Reporting statistical results for LVC

Gregory R. Guy
Rena Torres Cacoullos

NWAV 47 - NYU
19 October 2018

Outline

- Best practice basics for quantitative data
- Modeling
- Main effects
- Random effects
- Interactions
- Tables and figures

Basics

- Clarity
  - Any quantitative data presented should aim to be easily interpretable. Design your tables and figures so that the informed reader can figure out the points they illustrate
- Efficiency
  - Put in the right amount of information, what is needed to make your point, without a flood of material which is not necessary, or omissions that require extra work from the reader.
  - Wherever possible, a table or figure should contain all the information necessary to interpret it without having to go read something in the text of the paper.

- Transparency
  - Fully explain everything you did, the model that you used, the software you used, the assumptions of the model...

- Recoverability
  - Present enough information so that readers can reconstruct what you did and evaluate your analysis. Maximum recoverability is achieved by open source corpora, but this is not feasible for some research, or permitted by some IRB protocols.

For a statistical analysis to be credible, something needs to be shown that characterizes the data. Minimally required are the total number of observations (tokens), the number of observations per speaker (not as an average), and the distribution of speakers across social categories, so that the reader can appreciate the balance (or imbalance) in the research design.

Paolillo 2017

Data quantity

- The most basic statistic for any quantitative analysis is N: the number of observations, tokens, etc.
- Report your Ns prominently wherever relevant. This includes subdivisions of the data set,
  - by external variables: Ns per speaker, per group of speakers, per speech style or setting, etc.
  - by internal variables: Ns for each factor in a factor group, each level of a predictor variable, etc.

Data quantity and significant digits

- Precision of all statistical information depends directly on N – the number of data points, tokens, observations, etc.
  - -more data = more precise statistical results, more accurate estimates, higher levels of significance
- Therefore, the order of magnitude of all your reported statistics should not exceed the order of magnitude of your N.
  - Otherwise, you are implicitly claiming greater precision than your data support
- Rule of thumb: don't give more significant digits in your stats than you have in your N.
  - with 10's of tokens, give 2 significant digits
  - with 100's of tokens, give 3 significant digits
  - with 1000's, give 4

Yiddish verb agreement (Bleaman 2018)

	Estimate	Std. error	*z*-value	*p*-value		N
(Intercept)	3.20	0.59	5.41	<0.001	***	806
Adjacency (vs. adjacent)						132
nonadjacent	1.15	0.36	3.19	0.001	**	674
First position filler (vs. expletive)						476
adverb	-2.25	0.31	-7.24	<0.001	***	229
deictic	-1.03	0.55	-1.87	0.062	.	52
wh	-1.19	0.66	-2.90	0.004	**	20
zero	-1.95	0.68	-2.87	0.004	**	29
Subject humanness (vs. human)						409
nonhuman	0.65	0.26	2.49	0.013	*	397
Verb type (vs. existential)						590
copula	-2.52	0.64	-3.96	<0.001	***	93
be-verb	-2.33	0.66	-3.53	<0.001	***	55
have-verb	-4.76	0.71	-6.69	<0.001	***	68
Speaker gender (vs. female)						336
male	0.65	0.26	2.50	0.012	*	470
Community (vs. Hasidic)						467
Yiddishist	-2.46	0.45	-5.52	<0.001	***	339
Community : Verb type						
Yiddishist : copula	2.20	0.64	3.41	<0.001	***	
Yiddishist : *be*-verb	1.30	0.83	1.58	0.115		
Yiddishist : *have*-verb	3.75	0.88	4.25	<0.001	***	

Table 5.4: Estimates for effects from best-fit logistic regression model of variable number marking (n = 806), where positive estimates favor the singular variant; significance codes: *** = <0.001, ** = <0.01, * = <0.05, . = <0.1. Token counts for the interaction of community and verb type is provided in Table 5.2.

Tamminga, LVC 2016

	Estimate	SE	$Pr(>\|z\|)$	N	% retained
Predictors of interest					
Prime variant (vs. deletion)				(del.) 2839	46
Retention prime	-0.083	0.14	.56	2942	55
Morphological match (vs. mismatch)				(mismatch) 2075	58
Match	-0.43	0.13	.0011	3706	47
Log lag	-0.05	0.069	.50	N/A	N/A
Interaction terms					
Prime x matched	0.64	0.18	$< .001$		
Prime x lag	0.17	0.09	.065		
Matched x lag	0.20	0.08	.020		
Prime x matched x lag	-0.39	0.12	.0013		
Control predictors					
Target morphological category (vs. monomorph)				(mono) 4075	40
Polymorpheme	-0.43	0.13	.0011	1706	75
Preceding segment (vs. liquid)				(liquid) 540	59
Nasal	-0.61	0.12	$< .001$	2032	46
Obstruent	0.20	0.13	.13	1181	76
Sibilant	-0.58	0.12	$< .001$	2028	38
Following segment (vs. approximant)				(appx.) 1070	40
Obstruent	-1.36	0.11	$< .001$	1286	16
Pause	0.82	0.090	$< .001$	1568	62
Vowel	1.09	0.088	$< .001$	1857	70
Log frequency	-0.13	0.015	$< .001$	N/A	N/A
Speaker gender (vs. female)				(F) 3202	53
Male	-0.29	0.087	$< .001$	2579	47
Intercept	1.26	0.23	$< .001$		

Table 3: GLMM predicting retention by morphological match in TD data, N=5781. AIC = 6251.5.

Summary statistics: means, proportions, percentages

- Quantitative modeling aims to use independent variables to predict or explain the value of some number characterizing the dependent variable
- The simplest value that characterizes the dependent variable is
  - for continuous variables: the MEAN. e.g. mean formant frequency
  - for nominal variables: the PERCENTAGE or PROPORTION of occurrences. e.g. % deletion or retention of a segment; % overt vs. null subjects
- Wherever possible, give these simple values in addition to the weights, estimates, parameter values etc. that are produced by your model. This provides a simple check on the validity of your model.

The significance of significance

- Significance in statistics is the likelihood of the null hypothesis being true.
- But what is the null hypothesis? It is usually comparative: what's the likelihood of the data characterized by predictor X being randomly different from the data characterized by predictor Y.
- So what are you comparing?
- What test of significance are you using?
- Reporting: it is usually adequate to give asterisks (*p<.05, **p<.01, etc.), or p values with 2-3 significant digits. For small p values (<.001) there is no need to give the number of zeros preceding a non-zero value.
  - e.g. p = 3.45 E13

Hypotheses

- All choices of models imply hypotheses about what affects what. Every independent variable included in the model entails a hypothesis that it has some effect on the dependent variable.
- This includes distinctions between levels of a predictor (factors in a factor group).
  - e.g. specifying possible phonological contexts as {Obstruent, Glide, Liquid, Vowel} vs. {Obstruent, Sonorant, Vowel} vs. {Consonant, Vowel}. Why choose one of these instead of other possible classifications?
- A paper submitted for publication must be clear on what those hypotheses are, and how the quantitative results bear on them.

Modelling

- Most linguistic variation is affected by multiple independent variables, whether language-internal or external.
- The distribution of linguistic variables across the independent variables is always lumpy; balanced cells in a cross tab are unlikely to occur.
- Therefore, a **multivariate** regression analysis is almost always necessary. Separate univariate analyses are rarely appropriate.
- Different multivariate models have different virtues, and make different assumptions.
- **The model chosen for any analysis must be explicitly described and justified, and the analyst should show awareness of the assumptions behind the model.**

Mixed effects models

- Multilevel or mixed effects models are used for data where some observations are clustered into larger groups: a hierarchical, not orthogonal data structure.
  - e.g. a set of speakers comprising males and females. Pete and Harry form the male group, Jane and Sally the female group. These are not orthogonal analyses: Pete can never be observed in both the male and female condition.
- The higher level groups are treated as **fixed effects**, i.e. correlated with the dependent variable, while the data points within those groups are treated as randomly varying around the central value of the group, hence **random effects**, assumed to be uncorrelated with the dependent variable.
- Goldvarb is a fixed effects model, assuming all factor groups are orthogonal. Individual speakers cannot be included in the same model as social groups.

Reporting main effects

- All independent variables tested should be reported, whether significant or not. Non-significance of a potential predictor is an important finding!
- Reports of effects should be readily interpretable. Specify:
  - magnitude of effect (estimate, factor weight)
  - **direction of effect (sign, whether it favors or disfavors the predicted variant)**
  - significance.
- The effects (estimates, factor weights, etc.) of every level of every predictor (every factor in every factor group) should be reported, whether significant or not.

Convergence

- Regression analyses proceed iteratively, refining estimates at each successive iteration, testing whether that iteration has significantly improved a goodness of fit measure, continuing until a best fit is achieved. This is called convergence.
- But with a complex model, there are very many possible adjustments to parameter estimates that can refine the model. Sometimes it is impossible, within practical limits of computational power, to find a unique best fit model. This is nonconvergence.
- Nonconvergence is typically due to an imbalanced distribution of data across predictors, such as a collinearity or non-othogonality of predictors. (most tokens of x are also tokens of y).
- **If your model fails to converge, explore why! Restructure the model by eliminating the imbalance (e.g. eliminate predictors, collapse levels, etc.)**
- **Report nonconvergence!**

Reference levels in LME models

- Effect estimates are calculated with reference to some level of each independent variable.
- **The reference level should always be specified, and relevant summary statistics provided for it (e.g. N, %).**
- **The reference level should be chosen purposefully and sensibly, not by default.** The default in R packages is often alphabetical order; this is supremely irrelevant and ill-advised.
- Good motivations for reference level:
  - the level with the largest N (which is likely to be the most accurate estimate of the effect)
  - a level that is at one end of a scale (e.g. the youngest or oldest age group, the highest or lowest social class group, the most or least sonorant adjacent segment)

Links, van Kemenade & Grondelaers, LVC 2017

TABLE 4. *Logistic regression estimates: Individual resumptive adverb predictors, in three language stages*

	850–1050 (O2&O3) ($n = 4148$)			1150–1250 (M1) ($n = 435$)			1350–1639 (M3,M4&E2) ($n = 4406$)		
	Odds Ratio	*n*	%	Odds Ratio	*n*	%	Odds Ratio	*n*	%
Mood main clause									
Indicative	Reference category	2810	68	Reference category	333	77	Reference category	4053	92
Other	**.10*****	1338	32	.34	102	23	1.30	353	8
Mood subclause									
Indicative	Reference category	3290	79	Reference category	435	100	Reference category	4406	100
Other	1.17	858	21	—	—	—	—	—	—
Subclause-internal particle									
No	Reference category	3604	87	Reference category	432	99	Reference category	4392	100
Yes	**1.70*****	544	13	2.22	3	1	.30	14	0
Negation main clause									
No	Reference category	3518	85	Reference category	349	80	Reference category	3893	88
Yes	**.47*****	630	15	.86	86	20	.76	513	12
Negation subclause									
No	Reference category	3628	87	Reference category	360	83	Reference category	4026	97
Yes	**.61*****	520	13	.69	75	17	1.09	380	3
Subject main clause									
New	Reference category	2889	70	Reference category	312	72	Reference category	3552	81
Given	**.58*****	306	7	1.47	26	6	**.66***	278	6
Other	**.13*****	851	21	.59	88	20	**.50***	443	10
Proper	1.62	102	2	4.09	9	2	.57	133	3
Subject subclause									
New	Reference category	3141	76	Reference category	349	80	Reference category	3328	76
Given	1.18	704	17	.83	64	15	.91	602	14
Other	**2.26****	94	2	.00	10	2	.89	102	2
Proper	**2.52*****	209	5	.99	12	3	**.50****	374	8

Reference levels with binary and multi-level predictors

- A predictor with only two levels will return only one effect estimate: how different the reference level is from the other level.
- A predictor with three or more levels will return effect estimates for all but the reference level. The significance statistics will indicate whether the estimate is significantly different FROM THE REFERENCE LEVEL.
- This does not tell you whether the differences between the other estimates is significant or not. Hence these results may not be particularly useful or illuminating.
- **Explore and explain differences among levels in multi-level predictors.**

Forrest, LVC 2017

TABLE 1. *Regression results for best-fitting model predicting* –in

Observations: 13,167 (overall percentage – *in* 44.0%)
Log likelihood: – 5,106.1
Akaike information criterion: 10,272.1
Bayesian information criterion: 10,496.7

	Estimates	Number of tokens and (percentage –*in*) for factors
Lexical category (vs. adjective)		681 (13)
Gerund	0.638* (.296)	5039 (37)
Verbal	1.049*** (.302)	6661 (54)
Noun	.242 (.332)	786 (30)
Preceding place of articulation (vs. other)		5512 (55)
Coronal	– .805*** (.137)	5574 (30)
Velar	.372 (.208)	1911 (53)
Following place of articulation (vs. other)		6665 (44)
Pause	– .812*** (.112)	1525 (31)
Coronal	.165* (.073)	4240 (50)
Velar	– .501** (.170)	737 (32)
Social factors		
Male	.615 (.345)	5933 (49)
(vs. Female)	N/A[a]	7234 (40)
Year of birth (scaled)[b]	– 1.099*** (.220)	
College	– 2.863*** (.628)	7635 (36)
Graduate	– 3.509*** (.732)	2146 (34)
No college	– 1.509* (.719)	2235 (65)
(vs. two-year-college)	N/A[a]	1151 (74)
Frequency measures		
SUBTLEX frequency (logged)	.482*** (.064)	
percentage of high -*in* grammatical (scaled)[c]	.362*** (.058)	
Percentage of low -*in* phonological (scaled)[d]	– .131 (.959)	
Frequency × year of birth	.066* (.033)	
Frequency × FDC_{phon}	– .128** (.046)	
(Intercept)	– 1.386 (.800)	

Notes: *$p < .05$; **$p < .01$; ***$p < .001$.
[a]No estimates are calculated for reference levels.
[b]Mean year of birth = 1961, SD = 18 years.
[c]Mean percentage of high -*in* grammatical = 50.6%, SD = 30.5%.
[d]Mean percentage of low -*in* phonological = 17.2%, SD = 16.1%.

Random effects in variationist linguistics

- The main usage of mixed effects models in variationist research has been
  - to treat individuals as random variables in the social dimension, and
  - to treat words as random variables wrt phonological or syntactic variation.
- It is always reasonable to assume random differences between individual speakers
- Assuming random differences between words, in their relationship to phonological variables, raises the hoary debate in linguistics about whether ‘each word has its own history’, or phonological processes are systematic (‘exceptionless’) across the lexicon.

Random effects: issues

- Grammatical consistency of the speech community
  - Speakers in a community have generally been found, when data are sufficient, to share constraints on variables with a remarkable level of consistency.
  - This suggests that idiosyncratic differences are likely to be modest, for speakers whose social characteristics are similar,
- Autonomy of phonology and syntax
  - The several structural levels of language (syntax, phonology, lexicon) are fairly autonomous, so that words behave fairly consistently wrt phonological and syntactic properties and processes.
  - Documented lexical differences in phonological or syntactic variation are relatively modest, and involve a small number of mostly high frequency words

Random effects: data quantity

- In any multivariate analysis, the validity and reliability of an estimate of the effect of any predictor or factor is heavily dependent on the N of observations relevant to that predictor.
- This includes random effects: each random predictor in the model, e.g. individual speakers or individual words, needs an adequate N to yield a valid, reliable estimate.
- What N is adequate? A common rule of thumb is a minimum of 30 tokens.
- A good N for individuals may need to be higher, to have sufficient data to see if they have the same effects for a number of other constraints. (Guy 1980 indicates c. 100 tokens are needed per individual to get good estimates.)

Random effects: statistical assumptions

- Most models assume that random effects will have a normal distribution (being **random**, after all!)
- This is likelier to be true with a large number of cases, but with smaller samples, not so much.
- With a modest number of speakers, it's very possible to have a skewed distribution (with a preponderance of speakers at one side) or a bimodal distribution (with two separate clusters).
- These deviations from normality will impair the validity of your analysis. Therefore, you probably need at least 20-30 random factors for a valid analysis (20-30 speakers, or 20-30 different words)

Random effects for lexical items

- Using random effects for words encounters serious data quantity problems.
- The Zipfian distribution of lexical frequency means that a few words will have large Ns and the vast majority of words will have low frequencies.
  - Words with N below 30 should not be analyzed separately; rather, they should be pooled, which undermines the point of a mixed effects model for word.
- High-frequency words are shown to behave differently in a number of studies.
  - So, if they show distinctive random effects, is this due to their lexical identity or their lexical frequency.
- Hence the use of random effects for lexical item is of dubious utility. If you really expect lexical idiosyncrasy, make this the focus of your study, not a collateral feature of your model.

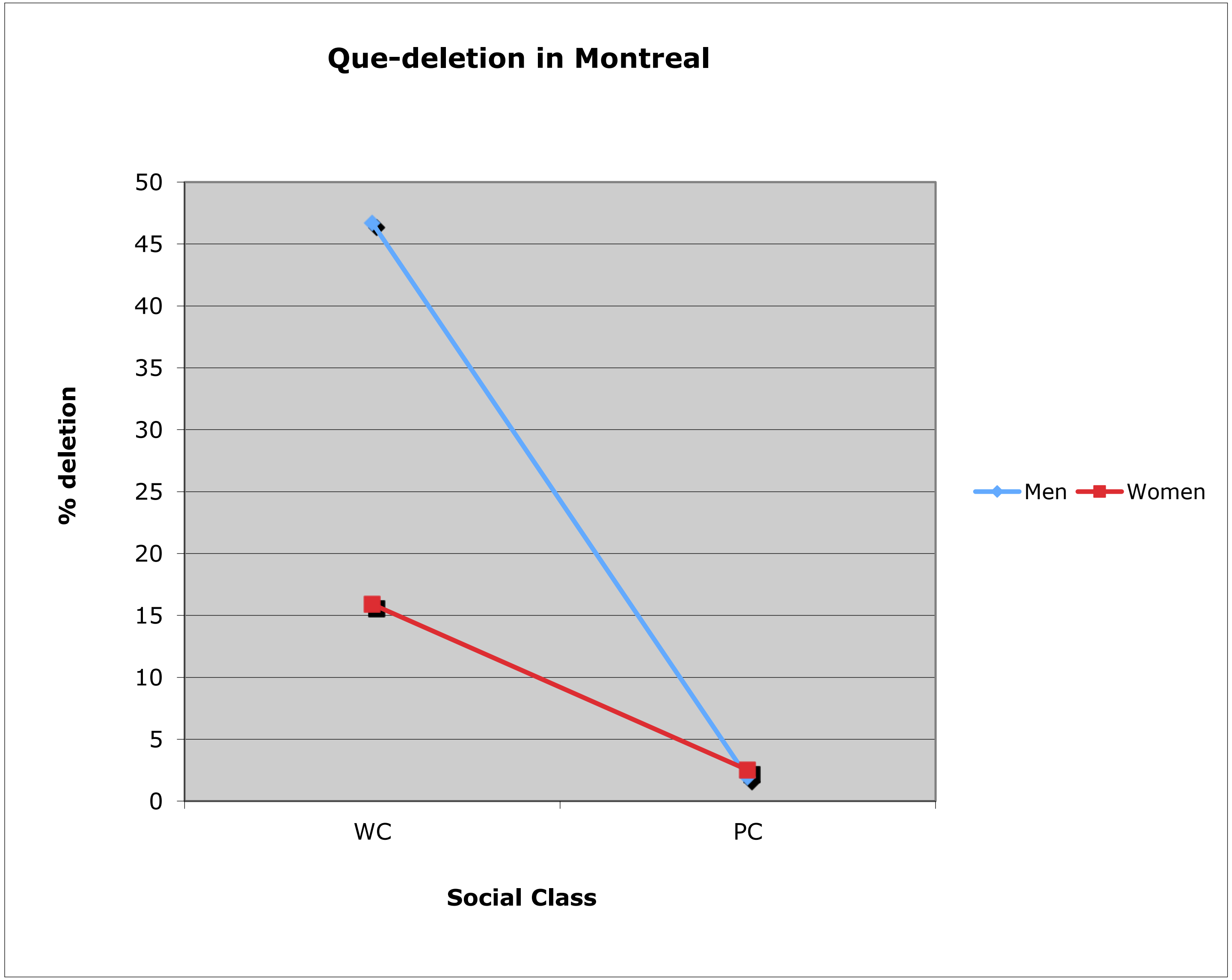
Reporting random effects

- Strive for N ≥ 30 for every random factor (word or speaker)
- Always report Ns and summary statistics (means or percentages) for each random effect. (Paolillo 2017: "At an absolute minimum, the standard deviations of all random effects should be reported.")
- Inspect the results for random factors, and report:
  - are they meaningful?
  - are they fairly normally distributed?
  - are they significant?
  - are some words or speakers really substantially different from others?
  - if so, which ones, and why?
  - How much of the variance do they explain? If it's a lot, the fixed effects aren't doing much (your variability isn't very orderly); if it's very little, there's not much point in using a mixed model
  - Occam's Razor: use the fewest explanatory principles necessary!

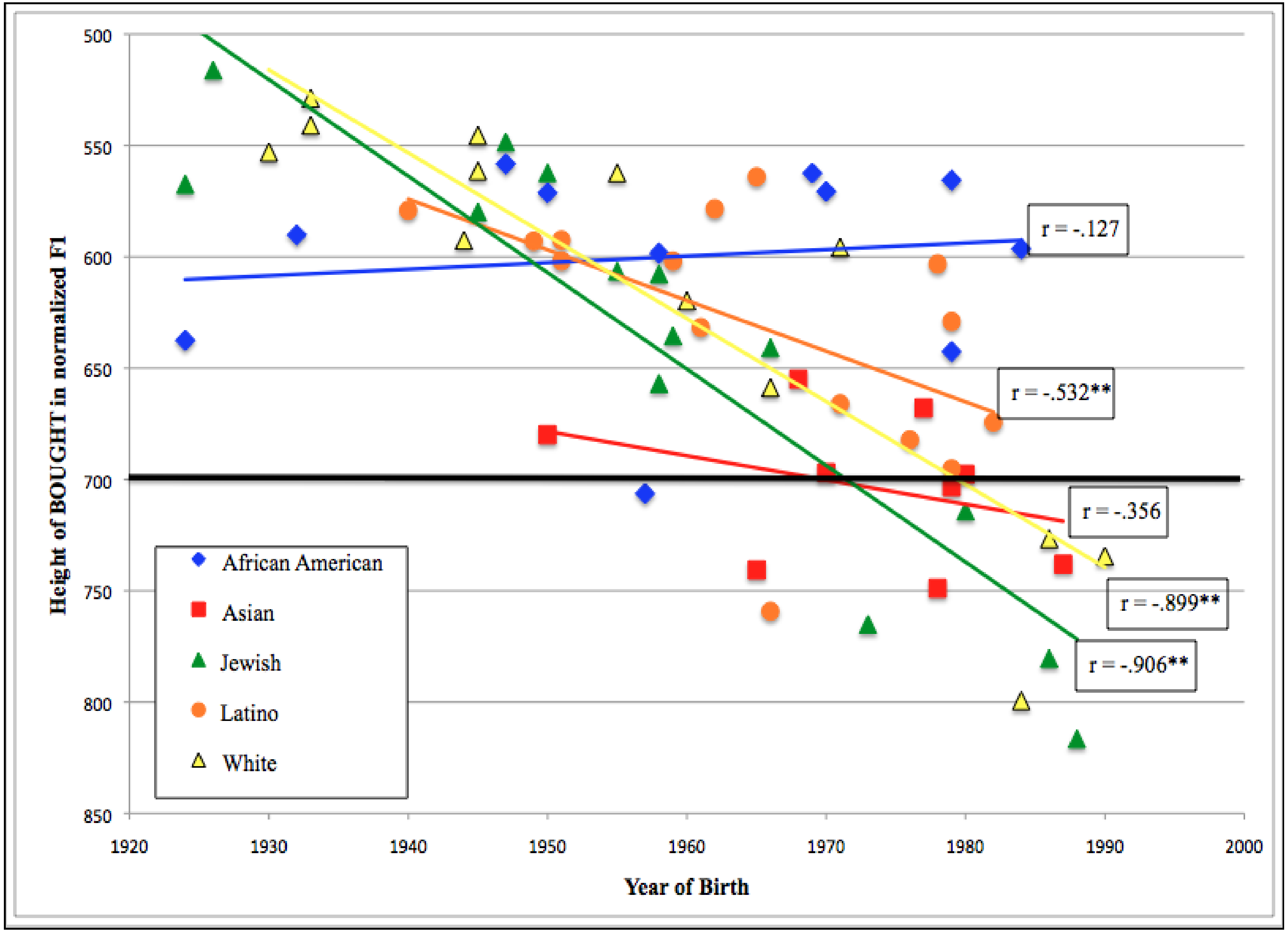
Interactions

- Interactions are important to identify but hard to interpret and understand
- By all means, test for them, but purposefully: **what is the implicit hypothesis for each interaction you test for?**
- Three-way interactions are very hard to interpret.
  - What hypothesis motivated you to test for it?
  - Do you understand what it means?
  - Can you explain it?
- Testing lots of possible interactions increases the likelihood of getting a spurious ‘significant’ effect (at the $p = .05$ level, you should get one false positive for every 20 tests.)

Interaction: Class and gender



Interaction between age and ethnicity: BOUGHT in NYC: Becker 2010



Examples

Bouchard, 2017

Table 1. Significant linguistic factors for the use of null subject

(Intercept=1.48; N=4512; [Ø]=68.5%)

	Estimate	p-value	Factor weight	%[Ø]	N-total
Type of clause (vs. coordinate clause)			0.52	68.7	470
main clause	0.08	0.50	0.54	71.2	3170
subordinate clause	-0.35	<0.01**	0.44	58.4	872
			range: 0.10		
Priming effects (vs. no priming)			0.47	64.8	1009
full NP	0.17	0.17	0.52	63.1	604
null subject	0.43	<0.001***	0.58	76.9	1975
overt subject	-0.18	0.08	0.43	58.0	924
			range: 0.15		
Morphological regularity (vs. irregular)			0.54	71.1	1822
regular	-0.29	<0.001***	0.46	66.7	2690
			range: 0.08		
Semantic content (vs. external activity verb)			0.55	72.1	2773
mental activity verb	-0.08	0.46	0.53	64.3	603
stative activity verb	-0.49	<0.001***	0.43	61.8	1136
			range: 0.12		
Person and number (vs. 1st person singular)			0.46	64.5	1595
2nd person singular	-0.06	0.86	0.45	58.7	46
3rd person singular	0.31	<0.001***	0.54	73.5	2037
1st person plural	-0.02	0.86	0.46	54.7	406
3rd person plural	0.50	<0.001***	0.59	72.4	428
			range: 0.14		
Animacy (vs animate)			0.30	67.4	4309
inanimate	1.73	<0.001***	0.70	91.1	203
			range: 0.40		
Coreferentiality (vs coreference with subject)			0.72	77.8	2491
Switch, coreference with IO	-2.04	<0.001***	0.25	33.3	21
Switch, coreference with DO	-0.82	0.01*	0.54	61.7	47
Switch, coreference with OO	-0.90	0.05*	0.52	60.9	23
Complete switch	-1.03	<0.001***	0.48	57.1	1930
			range: 0.47		

Bouchard 2017

Tagliamonte et al., 2017

YEAR	TORONTO 2002–2004			VICTORIA 2012			CHRISTCHURCH 2000–2005			PERTH 2011–2013		
input	*.75*			*.65*			*.23*			*.71*		
total N	*3,024*			*590*			*285*			*341*		
	FW	%	*N*	FW	%	*N*	FW	%	*N*	FW	%	*N*
TENSE												
HP	.71	87.9	960	.73	81.8	181	.84	57.6	33	.68	81.2	69
present	.44	66.5	412	.40	63.0	27	.34	11.1	36	.25	36.8	38
past	.29	58.7	905	.31	49.3	229	.44	26.7	150	.48	71.2	163
range	*42*			*42*			*50*			*43*		
PERSON												
first	.56	73.2	1,154	.54	61.4	259	.58	31.4	137	.64	78.8	156
third	.44	64.1	1,396	.45	55.1	254	.39	19.1	110	.33	52.7	131
range	*12*			*9*			*19*			*31*		
CONTENT												
thought	.60	64.0	583	.56	58.2	158	.68	33.3	66	[.48]	73.5	49
speech	.47	57.8	2,325	.47	50.7	420	.43	19.6	204	[.50]	54.4	287
range	*13*			*9*			*25*					
GENDER												
female	.54	64.1	2,220	.68	72.1	283	.57	27.5	193	.46	54.5	143
male	.37	43.4	804	.33	34.5	307	.35	12.0	92	.52	57.6	198
range	*17*			*35*			*22*			*6*		

TABLE 3. Predictors constraining *be like* among speakers born in the 1980s in Toronto, Victoria, Christchurch, and Perth.

Tagliamonte et al., 2017

AIC	BIC	logLik	deviance
3,556	3,664	–1761	3,522

RANDOM EFFECTS

Groups	variance	*SD*
speaker (intercept)	1.826	1.3513

Number of observations: 4,095; Groups: individual, 191

FIXED EFFECTS

	ESTIMATE	*SE*	*z*-VALUE	Pr(<|*z*|)
(intercept)	−0.2876	0.3877	−0.742	0.458213
content.thought	0.8001	0.3811	2.099	0.035790 *
person.third	1.2285	0.3298	3.725	0.000195 ***
tense.HP	−1.3854	0.3319	−4.174	3.00e-05 ***
city.PTH				
city.TO	−0.2805	0.4287	−0.654	0.512887
city.CHCH	2.8362	0.5028	5.641	1.70e-08 ***
city.VIC	0.9113	0.5299	1.720	0.085468 .
content.thought : city.PTH				
content.thought : city.TO	−1.3437	0.4152	−3.237	0.001209 **
content.thought : city.CHCH	−1.3429	0.4606	−2.916	0.003549 **
content.thought : city.VIC	−2.0199	0.4765	−4.239	2.25e-05 ***
person.third : city.PTH				
person.third : city.TO	−0.5698	0.3543	−1.608	0.107819
person.third : city.CHCH	−0.3271	0.4248	−0.770	0.441275
person.third : city.VIC	−0.4743	0.4221	−1.124	0.261185
tense.HP : city.PTH				
tense.HP : city.TO	−0.3346	0.3610	−0.927	0.354066
tense.HP : city.CHCH	0.3331	−0.4122	0.425	−0.96800
tense.HP : city.VIC	−0.6918	0.4300	−1.609	0.107643

Signif. codes: 0 '***', 0.001 '**', 0.01 '*', 0.05 '.', 0.1 ' '

TABLE 4. Mixed-effects model for *be like*, speakers born 1970–1989.

Clark & Watson, LVC 2016.
Realization of prepausal –t as /h/

TABLE 3. *Fixed effects for final model of polysyllabic words in Liverpool conversational and read speech in two age groups (older and younger speakers), n = 940*

	Estimate	Std. Error	*z* Value	*Pr*(>|z|)	Sig.	*n*	Percentage [h]	
(Intercept)	−9.67	1.96	−4.935	8.00E-07	***			
Age group (younger)	2.78	1.32	2.105	3.53E-02	*	Older: 416 Younger: 524	Older: 3.6 Younger: 39.1	
Preceding vowel stress (unstressed)	0.93	1.61	0.577	5.64E-01		Stressed: 326 Unstressed: 614	Stressed: 4.9 Unstressed: 33.2	
Lexical frequency (log)	1.28	0.39	3.227	1.25E-03	**	NA	NA	
Age group (younger) *							Older (%[h])	Younger (%[h])
Preceding vowel stress (unstressed)	3.89	1.41	2.758	5.81E-03	**	Stressed Unstressed	145 (1.4) 271 (4.8)	181 (7.7) 343 (55.7)

Links, van Kemenade & Grondelaers, LVC 2017

TABLE 4. *Logistic regression estimates: Individual resumptive adverb predictors, in three language stages*

	850–1050 (O2&O3) ($n = 4148$)			1150–1250 (M1) ($n = 435$)			1350–1639 (M3,M4&E2) ($n = 4406$)		
	Odds Ratio	n	%	Odds Ratio	n	%	Odds Ratio	n	%
Mood main clause									
Indicative	Reference category	2810	68	Reference category	333	77	Reference category	4053	92
Other	**.10*****	1338	32	.34	102	23	1.30	353	8
Mood subclause									
Indicative	Reference category	3290	79	Reference category	435	100	Reference category	4406	100
Other	1.17	858	21	—	—	—	—	—	—
Subclause-internal particle									
No	Reference category	3604	87	Reference category	432	99	Reference category	4392	100
Yes	**1.70*****	544	13	2.22	3	1	.30	14	0
Negation main clause									
No	Reference category	3518	85	Reference category	349	80	Reference category	3893	88
Yes	**.47*****	630	15	.86	86	20	.76	513	12
Negation subclause									
No	Reference category	3628	87	Reference category	360	83	Reference category	4026	97
Yes	**.61*****	520	13	.69	75	17	1.09	380	3
Subject main clause									
New	Reference category	2889	70	Reference category	312	72	Reference category	3552	81
Given	**.58*****	306	7	1.47	26	6	**.66***	278	6
Other	**.13*****	851	21	.59	88	20	**.50***	443	10
Proper	1.62	102	2	4.09	9	2	.57	133	3
Subject subclause									
New	Reference category	3141	76	Reference category	349	80	Reference category	3328	76
Given	1.18	704	17	.83	64	15	.91	602	14
Other	**2.26****	94	2	.00	10	2	.89	102	2
Proper	**2.52*****	209	5	.99	12	3	**.50****	374	8

TABLE 4. *Continued*

	850–1050 (O2&O3) ($n = 4148$)			1150–1250 (M1) ($n = 435$)			1350–1639 (M3,M4&E2) ($n = 4406$)		
	Odds Ratio	*n*	%	Odds Ratio	*n*	%	Odds Ratio	*n*	%
Subclause weight	**1.11****	Continuous variable		1.05	Continuous variable		1.03	Continuous variable	
No constituents in subclause	**.91***	Continuous variable		.93	Continuous variable		1.02	Continuous variable	
Text type									
Moral	Reference category	1234	30	Reference category	373	86	Reference category	1114	25
Law	.73	519	12	—	—	—	1.08	347	8
Narrative	1.06	708	17	1.81	62	14	**.57*****	**1267**	**29**
Science	**1.39***	646	16	—	—	—	**1.67*****	**721**	**16**
Ego document	—	—	—	—	—	—	**.46****	**452**	**10**
Bible	**.27*****	1041	25	—	—	—	**.56***	**423**	**10**
Spoken	—	—	—	—	—	—	**.21***	**82**	**2**
Translation									
No	Reference category	1438	35	Reference category			Reference category	3289	75
Yes	**1.36****	2710	65	—	—	—	1.18	1117	25
Intercept	**1.54***			**.23***			**.15*****		
Model's chi-square	1570.95 (*df* = 18)***			34.23 (*df* = 13)**			269.09 (*df* = 19)***		
Nagelkerke's R^2	.42			.12			.12		
Amount of potentials (correlatives)	4148 (1764)			435 (81)			4406 (498)		

Note: Predicted odds are for the use of a resumptive adverb. Bold values are statistically significant. *$p < .05$, **$p < .01$, ***$p < .001$.

Burnett, Koopman & Tagliamonte, LVC 2018

TABLE 6. *Binomial mixed-effect regression model predicting "any" negation*

AIC: 215
BIC: 2476
Log likelihood: –100
Deviance: 199
df: 392

Random effects					
Groups	Name	Variance	SD		
Speaker	(Intercept)	1.05	1.02		
Model information					
Observations, *n*	400	Individuals, *n*	81	Overall proportion	12% *any* negation
Fixed effects	Estimate	SE	Pr(>|z|)	*n*/cell	% *any* negation
(Intercept)	–.653	.683	.3393		
Predictors					
Verb					
be (reference level)				66	8
Existential	1.516	.464	.0011**	334	32
Syntactic domain					
Higher than VP (reference level)				31	58
Lower than VP	3.620	.618	4.5e–09***	351	6
Widening	.806	.753	.2845	18	39
Gender					
Female (reference level)				242	12
Male	.196	.521	.7072	158	11
Age					
Older	–1.734	.645	–.0017**	137	8
Middle aged (reference level)				165	14
Young	–.893	.667	.1877	98	14

Note: Significance codes: *$p < .01$; **$p < .001$; ***$p < .0001$. AIC, Akaike information criterion; BIC, Bayesian information criterion.

One more thing: Graphics

- While tabular presentations of data have great virtues, notably digital specificity, graphical representations are often more effective at communicating a finding.
  - They are literally 'graphic'!
- Examples: vowel spaces and other acoustic measures, temporal trends, geographic trends, proportions, box-plots for means and standard deviations...
  - Becker's BOUGHT data: interaction between ethnicity and time
- Consider carefully what the most effective representation of your data is.
  - It may mean using both tables and figures.

Interaction between age and ethnicity: BOUGHT in NYC: Becker 2010