Judging Formulaic Language: Training, Linguistics and Judging the Judgments*

Russell Hugo

Department of Linguistics, University of Washington,
Guggenheim Hall 4th Floor, Box 352425, Seattle, WA 98195-2425
rlhugo@uw.edu

Summary: Formulaic language generally refers to a group of more than two words that are stored as a chunk in the lexicon rather than being generated by a grammar or similar system. This paper describes a project which examines the consistency and reliability of individuals' judgments. In particular, what attributes of the judges might play a role or be related to bias? To answer these questions, an online survey was constructed comprising twenty chunks with their sentential context. Respondents with a background in linguistics or English language academics were sought and make up the majority of the total group. Results seem to suggest that disagreement is fairly prevalent and the educational background of the individual plays a role. This study highlights the difficulties with subjective/human language judgments and some of the biases that may affect them, regardless of substantial training or strict metrics.

Keywords: Formulaic, chunk, bias, questionnaire, judgements

* Thank you to Prof. Amy Snyder Ohta for substantial support and guidance on this project, Allison Germain for help with the syntactic groupings, the various statistics consultants in the UW Statistics Department (special thanks to Jenn Kirk for exceptional R debugging skills) as well as the Center For Statistics And The Social Sciences, and all of the participants for their time.

University of Washington Working Papers in Linguistics (UWWPL). Volume 33 (2015).

"Formulaic language is when two or more lexemes are co-entrenched in the internal lexico-grammars of cross-section of members of a language community."

***

"It is a group of words which is stored in the mind of the speaker and is used like a formula in certain context"

***

"Formulaic language is indefinable."

- Comments from different anonymous respondents when asked how they would define the term 'formulaic language'.

1 Introduction

Contrary to the quotes above, taken from one of the studies described in this paper, determining a broad definition for formulaic language is not hotly debated with in the field. The term *formulaic language* generally refers to groups of more than two words that are not generated by a grammar or another system that combines the smaller pieces when produced. Instead, the group of words, or 'chunk', is stored in the mental lexicon or elsewhere in the brain for rapid recall, hence the term 'formulaic'. The string of words acts as a ready-made formula for what might otherwise require a greater processing effort. One of the arguments for the existence and utility of formulaic chunks is that it reduces the processing load for both native and non-native speakers (Conklin & Schmitt, 2008; Jiang & Nekrasova, 2007; Pawley & Syder, 1983). Some researchers suggest that formulaic chunks can be beneficial for L2 acquisition (Millar, 2011), while others argue that non-natives may not enjoy noticeable benefits (Siyanova-Chanturia, Conklin, & Schmitt, 2011). Wray (2008) uses the term Morpheme Equivalent Unit (MEU), which is: "a word or word string, whether incomplete or including gaps for inserted variable items, that is processed like a morpheme, that is, without recourse to any form-meaning matching of any sub-parts it may have." (pg. 12)

However, determining *what actual language is formulaic* can be more problematic. Some of the more commonly held definitions include idioms (e.g., *kick the bucket*), but the range and scope of what is believed to be formulaic is much more complicated, and can be linked to larger theories of linguistics (e.g., generativism). Researchers have claimed that analyzed texts have been comprised of anywhere from approximately 20-60% of formulaic language (Biber, 1999; Erman & Warren, 2000; Nattinger & DeCarrico, 1992). In the context of processing, is a certain string of language that has a lighter processing load formulaic or merely a common string that is light due to frequency effects (Arnon & Snider, 2010; Ellis, Simpson-Vlach, & Maynard, 2008)? How does function and discourse usage (Nekrasova, 2009) relate to frequency? This question of what is actually formulaic is left to other researchers, including those working with other models and approaches (e.g., corpus linguistics, ERPs, psycholinguistics, etc).

Instead, the focus of this paper is on judgments made by individuals. One question is how consistent and reliable are judgments and what factors associated with the judges might play a role? Are linguists or native speakers of the language in question likely to judge certain strings different than non-native speakers of that language? Does training play a role? Are certain categories of strings more likely to be judged as formulaic? This project was motivated by an interest in further examining the reliability of developing diagnostics for identifying formulaicity that rely on human judgments. Is it possible to develop a series of metrics that can be used consistently by a group of individuals, or are their biases that play a large role in the process that render the metrics problematic in certain cases. Perhaps, the data may also highlight some subtle biases from certain groups (e.g., semanticists) that could affect other areas of research that rely on judgments.

This paper describes two pilot studies and a larger final study that aim to provide some evidence for these questions. In order to get a better understanding of how certain factors related to judges or formulaicity might affect the judgments, an initial pilot study was carried out by a professor and graduate students taking a seminar dedicated to formulaic language.[1] While describing an interesting personal event and retelling a silent video click, a mixture of L1 and non-L1 English speakers were recorded and then later asked to review a transcript of their speech and highlight chunks they believed to be formulaic. Four trained[2] judges later reviewed each of the transcribed recordings and marked strings they believed to be formulaic. Later, a second pilot study was conducted online whereby a survey was presented to 20 self-selected subjects that asked them to judge certain chunks selected by the judges in Pilot Study 1 as being formulaic or not. Overall, the judgments made in Pilot Study 2 patterned with those in Pilot Study 1 in that the strings that received the most selections in the first study were judged as being more likely to be formulaic by participants in Pilot Study 2, but there were some noticeable patterns of disagreement. Building on these two pilot studies a new set of tokens were selected and presented online for people to rate. People with academic backgrounds in English or linguistics were targeted for recruitment although anyone was welcome to take part in the study. Participants were asked to provide detailed educational background information to see if some of the differences seen in the pilot results would also be present in a larger data set. Later, participants were recruited using Amazon's Mechanical Turk (AMT) system. This was done as a way to increase the number of controls while also testing the feasibility of using the AMT for similar linguistic projects. This paper is a work-in-progress and presents only rough analyses of the final data. All of the relevant data is included in the hopes that it might be useful or of interest to other researchers working on formulaicity.

2 Methods

Research questions for the pilot studies:[3]

1. Is there evidence supporting the idea that the level of linguistic training appear has an effect on the judgment types or rates?

2. How do non- or minimally trained judges compare to the 4 quarter-long trained judges and speaker judgments of the first study? In other words, does training increase the agreement of the judges?

Pilot Study 1

Participants

For this study 6 subjects were interviewed and recorded. All participants were between the ages of 18-35. 2 were native speakers of English, one male and one female. 2 were female native speakers of Korean, yet proficient enough in English to be teaching assistants at a large public university in the United States, where one teaches political science and the other, Korean.[4] The last 2 participants

[1] This study was conducted by Prof. Amy Snyder Ohta, Hyunjung Ahn, Brent Carey and the author.
[2] The judges were very familiar with the diagnostics provided by Wray (2008).
[3] Additional questions that were unanswerable do to the limited subject pool for study 2 included: Does the level of training on judging formulaic language have an effect on the judgment types or rates? Does the L1 or level of proficiency with English have an effect on the judgment types or rates?
[4] To address a question raised by an anonymous reviewer who asked about the English proficiency requirement for a Korean language instructor. At the UW, many language TAs are required to take an English proficiency test as most courses do not use an immersion format. The Korean program at UW emphasizes a substantial amount of explicit metalinguistic grammar teaching in English, in addition to cultural and historical information in the first two years of study.

also had Korean as an L1 and were currently international students at the same university enrolled in the English Language Program.

The native language of the subjects was restricted to either Korean or English for consistency, availability of subjects, and because one of the researchers was a native speaker of Korean and could develop materials in Korean that could clearly communicate the tasks which were presented to the participants.

The four judges included one professor (English L1) and three linguistics graduate students (2 English L1s & 1 Korean L1) who were taking part in a seminar on formulaic language.

Procedure

Each subject was interviewed separately and was given two tasks. For the first task (*conversation*) the subject was given a list of topics (Favorite vacation, Tell me about your self (in general), Hobbies, Family) all having to do with something they were familiar with (Skehan & Foster, 1997) and were asked to pick one that they felt they could comfortably talk about in English for about 5 minutes. For the second task the researcher informed that they would show the participant a short clay animation video clip[5], that the researcher had not seen before, and leave the room. After the clip was over the researcher would return and ask the subject to tell them about it. (e.g., What happened? What did you think about it? Did you like it? Why or Why not?)

After the recording, the subjects were asked to return the next day. In the interim their speech was transcribed and each task was trimmed to approximately 200 words. The next day the participants were given a brief training on identifying and judging formulaic language (Wray, 2008). The subjects then highlighted any word strings they believed to be formulaic. The subject's truncated transcript was later reviewed by the four trained judges. Each judge reviewed the transcript individually noting each chunk they though was formulaic, as well as noting which of the criteria outlined by (Wray, 2008) it fit. The judgments of each subject were not disclosed until each judge had completed their judgments for each of the transcripts.

Analysis

Each chunk that was marked by at least one person (the speaker as judge, the four trained judges or one of the additional TA L2 judges[6]) was analyzed (both per-token and per-task) based on two criteria:

- Was it marked as being formulaic by all, the majority, or few of the judges?
- Was it marked as being formulaic by only native English speakers, only non-native English speakers or at least one judge from each group?

In addition to the quantity of judgments per-speaker, a per-task calculation was made for each judge. In general, there was a very low degree of agreement among the three trained judges, and no apparent strong effect of having an L1 of English. The detailed results for pilot study 1 are presented in comparison with the results from the second pilot in section 0.

[5] Participants were shown one of two different clips: Pingu Runs Away, or Pingu Goes Fishing.
[6] Data collected by Hyunjung Ahn.

Pilot Study 2

Participants

The second pilot study[7] was offered online as a questionnaire built with Catalyst WebQ. The questionnaire was distributed online primarily via Facebook. People were informed that it was a pilot study upfront and no personal identifying information would be published[8], and that all people were welcome to participate, regardless of native language or linguistic training. 20 participants completed the survey. 17 self-identified as being native speakers of English. The 3 English L2 participants stated that they had been learning English for more than 10 years. The participants were asked "Have you had any linguistics training or taken linguistics or grammar classes?" 7 responded with "Yes, many", 3 with "Yes, some", 3 "Yes, a little", and 7 "No".

Procedure

After the brief biographical questions, the participants were given an introduction to what formulaic language is thought to be with some examples. Once they were finished with the introductory training, the subjects were given the option of continuing their training ("learning more") or starting the quiz immediately. It was hoped that there would be enough participants who chose the option of learning more that there might be some effect seen in the data if it had an effect. Unfortunately, only 2 of the 18 participants chose to read the extra training.

The rest of the questionnaire consisted of 20 questions. Each question took one string of words ('chunk') from one of the transcripts from the first study that was judged by someone (the speaker themselves, or one of the judges) to be formulaic. This chunk was first presented to the participant in its sentential context. Below the sentence, the chunk is singled out and the participant is asked if the chunk is formulaic. They were given the following choices[9]:

> **It is definitely formulaic, It is probably formulaic, It isn't formulaic, I don't know / I'm not sure, or give comments.**

All 20 chunks that comprised the questions were pulled from the conversation task transcripts from the first study. 6 tokens were taken from the male native English speaker, 7 from the female political science TA, and 7 from the female ELP student.

Analysis

Using a modified scale that was employed for the judgments of how strings fit Wray's (2008) formulaic criteria for Pilot Study 1, the results for Pilot Study 2 were calculated with the following values for each selection (0,1,2):

It is definitely formulaic = 2
It is probably formulaic = 1

[7] Due to time and other constraints this study was not constructed in a statistically reliable way. For example, it was not a true random sample nor was the pool large enough. In addition the final judgments from the first study had not yet been compiled so the selection of tokens/prompts is not as balanced as it could be. As such, the results of this study can at best only be used to hint at possibilities for further studies and that all arguments made in the discussion section should be tempered by this knowledge. Thus, no claims of statistical significance can be made.
[8] This and the official study were reviewed and cleared by the Human Subjects Division of UW: application #43774.
[9] See section 0 in the appendices for an example of the formatting for each question.

It isn't formulaic = 0
I don't know / I'm not sure = null

As such, there was a total of 20 tokens with a possible judgment total of 40 for one participant if they selected everything as being *definitely* formulaic. A participant's total judgment sum for each prompt was divided by the total number of non-null prompts. If a participant selected *I don't know* for two prompts, their mean for the entire questionnaire was calculated by dividing their total by 18 (20-2 nulls).

Each token/prompt was calculated individually as well. Each participant's judgment values were totaled and divided by the number of participants minus the nulls. This was also done at the level of each ling/grammar group. Finally, each speaker set of prompts (Native = 6, TA = 7, ELP = 7) were analyzed as well by group.

General results for the pilot studies

The average judgment value for the entire questionnaire and all participants was 0.74 per token, out of a possible 2.0. The mean of tokens answered with something other than *I don't know* was 18.65 out of the total 20.

The data from study 2 was divided into two groups: those who self-reported as having ***some*** **or more linguistics training (from now on to be referred to as: T+)** (N = 10) and those with ***little*** **or *no* training (from now on to be referred to as: T-)** (N = 10) (**Error! Reference source not found.**). All three of the English L2 participants were part of the group with more linguistics training (T+). While there were only two respondents who chose to do the extra training[10] they had a mean of .973 compared to the remaining 18 participants average of .719.

The L2 group's (N = 3) average was very similar to the rest of the T+ participants (.628). On average those in the T- group judged the prompts as being more formulaic (.890) than those in the T+ group (.599). For only 3 of the 20 tokens did the T+ group have a higher mean than the T- group, by a difference of 0.10 or more: "main actor", "take picture" (produced by ELP "Francine[11]") (Table 3) & "human beings" (produced by TA "Marcie") (Table 2).
The tables below show the results for the tokens presented in Pilot Study 2 along with the comparative judgments from Pilot Study 1. Table 2 gives the results for tokens produced by the TA "Marcie" during the conversation task. Table 3 gives the results related to the ELP participant "Francine", and

Table 4 is for the native English-speaking male.

Looking at Table 1, Marcie's tokens were judged as the least formulaic (0.51) of the three speakers on average by the participants of the online questionnaire (Pilot Study 2), while the native male's tokens were judged as being the most formulaic on average (1.14). However both of the non-native speaker's tokens were judged to be substantially less formulaic when compared with the native speaker. This is noted in order to add to the discussion of whether having an English L1 affects judgments since it seems to affect production rates. The idea of an L1 and L2 affect is examined more closely in the final study.

[10] This might be rectified in future studies by designing two separate studies with different training modules that are randomly assigned to subjects.
[11] All names are pseudonyms.

Table 1: (Pilot Study 2) Judgment means by speaker & linguistic training groups (T-/T+), and (Pilot Study 1) # of judges and judge type.

				# of judges % per token			type of judges % per token		
Speaker and # of tokens	T- mean	T+ mean	Total mean	All	Most	Few	L2	N	Both
TA Marcie – 7	0.63	0.40	0.51	14%	14%	71%	14%	57%	29%
ELP Francine – 7	0.82	0.49	0.65	0%	57%	43%	14%	14%	71%
Native male - 6	1.24	1.03	1.14	0%	67%	33%	0%	33%	66%

Turning to Table 2, the word string that received the most selections in Pilot Study 1 (of the tokens used for Pilot Study 2), "real political change" received little emphasis by the participants in Pilot Study 2. This may be due to the lack of the larger context of this speaker's background and will be discussed further later in this paper.

Table 2: Results for tokens from TA "Marcie's" conversation task

Chunk	Speaker Judged as formulaic	Trained-judges	TA Judge* (when appropriate)	Mean of all Pilot Study 2 judges	Pilot Study 2 (T- Mean - T+ Mean)	Judged by: All/Most/Few	Judged by: Only Native/Only Non-native/Both
(real*) political change	X	A* B R*S		0.50	0.20	All	Both
social changes		B		0.16	0.12	Few	Only N
dramatic relationship		A		0.25	0.10	Few	Only N

human beings[12]	X	A B S		1.21	-0.40	Most	Both
authors articulate	X			0.22	0.00	Few	Only L2
articulate emotions		A B		0.28	0.33	Few	Only N
I wouldn't say it's necessarily		A		1.00	1.20	Few	Only N

Table 3: Results for tokens from ELP "Francine's" conversation task

Chunk	Speaker Judged as formulaic	Trained-judges	TA Judge* (when appropriate)	Mean of all Pilot Study 2 judges	Pilot Study 2 (T- Mean - T+ Mean)	Judged by: All/Most/Few	Judged by: Only Native/Only Non-native/Both
before I came here	X	A R		0.60	0.20	Most	Both
because this time		R S		0.41	0.54	Few	Both
main actor		A B R		0.56	-0.10	Most	Only N
he's so nice	X	A R		0.78	0.67	Most	Both
he's very kind	X	A R		0.37	0.57	Most	Both
comes to	X			0.70	1.00	Few	Only L2
take picture (together *)		A*B		1.16	-0.54	Few	Only N

[12] A comment from a participant in the T+ group: "I would say formulaic for sure, especially since good crosslinguistic evidence points to this being the same thing worldwide. Many languages have a single word for this concept whereas we need these two words in conjunction."

Table 4: Results for tokens from the native male speaker's conversation task

Chunk	Speaker Judged as formulaic	Trained-judges	TA Judge* (when appropriate)	Mean of all Pilot Study 2 judges	Pilot Study 2 (T- Mean - T+ Mean)	Judged by: All/Most/Few	Judged by: Only Native/Only Non-native/Both
my favorite __________		R B		0.53	0.06	Few	Only N
one day		A R B S		1.26	0.29	Most	Both
got in early		A		1.22	0.63	Few	Only N
you know		A R B S	2	1.39	0.33	Most	Both
over and over and over[13]	X	A R B S	1	1.25	0.00	Most	Both
every single	X	A R B S	1	1.22	-.05	Most	Both

Table 5 and Table 7 on the other hand, shows a fairly clear pattern where the tokens that were selected by more judges in Pilot Study 1 had a higher judgment mean in Pilot Study 2. This needs further study as some of the tokens picked initially for Pilot Study 2 had not yet received all of the judgments for Pilot Study 1 and after the judgments were made and analyzed, the tokens for the native speaker had a higher number of judges selecting them on average, when compared to the two non-native subject's conversation task tokens.

In

Table 5, the values are compared based on the judge type: This looks at whether the judges that selected the token/string as being formulaic are native speakers of English (N = 3-4), non-natives (N = 2-4) or if at least one member of each group selected it (*Both*). Those that received judgments from both received the highest mean from Pilot Study 2. While there is a possibility that the L2 judges select differently from what native speakers (17/20 for Pilot Study 2) select, there is no reasonable way to conclude either way with these results, but they are provided for those readers who are curious.

[13] Comment from a participant from the T+ group: "The structure can be used with different words in between the "ands", but strangely not "under", so potentially formulaic."

Table 5: Comparison of Pilot Study 1 and Pilot Study 2 by judge type

Tokens sorted by judgments from Pilot Study 1	Mean of judgments from Pilot Study 2
Both L2 and Native	0.899
Only Native	0.645
Only L2	0.46

Table 6 on the other hand shows a fairly clear pattern where the tokens that were selected by more judges in Pilot Study 1 had a higher judgment mean in Pilot Study 2. This needs further study as some of the tokens picked initially for Pilot Study 2 had not yet received all of the judgments for Pilot Study 1 and after the judgments were made and analyzed, the tokens for the native speaker had a higher number of judges selecting them on average, when compared to the two non-native subject's conversation task tokens.

Table 6: Comparison of Pilot Study 1 and Pilot Study 2 by judge quantity

Tokens sorted by judgments from Pilot Study 1	Mean of judgments from Pilot Study 2
Judged as formulaic by most or all judges	0.914
Judged as formulaic by few judges	0.593

Discussion & Summary of the pilot studies

To begin, it should be reiterated that there is a crucial difference between Pilot Study 1 and Pilot Study 2. In Pilot Study 1, the judges selected chunks from a transcription that they believed to be formulaic, while in Pilot Study 2, the judges were asked if they thought that those selections were in fact formulaic, and how sure they were of their judgment. Devising a testing procedure that would allow the participants for Pilot Study 2 to judge raw text at such a short notice was not feasible, due to the statistical problems and likely reduction in the response rate. The method that was selected was chosen because it was arguably simpler to understand, easier to complete and required less time for analysis. So, while it wasn't a close extension of Pilot Study 1's methods, it did provide another perspective on its results and the relationship between the judges and their judgments.

While strong conclusions of any kind can not be drawn from these pilots they provided some guidance for the further study. One important factor that appears to play a role in judging formulaic language is genre and context. The TA Marcie's tokens were judged as being the least formulaic by the participants for Pilot Study 2. These tokens were present in sentential, but not paragraphical contexts. In addition, the four trained judges, as well as the speaker herself knew that she was speaking in a register that both was strongly genre-specific, as well as falling under the category K, provided by Wray (2008), which states:

> "By my judgment, this word string contains linguistic material that is too sophisticated, or not sophisticated enough, to match the speaker's general grammatical and lexical competence." (pg., 121)

This issue of genre is something that is largely absent from Wray's judgment model, but was included along diagnostic "C" in the additional training on the questionnaire that few participants read.

The results also seemed to suggest that the level of training a judge has may not be a strong predictor for what is judged as formulaic. While the judges for Pilot Study 1 were not consistently in agreement, the judgments from Pilot Study 2 show that non-trained (regardless of linguistic background) judges seem to pattern roughly with the trained judges. The fewer the number of judges that selected the token on average from Pilot Study 1 (S1) the lower the judgment mean from Pilot Study 2 (Table 1). Also, when more tokens were marked as formulaic by both native and non-native (L2) speakers (in Pilot Study 1) the means for those tokens appear to be higher for the participants in Pilot Study 2 as well. Linguistic training, which did pattern in this case with more conservative judgment (judging things as formulaic less often), did not appear to correlate with the choices of the 4 trained judges.

Summarizing, as discussed earlier, more data, an improved statistical model and a more rigorous instrument were needed in order to support any possible hypotheses. Regardless, without the benefit of substantial training or further context, the respondents for Pilot Study 2 largely patterned with the judgments made by the judges in Pilot Study 1. The more judges that selected the token, the more likely it was believed to be formulaic by the participants for Pilot Study 2, yet overall the amount of variation and potential patterns of disagreement needed a closer examination.

3 Final Study

Background

The two pilot studies, discussed above, found substantial variation in judgments, even among judges who had similar training on identifying formulaicity. This final study builds on the pilot studies and examines the consistency and reliability of individuals' judgments. In particular, what attributes of the judges might play a role or be related to bias? Also, what is the general level of agreement among judgments for certain types of chunks?

Methods

To answer these questions, an online survey was constructed comprising twenty chunks along with their sentential context. The chunks were selected by the three linguists (two PhD students and one professor) who had just completed the seminar using Wray's (2008) book discussed in the pilot studies. Each person was given excerpts from transcripts of interviews with L1 and L2 English speakers. The content of these transcripts was new to all of them. The linguists were asked to identify chunks they believed to be formulaic, simply by highlighting or underlining a string. Each of their selections were noted.

These selections were then assigned syntactic categories by a syntactician in ways that could best describe the selections with a minimal amount of categories (i.e., 3-4)[14]. Three systems for dividing the tokens were provided, which were: **syntactic structure** (e.g., *DP Incomplete, Conj TP, CP, ADJ Categorical, VP Incomplete, ADV + VP*), **by function** (e.g., *argument, non-phrase, adjunct/conjunct, predicate*), by the **category of head** (e.g., *nominal, sentential, verbal, prepositional*). Of these, 'by function' was the most amenable categorization to the proposed statistical model, and potentially a more general syntactic description.

Twenty tokens were selected using two criteria: type of syntactic structure ('by function') and the number of judges that selected them, with the former metric having more weight. In other words, the goal was to have a balanced mixture of tokens selected by all three judges, by two of the

[14] The minimal groupings was an idea suggested by the statistics consultants in order to improve the utility of the linear regressions ran on the data.

judges (also balanced between pairs of judges (i.e., judges A & B, B & C, A & C), by one of the judges (e.g., A, B, C), and a few by none of the judges that had structures underrepresented in the token set. Again, primary weight was given such that each linguist would have their judgments equally represented, as well as representing each combination and amount of agreement, with secondary weight being given to balance according to syntactic category.

An online questionnaire was constructed using Catalyst WebQ. After agreeing to take the questionnaire, participants were asked some biographical questions, including education level, field of degree, occupation, L1 and L2s, age and whether they knew anything about formulaic language.

Many of these questions were tailored specifically to English and linguistics majors. People who described themselves as studying either subject were asked to report how much they had studied of certain subfields (e.g., generative linguistics, English literature, etc.) and what their primary field of interest was.

After completing the biographical info, participants were presented with a short description of formulaic language modeled after Wray (2008). Then, they were given the option to learn more. If they selected this option, they were presented with six criteria taken from Wray (2008) which were provided as follows:

> Some criteria for formulaic language:
>
> (This criteria comes from Wray (2008), but I am interested in your opinion. If you think something is formulaic but doesn't fit one of these models, **please trust your instincts**.)
>
> **1: There is something grammatically unusual about the word string.**
> *e.g., holier than thou, if I were you*
>
> **2: Part or all of the word string lacks semantic transparency.**
> *e.g., run amok, beat around the bush, by and by*
>
> **3: This word string is associated with a specific situation and/or register (or genre).**
> *e.g., Happy birthday / Excuse me, I wonder if you would mind*
>
> **4: By my judgment, although this word string is novel, it is a clear derivation, deliberate or otherwise, of something that can be demonstrated to be formulaic in its own right.**
> *e.g., "I slept like a twig"... comes from "I slept like a log" or "Somewhere over the raincoat..."*
>
> **5: The word string contains linguistic material that is too sophisticated, or not sophisticated enough, to match the speaker's general grammatical and lexical competence.**
> *e.g., "iechyd da" in Welsh means 'good health' but the speaker may only know the phrase and not actually know how to construct it.*
>
> **6: An underlying frame for a phrase:**
> *e.g., Slept like a ________ (fish, log, etc)*

Participants were then told that they "will be presented with some sentences from spontaneous speech made by many different speakers. Please judge them as you see fit. You do not need to follow the previous training if you disagree with the model. There are no right or wrong answers. We want your best judgment and honest opinion." Each token was presented in its full original sentential context and participants were asked to rate its formulaicity using a five-point scale[15].

[15] See section **Error! Reference source not found.** in the appendices for an example of a single question.

Participants

For this study, there was a total of 211 participants. Self-selected (i.e., volunteer) respondents with a background in linguistics or English language academics were sought and make up the majority of the total group. Primary recruitment took place via various department and academically-related email lists.

Amazon Mechanical Turk and Crowdsourcing

However, after more than two years of passive, self-selected recruitment procedures, it was decided that additional control data should be collected using Amazon's Mechanical Turk (AMT). A side question this study examines is the reliability of AMT workers for linguistic instruments like this project. To help gauge the reliability of both groups, an 'are you paying attention' question was included in the token set and response time was monitored. Participants that completed the questionnaire below a certain time threshold and incorrectly answered the 'paying attention' question were omitted from the final results. Also, approximately half of the AMT participants were designated as 'Masters'[16] and the other half were normal workers.

All workers had an approval rate of 97% or higher with at least 5000 hits approved. Maintaining a fair pay scale was complicated due to the varying time it took for people to complete the questionnaire. According to the system, $2.00 equated to a minimum wage for the 20 minute estimated time it would take. Adding the Amazon fee, it became $2.60 per worker for the Human Intelligence Task (HIT).

For the first trial batch of 10 (Masters) workers, some did the survey in as little as 6.18 minutes, while the average time was 12 minutes 5 seconds, which equates to roughly $9.80 per hour. Workers that left thorough responses to all of the biographical questions were given a $0.50 bonus. While the ethics of AMT are better debated elsewhere, some sources on the web for social scientists don't clearly estimate the actual cost of paying a 'minimum wage'[17] as the people who quickly complete the HIT skew the totals and those who do a slower, and possibly more thorough job, could be underpaid if the estimated time and payment is too low. One possible option for smaller batches would be to maintain a lower payment rate and pay bonuses to those who take longer, but this method could discourage people from accepting the HIT, or encourage them to abuse the bonus system by rushing through the HIT but leaving the window open to give the system the appearance that they spent a longer time than they actually gave their attention to.

One interesting result was that the 'master' workers who did not get the 'paying attention' question correct also took the most time to complete the task. This pattern was reversed for the non-'master'/normal workers, as the faster they completed the task, the more likely they were to incorrectly respond to the 'paying attention' question. Non-English L1s were not any more likely to get the 'paying attention' question wrong.

The following two tables provide details of the participant totals.

Table 7: Breakdown of recruitment for participants

144	Self-selected
35	AMT Masters
32	AMT Normal

[16] This designation is assigned by AMT's system after participants meet certain requirements.
[17] e.g., http://www.huffingtonpost.com/julian-dobson/mechanical-turk-amazons-underclass_b_2687431.html

Table 8: Breakdown of Educational background of participants

89	None
15	English
27	Linguistics Undergrad
46	Linguistics Graduate
34	Linguistics Instructor

Results

The patterns that were hinted at in the pilot studies appear to be supported in this larger study, such that there is evidence of substantial disagreement for certain types of language chunks and that there is evidence that bias may, to some degree, be predictable based on the educational background and other factors of the judges.

To review, this project sought to understand:

- What the variability of formulaic judgments is in general.
- Whether there is any evidence of bias and what factors may correlate with or influence that bias.

This section will begin with an overview of some of the results utilizing box plots. The primary analyses for the project involve linear regressions with targeted robust regressions for comparison checks. Because the amount of data that was collected is limited in comparison with the number of variables under the lens, the results cannot scientifically confirm any hypotheses. Instead, the following coefficient inferences are descriptive. Potential statistical significance will be highlighted, but again, this does not mean it is, necessarily, significant.

In the following sections, the term ***liberal*** refers to a tendency towards judging a chunk as being formulaic, while ***conservative*** refers to a tendency to judge a chunk as not being formulaic, and ***agreement*** refers to a tendency for a group of participants to judge tokens similarly (i.e., person A = 3 & person B = 4 shows more agreement than A = 2 & person B = 5).

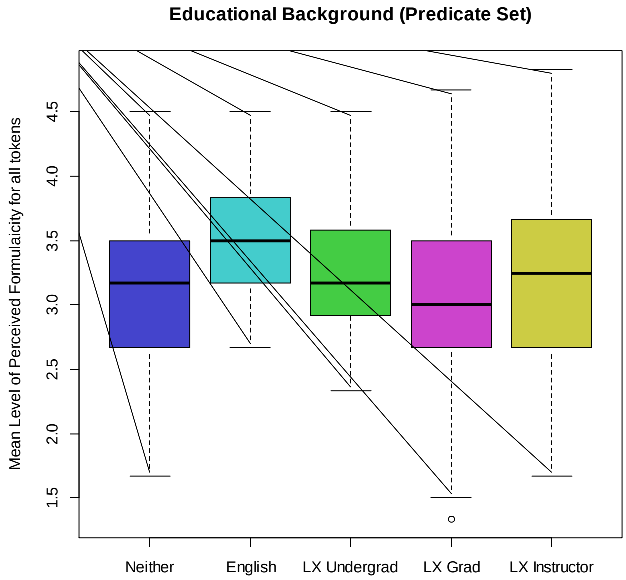
Overall Results[18]

Educational Background Boxplots

Each boxplot below is followed by the chunks, in their sentential context that belong to that particular token set. The chunks that the participants were asked to judge are in **bold**. Figure 1 below presents a box plot of the judgments made for all of the tokens belonging to the ‘predicate set’ according to the educational background of the respondents. The category, ‘Neither’, refers to a participant who did not study linguistics or English as a major or minor at the undergraduate level or beyond. In this plot, the English and linguistics undergraduate groups are the most liberal in their judgments, with linguistics graduate students being more variable and more conservative. Overall, tokens belonging to the predicate set were found to be only slightly formulaic, with the English and linguistic instructor groups being a small amount more liberal.

[18] Readers are asked to please pardon the rough formatting and organization of the following data. This UW Linguistics Working Papers version of this project was finished in order to make sure the data would be publically available in a reasonable time for interested parties. As many people had contributed to this project by participating over many years, the author felt it better to present the data, even if it was less polished than it could be.

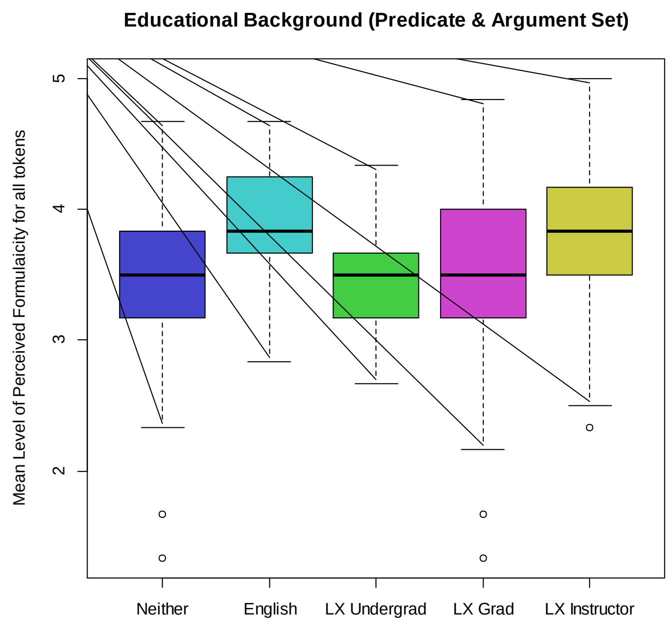
Figure 1: Educational Background (Predicate Set)



> "I never really felt uncomfortable even when I didn't understand everything they were saying, and um, I think that was **a big part of it.**"
>
> "Western communication- it's **kind of** 'attack mode' communication, and I've never been comfortable with that."
>
> "But, somehow, when I **tried to** speak Chinese I just couldn't."
>
> "It was a **good chance**. So, when I started college, I took Japanese classes for one year and then went to Japan."
>
> "And, so, you go in these bizarre translation classes- it was **really very boring** and I felt that I had just not learned enough..."
>
> "But the pamphlets all said you know if you're **on financial aid** that should not be something detracts-distracts you from trying."

The 'predicate & argument set' (Figure 2), differs little from the purely 'predicate set' (which is understandable), but received slightly more liberal judgments and, on average, shows more agreement.

Figure 2: Educational Background (Predicate & Argument Set)



"I spent a lot of times cultivating friendships with people and it wasn't a conscious effort."

"And I really love Japanese for the fact that it is a much more, um, passive language, so to speak. And I **just ate it up**- I loved that."

"Um, I think it **made a big difference** that I learned quickly and well while I was in Japan that I had had two years before I went."

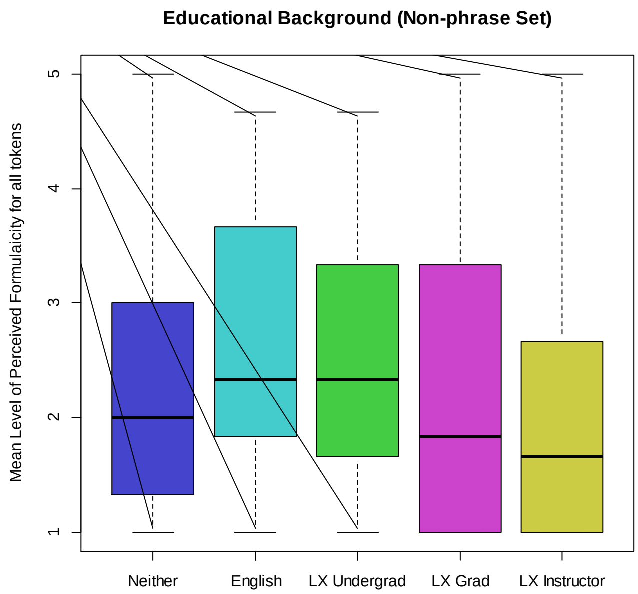
" I think that, to bring it up again that someone here **trying to learn English**, um, would have a little bit of a different situation."

"The teacher saw me and realized my interest, and **did her best** to teach me and encourage me."

"I entered the class figuring I'd try it as an experiment for about two weeks, but I **had no idea** that I'd become a Japanese major!"

The non-phrase set (Figure 3) was by far the most conservatively judged and had substantially less agreement than the other three token sets. The least agreement found in all of the token set plots, were those of the linguistics graduate student group for this set.

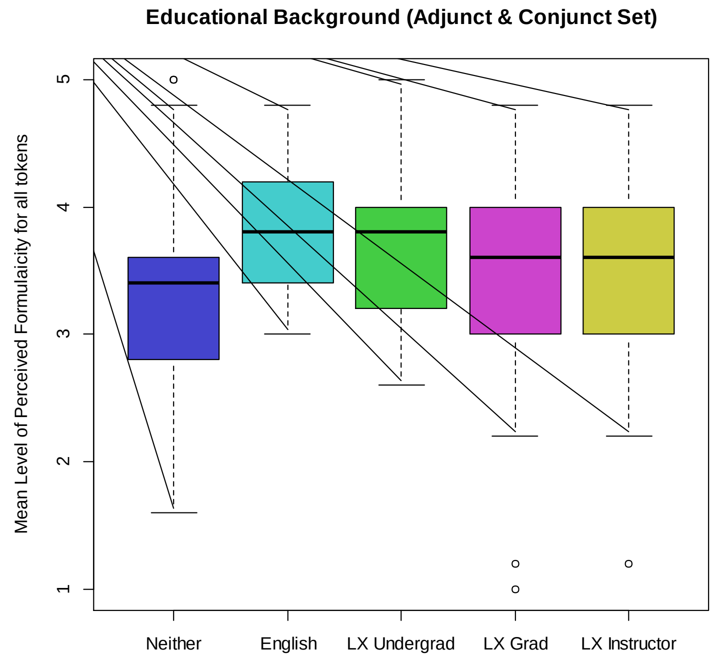
Figure 3: Educational Background (Non-phrase Set)



> "I mean **I've** recommended that to friends that are there trying to learn Japanese..."
>
> "...even though some of it may be superficial you enjoy it, you know, so **there are** cultural things that help..."
>
> "I work for a Japanese company so we do it at times but it's very American, even at a Japanese company, so **we'll** all go out at times…"

The 'adjunct & conjunct set' returns to a slightly more liberal and higher agreement judgment pattern.

Figure 4: Educational Background (Adjunct & Conjunct Set)



> "...they always ask, you know, 'What do you think I should do?' and I say, 'Don't hang out with Americans, **if you can possibly avoid it.**'"
>
> "So, **by the time** I entered college, I made as many Japanese friends as I could."
>
> "So, by making friends with just Japanese, talking just about Japanese things, eating just Japanese food- now **when I look at it** I realize that wasn't very good for them."
>
> "But the pamphlets all said you know if you're on financial aid that should not be something detracts-distracts you **from trying**."[19]
>
> "Because I'm from Los Angeles, **at first** I wanted to study not Japanese but Chinese."

In all of the four token sets plotted based on educational background, the English group was the most liberal and the 'neither' group consistently fell in the middle, which is possible evidence that they were successful in their role as controls.

Additional Boxplot Highlights

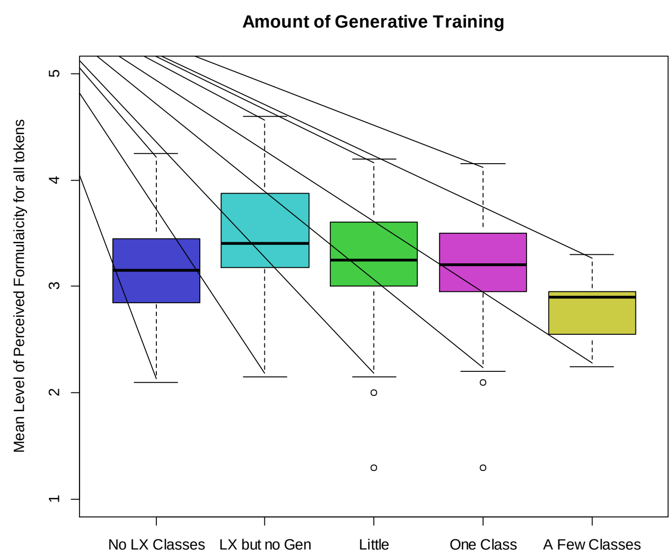
This section will discuss some of the analyses of the entire token set. While there may be patterns and potential effects based on the type of token (i.e., which set the token belongs to), these analyses

[19] Participants were asked whether **"from ____ing"**, not "trying" is formulaic with the gap, as one of the original judges described the chunk as 'from x ing' with x being a variable, or gap that is filled in to complete the chunk.

provide a different perspective. Interestingly, there was not even a slightly significant difference in responses (whether for agreement or liberality) between English L1 and non-English L1 participants. Overall, the short training which presented some of Wray's (2008) criteria shows only the slightest increase in agreement and liberality for the 69 participants who chose to read it.

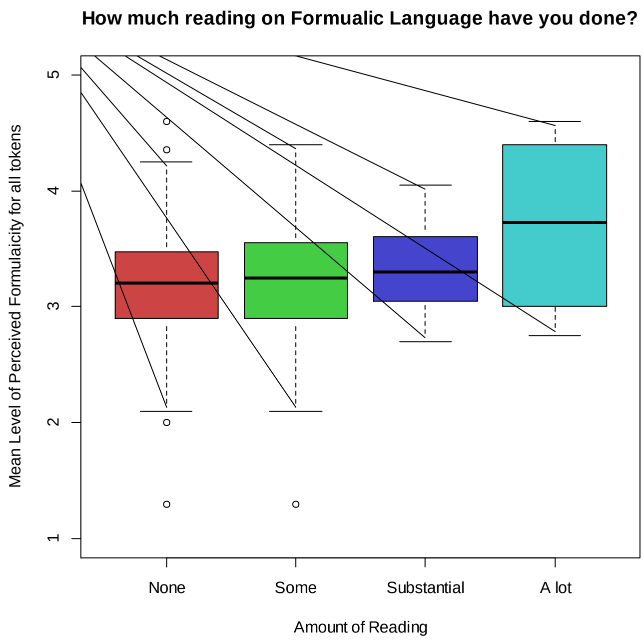
A potential source of bias that came out of the pilots was the issue of syntactic training and theoretical perspectives. One of the educational background questions that showed a fairly broad spread of exposure from participants was generative syntax training. Figure 5, below, shows a plot of responses based on the amount of exposure to generative theory, which seems to suggest that an increase in exposure patterns with more conservative judgments. Since conservative judgments would imply that the chunks were not formulaic, but instead generated, this patterns seems to follow that logic and the evidence found in the two pilots.

Figure 5: Generative Training Boxplot



The following boxplot (Figure 6) shows a slight increase towards liberal judgments patterned with an increase in reading about formulaic language. As would seem natural, the amount of respondents familiar with the topic was low, and so the N values were a bit lower than most other analyses ('Some' (N=54), 'Substantial' (N=14), and 'A lot' (N=4)). While it seems that reading about the topic might increase agreement, further study is required.

Figure 6: Amount of Reading About Formulaic Language[20]



Linear Regressions

This section will present some of the key results from the linear regressions. The formula used in R was as follows:

> **lm(formula = predicate ~ ED.Coding + reading.on.formulaicity + non.EN.lang + training + age + L1.EN + reading.on.formulaicity:training, data = token.dat)**

'Predicate' is the token group being analyzed with the six following covariates and 'reading.on.forumulaicity:training' is a check for whether there is a significant interaction effect between the amount of reading on formulaic language and training. Certain comparisons were ran on a subset of those participants belonging to either the English or one of the three linguistics educational backgrounds.[21]

Predicate Token Set

> **Residual standard error: 0.6176 on 193 degrees of freedom**
> **Multiple R-squared: 0.1328, Adjusted R-squared: 0.05644**
> **F-statistic: 1.739 on 17 and 193 DF, p-value: 0.03889**
> **KEY: Estimate Std. Error t value Pr(>|t|)**
> **ED.CodingEnglish 0.33843 0.18084 1.871 0.0628 .**

[20] Admittedly, the distinction between 'substantial' and 'a lot' is not very precise and in hindsight, these were not ideal terms. Regardless, the options were laid out similar to a common Likert scale, clearly implying that left-to-right there was in increase.

[21] Additional plots related to these linear regressions (e.g., QQ) can be found in the appendix.

English majors were the most liberal of the educational background groups with a P value that seems to be near significance.

reading.on.formulaicity1	**0.07307**	**0.11942**	**0.612**	**0.5413**
reading.on.formulaicity2	**0.20713**	**0.19744**	**1.049**	**0.2955**
reading.on.formulaicity3	**0.86830**	**0.38544**	**2.253**	**0.0254 ***

There data also seems to suggest that an increase in the liberality corresponds to an increase in reading done on the subject with a potentially significant P value for the group who had read the most. Note, again, this group had a fairly low N value and this should not be taken as strong evidence.

non.EN.langYes	**0.26652**	**0.11236**	**2.372**	**0.0187 ***

While the boxplots for the overall token set did not show a visible difference between English L1 and non-English L1 participants, the predicate token set shows that there may be a significant effect for people who know a second language (in addition to English) being more liberal in their judgments (+0.26652).

Non-Phrase Token Set

Residual standard error: 1.098 on 193 degrees of freedom
Multiple R-squared: 0.07441,   Adjusted R-squared: -0.00712
F-statistic: 0.9127 on 17 and 193 DF, p-value: 0.5599

age25-35	**-0.52434**	**0.26405**	**-1.986**	**0.0485 ***
age36-45	**-0.77468**	**0.29968**	**-2.585**	**0.0105 ***
age46-60	**-0.59431**	**0.30689**	**-1.937**	**0.0543 .**
age60+	**-0.53006**	**0.34160**	**-1.552**	**0.1224**

The non-phrase set shows potential for an effect on increased age (older than 24[22]) on conservative judgments.

trainingYes	**-0.27612**	**0.19723**	**-1.400**	**0.1631**

While not significant, the non-phrase set is the only one out of the four where taking the extra training showed an increase in conservative judgments.

[22] Note: The under 24 age group is the baseline statistic that the values of the other groups are compared to, that is why it, and similar baseline variables, are absent from other data sets.

Predicate & Argument Token Set

Residual standard error: 0.6071 on 193 degrees of freedom
Multiple R-squared: 0.1194, Adjusted R-squared: 0.04185
F-statistic: 1.54 on 17 and 193 DF, p-value: 0.08448

ED.CodingEnglish	**0.362895**	**0.175**	**2.041**	**0.0426**	*****
ED.CodingLX Undergrad	**-0.002758**	**0.148870**	**-0.019**	**0.9852**	
ED.CodingLX Grad	**-0.006278**	**0.125395**	**-0.050**	**0.9601**	
ED.CodingLX Instructor	**0.173981**	**0.153224**	**1.135**	**0.2576**	

The predicate & argument token set provides the second strongest support for the idea that there may be an pattern with English education backgrounds and increased liberality, second only to the Adjunct and Conjunct set.

reading.on.formulaicity1	**0.043294**	**0.117393**	**0.369**	**0.7127**	
reading.on.formulaicity2	**0.252018**	**0.194092**	**1.298**	**0.1957**	
reading.on.formulaicity3	**0.817340**	**0.378894**	**2.157**	**0.0322**	*****
reading.on.formulaicity1:trainingYes	**-0.053746**	**0.244900**	**-0.219**	**0.8265**	
reading.on.formulaicity2:trainingYes	**0.270771**	**0.486693**	**0.556**	**0.5786**	
reading.on.formulaicity3:trainingYes	**-1.850441**	**0.720012**	**-2.570**	**0.0109**	*****

The predicate & argument token set was the only one to see an interaction effect for reading and training on formulaicity, but it seems to be statistically significant for the group that both did most amount of reading and did the training.

Adjunct & Conjunct Token Set

Residual standard error: 0.7012 on 193 degrees of freedom
Multiple R-squared: 0.1427, Adjusted R-squared: 0.06714
F-statistic: 1.889 on 17 and 193 DF, p-value: 0.02091

ED.CodingEnglish	**0.474586**	**0.205330**	**2.311**	**0.02187**	*****
ED.CodingLX Undergrad	**0.199384**	**0.171945**	**1.160**	**0.24765**	
ED.CodingLX Grad	**0.123041**	**0.144831**	**0.850**	**0.39663**	
ED.CodingLX Instructor	**0.002849**	**0.176973**	**0.016**	**0.98717**	

In this set, English shows the strongest effect and significance.

non.EN.langYes	**0.254942**	**0.127576**	**1.998**	**0.04708**	*****

While not as significant as the predicate token set, there is also evidence here of an increase in liberality for people who know a language other than English.

age25-35	**-0.353342**	**0.168569**	**-2.096**	**0.03737**	*****
age36-45	**-0.212359**	**0.191316**	**-1.110**	**0.26838**	
age46-60	**-0.139284**	**0.195913**	**-0.711**	**0.47798**	
age60+	**-0.609960**	**0.218076**	**-2.797**	**0.00568**	******

Like the predicate & argument set, the 60+ age group is the most conservative, and here there seems to be a much more significant effect. Also, similar to the predicate set, the 46-60 group is less conservative than all of the groups older than 24 years of age.

English and Linguistics Subfield Comparisons

This section looks at possible effects of training in various subfields. Only the subset of people with English and linguistics educational backgrounds were analyzed.

Predicate Token Set (Subfield)

```
lm(formula = predicate ~ ED.Coding + Generative.LX + Func.LX +
Comp.LX + Applied.LX + Anth.LX + General.LX + EN.Lit + General.EN +
reading.on.formulaicity + syntax + semantics + phonology + sociolinguistics
+ lang.pedagogy + theoretical.lang.acq + comp.LX + EN.grammar + L1.EN
+ training, data = token.LingEN)

Residual standard error: 0.7044 on 56 degrees of freedom
Multiple R-squared:  0.4971,    Adjusted R-squared:  -0.07764
F-statistic: 0.8649 on 64 and 56 DF,  p-value: 0.7139

ED.CodingLX Undergrad    -0.8162484  0.4120697  -1.981   0.0525 .
ED.CodingLX Grad         -0.8366526  0.3965538  -2.110   0.0394 *
ED.CodingLX Instructor   -0.9646435  0.4272395  -2.258   0.0279 *
```

With this analysis, the English majors are much more, and potentially significantly, liberal than linguists, with linguists becoming increasingly conservative with more time in the field.

```
Anth.LX2                 -0.3361702  0.1958165  -1.717   0.0915 .
```

Those with minimal exposure to anthropological linguistics were more conservative. This effect was not significant, and what effect there was, tended towards liberality for those with more training.

```
theoretical.lang.acq2     0.7494021  0.3620062   2.070   0.0431 *
theoretical.lang.acq3     0.7509608  0.3750536   2.002   0.0501 .
theoretical.lang.acq4     0.9874562  0.4540986   2.175   0.0339 *
theoretical.lang.acq5     1.0880554  0.5517428   1.972   0.0536 .
```

For this token set, exposure to theoretical language acquisition studies showed the greatest potential effect, with a seemingly significant increase in liberality.

Non-Phrase Token Set (Subfield)

```
Residual standard error: 1.112 on 56 degrees of freedom
Multiple R-squared:  0.5595,    Adjusted R-squared:  0.05617
F-statistic: 1.112 on 64 and 56 DF,  p-value: 0.3443
trainingYes              -0.584202   0.288872  -2.022   0.0479 *
```

For the non-phrase set, taking the training showed a potentially significant increase in conservatism.

```
ED.CodingLX Grad         -1.290962   0.626225  -2.061   0.0439 *
ED.CodingLX Undergrad    -1.597164   0.650727  -2.454   0.0172 *
ED.CodingLX Instructor   -1.657160   0.674683  -2.456   0.0172 *
```

Once again, there is evidence that linguists are more (and increasing with time in the field) conservative than English majors.

Anth.LX2	**0.347162**	**0.309227**	**1.123**	**0.2664**
Anth.LX3	**1.749755**	**0.705033**	**2.482**	**0.0161 ***
Anth.LX4	**1.940635**	**0.816638**	**2.376**	**0.0209 ***

Contrary to the predicate set, anthropological linguistics show a significant increase in liberality. Note that this trend is much more significant than the predicate set, and it also follows an increase in exposure, while the limited exposure (LX2) in the predicate set deviated from the greater exposure (LX3) and no-exposure.

Func.LX2	**0.561386**	**0.391884**	**1.433**	**0.1576**
Func.LX3	**0.994165**	**0.509072**	**1.953**	**0.0558 .**
Func.LX4	**0.005941**	**1.007723**	**0.006**	**0.9953**
Generative.LX2	**-0.876120**	**0.418502**	**-2.093**	**0.0408 ***
Generative.LX3	**-0.731246**	**0.567652**	**-1.288**	**0.2030**
Generative.LX4	**-0.810229**	**0.931492**	**-0.870**	**0.3881**

While there is only limited significance for functionalism training, it seems to contrast the liberality of those with generative training.

sociolinguistics2	**-0.387858**	**0.689656**	**-0.562**	**0.5761**
sociolinguistics3	**-0.593569**	**0.670559**	**-0.885**	**0.3798**
sociolinguistics4	**-0.962136**	**0.694013**	**-1.386**	**0.1711**
sociolinguistics5	**-1.444183**	**0.791874**	**-1.824**	**0.0735 .**

While only having some limited (potential) significance, an increased exposure to sociolinguistics seems to pattern with an increase in conservatism.

Predicate & Argument Token Set (Subfield)

Residual standard error: 0.6344 on 56 degrees of freedom
Multiple R-squared: 0.5203, Adjusted R-squared: -0.02798
F-statistic: 0.949 on 64 and 56 DF, p-value: 0.5822

phonology2	**-0.77110**	**0.59978**	**-1.286**	**0.2039**
phonology3	**-0.73606**	**0.52597**	**-1.399**	**0.1672**
phonology4	**-1.00378**	**0.54186**	**-1.852**	**0.0692 .**
phonology5	**-1.08477**	**0.52053**	**-2.084**	**0.0417 ***

While phonology tends toward conservatism in the other sets, here it seems to be significant for the more advanced phonologists.

sociolinguistics2	**0.61737**	**0.39329**	**1.570**	**0.1221**
sociolinguistics3	**0.90237**	**0.38240**	**2.360**	**0.0218 ***
sociolinguistics4	**0.80480**	**0.39577**	**2.033**	**0.0468 ***
sociolinguistics5	**0.58283**	**0.45158**	**1.291**	**0.2021**

Sociolinguistics patterns with an increase in liberality in general. Here it seems to be significant. Only with the non-phrase set is there a tendency towards conservatism for sociolinguistic training.

lang.pedagogy2	**-0.61794**	**0.27799**	**-2.223**	**0.0303 ***

Minimal exposure to language pedagogy (e.g., TESOL) hints at an increase in conservatism. Higher levels of exposure do not show significance.

theoretical.lang.acq2	**0.75136**	**0.32600**	**2.305**	**0.0249 ***
theoretical.lang.acq3	**0.46597**	**0.33775**	**1.380**	**0.1732**
theoretical.lang.acq4	**0.68528**	**0.40893**	**1.676**	**0.0994 .**
theoretical.lang.acq5	**0.89773**	**0.49687**	**1.807**	**0.0762 .**

Like the predicate set, theoretical language acquisition here shows a potentially significant increase in liberality.

comp.LX2	**-0.40343**	**0.25209**	**-1.600**	**0.1151**
comp.LX3	**-0.47724**	**0.28301**	**-1.686**	**0.0973 .**
comp.LX4	**-0.22678**	**0.32861**	**-0.690**	**0.4930**
comp.LX5	**-0.13488**	**0.45729**	**-0.295**	**0.7691**

Except for the non-phrase set, computational linguistics training trends with more conservative judgments. This set shows the only instance where it is even remotely close to significance, however.

Adjunct or Conjunct Token Set (Subfield)

Residual standard error: 0.7336 on 56 degrees of freedom
Multiple R-squared: 0.5457, Adjusted R-squared: 0.0265
F-statistic: 1.051 on 64 and 56 DF, p-value: 0.4264

ED.CodingLX Undergrad	**-0.633840**	**0.429128**	**-1.477**	**0.14527**
ED.CodingLX Grad	**-0.771171**	**0.412970**	**-1.867**	**0.06709 .**
ED.CodingLX Instructor	**-0.940782**	**0.444926**	**-2.114**	**0.03894 ***

This follows the trend of the previous sets for the English category being more liberal than linguistics.

Func.LX4	**-1.353911**	**0.664552**	**-2.037**	**0.04635 ***

Only the highest amount of training showed significance for functionalism and increased conservatism.

sociolinguistics2	**0.987693**	**0.454800**	**2.172**	**0.03413 ***
sociolinguistics3	**1.062919**	**0.442206**	**2.404**	**0.01957 ***
sociolinguistics4	**0.821309**	**0.457673**	**1.795**	**0.07813 .**
sociolinguistics5	**0.963343**	**0.522209**	**1.845**	**0.07037 .**

Sociolinguistics shows possible effects on liberality for all levels in this set.

theoretical.lang.acq2 0.865254 0.376992 2.295 0.02549 *
theoretical.lang.acq3 1.009076 0.390580 2.584 0.01242 *
theoretical.lang.acq4 1.306035 0.472897 2.762 0.00776 **
theoretical.lang.acq5 1.164674 0.574583 2.027 0.04743 *

This set shows another potential for a significant effect on liberality for theoretical language acquisition training.

L1.ENYes 0.433600 0.249924 1.735 0.08826 .

This is the only token set with a possible, but weak, significance, for an effect of having English as an L1 patterning with increase liberality.

Primary Interest

This section examines the effect of a primary interest. As a reminder, participants were asked their amount of exposure to each subfield but were then instructed to select a single primary interest.

Primary.InterestGenerative LX -1.02619 0.48732 -2.106 0.0379 *

Generative linguistics was twice as conservative as any other primary interest for the non-phrase set and had the most significant effect. While not significant, functionalists were generally more conservative than generativists (non-phrase was the deviation from the other sets). Computational linguists were also the most liberal of all of the primary interests.

Amazon Mechanical Turk

Here, the focus will briefly return to the question of whether there is any reasonable difference between AMT master (N=35) and non-master (N = 32) workers as respondents. Note that both groups received the same payment and had to be above the same threshold of ratings and completed jobs, and the only difference in the settings was the title of ‘master’ and the qualifications that may come with it.

Predicate
AMTAMT -0.083075 0.132863 -0.625 0.5325
AMTAMT(M) -0.018048 0.130650 -0.138 0.8903

Non-phrase
AMTAMT 0.497338 0.227560 2.186 0.030 *
AMTAMT(M) -0.192797 0.223770 -0.862 0.390

Predicate & Argument
AMTAMT -0.435462 0.126292 -3.448 0.000686 ***
AMTAMT(M) -0.144493 0.124189 -1.163 0.245994

Adjunct or Conjunct
AMTAMT -0.44510 0.14631 -3.042 0.00266 **
AMTAMT(M) -0.27181 0.14387 -1.889 0.06029 .

Except for the non-phrase set, non-masters were at least twice as conservative as master workers. Also, non-masters significantly deviated from the non-AMT (i.e., self-selected) participants in three out of the four token sets. This data suggests that master workers may be better suited for similar types of linguistic and social science research.

Predicate
paying.attention1 -0.915377 0.322066 -2.842 0.00606 **

Predicate & Argument
paying.attention1 -0.9422 0.3488 -2.702 0.00889 **

As described above, a dummy question was included for the AMT participants to see if they were paying attention. The question had the same format as the others, but instead of presenting a chunk and a sentence it instructed the participants to select 3 ‘I’m not sure’ if they are paying attention. Participants who did not select the correct answer were separated from the main data, their responses were later compared to the rest of the data. While there were a few people who did not select the correct answer (N=15), none of them selected 5 ‘It **is** formulaic’, and strangely, only those that selected 1 ‘It is **not** formulaic’ showed a significant effect. The effect showed an increase in conservatism and the effect was only significant for the predicate and predicate & argument sets. Whether this shows the inherent variability in formulaic judgments or a tendency for random clickers to stick between the middle three selections in a five point scale, or another reason entirely, is not clear based on this data unfortunately.

Summary of Linear Regressions

- Multilingualism appears to pattern with an increase in liberal judgments.
- The 18-24 age group is by far the most liberal. However, as a possible explanation, they have had less time to pursue educational training in English and linguistics via higher education.
- One of the most consistently seen effects in the data is that English majors are more liberal than linguists and the conservativism of linguists increased with time spent in the field.
- Of the subfields:
    - Theoretical language acquisition showed a substantial, and potentially significant tendency to increase liberality with an increase in exposure to the subfield.
    - A background in anthropological linguistics patterned with a significant increase in liberality for some sets.
    - Sociolinguistics may pattern with a significant increase in liberality. Only with the non-phrase set did sociolinguistics show a tendency towards conservatism.
    - Generative linguists were twice as conservative as any other primary interest for the non-phrase set and that group of participants had the most significant effect.
    - While not significant, computational linguistics seemed to be the most liberal primary interest overall and functionalism was the most conservative.
    - While it is reasonable that the non-phrase set received the most conservative judgments, it also had far less agreement than the other three categories. Perhaps tokens that could belong to this set could be studied further on their own as they seem to be the more contentious chunks.

Conclusion

While the strongest interpretation of this study can only claim evidence of statistical, and not scientific, significance, it is hoped that the study will be useful for those interested in issues of language judgments and the limits of metrics and training when it comes to human biases. Crucially, this paper is not intended as a criticism of the excellent work being done by formulaic language researchers, such as Nattinger & DeCarrico (1992) and Wray (2008), the latter whose work in particular contributed greatly to this project. It is clear that the diagnostics and metrics that have been devised can be very useful. Instead, one takeaway from this project is that the seemingly reasonable strength of the biases and the general amount of disagreement when it comes to judgments like these simply highlight a need for greater caution and clarity.

For future studies, it may be worth investigating other potential factors, such as sentential vs. paragraph vs. biographical contexts. In addition, other formulaic language types may also uncover interesting evidence (e.g., idioms, semantic opacity, prep + verb chunks, etc.), as well as the further exploration of 'non-phrase' chunks.

4 References

Arnon, Inbal, & Snider, Neal (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory & Language*, *62*(1).

Biber, Douglas (1999). *Longman grammar of spoken and written English*. Harlow, England; [New York]: Longman.

Conklin, Kathy, & Schmitt, Norbert (2008). Formulaic Sequences: Are They Processed More Quickly than Nonformulaic Language by Native and Nonnative Speakers? *Applied Linguistics, 29*(1).

Ellis, Nick C., Simpson-Vlach, Rita, & Maynard, Carson (2008). Formulaic Language in Native and Second Language Speakers: Psycholinguistics, Corpus Linguistics, and TESOL. *TESOL Quarterly: A Journal for Teachers of English to Speakers of Other Languages and of Standard English as a Second Dialect, 42*(3), 375-396.

Erman, Britt, & Warren, Beatrice (2000). The idiom principle and the open choice principle. *Text., 20*(1), 29.

Jiang, Nan, & Nekrasova, Tatiana M. (2007). The processing of formulaic sequences by second language speakers. *Modern Language Journal, 91*(3), 433-445.

Millar, Neil (2011). The Processing of Malformed Formulaic Language. *Applied Linguistics, 32*(2), 129-148.

Nattinger, James R., & DeCarrico, Jeanette S. (1992). *Lexical phrases and language teaching*. Oxford [England]; New York: Oxford University Press.

Nekrasova, Tatiana M. (2009). English L1 and L2 Speakers' Knowledge of Lexical Bundles. *Language Learning, 59*(3).

Pawley, A., & Syder, F. H. (1983). *Two puzzles for linguistic theory: native-like selection and native-like fluency*.

Siyanova-Chanturia, Anna, Conklin, Kathy, & Schmitt, Norbert (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research, 27*(2), 251-272.

Skehan, Peter, & Foster, Pauline (1997). Task Type and Task Processing Conditions as Influences on Foreign Language Performance. *LTR Language Teaching Research, 1*(3), 185-211.

Wray, Alison (2008). *Formulaic language : pushing the boundaries*. Oxford; New York: Oxford University Press.

5 Appendices

Example question from pilot study 2

"I mean, how those **authors articulate** those emotions and the ideas of those people living in that period was really impressive."

Do you think "***authors articulate***" in this sentence is formulaic?

- ○ It is definitely formulaic
- ○ It is probably formulaic
- ○ It isn't formulaic
- ◉ I don't know / I'm not sure
- ○ Comments:

Example question from the main study

"I never really felt uncomfortable even when I didn't understand everything they were saying, and um, I think that was a big part of it."

Do you think "a big part of it" is formulaic?

1	2	3	4	5
It is **not** formulaic	It is **probably not** formulaic	I'm not sure	It **is probably** formulaic	It **is** formulaic

Responses to the question: "In a sentence or two, please give your definition of formulaic language."

(These quotes are presented as submitted by the respondents. No corrections have been made.)

- Formulaic language is indefinable.
- Formulaic language refers to commonly-used fixed expressions such as idioms.
- Follows a certain word order
- Akin to "lexical bundles" or the lexical approach to ESL/EFL, it views language as used largely in chunks larger than individual words.
- Language that has a ready-made shape and is used even though some other shape could, in principle, express the same or a similar meaning.
- Formulaic langauge covers relatively high frequency 'set' expressions which may or may not be idiomatic (semantically transparent or opaque), which may or may not include open slots (e.g. to X one's way to Z), and in some formulations may cover highly frequent sequences which do not constitute syntactic constituents (e.g. "in the middle of".
- language in which there is a "default" or expected phrase or term to express the event/feeling/moment/circumstance
- Chunks of language (part of a phrase, phrase or whole sentence) that allow no or little variation and which express an idiomatic, non-literal meaning. I assume that formulaic language is processed as a chunk, rather than an analyzed grammatical structure.
- i'd go with wray's definition, i.e. a sequence, continuous or discontiuous that is stored and retrieved as a whole, ...

- Fixed expressions of three or more words
- a conventional expression which may or may not be an idiom (that usually has some semantic opacity) and which usually has some degree of formal rigidity (lexico-grammatical fixedness).
- Phrases or sentences that are structured in a consistent way among speakers of a given language, including idiomatic use and generic features of texts
- Language chunks
- Formulaic language consists of fixed expressions (with transparent meaning, so no idioms) that are used in a specific context according to a relatively fixed and describable "script". There may, of course, be "gaps" in both the individual expressions ("That'll be $___, please.") and in the scripts.
- To me, formulaic language refers to prefabricated chunks of language, rady-made chunks of language.
- frozen expressions used in poetry to fill metrically difficult positions
- Strings of language that are learned and produced as "chunks" like individual lexical units. They are longer than these, but do not always conform to grammar.
- multi-word units that constitue one semantic unit
- A string or plug-and-go linguistic structure stored in the brain directly (usually b/c of frequency effect), rather than being constructed with each use.
- A chunk of linguistic material that gets processed as a single non-compositional unit
- chunks/collocations/frozen phrases
- The use of fixed or repetetive phrases in a particular context.
- Language that belongs to a certain genre and is used to signal certain things. I am most familiar with formulaic language in Vedic and Sanskrit, though.
- Use of set phrases to express certain ideas under certain conditions.
- Ready-made expressions, not necessarily constituting grammatically complete sentences.
- Formulaic language can be seen as a set of words which are usually found together in the same order and in the same context.
- frequently recurrent lexico-syntactic strings which can have different degrees of semantic opacity and structural fixedness
- Formualic language refers to items of a language (usually words) that often co-occur or collocate. They are some sort of "prefabricated language chunks" which facilitate communication (at least between native speakers) because they constitute preexisting patterns that the speakers are used to and expect when communicating certain meanings/functions so that the speakers can focus on the (new) meaning/content rather than having to constantly invent/decode new forms for certain functions.
- The use of (semi-fixed) multi-word units in communication
- groups of lexical items that occur together very frequently
- Phrases that are retrieved from the lexicon as a "set phrase", rather than being generated LI by LI.
- Formulaic language is when two or more lexemes are co-entrenched in the internal lexico-grammars of cross-section of members of a language community.
- Sequences of words that frequently co-occur and/or have high cloze probability, and which have varying degrees of idiomaticity.
- ready-made chunks of language
- a set of prefabricated words or phrases found in proverbs, idioms, fixed events such greetings, etc.

- Syntactic segments probably stored as single lexical items and used in response to frequent/predictable situations.
- Phrases that one have to learn as a whole.
- A tendency towards pairing or grouping certain words together. Once the beginning of a phrase is used, speakers tend towards use of the related words, or may be promted to use a certain kind of response.
- Formulaic language often consists of patterns and "formulas" which we use as a basis for use in certain social, political, cultural, or religious contexts.
- Phrases which are habitual, where making a single word change is possible without losing the sense of fomulaic, but in which in general the entire phrase is a single element of speech/writing
- Formulaic language is organized to include oft heard terms, expressions, or idioms which may not be generated under the laws of optimality theory.
- culturally-specific language use
- predictable, repeated language such as in greetings/small talk
- Strings or constructions that speakers store & reuse. Some are non-compositional, others could have been created from the rules of the language but are frequent enough to be stored anyway.
- Formulaic language is language that is non-compositional, i.e. it's chunked as a phrase rather than a word. An example might be the greeting "How's it going?". The speaker is not usually actually interested in how it's going, they are using the phrase as a equivalent to "Hey".
- A sequence of linguistic elements that are produced and/or interpreted as a unit without being constructed or deconstructed at the time of production or perception.
- Using a group of words together out of habit, even when there are other ways to say the same thing.
- a conventional, frequently repeated phrase used in a specific circumstance to indicate a specific feeling/idea/etc.
- Formulaic language is a language with a correct grammatical structure.
- Words or phrases put together to tell about something and is put forth from memory and with fewer words rather than trying to sound more fancy with bigger words.
- different types of language that have structures with strict rules and definitions.
- Patterns in phrases and languages.
- Fixed or prefabricated combinations of words that make language easier to understand.
- ITS DIFFICULT TO LEARN
- which is very good english
- Fixed combination of words
- forms a structure
- A segment of language made up of several morphemes or words which are learned together and used as if they were a single item
- Formulaic language describes a unit of language actually composed of multiple words that are commonly used as if they were a single unit.
- A perfect language without any error
- It is a group of words which is stored in the mind of the speaker and is used like a formula in certain context
- It may be considered a lexical unit that is more than one word long.
- A generic/standard figure of speech that is commonly used
- Fixed combinations of words that could work together.
- combination of words that make the sentence more fluid.

- Formulaic Language is a segment of language made up of several morphemes or words which are learned together and used as a single item.

Responses to the question: "If the training has changed or confirmed your understanding of formulaic language, please write a brief explanation how below."

(This question was asked after people took the optional training where they were presented with some of Wray's (2008) criteria. These quotes are presented as submitted by the respondents. No corrections have been made.)

- I would, before seeing your training, have included as "formulaic language" strings that are perfectly well generable by the same rules as the rest of the language is, but which are used in preference to other possibly generable strings to express a particular meaning or as something that one says in a particular situation.
- confirmed except that it lacks n-grams for semantically transparent lexical bundles e.g. "in the middle of___"
- It confirmed for me that we are looking for a "formula". One doesn't necessarily understand the formula but one understands how to correctly apply it.
- It convinced me that there are many varieties of formulaic language and treating all of them in a unitary fashion may be neither possible nor desirable.
- I had a vague idea about it, then it became a bit firm.
- Provided more examples that helped confirm my understanding
- it hasn't
- The idea that we don't build formulaic language word-by-word but have the phrases memorized as a chunk is interesting and clarifying.
- Good examples were shown.
- I can see you're not just interested in idioms, which is good
- It has confirmed it but it also presents a wider notion of what I would consider formulaic (even though I am awware that the boundaries are always fuzzy here).
- Confirmed my understanding
- The basics, but it seems that formulaic language can be practially anything.
- Hadn't thought about formulaic language including a unit in which substitutions might be made. Source of some humour in lnguage?
- I think I get it
- I believe it has some similarities with the concept of specific semantical unit in specialized languages.
- examples very helpful
- I was unfamiliar with this concept, but I think I understand enough to do the quiz.
- confirmed
- Is it a study of local "isms"? In my house we'll say we 'slept like a baby rock' if we'd had a great nights sleep.
- I dont fully understand the purpose of the subject but i get the basic understanding now.
- Now makes sense the things I often say that I get chipped for saying by my ex-english teacher partner... I can now explain they're formulaic language so leave me alone!
- I was unaware of the term Formulaic Launguage, but after reading the examples it makes sense.
- I feel as though I am not quite grasping formulaic languauge. Did I just use formulaic language?
- Mostly confirmed, as I'm not familiar with the literature on it.
- It clarified it a bit, to extend beyond simple idioms.

- the training confirmed my intuition that formulaic language is something chunk-like
- The training gave me a general idea about what kinds of phrases are considered to be formulaic language
- The training confirmed what had been presented on the previous page and made it slightly more explicit.
- recognized phrases
- Still not perfectly clear but interested in seeing how I do.
- Its a word string that can be used with different wods substituting for a word but it still comes from the known phrase, such as I slept like a twig coming from I slept like a log.
- Formulaic language is phrases that we memporze where we go not indeprendently structure each word
- A strng of words that is well-travelled.
- It is not how individual words make up a sentence, rather it's how a chunk of words formulate it.
- a segment of language made up of several morphemes or words which are learned together and used as if they were a single item
- I'm going to give it a try
- made it feel a little more clear
- not sure
- the extra examples helped
- that is more than one word long.
- There is no single satisfactory definition of formulaic language, and researchers differ in what they consider formulaic.
- I believe my understanding of formulaic language has been confirmed through reading the additional information.
- Before this I didn't know what formulaic language was, now I think I might understand.
- I have a slightly stronger understanding now because examples
- I didn't really know the term but it makes sense now
- I think the additional examples have given me a better understanding of formulaic language.
- i think it describes what i previously thought of as "stock phrases"

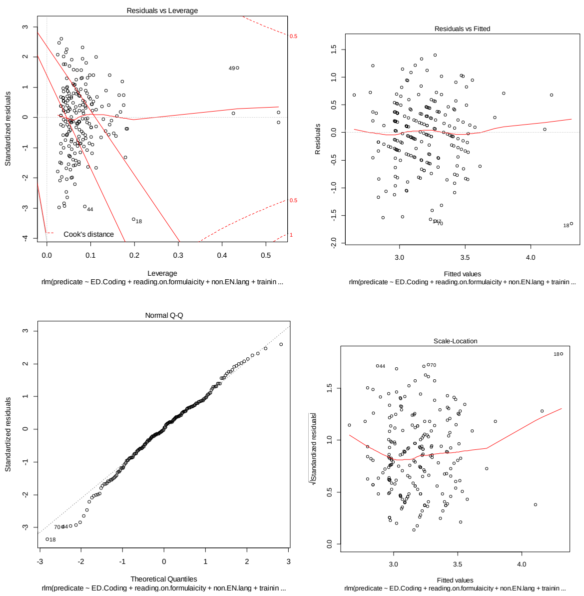
Statistics

General Comparisons (Robust Regressions)

Predicate

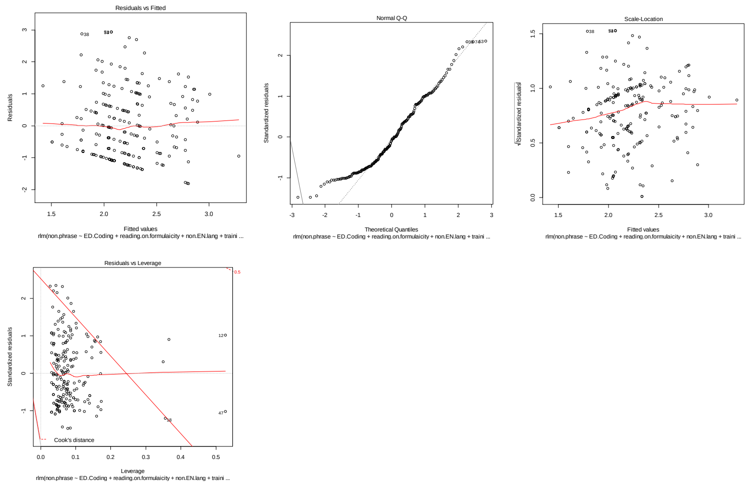
```
rlm(formula = predicate ~ ED.Coding + reading.on.formulaicity +
non.EN.lang + training + age + L1.EN + reading.on.formulaicity:training,
data = token.dat)

Residuals:
     Min       1Q    Median       3Q      Max
-1.642681 -0.370928  0.009718  0.368352  1.396927

Coefficients:
                                  Value   Std. Error
(Intercept)                        3.3304  0.2039
ED.CodingEnglish                   0.3147  0.1842
ED.CodingLX Undergrad              0.0319  0.1543
ED.CodingLX Grad                  -0.1415  0.1300
ED.CodingLX Instructor            -0.1831  0.1588
reading.on.formulaicity1           0.0899  0.1217
reading.on.formulaicity2           0.1562  0.2011
reading.on.formulaicity3           1.1388  0.3927
non.EN.langYes                     0.2437  0.1145
trainingYes                        0.0176  0.1130
age25-35                          -0.1795  0.1513
age36-45                          -0.1854  0.1717
age46-60                           0.0232  0.1758
age60+                            -0.3684  0.1957
L1.ENYes                          -0.1861  0.1526
reading.on.formulaicity1:trainingYes -0.0561  0.2538
reading.on.formulaicity2:trainingYes  0.2602  0.5044
reading.on.formulaicity3:trainingYes -1.0513  0.7462
                                  t value
(Intercept)                       16.3358
ED.CodingEnglish                   1.7083
ED.CodingLX Undergrad              0.2069
ED.CodingLX Grad                  -1.0889
ED.CodingLX Instructor            -1.1529
reading.on.formulaicity1           0.7389
reading.on.formulaicity2           0.7767
reading.on.formulaicity3           2.9002
non.EN.langYes                     2.1292
trainingYes                        0.1562
age25-35                          -1.1868
age36-45                          -1.0798
age46-60                           0.1318
age60+                            -1.8830
L1.ENYes                          -1.2192
reading.on.formulaicity1:trainingYes -0.2209
reading.on.formulaicity2:trainingYes  0.5158
reading.on.formulaicity3:trainingYes -1.4089
```

Residual standard error: 0.5463 on 193 degrees of freedom



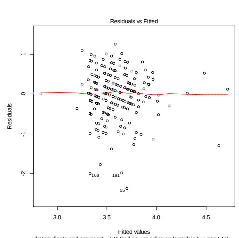
Non-Phrase

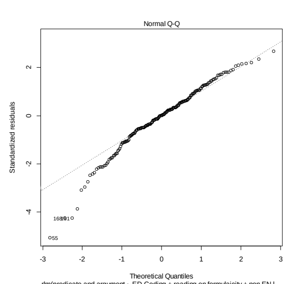
```
rlm(non.phrase ~ ED.Coding + reading.on.formulaicity +  non.EN.lang +
training + age + L1.EN + reading.on.formulaicity:training, data = token.dat)

Residuals:
    Min       1Q   Median       3Q      Max
-1.80463 -0.86915 -0.06768  0.75876  2.93232

Coefficients:
                                 Value   Std. Error
(Intercept)                       2.5817  0.3577
ED.CodingEnglish                  0.4596  0.3232
ED.CodingLX Undergrad            -0.1332  0.2707
ED.CodingLX Grad                  0.0309  0.2280
ED.CodingLX Instructor           -0.0565  0.2786
reading.on.formulaicity1         -0.0844  0.2134
reading.on.formulaicity2          0.2200  0.3529
reading.on.formulaicity3          0.7359  0.6889
non.EN.langYes                    0.0176  0.2008
trainingYes                      -0.2687  0.1982
age25-35                         -0.5733  0.2654
age36-45                         -0.8389  0.3012
age46-60                         -0.6959  0.3084
age60+                           -0.6292  0.3433
L1.ENYes                          0.3073  0.2678
reading.on.formulaicity1:trainingYes  0.4234  0.4453
reading.on.formulaicity2:trainingYes  0.2617  0.8849
reading.on.formulaicity3:trainingYes  0.0453  1.3091
                                 t value
(Intercept)                       7.2181
ED.CodingEnglish                  1.4219
ED.CodingLX Undergrad            -0.4922
ED.CodingLX Grad                  0.1357
ED.CodingLX Instructor           -0.2028
reading.on.formulaicity1         -0.3952
reading.on.formulaicity2          0.6235
reading.on.formulaicity3          1.0682
non.EN.langYes                    0.0874
trainingYes                      -1.3556
age25-35                         -2.1606
age36-45                         -2.7854
age46-60                         -2.2565
age60+                           -1.8329
L1.ENYes                          1.1476
reading.on.formulaicity1:trainingYes  0.9508
reading.on.formulaicity2:trainingYes  0.2958
reading.on.formulaicity3:trainingYes  0.0346

Residual standard error: 1.276 on 193 degrees of freedom
```

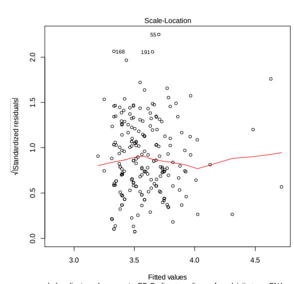

Residuals vs Fitted
Residuals
Fitted values
rlm(non.phrase ~ ED.Coding + reading.on.formulaicity + non.EN.lang + traini ...
Normal Q-Q
Standardized residuals
Theoretical Quantiles
rlm(non.phrase ~ ED.Coding + reading.on.formulaicity + non.EN.lang + traini ...
Scale-Location
Fitted values
rlm(non.phrase ~ ED.Coding + reading.on.formulaicity + non.EN.lang + traini ...
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage
rlm(non.phrase ~ ED.Coding + reading.on.formulaicity + non.EN.lang + traini ...

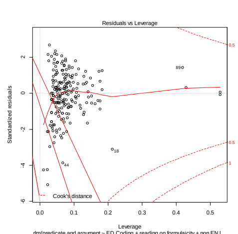
Predicate and Argument

```
rlm(predicate.and.argument ~ ED.Coding + reading.on.formulaicity +
non.EN.lang + training + age + L1.EN + reading.on.formulaicity:training,
data = token.dat)


Residuals:
    Min       1Q   Median       3Q      Max
-2.36592 -0.30669  0.00488  0.30454  1.24880

Coefficients:
                                      Value   Std. Error t value
(Intercept)                           3.4690  0.1915    18.1162
ED.CodingEnglish                      0.3816  0.1730     2.2053
ED.CodingLX Undergrad                -0.0149  0.1449    -0.1031
ED.CodingLX Grad                      0.0012  0.1221     0.0095
ED.CodingLX Instructor                0.1695  0.1491     1.1364
reading.on.formulaicity1              0.0655  0.1143     0.5734
reading.on.formulaicity2              0.2123  0.1889     1.1235
reading.on.formulaicity3              0.9640  0.3688     2.6138
non.EN.langYes                        0.0684  0.1075     0.6359
trainingYes                           0.1633  0.1061     1.5394
age25-35                             -0.1208  0.1421    -0.8502
age36-45                              0.0658  0.1612     0.4080
age46-60                              0.0275  0.1651     0.1664
age60+                               -0.2706  0.1838    -1.4725
L1.ENYes                             -0.0197  0.1434    -0.1377
reading.on.formulaicity1:trainingYes  0.0032  0.2384     0.0135
reading.on.formulaicity2:trainingYes  0.2307  0.4737     0.4870
reading.on.formulaicity3:trainingYes -2.0085  0.7008    -2.8659

Residual standard error: 0.472 on 193 degrees of freedom
```



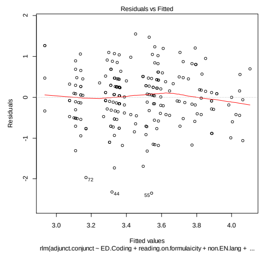


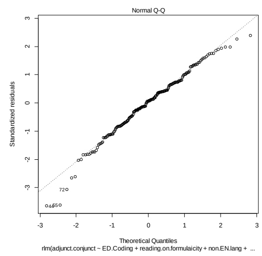


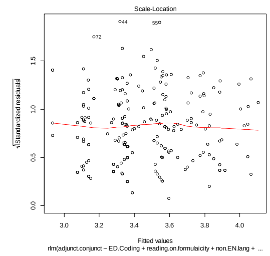


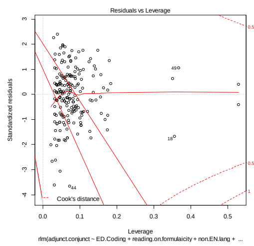
Adjunct and Conjunct

```
rlm(adjunct.conjunct ~ ED.Coding + reading.on.formulaicity +  non.EN.lang
+ training + age + L1.EN + reading.on.formulaicity:training, data =
token.dat)

Call: rlm(formula = adjunct.conjunct ~ ED.Coding +
reading.on.formulaicity +
   non.EN.lang + training + age + L1.EN + reading.on.formulaicity:training,
   data = token.dat)
Residuals:
     Min       1Q   Median       3Q      Max
-2.34412 -0.44518  0.04096  0.43402  1.54678

Coefficients:
                                  Value   Std. Error t value
(Intercept)                        3.4635  0.2289    15.1300
ED.CodingEnglish                   0.4657  0.2069     2.2511
ED.CodingLX Undergrad              0.2130  0.1732     1.2298
ED.CodingLX Grad                   0.1940  0.1459     1.3292
ED.CodingLX Instructor             0.0173  0.1783     0.0969
reading.on.formulaicity1          -0.0347  0.1366    -0.2541
reading.on.formulaicity2          -0.0176  0.2259    -0.0780
reading.on.formulaicity3           0.4992  0.4409     1.1322
non.EN.langYes                     0.2485  0.1285     1.9334
trainingYes                        0.0293  0.1269     0.2311
age25-35                          -0.3129  0.1698    -1.8426
age36-45                          -0.1427  0.1927    -0.7402
age46-60                          -0.0933  0.1974    -0.4727
age60+                            -0.5193  0.2197    -2.3634
L1.ENYes                          -0.0396  0.1714    -0.2312
reading.on.formulaicity1:trainingYes  0.1841  0.2850     0.6462
reading.on.formulaicity2:trainingYes  0.6340  0.5663     1.1195
reading.on.formulaicity3:trainingYes -0.1249  0.8378    -0.1491

Residual standard error: 0.6594 on 193 degrees of freedom
```


Residuals vs Fitted
Residuals
Fitted values
rlm(adjunct.conjunct ~ ED.Coding + reading.on.formulaicity + non.EN.lang + ...


Normal Q-Q
Standardized residuals
Theoretical Quantiles
rlm(adjunct.conjunct ~ ED.Coding + reading.on.formulaicity + non.EN.lang + ...

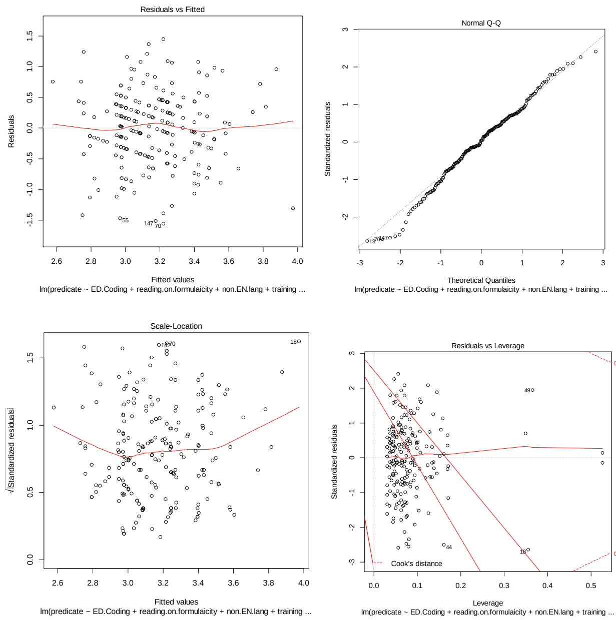

Scale-Location
√|Standardized residuals|
Fitted values
rlm(adjunct.conjunct ~ ED.Coding + reading.on.formulaicity + non.EN.lang + ...


Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage
rlm(adjunct.conjunct ~ ED.Coding + reading.on.formulaicity + non.EN.lang + ...

General Comparisons (Linear Regressions)

Predicate

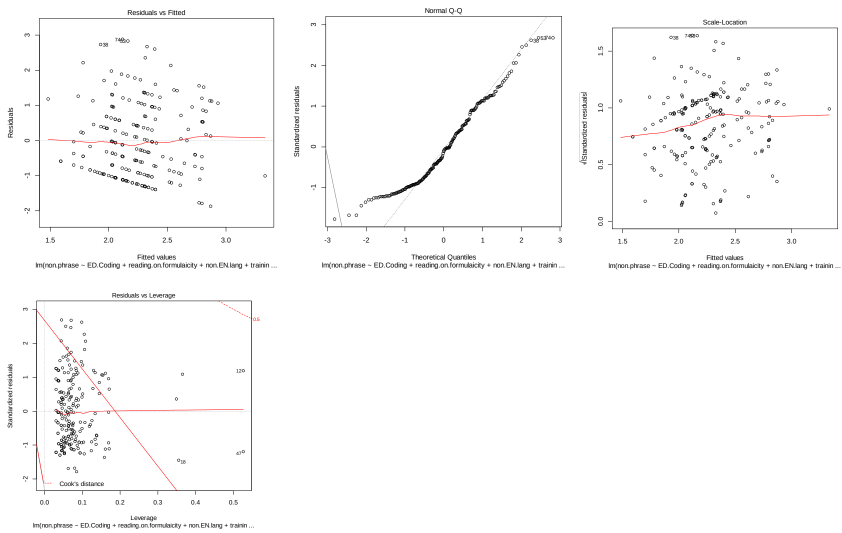
```
lm(formula = predicate ~ ED.Coding + reading.on.formulaicity +
   non.EN.lang + training + age + L1.EN, data = token.dat)

Residuals:
     Min       1Q   Median       3Q      Max
-1.55364 -0.35188  0.02257  0.37513  1.44636

Coefficients:
                                      Estimate Std. Error t value Pr(>|t|)
(Intercept)                            3.27902    0.20012  16.385   <2e-16 ***
ED.CodingEnglish                       0.33843    0.18084   1.871   0.0628 .
ED.CodingLX Undergrad                  0.03044    0.15144   0.201   0.8409
ED.CodingLX Grad                      -0.18136    0.12756  -1.422   0.1567
ED.CodingLX Instructor                -0.20577    0.15587  -1.320   0.1883
reading.on.formulaicity1               0.07307    0.11942   0.612   0.5413
reading.on.formulaicity2               0.20713    0.19744   1.049   0.2955
reading.on.formulaicity3               0.86830    0.38544   2.253   0.0254 *
non.EN.langYes                         0.26652    0.11236   2.372   0.0187 *
trainingYes                            0.02477    0.11089   0.223   0.8235
age25-35                              -0.12556    0.14847  -0.846   0.3988
age36-45                              -0.15507    0.16850  -0.920   0.3586
age46-60                               0.03216    0.17255   0.186   0.8523
age60+                                -0.36999    0.19207  -1.926   0.0555 .
L1.ENYes                              -0.17603    0.14982  -1.175   0.2415
reading.on.formulaicity1:trainingYes -0.07630    0.24913  -0.306   0.7597
reading.on.formulaicity2:trainingYes  0.24547    0.49510   0.496   0.6206
reading.on.formulaicity3:trainingYes -0.75563    0.73244  -1.032   0.3035
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6176 on 193 degrees of freedom
Multiple R-squared:  0.1328,    Adjusted R-squared:  0.05644
F-statistic: 1.739 on 17 and 193 DF,  p-value: 0.03889
```



Non Phrase

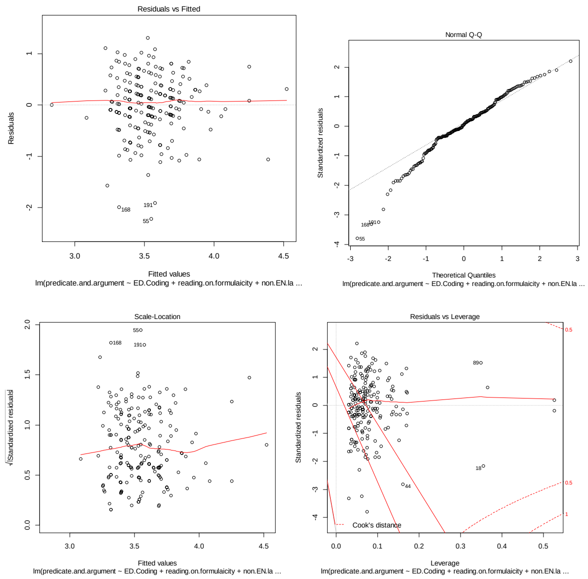
```
lm(formula = non.phrase ~ ED.Coding + reading.on.formulaicity +
   non.EN.lang + training + age + L1.EN, data = token.dat)

Residuals:
   Min      1Q  Median     3Q    Max
-1.8690 -0.9276 -0.1158  0.7768  2.8829

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                       2.60991    0.35592   7.333 6.05e-12 ***
ED.CodingEnglish                  0.50154    0.32164   1.559   0.1206
ED.CodingLX Undergrad            -0.09766    0.26934  -0.363   0.7173
ED.CodingLX Grad                  0.03021    0.22687   0.133   0.8942
ED.CodingLX Instructor           -0.07682    0.27722  -0.277   0.7820
reading.on.formulaicity1         -0.05885    0.21239  -0.277   0.7820
reading.on.formulaicity2          0.16664    0.35116   0.475   0.6357
reading.on.formulaicity3          0.67074    0.68551   0.978   0.3291
non.EN.langYes                    0.06102    0.19984   0.305   0.7604
trainingYes                      -0.27612    0.19723  -1.400   0.1631
age25-35                         -0.52434    0.26405  -1.986   0.0485 *
age36-45                         -0.77468    0.29968  -2.585   0.0105 *
age46-60                         -0.59431    0.30689  -1.937   0.0543 .
age60+                           -0.53006    0.34160  -1.552   0.1224
L1.ENYes                          0.22085    0.26646   0.829   0.4082
reading.on.formulaicity1:trainingYes  0.34467    0.44308   0.778   0.4376
reading.on.formulaicity2:trainingYes  0.25451    0.88054   0.289   0.7729
reading.on.formulaicity3:trainingYes  0.05141    1.30267   0.039   0.9686
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Residual standard error: 1.098 on 193 degrees of freedom
Multiple R-squared: 0.07441, Adjusted R-squared: -0.00712
F-statistic: 0.9127 on 17 and 193 DF, p-value: 0.5599



Predicate and Argument

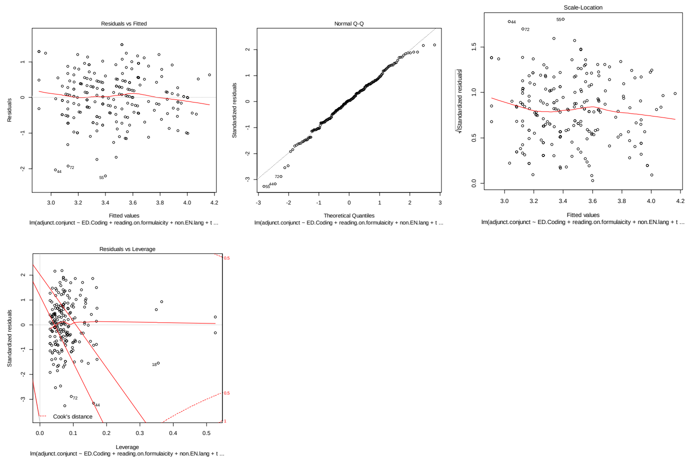
```
lm(formula = predicate.and.argument ~ ED.Coding +
reading.on.formulaicity +
   non.EN.lang + training + age + L1.EN, data = token.dat)

Residuals:
    Min       1Q   Median       3Q      Max
-2.21593 -0.22716  0.02724  0.35602  1.30761

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                      3.406117   0.196724  17.314   <2e-16 ***
ED.CodingEnglish                 0.362895   0.177775   2.041   0.0426 *
ED.CodingLX Undergrad           -0.002758   0.148870  -0.019   0.9852
ED.CodingLX Grad                -0.006278   0.125395  -0.050   0.9601
ED.CodingLX Instructor           0.173981   0.153224   1.135   0.2576
reading.on.formulaicity1         0.043294   0.117393   0.369   0.7127
reading.on.formulaicity2         0.252018   0.194092   1.298   0.1957
reading.on.formulaicity3         0.817340   0.378894   2.157   0.0322 *
non.EN.langYes                   0.078155   0.110455   0.708   0.4801
trainingYes                      0.158576   0.109012   1.455   0.1474
age25-35                        -0.144257   0.145947  -0.988   0.3242
age36-45                        -0.009151   0.165641  -0.055   0.9560
age46-60                        -0.007273   0.169622  -0.043   0.9658
age60+                          -0.315478   0.188811  -1.671   0.0964 .
L1.ENYes                         0.056880   0.147275   0.386   0.6998
reading.on.formulaicity1:trainingYes -0.053746   0.244900  -0.219   0.8265
reading.on.formulaicity2:trainingYes  0.270771   0.486693   0.556   0.5786
reading.on.formulaicity3:trainingYes -1.850441   0.720012  -2.570   0.0109 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6071 on 193 degrees of freedom
Multiple R-squared:  0.1194,    Adjusted R-squared:  0.04185
F-statistic:  1.54 on 17 and 193 DF,  p-value: 0.08448
```



Adjunct or Conjunct

```
lm(formula = adjunct.conjunct ~ ED.Coding + reading.on.formulaicity +
   non.EN.lang + training + age + L1.EN, data = token.dat)

Residuals:
   Min      1Q  Median      3Q     Max
-2.1993 -0.4379  0.0256  0.4530  1.4808

Coefficients:
                                      Estimate Std. Error t value Pr(>|t|)
(Intercept)                           3.445147   0.227216  15.162  < 2e-16 ***
ED.CodingEnglish                      0.474586   0.205330   2.311  0.02187 *
ED.CodingLX Undergrad                 0.199384   0.171945   1.160  0.24765
ED.CodingLX Grad                      0.123041   0.144831   0.850  0.39663
ED.CodingLX Instructor                0.002849   0.176973   0.016  0.98717
reading.on.formulaicity1             -0.041731   0.135588  -0.308  0.75858
reading.on.formulaicity2              0.041505   0.224176   0.185  0.85331
reading.on.formulaicity3              0.558191   0.437623   1.276  0.20366
non.EN.langYes                        0.254942   0.127576   1.998  0.04708 *
trainingYes                           0.043451   0.125909   0.345  0.73039
age25-35                             -0.353342   0.168569  -2.096  0.03737 *
age36-45                             -0.212359   0.191316  -1.110  0.26838
age46-60                             -0.139284   0.195913  -0.711  0.47798
age60+                               -0.609960   0.218076  -2.797  0.00568 **
L1.ENYes                              0.031480   0.170102   0.185  0.85337
reading.on.formulaicity1:trainingYes  0.072666   0.282859   0.257  0.79753
reading.on.formulaicity2:trainingYes  0.634995   0.562130   1.130  0.26004
reading.on.formulaicity3:trainingYes -0.196777   0.831613  -0.237  0.81320
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7012 on 193 degrees of freedom
```

```
Multiple R-squared:  0.1427,    Adjusted R-squared:  0.06714
F-statistic: 1.889 on 17 and 193 DF,  p-value: 0.02091
```



English and Linguistics Subfield Comparisons

Predicate

```
lm(formula = predicate ~ ED.Coding + Generative.LX + Func.LX +
   Comp.LX + Applied.LX + Anth.LX + General.LX + EN.Lit + General.EN
+
   reading.on.formulaicity + syntax + semantics + phonology +
   sociolinguistics + lang.pedagogy + theoretical.lang.acq +
   comp.LX + EN.grammar + L1.EN + training, data = token.LingEN)

Residuals:
    Min       1Q   Median       3Q      Max
-1.32391 -0.30522 -0.01886  0.35370  1.09400

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)            3.8108949  0.6406875   5.948 1.85e-07 ***
ED.CodingLX Grad      -0.8366526  0.3965538  -2.110   0.0394 *
ED.CodingLX Undergrad -0.8162484  0.4120697  -1.981   0.0525 .
ED.CodingLX Instructor -0.9646435  0.4272395  -2.258   0.0279 *
Generative.LX2        -0.1868664  0.2650139  -0.705   0.4837
Generative.LX3        -0.0001702  0.3594629   0.000   0.9996
Generative.LX4        -0.4206319  0.5898624  -0.713   0.4787
Func.LX2               0.2572325  0.2481586   1.037   0.3044
Func.LX3               0.4196293  0.3223675   1.302   0.1983
Func.LX4              -0.2867582  0.6381354  -0.449   0.6549
Comp.LX2               0.0525710  0.2427781   0.217   0.8294
Comp.LX3               0.2920396  0.3154485   0.926   0.3585
Comp.LX4               0.1531240  0.6494700   0.236   0.8145
Applied.LX2           -0.0867139  0.2464718  -0.352   0.7263
Applied.LX3           -0.2013568  0.3921473  -0.513   0.6096
Applied.LX4           -0.3808189  0.6285822  -0.606   0.5471
Anth.LX2              -0.3361702  0.1958165  -1.717   0.0915 .
Anth.LX3               0.1169338  0.4464586   0.262   0.7943
```

```
Anth.LX4               0.4512729  0.5171316   0.873   0.3866
General.LX2            0.2450521  0.4519202   0.542   0.5898
General.LX3            0.3689939  0.4811233   0.767   0.4463
General.LX4            0.7833068  0.4949632   1.583   0.1192
EN.Lit2               -0.0569525  0.2397656  -0.238   0.8131
EN.Lit3               -0.1912961  0.3453202  -0.554   0.5818
EN.Lit4               -0.1639583  0.4554018  -0.360   0.7202
General.EN2            0.0578846  0.2622727   0.221   0.8261
General.EN3            0.0493841  0.3653948   0.135   0.8930
General.EN4            0.0455784  0.5192425   0.088   0.9304
reading.on.formulaicity1  0.0374613  0.2179913   0.172   0.8642
reading.on.formulaicity2  0.3716165  0.3696550   1.005   0.3191
reading.on.formulaicity4  0.7848271  0.5691724   1.379   0.1734
syntax2                0.0599963  0.6525991   0.092   0.9271
syntax3               -0.2861968  0.7342523  -0.390   0.6982
syntax4               -0.0294635  0.7112909  -0.041   0.9671
syntax5               -0.3396462  0.7657341  -0.444   0.6591
semantics2             0.1246276  0.4324511   0.288   0.7743
semantics3            -0.2995068  0.4127174  -0.726   0.4710
semantics4             0.1609665  0.3671809   0.438   0.6628
semantics5             0.4117518  0.4725547   0.871   0.3873
phonology2            -0.6774228  0.6660191  -1.017   0.3135
phonology3            -0.4009027  0.5840635  -0.686   0.4953
phonology4            -0.4410779  0.6016998  -0.733   0.4666
phonology5            -0.4873902  0.5780150  -0.843   0.4027
sociolinguistics2      0.4942316  0.4367211   1.132   0.2626
sociolinguistics3      0.3086729  0.4246280   0.727   0.4703
sociolinguistics4      0.3149852  0.4394800   0.717   0.4765
sociolinguistics5      0.2054600  0.5014501   0.410   0.6836
lang.pedagogy2        -0.2632589  0.3086909  -0.853   0.3974
lang.pedagogy3        -0.0813619  0.3312849  -0.246   0.8069
lang.pedagogy4        -0.1526978  0.3371404  -0.453   0.6524
lang.pedagogy5        -0.0166161  0.4780408  -0.035   0.9724
theoretical.lang.acq2  0.7494021  0.3620062   2.070   0.0431 *
theoretical.lang.acq3  0.7509608  0.3750536   2.002   0.0501 .
theoretical.lang.acq4  0.9874562  0.4540986   2.175   0.0339 *
theoretical.lang.acq5  1.0880554  0.5517428   1.972   0.0536 .
comp.LX2              -0.2168076  0.2799257  -0.775   0.4419
comp.LX3              -0.4874045  0.3142627  -1.551   0.1265
comp.LX4              -0.6015526  0.3649062  -1.649   0.1048
comp.LX5              -0.1457436  0.5077968  -0.287   0.7752
EN.grammar2           -0.2543673  0.4851161  -0.524   0.6021
EN.grammar3           -0.3888059  0.4813171  -0.808   0.4226
EN.grammar4           -0.4149610  0.4673564  -0.888   0.3784
EN.grammar5           -0.4781660  0.4791935  -0.998   0.3226
L1.ENYes               0.0761078  0.2399896   0.317   0.7523
trainingYes           -0.2617932  0.1829267  -1.431   0.1579
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7044 on 56 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.4971,    Adjusted R-squared:  -0.07764
F-statistic: 0.8649 on 64 and 56 DF,  p-value: 0.7139
```

Non-Phrase

```
lm(formula = non.phrase ~ ED.Coding + Generative.LX + Func.LX +
   Comp.LX + Applied.LX + Anth.LX + General.LX + EN.Lit + General.EN
+
   reading.on.formulaicity + syntax + semantics + phonology +
   sociolinguistics + lang.pedagogy + theoretical.lang.acq +
   comp.LX + EN.grammar + L1.EN + training, data = token.LingEN)

Residuals:
    Min       1Q   Median       3Q      Max
-1.88965 -0.52593 -0.00379  0.41997  2.00643

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)            1.759421   1.011753   1.739   0.0875 .
ED.CodingLX Grad      -1.290962   0.626225  -2.061   0.0439 *
ED.CodingLX Undergrad -1.597164   0.650727  -2.454   0.0172 *
ED.CodingLX Instructor -1.657160  0.674683  -2.456   0.0172 *
Generative.LX2        -0.876120   0.418502  -2.093   0.0408 *
Generative.LX3        -0.731246   0.567652  -1.288   0.2030
Generative.LX4        -0.810229   0.931492  -0.870   0.3881
Func.LX2               0.561386   0.391884   1.433   0.1576
Func.LX3               0.994165   0.509072   1.953   0.0558 .
Func.LX4               0.005941   1.007723   0.006   0.9953
Comp.LX2               0.622111   0.383387   1.623   0.1103
Comp.LX3               0.418595   0.498146   0.840   0.4043
Comp.LX4               1.404551   1.025622   1.369   0.1763
Applied.LX2           -0.108531   0.389220  -0.279   0.7814
Applied.LX3            0.386760   0.619267   0.625   0.5348
Applied.LX4            0.796262   0.992637   0.802   0.4258
Anth.LX2               0.347162   0.309227   1.123   0.2664
Anth.LX3               1.749755   0.705033   2.482   0.0161 *
Anth.LX4               1.940635   0.816638   2.376   0.0209 *
General.LX2            1.153391   0.713658   1.616   0.1117
General.LX3            1.159285   0.759774   1.526   0.1327
General.LX4            1.553522   0.781630   1.988   0.0518 .
EN.Lit2               -0.036715   0.378630  -0.097   0.9231
EN.Lit3               -0.413064   0.545319  -0.757   0.4519
EN.Lit4               -0.088393   0.719156  -0.123   0.9026
General.EN2           -0.075567   0.414173  -0.182   0.8559
General.EN3            0.176621   0.577020   0.306   0.7607
General.EN4           -0.813037   0.819971  -0.992   0.3257
reading.on.formulaicity1  0.172065   0.344245   0.500   0.6192
reading.on.formulaicity2 -0.577684   0.583747  -0.990   0.3266
reading.on.formulaicity4  0.636630   0.898819   0.708   0.4817
syntax2               -0.327573   1.030564  -0.318   0.7518
syntax3               -0.352818   1.159508  -0.304   0.7620
syntax4                0.373737   1.123248   0.333   0.7406
syntax5                0.217391   1.209223   0.180   0.8580
semantics2            -0.185674   0.682913  -0.272   0.7867
semantics3            -0.987083   0.651750  -1.515   0.1355
semantics4            -0.491433   0.579840  -0.848   0.4003
```

```
semantics5              -0.845285   0.746243  -1.133   0.2622
phonology2              -0.157535   1.051756  -0.150   0.8815
phonology3               0.042807   0.922334   0.046   0.9631
phonology4               0.436593   0.950185   0.459   0.6477
phonology5               0.746416   0.912783   0.818   0.4170
sociolinguistics2       -0.387858   0.689656  -0.562   0.5761
sociolinguistics3       -0.593569   0.670559  -0.885   0.3798
sociolinguistics4       -0.962136   0.694013  -1.386   0.1711
sociolinguistics5       -1.444183   0.791874  -1.824   0.0735 .
lang.pedagogy2          -0.507400   0.487475  -1.041   0.3024
lang.pedagogy3           0.114456   0.523154   0.219   0.8276
lang.pedagogy4          -0.154882   0.532401  -0.291   0.7722
lang.pedagogy5          -0.516141   0.754907  -0.684   0.4970
theoretical.lang.acq2    0.773978   0.571669   1.354   0.1812
theoretical.lang.acq3    0.619311   0.592273   1.046   0.3002
theoretical.lang.acq4    0.573775   0.717098   0.800   0.4270
theoretical.lang.acq5    0.620700   0.871295   0.712   0.4792
comp.LX2                 0.148752   0.442050   0.337   0.7377
comp.LX3                -0.036838   0.496274  -0.074   0.9411
comp.LX4                 0.046550   0.576248   0.081   0.9359
comp.LX5                 0.079348   0.801897   0.099   0.9215
EN.grammar2              1.184624   0.766080   1.546   0.1277
EN.grammar3              0.617804   0.760081   0.813   0.4198
EN.grammar4              0.740260   0.738034   1.003   0.3202
EN.grammar5              0.877338   0.756727   1.159   0.2512
L1.ENYes                 0.165249   0.378984   0.436   0.6645
trainingYes             -0.584202   0.288872  -2.022   0.0479 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.112 on 56 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.5595,    Adjusted R-squared:  0.05617
F-statistic: 1.112 on 64 and 56 DF,  p-value: 0.3443
```

Predicate and Argument

```
lm(formula = predicate.and.argument ~ ED.Coding + Generative.LX +
   Func.LX + Comp.LX + Applied.LX + Anth.LX + General.LX + EN.Lit +
   General.EN + reading.on.formulaicity + syntax + semantics +
   phonology + sociolinguistics + lang.pedagogy + theoretical.lang.acq +
   comp.LX + EN.grammar + L1.EN + training, data = token.LingEN)

Residuals:
    Min       1Q   Median       3Q      Max
-1.02853 -0.29415 -0.03898  0.30006  0.96001

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)              3.70737    0.57697   6.426 3.06e-08 ***
ED.CodingLX Grad        -0.28825    0.35711  -0.807   0.4230
ED.CodingLX Undergrad   -0.50648    0.37109  -1.365   0.1778
ED.CodingLX Instructor  -0.04711    0.38475  -0.122   0.9030
Generative.LX2          -0.10112    0.23866  -0.424   0.6734
Generative.LX3           0.09575    0.32371   0.296   0.7685
Generative.LX4           0.14001    0.53120   0.264   0.7931
Func.LX2                 0.04825    0.22348   0.216   0.8298
Func.LX3                 0.05434    0.29031   0.187   0.8522
Func.LX4                -0.65409    0.57467  -1.138   0.2599
Comp.LX2                 0.20960    0.21863   0.959   0.3418
Comp.LX3                -0.03772    0.28408  -0.133   0.8948
Comp.LX4                -0.12259    0.58488  -0.210   0.8347
Applied.LX2             -0.20006    0.22196  -0.901   0.3713
Applied.LX3             -0.39421    0.35314  -1.116   0.2691
Applied.LX4              0.02161    0.56607   0.038   0.9697
Anth.LX2                -0.23805    0.17634  -1.350   0.1825
Anth.LX3                 0.30505    0.40205   0.759   0.4512
Anth.LX4                 0.39916    0.46570   0.857   0.3950
General.LX2              0.21492    0.40697   0.528   0.5995
General.LX3              0.39181    0.43327   0.904   0.3697
General.LX4              0.85897    0.44574   1.927   0.0590 .
EN.Lit2                  0.23026    0.21592   1.066   0.2908
EN.Lit3                  0.02571    0.31098   0.083   0.9344
EN.Lit4                  0.03649    0.41011   0.089   0.9294
General.EN2              0.28065    0.23619   1.188   0.2398
General.EN3              0.04133    0.32905   0.126   0.9005
General.EN4             -0.55547    0.46760  -1.188   0.2399
reading.on.formulaicity1  0.12772   0.19631   0.651   0.5180
reading.on.formulaicity2  0.49924   0.33289   1.500   0.1393
reading.on.formulaicity4 -0.02093   0.51256  -0.041   0.9676
syntax2                  0.37780    0.58769   0.643   0.5229
syntax3                 -0.42770    0.66123  -0.647   0.5204
syntax4                 -0.02569    0.64055  -0.040   0.9682
syntax5                 -0.37526    0.68958  -0.544   0.5885
semantics2              -0.04293    0.38944  -0.110   0.9126
semantics3              -0.13512    0.37167  -0.364   0.7176
```

```
semantics4             -0.03309    0.33066  -0.100   0.9206
semantics5              0.24988    0.42556   0.587   0.5594
phonology2             -0.77110    0.59978  -1.286   0.2039
phonology3             -0.73606    0.52597  -1.399   0.1672
phonology4             -1.00378    0.54186  -1.852   0.0692 .
phonology5             -1.08477    0.52053  -2.084   0.0417 *
sociolinguistics2       0.61737    0.39329   1.570   0.1221
sociolinguistics3       0.90237    0.38240   2.360   0.0218 *
sociolinguistics4       0.80480    0.39577   2.033   0.0468 *
sociolinguistics5       0.58283    0.45158   1.291   0.2021
lang.pedagogy2         -0.61794    0.27799  -2.223   0.0303 *
lang.pedagogy3          0.15932    0.29834   0.534   0.5954
lang.pedagogy4         -0.05544    0.30361  -0.183   0.8558
lang.pedagogy5          0.00423    0.43050   0.010   0.9922
theoretical.lang.acq2   0.75136    0.32600   2.305   0.0249 *
theoretical.lang.acq3   0.46597    0.33775   1.380   0.1732
theoretical.lang.acq4   0.68528    0.40893   1.676   0.0994 .
theoretical.lang.acq5   0.89773    0.49687   1.807   0.0762 .
comp.LX2               -0.40343    0.25209  -1.600   0.1151
comp.LX3               -0.47724    0.28301  -1.686   0.0973 .
comp.LX4               -0.22678    0.32861  -0.690   0.4930
comp.LX5               -0.13488    0.45729  -0.295   0.7691
EN.grammar2            -0.07349    0.43687  -0.168   0.8670
EN.grammar3            -0.23652    0.43345  -0.546   0.5875
EN.grammar4            -0.39346    0.42087  -0.935   0.3539
EN.grammar5            -0.19909    0.43153  -0.461   0.6463
L1.ENYes                0.05267    0.21612   0.244   0.8084
trainingYes            -0.01572    0.16473  -0.095   0.9243
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6344 on 56 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.5203,    Adjusted R-squared:  -0.02798
F-statistic: 0.949 on 64 and 56 DF,  p-value: 0.5822
```

Adjunct or Conjunct

```
lm(formula = adjunct.conjunct ~ ED.Coding + Generative.LX + Func.LX +
    Comp.LX + Applied.LX + Anth.LX + General.LX + EN.Lit + General.EN
+
    reading.on.formulaicity + syntax + semantics + phonology +
    sociolinguistics + lang.pedagogy + theoretical.lang.acq +
    comp.LX + EN.grammar + L1.EN + training, data = token.LingEN)

Residuals:
     Min       1Q   Median       3Q      Max
-1.73235 -0.28142  0.02604  0.30597  1.11046

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               3.508935   0.667210   5.259 2.35e-06 ***
ED.CodingLX Grad         -0.771171   0.412970  -1.867  0.06709 .
ED.CodingLX Undergrad    -0.633840   0.429128  -1.477  0.14527
ED.CodingLX Instructor   -0.940782   0.444926  -2.114  0.03894 *
Generative.LX2           -0.310720   0.275985  -1.126  0.26503
Generative.LX3           -0.096032   0.374344  -0.257  0.79848
Generative.LX4           -0.669281   0.614281  -1.090  0.28058
Func.LX2                 -0.091647   0.258432  -0.355  0.72420
Func.LX3                  0.188526   0.335713   0.562  0.57665
Func.LX4                 -1.353911   0.664552  -2.037  0.04635 *
Comp.LX2                  0.379511   0.252828   1.501  0.13896
Comp.LX3                  0.494860   0.328507   1.506  0.13759
Comp.LX4                  0.269900   0.676356   0.399  0.69137
Applied.LX2               0.102761   0.256675   0.400  0.69042
Applied.LX3              -0.150563   0.408381  -0.369  0.71375
Applied.LX4              -0.236451   0.654604  -0.361  0.71930
Anth.LX2                  0.130145   0.203923   0.638  0.52594
Anth.LX3                 -0.135579   0.464941  -0.292  0.77167
Anth.LX4                  0.393472   0.538539   0.731  0.46805
General.LX2              -0.331096   0.470628  -0.704  0.48465
General.LX3              -0.368549   0.501040  -0.736  0.46506
General.LX4               0.365296   0.515453   0.709  0.48146
EN.Lit2                  -0.022042   0.249691  -0.088  0.92997
EN.Lit3                  -0.091024   0.359616  -0.253  0.80111
EN.Lit4                  -0.476206   0.474254  -1.004  0.31964
General.EN2               0.121343   0.273130   0.444  0.65856
General.EN3              -0.002325   0.380521  -0.006  0.99515
General.EN4              -0.182105   0.540738  -0.337  0.73755
reading.on.formulaicity1  0.134653   0.227016   0.593  0.55547
reading.on.formulaicity2  0.564125   0.384958   1.465  0.14840
reading.on.formulaicity4  0.498898   0.592735   0.842  0.40354
syntax2                   0.442521   0.679615   0.651  0.51762
syntax3                  -0.435156   0.764648  -0.569  0.57157
syntax4                   0.063010   0.740736   0.085  0.93251
syntax5                   0.034381   0.797433   0.043  0.96576
semantics2                0.324223   0.450353   0.720  0.47456
semantics3                0.152228   0.429803   0.354  0.72453
semantics4                0.105664   0.382381   0.276  0.78331
```

```
semantics5              0.341565   0.492117   0.694  0.49051
phonology2             -0.570000   0.693590  -0.822  0.41467
phonology3             -0.566243   0.608242  -0.931  0.35588
phonology4             -0.552894   0.626609  -0.882  0.38135
phonology5             -0.843237   0.601943  -1.401  0.16677
sociolinguistics2       0.987693   0.454800   2.172  0.03413 *
sociolinguistics3       1.062919   0.442206   2.404  0.01957 *
sociolinguistics4       0.821309   0.457673   1.795  0.07813 .
sociolinguistics5       0.963343   0.522209   1.845  0.07037 .
lang.pedagogy2         -0.211681   0.321470  -0.658  0.51293
lang.pedagogy3         -0.133928   0.344999  -0.388  0.69934
lang.pedagogy4         -0.372145   0.351097  -1.060  0.29372
lang.pedagogy5          0.037467   0.497830   0.075  0.94027
theoretical.lang.acq2   0.865254   0.376992   2.295  0.02549 *
theoretical.lang.acq3   1.009076   0.390580   2.584  0.01242 *
theoretical.lang.acq4   1.306035   0.472897   2.762  0.00776 **
theoretical.lang.acq5   1.164674   0.574583   2.027  0.04743 *
comp.LX2               -0.571545   0.291514  -1.961  0.05491 .
comp.LX3               -0.439943   0.327272  -1.344  0.18428
comp.LX4               -0.079709   0.380012  -0.210  0.83462
comp.LX5               -0.180940   0.528818  -0.342  0.73351
EN.grammar2            -0.615573   0.505199  -1.218  0.22815
EN.grammar3            -0.632918   0.501242  -1.263  0.21193
EN.grammar4            -0.610161   0.486704  -1.254  0.21517
EN.grammar5            -0.342706   0.499031  -0.687  0.49508
L1.ENYes                0.433600   0.249924   1.735  0.08826 .
trainingYes            -0.027901   0.190499  -0.146  0.88408
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7336 on 56 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.5457,    Adjusted R-squared:  0.0265
F-statistic: 1.051 on 64 and 56 DF,  p-value: 0.4264
```

Primary Interest

Predicate

```
lm(formula = predicate ~ ED.Coding + Primary.Interest + training,
    data = token.ling)

Residuals:
     Min       1Q   Median       3Q      Max
-1.77456 -0.36484  0.03716  0.42655  1.50695

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                  2.94628    0.25735  11.449   <2e-16 ***
ED.CodingLX Undergrad        0.22768    0.16885   1.348    0.181
ED.CodingLX Instructor       0.07852    0.16670   0.471    0.639
Primary.InterestApplied LX   0.35480    0.30639   1.158    0.250
Primary.InterestComp LX      0.30159    0.28852   1.045    0.299
Primary.InterestFUNC LX      0.01495    0.35593   0.042    0.967
Primary.InterestGeneral LX   0.06850    0.26997   0.254    0.800
Primary.InterestGenerative LX  0.15654    0.29574   0.529    0.598
trainingYes                 -0.13998    0.15167  -0.923    0.358
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6809 on 95 degrees of freedom
  (3 observations deleted due to missingness)
Multiple R-squared:  0.07378,   Adjusted R-squared:  -0.004223
F-statistic: 0.9459 on 8 and 95 DF,  p-value: 0.483
```

Non-Phrase

```
lm(formula = non.phrase ~ ED.Coding + Primary.Interest + training,
    data = token.ling)

Residuals:
    Min      1Q  Median      3Q     Max
-1.5090 -0.9507 -0.2557  0.7770  2.4531

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                  2.84116    0.42406   6.700 1.47e-09 ***
ED.CodingLX Undergrad        0.20020    0.27824   0.720   0.4736
ED.CodingLX Instructor      -0.01251    0.27468  -0.046   0.9638
Primary.InterestApplied LX  -0.55561    0.50487  -1.100   0.2739
Primary.InterestComp LX     -0.28170    0.47541  -0.593   0.5549
Primary.InterestFUNC LX     -0.53318    0.58650  -0.909   0.3656
Primary.InterestGeneral LX  -0.55828    0.44486  -1.255   0.2126
Primary.InterestGenerative LX -1.02619    0.48732  -2.106   0.0379 *
trainingYes                 -0.33217    0.24992  -1.329   0.1870
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.122 on 95 degrees of freedom
  (3 observations deleted due to missingness)
Multiple R-squared:  0.08793,   Adjusted R-squared:  0.01113
F-statistic: 1.145 on 8 and 95 DF,  p-value: 0.341
```

Predicate and Argument

```
lm(formula = predicate.and.argument ~ ED.Coding + Primary.Interest +
    training, data = token.ling)

Residuals:
     Min       1Q   Median       3Q      Max
-2.13422 -0.34805  0.01174  0.40378  1.28862

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                  3.616324   0.237956  15.197   <2e-16 ***
ED.CodingLX Undergrad        0.027483   0.156130   0.176   0.8607
ED.CodingLX Instructor       0.301097   0.154134   1.953   0.0537 .
Primary.InterestApplied LX  -0.072032   0.283303  -0.254   0.7998
Primary.InterestComp LX     -0.071612   0.266772  -0.268   0.7889
Primary.InterestFUNC LX     -0.228845   0.329105  -0.695   0.4885
Primary.InterestGeneral LX  -0.261538   0.249628  -1.048   0.2974
Primary.InterestGenerative LX -0.144420   0.273456  -0.528   0.5986
trainingYes                 -0.004347   0.140242  -0.031   0.9753
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6295 on 95 degrees of freedom
```

```
  (3 observations deleted due to missingness)
Multiple R-squared:  0.07328,    Adjusted R-squared:  -0.004756
F-statistic: 0.9391 on 8 and 95 DF,  p-value: 0.4884
```

Adjunct or Conjunct

```
lm(formula = adjunct.conjunct ~ ED.Coding + Primary.Interest +
    training, data = token.ling)

Residuals:
    Min      1Q  Median      3Q     Max
-2.5563 -0.4937  0.1090  0.5477  1.3102

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                   3.63228    0.29006  12.522   <2e-16 ***
ED.CodingLX Undergrad         0.16302    0.19032   0.857    0.394
ED.CodingLX Instructor        0.07629    0.18788   0.406    0.686
Primary.InterestApplied LX   -0.20314    0.34534  -0.588    0.558
Primary.InterestComp LX       0.03981    0.32519   0.122    0.903
Primary.InterestFUNC LX      -0.47286    0.40117  -1.179    0.241
Primary.InterestGeneral LX   -0.10547    0.30429  -0.347    0.730
Primary.InterestGenerative LX -0.22434    0.33333  -0.673    0.503
trainingYes                  -0.11577    0.17095  -0.677    0.500
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7674 on 95 degrees of freedom
  (3 observations deleted due to missingness)
Multiple R-squared:  0.0463,    Adjusted R-squared:  -0.03401
F-statistic: 0.5765 on 8 and 95 DF,  p-value: 0.7948
```

AMT

AMT vs. AMT(M) vs. None

Predicate

```
lm(formula = predicate ~ AMT + reading.on.formulaicity + non.EN.lang +
    training, data = token.dat)

Residuals:
     Min       1Q   Median       3Q      Max
-1.79235 -0.38712  0.04929  0.41346  1.45909

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               3.070983   0.101716  30.192   <2e-16 ***
AMTAMT                   -0.083075   0.132863  -0.625   0.5325
AMTAMT(M)                -0.018048   0.130650  -0.138   0.8903
reading.on.formulaicity1  0.053003   0.104866   0.505   0.6138
reading.on.formulaicity2  0.153058   0.184694   0.829   0.4082
reading.on.formulaicity4  0.607053   0.324261   1.872   0.0626 .
non.EN.langYes            0.143078   0.097405   1.469   0.1434
trainingYes               0.001699   0.096494   0.018   0.9860
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6345 on 203 degrees of freedom
Multiple R-squared:  0.03732,   Adjusted R-squared:  0.004125
F-statistic: 1.124 on 7 and 203 DF,  p-value: 0.3491
```

Non Phrase

```
lm(formula = non.phrase ~ AMT + reading.on.formulaicity + non.EN.lang +
    training, data = token.dat)

Residuals:
    Min      1Q  Median      3Q     Max
-1.8412 -0.9361 -0.1719  0.7849  2.6562

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               2.124776   0.174215  12.196   <2e-16 ***
AMTAMT                    0.497338   0.227560   2.186    0.030 *
AMTAMT(M)                -0.192797   0.223770  -0.862    0.390
reading.on.formulaicity1  0.008274   0.179609   0.046    0.963
reading.on.formulaicity2 -0.108919   0.316335  -0.344    0.731
reading.on.formulaicity4  0.455735   0.555378   0.821    0.413
non.EN.langYes            0.219032   0.166830   1.313    0.191
trainingYes              -0.093440   0.165269  -0.565    0.572
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Residual standard error: 1.087 on 203 degrees of freedom
Multiple R-squared:  0.04708,   Adjusted R-squared:  0.01422
F-statistic: 1.433 on 7 and 203 DF,  p-value: 0.1937
```

Predicate and Argument

```
lm(formula = predicate.and.argument ~ AMT + reading.on.formulaicity +
    non.EN.lang + training, data = token.dat)

Residuals:
     Min       1Q   Median       3Q      Max
-2.32927 -0.33185  0.00406  0.41168  1.28086

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               3.561652   0.096686  36.837  < 2e-16 ***
AMTAMT                   -0.435462   0.126292  -3.448 0.000686 ***
AMTAMT(M)                -0.144493   0.124189  -1.163 0.245994
reading.on.formulaicity1  0.041575   0.099680   0.417 0.677060
reading.on.formulaicity2  0.364488   0.175561   2.076 0.039141 *
reading.on.formulaicity4  0.417701   0.308226   1.355 0.176866
non.EN.langYes           -0.009182   0.092588  -0.099 0.921098
trainingYes               0.100951   0.091722   1.101 0.272363
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6031 on 203 degrees of freedom
Multiple R-squared:  0.08597,   Adjusted R-squared:  0.05445
F-statistic: 2.728 on 7 and 203 DF,  p-value: 0.01002
```

Adjunct or Conjunct

```
lm(formula = adjunct.conjunct ~ AMT + reading.on.formulaicity +
   non.EN.lang + training, data = token.dat)

Residuals:
   Min      1Q  Median      3Q     Max
-2.4825 -0.4105  0.0974  0.4002  1.4974

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)              3.43054    0.11201  30.627  < 2e-16 ***
AMTAMT                  -0.44510    0.14631  -3.042  0.00266 **
AMTAMT(M)               -0.27181    0.14387  -1.889  0.06029 .
reading.on.formulaicity1 -0.02007    0.11548  -0.174  0.86217
reading.on.formulaicity2  0.12593    0.20339   0.619  0.53649
reading.on.formulaicity4  0.35679    0.35708   0.999  0.31889
non.EN.langYes           0.18930    0.10726   1.765  0.07909 .
trainingYes              0.07205    0.10626   0.678  0.49849
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6987 on 203 degrees of freedom
Multiple R-squared:  0.1047,    Adjusted R-squared:  0.07383
F-statistic: 3.392 on 7 and 203 DF,  p-value: 0.001903
```

AMT Paying Attention

Predicate

```
lm(formula = predicate ~ AMT + paying.attention, data = token.subpaying)

Residuals:
    Min       1Q   Median       3Q      Max
-1.68430 -0.24924  0.08409  0.25076  1.41743

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        3.082574   0.111510  27.644  < 2e-16 ***
AMTAMT(M)         -0.001591   0.136456  -0.012  0.99073
paying.attention1 -0.915377   0.322066  -2.842  0.00606 **
paying.attention2  0.269985   0.196232   1.376  0.17382
paying.attention4  0.001024   0.239924   0.004  0.99661
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5382 on 62 degrees of freedom
Multiple R-squared:  0.1499,    Adjusted R-squared:  0.09506
F-statistic: 2.733 on 4 and 62 DF,  p-value: 0.03676
```

Non Phrase

```
lm(formula = non.phrase ~ AMT + paying.attention, data =
token.subpaying)

Residuals:
   Min      1Q  Median      3Q     Max
-1.4893 -0.8695 -0.2028  0.7972  2.1774

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)        2.4893     0.2122  11.732   <2e-16 ***
AMTAMT(M)         -0.6198     0.2596  -2.387    0.020 *
paying.attention1   0.2729     0.6128   0.445    0.658
paying.attention2   0.3788     0.3734   1.014    0.314
paying.attention4   0.4474     0.4565   0.980    0.331
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.024 on 62 degrees of freedom
Multiple R-squared:  0.1348,    Adjusted R-squared:  0.07894
F-statistic: 2.414 on 4 and 62 DF,  p-value: 0.05826
```

Predicate and Argument

```
lm(formula = predicate.and.argument ~ AMT + paying.attention,
   data = token.subpaying)

Residuals:
    Min       1Q   Median       3Q      Max
-1.59128 -0.42462  0.02272  0.45322  0.90872

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         3.2580     0.1207  26.981  < 2e-16 ***
AMTAMT(M)           0.2193     0.1478   1.484  0.14279
paying.attention1  -0.9422     0.3488  -2.702  0.00889 **
paying.attention2   0.0705     0.2125   0.332  0.74118
paying.attention4   0.1222     0.2598   0.470  0.63985
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5828 on 62 degrees of freedom
Multiple R-squared:  0.1483,    Adjusted R-squared:  0.09338
F-statistic: 2.699 on 4 and 62 DF,  p-value: 0.0386
```

Adjunct or Conjunct

```
lm(formula = adjunct.conjunct ~ AMT + paying.attention, data =
token.subpaying)

Residuals:
    Min       1Q   Median       3Q      Max
-1.50832 -0.30290  0.06805  0.35307  1.06805

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         3.1320     0.1164  26.908   <2e-16 ***
AMTAMT(M)           0.1150     0.1424   0.807    0.423
paying.attention1  -0.3703     0.3362  -1.101    0.275
paying.attention2  -0.1386     0.2048  -0.677    0.501
paying.attention4  -0.1178     0.2504  -0.470    0.640
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5618 on 62 degrees of freedom
Multiple R-squared:  0.04288,   Adjusted R-squared:  -0.01887
F-statistic: 0.6944 on 4 and 62 DF,  p-value: 0.5987
```