Collostructional analysis:

source("http://www.linguistics.ucsb.edu/faculty/stgries/teaching/groningen/coll.analysis.r")

**Select:**

2: (multiple) distinctive collocates or distinctive collexeme analysis

**Select:**

1: 2 alternatives

**Select:**

99

**Select:**

1: -log10 (Fisher-Yates exact, one-tailed) (= default)

**Select:**

4: collostruction strength

**Select:**

1: Raw list of all tokens

Rbrul regression:

source("http://www.danielezrajohnson.com/Rbrul.R")

rbrul()

Mixed-effects logistic regression:

library(lme4)

data$Genre <- as.factor(data$Genre)

data$Region <- as.factor(data$Region)

data$Relation <-as.factor(data$Relation)

data$Text\_ID <- as.factor(data$Text\_ID)

data$Year <- as.numeric(as.character(data$Year))

data$zYear <- scale(data$Year, center=TRUE, scale=TRUE)

data$Prep <- as.factor(data$Prep)

ANOVA:

model.fit <- glmer(Prep~(1|Text\_ID)+Region+zYear+Genre+Relation, family = binomial, data=data)

model.fit2 <- glmer(Prep~(1|Text\_ID)+Region+zYear+Genre+Relation+Relation:zYear, family = binomial, data=data)

anova(model.fit, model.fit2, test="Chisq")

LOESS plotting:

library(ggplot2)

data$PrepRecode <- ifelse(data$Prep == "MID",0,1)

ggplot(data, aes(Year, PrepRecode, color=Relation)) +

labs(y = "Proportion of WID", x = "\nYear") +

stat\_sum(aes(size=..n.., alpha=.1)) +

scale\_size\_area(max\_size=12) +

stat\_smooth() +

scale\_alpha\_continuous(guide="none", limits = c(0,.7)) +

scale\_color\_brewer(palette = "Set1") +

ylim(0,1)

Spline plotting:

library(splines)

library(MASS)

data$PrepRecode <- ifelse(data$Prep == "MID",0,1)

#This fits a cubic spline with **3** knots

ggplot(data, aes(Year, PrepRecode, color=Relation)) +

labs(y = "Proportion of WID", x = "\nYear") +

stat\_sum(aes(size=..n.., alpha=.1)) +

scale\_size\_area(max\_size=12) +

stat\_smooth(method= "lm", formula = y ~ ns(x,**3**)) +

scale\_alpha\_continuous(guide="none", limits = c(0,.7)) +

scale\_color\_brewer(palette = "Set1") +

ylim(0,1)

Logistic plotting:

data$PrepRecode <- ifelse(data$Prep == "MID",0,1)

ggplot(data, aes(Year, PrepRecode, color=Relation)) +

labs(y = "Proportion of WID", x = "\nYear") +

stat\_sum(aes(size=..n.., alpha=.1)) +

scale\_size\_area(max\_size=12) +

stat\_smooth(method= "glm", family = "binomial") +

scale\_alpha\_continuous(guide="none", limits = c(0,.7)) +

scale\_color\_brewer(palette = "Set1") +

ylim(0,1)