1	394	5.01	2.574450e−02
len	1	394	2.47	1.171585e−01
logTRTn	1	394	18.91	1.748319e−05

Model 2: Adding separate spillover slopes for subjects

$$y_{ij} = \mu + b_i + \beta_1 \mathsf{Sim}_{ij} + \beta_2 \mathsf{Len}_i + \beta_3 \mathsf{Freq}_i + (\beta_4 + \zeta_i)\mathsf{Spill}_i + \epsilon_{ij} \quad (10)$$

```
Random effects:
 Formula: ~1 + logTRTn | subject
            StdDev    Corr
(Intercept) 1.1921421 (Intr)
logTRTn     0.2076766 -0.958
Residual    0.5913787

> ranef(fm0a.NP3)
    (Intercept)      logTRTn
29  -0.84606792  0.102989271
30  -0.19850234 -0.002079418

Fixed effects: logTRT ~ similarity + lf + len + logTRTn
                     Value Std.Error  DF   t-value p-value
(Intercept)       5.111320 0.4080183 394 12.527183  0.0000
similaritydissim -0.204901 0.0665910 394 -3.077012  0.0022
lf                0.025584 0.0230279 394  1.111013  0.2672
len              -0.020878 0.0144536 394 -1.444454  0.1494
logTRTn           0.170734 0.0545716 394  3.128623  0.0019
```


Model 2 ANOVA

	numDF	denDF	F-value	p-value
(Intercept)	1	394	10165.61	0.000000e+00
similarity	1	394	4.66	3.138734e−02
lf	1	394	6.51	1.112292e−02
len	1	394	2.33	1.278449e−01
logTRTn	1	394	9.79	1.886893e−03

Comparing Models 1 and 2

```
Model df    AIC      p-value
  1    7   934.2326
  2    9   931.3991  0.0328
```

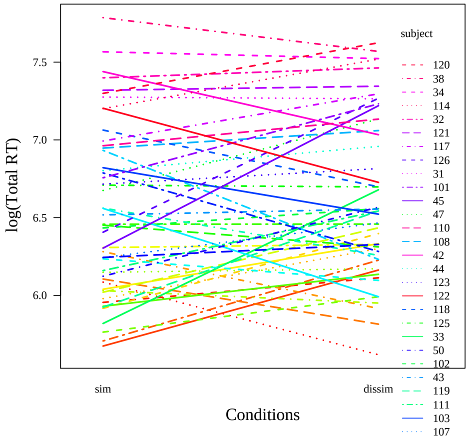
The model with the separate spillover slopes for each subject is a better fit.

The missing VP effect

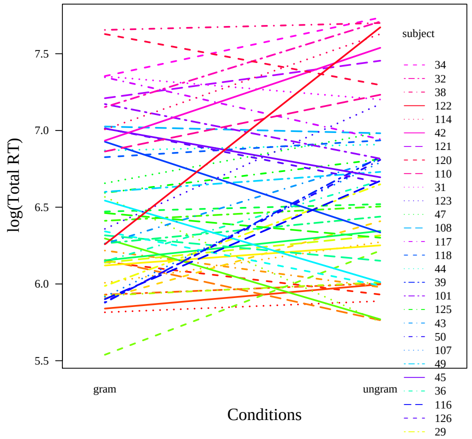
- If VP2 is forgotten, then no disturbance should occur at VP1 in the missing VP conditions (c, d) compared to the grammatical conditions.
- If VP2 is **not** forgotten, a disturbance is expected at VP1 in the missing VP conditions (c, d).

a. [NP1+ [NP2+ [NP3+ VP3] VP2] VP1]
b. [NP1+ [NP2- [NP3+ VP3] VP2] VP1]
c. [NP1+ [NP2+ [NP3+ VP3] -] VP1]
d. [NP1+ [NP2- [NP3+ VP3] -] VP1]

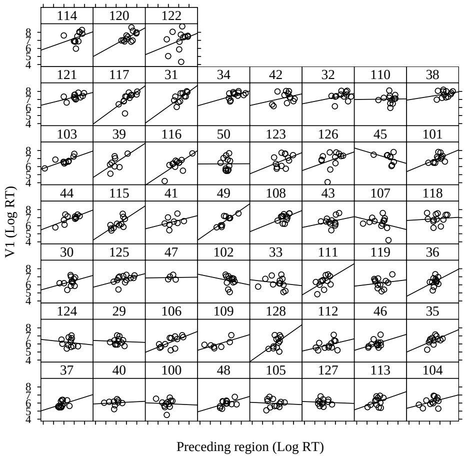
Subject-Similarity interaction at V1



Subject-Grammaticality interaction at V1



Subject-Spillover interaction at V1



Model 1: Spillover as fixed effect

$$y_{ij} = \mu + b_i + \beta_1 \mathsf{Sim}_{ij} + \beta_2 \mathsf{Gram}_{ij} + \beta_3 \mathsf{Spill}_i + \beta_4 \mathsf{Sim}_{ij} \times \mathsf{Gram}_{ij} + \epsilon_{ij} \quad (11)$$

```
Random effects:
 Formula: ~1 | subject
        (Intercept)  Residual
StdDev:   0.2700488 0.5804452

Fixed effects: logTRT ~ Sim * Gram + Spillover
             Value  Std.Error  DF    t-value  p-value
(Intercept)  3.54     0.278    522  12.77      0.00
Sim         -0.02     0.069    522  -0.30      0.76
Gram         0.08     0.069    522   1.21      0.23
Spillover    0.45     0.040    522  11.06      0.00
Sim:Gram     0.24     0.097    522   2.49      0.01
```

Model 1 ANOVA

	numDF	denDF	F-value	p-value
similarity	1	522	4.41	3.615563e−02
grammaticality	1	522	12.55	4.312170e−04
logTRTn	1	522	127.50	0.000000e+00
similarity:grammaticality	1	522	6.22	1.293493e−02

Model 2: Separate intercepts and slopes for spillover

$$y_{ij} = \mu + b_i + \beta_1 \mathsf{Sim}_{ij} + \beta_2 \mathsf{Gram}_{ij} + (\beta_3 + \zeta_i)\mathsf{Spill}_i + \beta_4 \mathsf{Sim}_{ij} \times \mathsf{Gram}_{ij} + \epsilon_{ij} \tag{12}$$

```
Random effects:
 Formula: ~1 + logTRTn | subject
            StdDev    Corr
(Intercept) 0.8052335 (Intr)
logTRTn     0.1548788 -0.988
Residual    0.5750767

Fixed effects: logTRT ~  Sim * Gram + Spillover
            Value  Std.Error  DF  t-value p-value
(Intercept) 3.51     0.29     522  12.04   0.00
Sim        -0.01     0.07     522  -0.17   0.86
Gram        0.096    0.07     522   1.4    0.16
Spillover   0.45     0.05     522   9.79   0.00
Sim:Gram    0.24     0.10     522   2.50   0.01
```


Model 2 ANOVA

	numDF	denDF	F-value	p-value
(Intercept)	1	522	25012.29	0.000000e+00
similarity	1	522	5.18	2.331227e−02
grammaticality	1	522	15.77	8.137455e−05
logTRTn	1	522	99.55	0.000000e+00
similarity:grammaticality	1	522	6.22	1.293434e−02

Model comparison

```
Model df    AIC    p-value
 1     7  1101.749
 2     9  1097.620 0.0172
```

Some practical details regarding R usage

An example of the shape of the data that was used for the analyses:

```
  subject   logTRT       lf grammaticality similarity
1     100 4.767884 4.077537           gram        sim
2     104 6.080139 4.077537           gram        sim
3     109 5.913125 4.077537           gram        sim
4     113 5.594674 4.077537           gram        sim
5     117 6.355778 4.077537           gram        sim
6     121 7.960833 4.077537           gram        sim
```

Several packages can be used for linear mixed effects models in R. Two are `nlme` and `lme4`. I show the basic usage in the following slides, with reference to the examples shown above.

Similarity effect at NP2: Model 1

$$y_{ij} = \mu + b_i + \beta_1 \mathsf{Sim}_{ij} + \beta_2 \mathsf{Len}_i + \beta_3 \mathsf{Freq}_i + \epsilon_{ij} \quad (13)$$

nlme syntax:

```
fm0.NP2 <- lme(logTRT~similarity+len+lf+logTRTn,
              random=~1|subject,
              data=pos5datafreqlen.gp,
              na.action=na.omit, method="REML")
```

lme4 syntax:

```
fm0.NP2.lmer <- lmer(logTRT~similarity+len+lf+(1|subject),
                    data=pos5datafreqlen.gp,
                    na.action=na.omit)
```


Similarity effect at NP2: Model 2

$$y_{ij} = \mu + b_i + b_{ij} + \beta_1 \mathsf{Sim}_{ij} + \beta_2 \mathsf{Len}_i + \beta_3 \mathsf{Freq}_i + \epsilon_{ij} \quad (14)$$

nlme syntax:

```
fm1.NP2 <- lme(logTRT~similarity+len+lf+logTRTn,
              random=~1|subject/similarity,
              data=pos5datafreqlen.gp,
              na.action=na.omit,method="REML")
```

lme4 syntax:

```
fm1.NP2.lmer <- lmer(logTRT~similarity+len+lf+(1|subject)+
                    (1|subject:similarity),
                    data=pos5datafreqlen.gp,
                    na.action=na.omit)
```


Region preceding NP3: Model 1

$$y_{ij} = \mu + b_i + \beta_1 \mathsf{Sim}_{ij} + \beta_2 \mathsf{Len}_i + \beta_3 \mathsf{Freq}_i + (\beta_4) \mathsf{Spill}_i + \epsilon_{ij} \quad (15)$$

nlme syntax:

```
fm0.NP3 <- lme(logTRT~similarity+lf+len+logTRTn,
              random=~1|subject,
              data=pos7datafreqlen,
              na.action=na.omit)
```

lme4 syntax:

```
fm0.NP3.lmer <- lmer(logTRT~similarity+lf+len+logTRTn+
                    (1|subject),
                    data=pos7datafreqlen,
                    na.action=na.omit)
```

Region preceding NP3: Model 2

$$y_{ij} = \mu+b_i+\beta_1\mathsf{Sim}_{ij}+\beta_2\mathsf{Len}_i+\beta_3\mathsf{Freq}_i+(\beta_4+\zeta_i)\mathsf{Spill}_i+\epsilon_{ij} \quad (16)$$

nlme syntax:

```
fm1.NP3 <- lme(logTRT~similarity+lf+len+logTRTn,
               random=~1+logTRTn|subject,
               data=pos7datafreqlen,
               na.action=na.omit)
```

lme4 syntax:

```
# Note: random intercept is implicit in (logTRTn|subject)
fm1.NP3.lmer <- lmer(logTRT~similarity+lf+len+logTRTn+
                    (logTRTn|subject),data=pos7datafreqlen,
                     na.action=na.omit)

# We can remove the random intercept term:
fm1.NP3.lmer <- lmer(logTRT~similarity+lf+len+logTRTn+
                    (logTRTn-1|subject),data=pos7datafreqlen,
                     na.action=na.omit)
```


Interaction plots in R

```
attach(pos5datafreqlen)
interaction.plot(similarity,itemnum,lf,las=1,
     fun = function(x) mean(x, na.rm=TRUE),
     ylab="log(Frequency)",
     xlab="Conditions",
     lwd=2,
     fixed=FALSE,
     cex.lab=1.5,
     col=rainbow(51))
detach(pos5datafreqlen)
```


XY plots in R

```
library(lattice)

ltheme <- canonical.theme(color = FALSE) ## in-built B&W theme
ltheme$strip.background$col <- "transparent" ## change strip bg
lattice.options(default.theme = ltheme) ## set as default
#some sensible defaults for scales:
scalelist <- list(x=list(alternating=0),
                   y=list(alternating=1),
                   tck=c(.5))
#function for plotting the regression lines:
drawfittedline <- function(x,y){
  panel.xyplot(x,y)
     panel.lmline(x,y,type="l",lwd=1,col="black")}

print(xyplot(logTRT~lf|subject,pos5datafreqlen,
       xlab="Log frequency",
       ylab="Log total reading times",
       panel=drawfittedline,
       scales=scalelist))
```


Selected bibliography

- Julian Faraway. Extending the linear model with R. Chapman and Hall,/CRC, Boca Raton, FL, 2006. *A somewhat sketchy presentation, but has many examples of the latest lme4 package's usage.*
- Stephen W. Raudenbush and Anthony S. Bryk. Hierarchical Linear Models: Applications and Data analysis methods. Quantitative techniques in the social sciences. Sage, second edition, 2002. *This is the most important reference of all; it provides a detailed and accessible treatment of the theory behind HLMs.*
- José C. Pinheiro and Douglas M. Bates. Mixed-Effects Models in S and S-PLUS. Springer-Verlag, New York, 2000. *A very useful (and expensive) user's manual for nlme, with a lot of technical detail.*
- W. N. Venables and B. D. Ripley. Modern Applied Statistics with S-PLUS. Springer, New York, 1999. *Terse and tight-lipped treatment of everything statistical. Will put hair on your chest.*

Acknowledgements

Tom Santner and Sumithra Mandrekar at the statistical consultancy service in Ohio State University introduced me to linear mixed-effects models in 2000. I have also learnt much from Reinhold Kliegl's Winter 2005-06 course at Potsdam on the use of lme in multiple regression research. Any errors in this document are of course my own responsibility.