

m15\_bhv\_202109\_lme\_rndint\_2fac\_naomit\_inv

## Morph 15 Behavioral Data Analysed via Linear Mixed Effects Model

These data are from subjects 8-31 (25 subjects) of the Morph15 Suffix productivity Project. The data were collected between 3/25/2015 and 25/9/2015 using E-Prime. The E-Primes files used were “Morph15Group1Correct” and “Morph15Group2Correct”. The excel file exported from E-data Aid is ‘m15\_bhv\_25’

The stimuli are read in from ‘m15\_stm\_frq.csv’.

## Load Packages

```
library(knitr)
library(markdown)
library(readr)
library(ez)
library(lme4)
library(stringr)
library(ggplot2)
library(dplyr)
library(tidyr)
library(gridExtra)
library(RColorBrewer)
```

## Define Standard error of the Mean Function

```
sem = function(x)
{
  sqrt(var(x)/length(x))
}
```

## Read in Data and Format Data

This chunk reads the dataset in long format into an R dataframe using the read.csv(file) function.

The stimuli are read in from “m15\_stm\_frq.csv”. This file contains stimuli divided into high and low parsability on the basis of affix frequency (afx\_frq), relative frequency determined via a median split based on the **stem frequency to whole-word frequency ratio** (med\_splt) and relative frequency determined via a split based on whether the **stem frequency to whole-word frequency ratio** is greater than 1 or less than 1 (ratio\_splt).

8/1/2021 New version of the analysis adds another split based on whether the stem to whole-word frequency is greater or less than 1.5.

\*\* In order to read in files, remember to 'set working directory to project directory'”

```
library(readr)
m15_bhv <- read_csv("M15_ALL_SUBJS_BHV_DATA_orig-3.csv")
```

Rows: 10000 Columns: 11

```
-- Column specification -----
Delimiter: ","
chr (3): Prime, List, Target
dbl (8): Subject, Trigger, StemID, Correct_Response, Target.RESP, Target.ACC...
```

i Use `spec()` to retrieve the full column specification for this data.  
i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

```
frq <- read_csv("stimuli/m15_stm_frq_rel_subset.csv")
```

Rows: 100 Columns: 21

-- Column specification -----

Delimiter: ","

chr (12): w\_word, stem.x, affix, transp, afx\_onset, prime.x, rel, afx\_frq, m...

dbl (9): whl\_word\_len, afx\_len, stem\_len, stem\_frq, stem\_logfrq, whl\_word\_f...

i Use `spec()` to retrieve the full column specification for this data.

i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

```
m15_bhv <- m15_bhv %>% mutate(Subject = replace(Subject, Subject == 241, 24)) # changes subj 241 to 24
```

This chunk changes removes extraneous digits from column names and makes them all lowercase, renames cols with .x after their name and removes rel column from frq dataframe

```
names(m15_bhv) <- tolower(names(m15_bhv))
```

```
names(frq) <- tolower(names(frq))
```

```
frq <- rename(frq, stemid = trigger)
```

```
frq <- rename (frq, prime = prime.x)
```

```
frq <- rename (frq, stem = stem.x)
```

```
frq$rel <- NULL
```

This chunk creates a separate dataframe for the words

```
m15_bhv_words<-filter(m15_bhv,correct_response == 6)
```

```
m15_bhv_words$relatedness <- ifelse(m15_bhv_words$trigger <= 100, "related", "unrelated")
```

We now combine the stimulus df with the data df

```
m15_bhv_frq_words <- left_join(m15_bhv_words, frq, by = "stemid")
```

```
# m15_bhv_frq_word <- dplyr::select(m15_bhv_frq_words, subject, prime, target, stem.x, stemid, relatedness, af
```

```
write.csv(m15_bhv_frq_words, "m15_bhv_frq_words.csv")
```

**Replaces NA rt values with the mean for the dataset in column new.rt. Also creates new dataframe (m15\_bhv\_frq\_words\_na.omit) with just the non-missing data**

```
m15_bhv_frq_words$new.rt<- replace_na(m15_bhv_frq_words$target.rt.trm.err,  
                                     mean(m15_bhv_frq_words$target.rt.trm.err,  
                                           na.rm=TRUE ))
```

```
missing_data<- filter(m15_bhv_frq_words, is.na(m15_bhv_frq_words$target.rt.trm.err))
```

```
m15_na.omit <- filter(m15_bhv_frq_words, !is.na(m15_bhv_frq_words$target.rt.trm.err))
```

```
xtab.missing.data <- xtabs(~relatedness+med_splt, data=missing_data)
```

```
m15_bhv_frq_words_rmna <-filter(m15_bhv_frq_words,!is.na(m15_bhv_frq_words$target.rt.trm.err))
```

This chunk creates two new dependent variables by doing a log transformation and an inverse transformation of the reaction times

```
m15_bhv_frq_words_rmna$log.rt <- log(m15_bhv_frq_words_rmna$target.rt.trm.err)
```

```
m15_bhv_frq_words_rmna$inverse.rt <- (1/(m15_bhv_frq_words_rmna$target.rt.trm.err))*1000
```

```
write.csv(m15_bhv_frq_words_rmna, "m15_bhv_frq_words_rmna.csv")
```

# Inverse RT

## Condition Means: Inverse RT

`summarise()` creates a new data frame. It will have one (or more) rows for each combination of grouping variables; if there are no grouping variables, the output will have a single row summarising all observations in the input. It will contain one column for each grouping variable and one column for each of the summary statistics that you have specified.

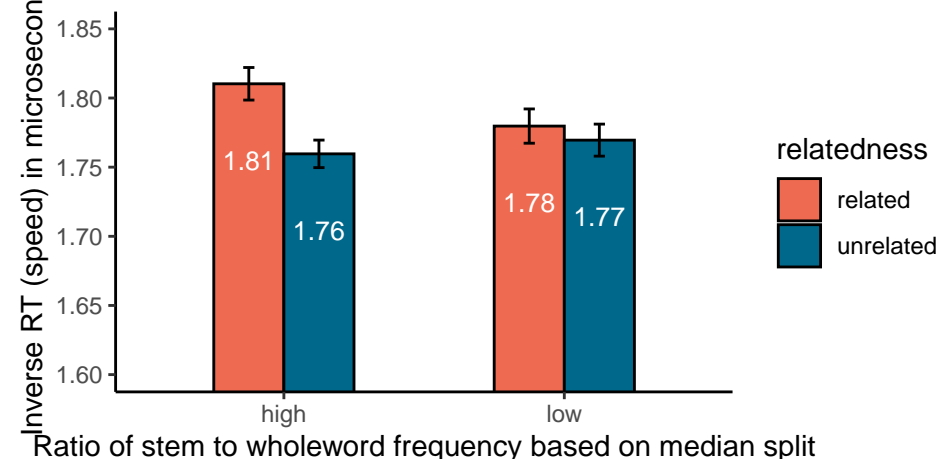
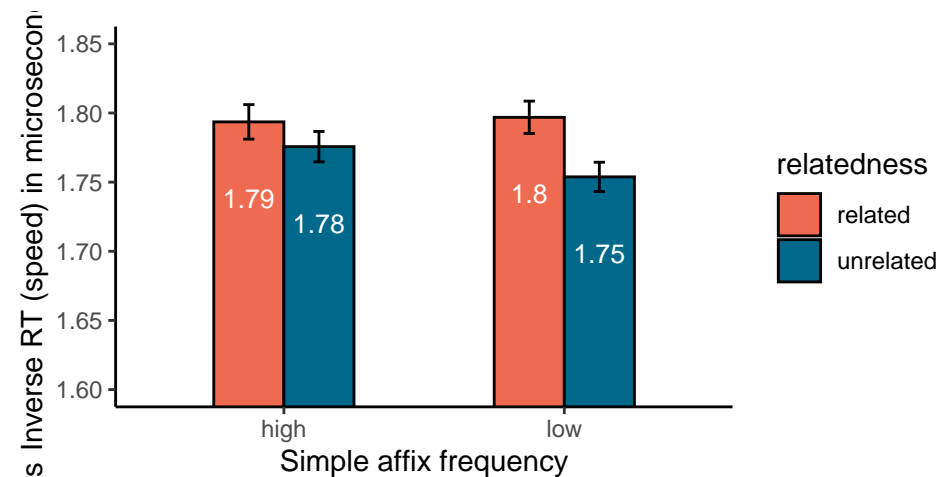
``summarise()`` has grouped output by 'afx\_freq'. You can override using the ``groups`` argument.

```
# A tibble: 4 x 5
# Groups:   afx_freq [2]
  afx_freq relatedness meanRT num_stim    se
  <chr>    <chr>      <dbl>   <int> <dbl>
1 high     related        1.79    1107 0.0125
2 high     unrelated        1.78    1131 0.0110
3 low      related        1.80    1161 0.0117
4 low      unrelated        1.75    1175 0.0106
```

``summarise()`` has grouped output by 'med\_splt'. You can override using the ``groups`` argument.

```
# A tibble: 4 x 5
# Groups:   med_splt [2]
  med_splt relatedness meanRT num_stim    se
  <chr>    <chr>      <dbl>   <int> <dbl>
1 high     related        1.81    1154 0.0118
2 high     unrelated        1.76    1165 0.00995
3 low      related        1.78    1114 0.0124
4 low      unrelated        1.77    1141 0.0116
```

## Barplots for Inverse RT



# Models

Reminder about goodness-of-fit criteria (from Wikipedia) Suppose that we have a statistical model of some data. Let  $k$  be the number of estimated parameters in the model. Let  $\hat{L}$  be the maximum value of the likelihood function for the model, where ‘likelihood’ is used to describe the plausibility of a value for the parameter, given some data. The likelihood function answers the question *What is the probability of observing the current dataset, when I assume a given set of parameters for my linear model?* Considering only  $\hat{L}$ , the likelihood  $L$  or its natural logarithm ( $\text{Log}L$ ) is maximum when  $\sum_{i=1}^n (y - \mu)^2$  is a minimum. Then the AIC value of the model is the following:

$$AIC = 2k - 2\ln(\hat{L})$$

Given a set of candidate models for the data, the preferred model is the one with the minimum AIC value. Thus, AIC rewards goodness of fit (as assessed by the likelihood function), but it also includes a penalty that is an increasing function of the number of estimated parameters.

From Petr Keil (<http://www.petrkeil.com/?p=836>) “Model selection is a process of seeking the model in a set of candidate models that gives the best balance between model fit and complexity (Burnham & Anderson 2002). I have always used AIC for that. But you can also do that by crossvalidation. Specifically, Stone (1977) showed that the AIC and leave-one out crossvalidation are asymptotically equivalent.” He also included the BIC (Bayesian Information Criterion) and found that all three are equivalent.

From (<https://www.scribbr.com/statistics/akaike-information-criterion/>) : If a model is more than 2 AIC units lower than another, then it is considered significantly better than that model.

## Inverse RT

### Run ANOVA Affix Frequency

Null Model: Affix Frequency: Inverse RT

```
afxfrq_inv_null.model = lmer(inverse.rt ~ 1 + (1|subject) + (1|stemid),  
                             data= m15_bhv_frq_words_rmna, REML=FALSE)  
summary(afxfrq_inv_null.model)
```

```
Linear mixed model fit by maximum likelihood ['lmerMod']  
Formula: inverse.rt ~ 1 + (1 | subject) + (1 | stemid)  
Data: m15_bhv_frq_words_rmna
```

AIC	BIC	logLik	deviance	df.resid
3085.7	3111.4	-1538.8	3077.7	4570

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.7210	-0.6352	0.0325	0.6270	6.7862

Random effects:

Groups	Name	Variance	Std.Dev.
stemid	(Intercept)	0.01301	0.1141
subject	(Intercept)	0.03130	0.1769
Residual		0.10782	0.3284

Number of obs: 4574, groups: stemid, 100; subject, 25

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	1.7687	0.0375	47.17

### Main Effects Model: Affix Frequency: Inverse RT

This model of the Inverse RT data contains fixed intercepts for *affix frequency* and *relatedness* as well as random intercepts for subjects and items (adjustments to the mean for each subject and item)

```
afxfreq_inv_main.model = lmer(inverse.rt ~ afx_freq + relatedness + (1|subject) + (1|stemid),
                             data= m15_bhv_freq_words_rmna, REML=FALSE)
summary(afxfreq_inv_main.model)
```

Linear mixed model fit by maximum likelihood ['lmerMod']  
 Formula: inverse.rt ~ afx\_freq + relatedness + (1 | subject) + (1 | stemid)  
 Data: m15\_bhv\_freq\_words\_rmna

AIC	BIC	logLik	deviance	df.resid
3078.9	3117.5	-1533.4	3066.9	4568

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.6801	-0.6222	0.0372	0.6364	6.7459

Random effects:

Groups	Name	Variance	Std.Dev.
stemid	(Intercept)	0.01303	0.1142
subject	(Intercept)	0.03131	0.1770
Residual		0.10756	0.3280

Number of obs: 4574, groups: stemid, 100; subject, 25

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	1.785736	0.039835	44.828
afx_freqlow	-0.001881	0.024842	-0.076
relatednessunrelated	-0.031853	0.009708	-3.281

Correlation of Fixed Effects:

	(Intr) afx_fr
afx_freqlow	-0.314
rltdnssnrlt	-0.123 0.001

### Interaction Model: Affix Frequency: Inverse RT

```
afxfreq_inv_inter.model = lmer(inverse.rt ~ afx_freq * relatedness + (1|subject) + (1|stemid),
                               data= m15_bhv_freq_words_rmna, REML=FALSE)
summary(afxfreq_inv_inter.model)
```

Linear mixed model fit by maximum likelihood ['lmerMod']  
 Formula: inverse.rt ~ afx\_freq \* relatedness + (1 | subject) + (1 | stemid)  
 Data: m15\_bhv\_freq\_words\_rmna

AIC	BIC	logLik	deviance	df.resid
3079.0	3124.0	-1532.5	3065.0	4567

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.7010	-0.6240	0.0372	0.6357	6.7682

Random effects:

Groups	Name	Variance	Std.Dev.
stemid	(Intercept)	0.01306	0.1143
subject	(Intercept)	0.03130	0.1769
Residual		0.10751	0.3279

Number of obs: 4574, groups: stemid, 100; subject, 25

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	1.77877	0.04015	44.304

afx_frqlow	0.01171	0.02672	0.438
relatednessunrelated	-0.01809	0.01387	-1.304
afx_frqlow:relatednessunrelated	-0.02695	0.01941	-1.388

Correlation of Fixed Effects:

	(Intr)	afx_fr	rltdns
afx_frqlow	-0.335		
rltdnssnrlt	-0.175	0.263	
afx_frqlw:r	0.125	-0.366	-0.715

### Compare Null and Main Models: Affix Frequency

```
anova(afxfrq_inv_null.model,afxfrq_inv_main.model)
```

Data: m15\_bhv\_frq\_words\_rmna

Models:

afxfrq\_inv\_null.model: inverse.rt ~ 1 + (1 | subject) + (1 | stemid)

afxfrq\_inv\_main.model: inverse.rt ~ afx\_frq + relatedness + (1 | subject) + (1 | stemid)

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
afxfrq_inv_null.model	4	3085.7	3111.4	-1538.8	3077.7			
afxfrq_inv_main.model	6	3078.9	3117.5	-1533.5	3066.9	10.758	2	0.004612

afxfrq\_inv\_null.model

afxfrq\_inv\_main.model \*\*

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### Compare Main and Interaction Models: Affix Frequency

```
anova(afxfrq_inv_main.model,afxfrq_inv_inter.model)
```

Data: m15\_bhv\_frq\_words\_rmna

Models:

afxfrq\_inv\_main.model: inverse.rt ~ afx\_frq + relatedness + (1 | subject) + (1 | stemid)

afxfrq\_inv\_inter.model: inverse.rt ~ afx\_frq \* relatedness + (1 | subject) + (1 | stemid)

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
afxfrq_inv_main.model	6	3078.9	3117.5	-1533.5	3066.9			
afxfrq_inv_inter.model	7	3079.0	3124.0	-1532.5	3065.0	1.926	1	0.1652

## Run ANOVA Median Split

Null Model: Median Split: Inverse RT

```
medsplnt_inv_null.model = lmer(inverse.rt ~ 1 + (1|subject) + (1|stemid),
                                data= m15_bhv_frq_words_rmna, REML=FALSE)
summary(medsplnt_inv_null.model)
```

Linear mixed model fit by maximum likelihood ['lmerMod']

Formula: inverse.rt ~ 1 + (1 | subject) + (1 | stemid)

Data: m15\_bhv\_frq\_words\_rmna

AIC	BIC	logLik	deviance	df.resid
3085.7	3111.4	-1538.8	3077.7	4570

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.7210	-0.6352	0.0325	0.6270	6.7862

Random effects:

Groups	Name	Variance	Std.Dev.
--------	------	----------	----------

```

stemid (Intercept) 0.01301 0.1141
subject (Intercept) 0.03130 0.1769
Residual          0.10782 0.3284
Number of obs: 4574, groups: stemid, 100; subject, 25

```

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	1.7687	0.0375	47.17

### Main Effects Model: Median Split: Inverse RT

This model of the Inverse RT data contains fixed intercepts for *affix frequency* and *relatedness* as well as random intercepts for subjects and items (adjustments to the mean for each subject and item)

```

medspltt_inv_main.model = lmer(inverse.rt ~ med_spltt + relatedness + (1|subject) + (1|stemid),
                               data= m15_bhv_frq_words_rmna, REML=FALSE)
summary(medspltt_inv_main.model)

```

Linear mixed model fit by maximum likelihood ['lmerMod']

Formula: inverse.rt ~ med\_spltt + relatedness + (1 | subject) + (1 | stemid)

Data: m15\_bhv\_frq\_words\_rmna

AIC	BIC	logLik	deviance	df.resid
3078.6	3117.2	-1533.3	3066.6	4568

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.6827	-0.6215	0.0358	0.6339	6.7432

Random effects:

Groups	Name	Variance	Std.Dev.
stemid	(Intercept)	0.01299	0.114
subject	(Intercept)	0.03132	0.177
Residual		0.10756	0.328

Number of obs: 4574, groups: stemid, 100; subject, 25

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	1.791644	0.039792	45.025
med_splttlow	-0.013733	0.024806	-0.554
relatednessunrelated	-0.031841	0.009708	-3.280

Correlation of Fixed Effects:

	(Intr) md_spl
med_splttlow	-0.311
rltdnssnrlt	-0.122 -0.002

### Interaction Model: Median Split: Inverse RT

```

medspltt_inv_inter.model = lmer(inverse.rt ~ med_spltt * relatedness + (1|subject) + (1|stemid),
                                data= m15_bhv_frq_words_rmna, REML=FALSE)
summary(medspltt_inv_inter.model)

```

Linear mixed model fit by maximum likelihood ['lmerMod']

Formula: inverse.rt ~ med\_spltt \* relatedness + (1 | subject) + (1 | stemid)

Data: m15\_bhv\_frq\_words\_rmna

AIC	BIC	logLik	deviance	df.resid
3076.2	3121.1	-1531.1	3062.2	4567

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.6547	-0.6239	0.0428	0.6259	6.7147

Random effects:

Groups	Name	Variance	Std.Dev.
stemid	(Intercept)	0.01300	0.1140
subject	(Intercept)	0.03131	0.1770
Residual		0.10745	0.3278

Number of obs: 4574, groups: stemid, 100; subject, 25

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	1.80176	0.04008	44.952
med_splttlow	-0.03437	0.02668	-1.289
relatednessunrelated	-0.05201	0.01362	-3.818
med_splttlow:relatednessunrelated	0.04093	0.01941	2.109

Correlation of Fixed Effects:

	(Intr)	md_spl	rltdns
med_splttlow	-0.331		
rltdnssnrlt	-0.171	0.256	
md_splttlw:r	0.120	-0.367	-0.702

### Compare Null and Main Models: Median Split

```
anova(medsplt_inv_null.model,medsplt_inv_main.model)
```

Data: m15\_bhv\_frq\_words\_rmna

Models:

medsplt\_inv\_null.model: inverse.rt ~ 1 + (1 | subject) + (1 | stemid)

medsplt\_inv\_main.model: inverse.rt ~ med\_splt + relatedness + (1 | subject) + (1 | stemid)

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
medsplt_inv_null.model	4	3085.7	3111.4	-1538.8	3077.7			
medsplt_inv_main.model	6	3078.6	3117.2	-1533.3	3066.6	11.059	2	0.003969

medsplt\_inv\_null.model

medsplt\_inv\_main.model \*\*

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### Compare Main and Interaction Models: Median Split

```
anova(medsplt_inv_main.model,medsplt_inv_inter.model)
```

Data: m15\_bhv\_frq\_words\_rmna

Models:

medsplt\_inv\_main.model: inverse.rt ~ med\_splt + relatedness + (1 | subject) + (1 | stemid)

medsplt\_inv\_inter.model: inverse.rt ~ med\_splt \* relatedness + (1 | subject) + (1 | stemid)

	npar	AIC	BIC	logLik	deviance	Chisq	Df
medsplt_inv_main.model	6	3078.6	3117.2	-1533.3	3066.6		
medsplt_inv_inter.model	7	3076.2	3121.2	-1531.1	3062.2	4.4464	1

Pr(>Chisq)

medsplt\_inv\_main.model

medsplt\_inv\_inter.model 0.03497 \*

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1



## Compare models using AIC

```
library(AICcmodavg)
```

Attaching package: 'AICcmodavg'

The following object is masked from 'package:lme4':

```
checkConv
```

```
afxfrq_models <- list(afxfrq_inv_null.model, afxfrq_inv_main.model, afxfrq_inv_inter.model)
medspl_t_models <- list(medspl_t_inv_null.model, medspl_t_inv_main.model, medspl_t_inv_inter.model)
model.names <- c('null.mod', 'maineffects.mod', 'interaction.mod')

aictab(cand.set = afxfrq_models, modnames = model.names)
```

Model selection based on AICc:

	K	AICc	Delta_AICc	AICcWt	Cum.Wt	LL
maineffects.mod	6	3078.92	0.00	0.50	0.50	-1533.45
interaction.mod	7	3079.00	0.08	0.48	0.98	-1532.49
null.mod	4	3085.66	6.75	0.02	1.00	-1538.83

```
aictab(cand.set = medspl_t_models, modnames = model.names)
```

Model selection based on AICc:

	K	AICc	Delta_AICc	AICcWt	Cum.Wt	LL
interaction.mod	7	3076.18	0.00	0.77	0.77	-1531.08
maineffects.mod	6	3078.62	2.44	0.23	0.99	-1533.30
null.mod	4	3085.66	9.49	0.01	1.00	-1538.83