



‘She said {that/∅} she couldn’t take a complement’:

Complementizer *that* omission in American English

**Isaac Bleaman, Daniel Duncan, Shelley Feuer, Gregory Guy,
Zachary Jagers, and Matthew Stuck
NWAV 43, 10/24/2014**

Introduction to the variable

Complementizer *that* omission:

1. If I ever felt *that* he was in jeopardy... (Buckeye/S01)
 2. I wish Ø they would celebrate every holiday. (Buckeye/S01)
- Focus on written language: Elsness 1984, Rohdenburg 1996
 - Focus on dialects spoken outside US: Tagliamonte and Smith 2005, Torres Cacoullos and Walker 2009

Overview

- Linguistic and nonlinguistic constraints documented in the literature
- Methodology
 - Corpus
 - Factor groups
- Analysis
- Conclusions

Previous analyses

- Kroch and Small 1978

That Deletion, Overall Results ($N = 506$)

	Speaker				
	Caller		Host/Guest		
	\emptyset	<i>that</i>	\emptyset	<i>that</i>	
<i>think</i>	$\frac{112}{126}$ (89%)	$\frac{14}{126}$ (11%)	$\frac{122}{155}$ (79%)	$\frac{33}{155}$ (21%)	$p < .05$
Other verbs	$\frac{140}{233}$ (60%)	$\frac{93}{233}$ (40%)	$\frac{64}{173}$ (37%)	$\frac{109}{173}$ (63%)	$p < .001$

Previous analyses

- Kroch and Small 1978

Effect of Verb Origin on *That* Deletion ($N = 502$)

Verb origin	All speakers	
	ϕ	<i>that</i>
Germanic (not including <i>think</i>)	$\frac{208}{442}$ (47%)	$\frac{234}{442}$ (53%)
Romance	$\frac{10}{60}$ (17%)	$\frac{50}{60}$ (83%)

$p < .001$

Previous analyses

- Thompson and Mulac (1991)

Occurrence of *that* with *I* vs. *you* vs. other subjects.

	- that		+ that		Total	
I	955	(90%)	110	(10%)	1065	(100%)
you	55	(91%)	6	(9%)	61	(100%)
other	102	(64%)	59	(36%)	161	(100%)

Chi-square (2, $N = 1287$) = 79.19, $p < 0.0001$

Occurrence of *that* with *think* vs. *guess* vs. all other verbs.

	- that		+ that		Total	
think	622	(91%)	61	(9%)	683	(100%)
guess	148	(99%)	2	(1%)	150	(100%)
other	342	(76%)	112	(24%)	454	(100%)

Chi-square (2, $N = 1287$) = 79.24, $p < 0.0001$

Previous analyses

- Rohdenburg (1996): “complexity principle”

Intervening material:

3. It didn't really yknow **register** in my mind **that** this will yknow not ruin...
4. They **know** full well **that** we had all the evidence...

- Tagliamonte and Smith's (2005) study of zero complementizer in English dialects spoken along the Irish Sea sought to test Rohdenburg's “complexity principle” and Thompson and Mulac's “epistemic parenthetical” account.

Our study: corpus and data

- Data taken from Buckeye Corpus (Pitt *et al.* 2007) of speakers from Columbus, OH area
 - 16 speakers chosen, balanced for sex and age (y<30, o>40)
- Extracted all relevant complement clause tokens (n>1200) and coded for overt *that* or zero complementizer
- Excluded tokens:
 - *yknow*
 - other complementizers (e.g., *I feel like he's a good guy.*)
 - data that would otherwise violate the “that-trace effect”
 - Who do you think (*that) he saw ____?

Our study: internal factors

- Matrix clause:
 - Person (1st, 2nd, 3rd -- regardless of number)
 - *say, mean, think*, other Germanic, Latinate
 - Copular constructions (e.g., *be happy [that]*);
aux/modal/infinitive in matrix verb
- Matrix verb frequency in data set (three categories)
- Matrix verb root length (monosyllabic vs. polysyllabic)
- Complexity factors (beyond the aforementioned)
 - Distance (adjacent of non-adjacent)
 - Coreferentiality of subjects
- Negation in matrix clause
- Perseveration (parallelism, persistence, etc.)

Overall results

- In general, omission is the favored variant.

Zero	Overt	Total
928	294	1222
75.9%	24.1%	

Multivariate Analysis

- Analysis using Goldvarb X (Sankoff, Tagliamonte and Smith 2005)
- Lexical item/lexical frequency more or less same, program had difficulty selecting between the two

Multivariate Analysis

Say	Mean	Think	Germanic	Romance
132	118	557	343	72
1	2	3	4	5
557	250	167	148	101

Multivariate Analysis

- Analysis using Goldvarb X (Sankoff, Tagliamonte and Smith 2005)
- Lexical item/lexical frequency more or less same, program had difficulty selecting between the two
- Seven total factors selected
- Age not significant by itself, but selected by model

Goldvarb Model (Lexical Effects)

<i>Input</i>	.803	Total N=1222	
Lexical Item	Weight	% Omitted	N
Think	.574	86.2%	557
Mean	.666	92.4%	118
Say	.572	75.8%	132
Germanic	.339	60.3%	343
Romance	.310	44.4%	72
<i>Range</i>	.356		
Syllable Count			
Monosyllabic	.518	79.1%	1118
Polysyllabic	.318	42.3%	104
<i>Range</i>	.200		

Goldvarb Model (Syntactic Effects)

Tense/Mood/Aspect	Weight	% Deleted	N
Present	0.544	81.9%	792
Past	0.513	67.1%	152
Do-support	0.484	80.6%	124
Aux/Modal/Infinitive	0.286	50.0%	154
<i>Range</i>	0.258		
Adjacency			
Adjacent	0.53	79.0%	1120
Distant	0.25	42.2%	102
<i>Range</i>	0.28		

Goldvarb Model (Syntactic Effects)

Coreferentiality			
Co-refer	0.662	83.7%	209
No co-refer	0.465	74.3%	1013
<i>Range</i>	0.197		
Matrix Subject			
1st	0.55	81.7%	915
2nd	0.54	68.3%	63
3rd	0.32	56.1%	244
<i>Range</i>	0.22		

Goldvarb Model (Social Effects)

Speaker Age			
Older	.440	73.5%	709
Younger	.582	79.3%	513
<i>Range</i>	.142		

A puzzle

- Perseveration significant as individual factor ($p=.0021$)
- Not selected as significant factor in multivariate analysis

	Isolated	Previous THAT	Previous Zero	Total
Zero	614	56	258	928
Overt	196	34	64	294
%Zero	75.8	62.2	80.1	75.9

Discussion

- Factor weights for common verbs strongly favor omission
 - *Mean* higher than *say, think*
 - Evidence of collocation?
- Adding complexity favors overt complementizer
- Results largely corroborate Tagliamonte and Smith 2005
 - Variation does *not* appear to be regional

Discussion

- What role does age play?
 - Younger speakers favor more than older
 - Do younger speakers use more epistemic parentheticals?
- Why doesn't perseveration seem to actually have an effect?
- Further research needed on both fronts

References

- Elsness, Johan. 1984. That or zero? A look at the choice of object clause connective in a corpus of American English. *English Studies* 65:519-533.
- Kaltenböck, Günther. 2009. Initial *I think*: Main or comment clause? *Discourse and Interaction* 2(1):49-70.
- Kroch, Anthony, and Cathy Small. 1978. Grammatical ideology and its effect on speech. In *Linguistic variation: Models and methods*, ed. David Sankoff, 44–55. New York: Academic Press.
- Pitt, M.A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E. and Fosler-Lussier, E. 2007. Buckeye Corpus of Conversational Speech (2nd release) [www.buckeyecorpus.osu.edu] Columbus, OH: Department of Psychology, Ohio State University (Distributor).
- Rohdenburg, G. (1996) Cognitive complexity and increased grammatical explicitness in English. *Cognitive Linguistics*, 7(2), 149-182.
- Sankoff, David, Sali A. Tagliamonte, and Eric Smith. 2005. *Goldvarb X: A variable rule application for Macintosh and Windows*. Department of Linguistics, University of Toronto.
- Scherre, Maria Marta Pereira, & Anthony J. Naro. 1991. Marking in discourse: 'Birds of a feather'. *Language Variation and Change* 3(1):23-32.
- Szmrecsányi, Benedikt M. 2004. On operationalizing syntactic complexity. *Journées internationales d'Analyse statistique des Données Textuelles* 7:1031-1038.
- Tagliamonte, Sali, and Jennifer Smith. 2005. No momentary fancy! The zero 'complementizer' in English dialects. *English Language and Linguistics* 9:289–309.
- Thompson, Sandra A., and Anthony Mulac. 1991. The discourse conditions for the use of the complementizer *that* in conversational English. *Journal of Pragmatics* 15:237-251.
- Torres Cacoullou, Rena, and James Walker. 2009. On the persistence of grammar in discourse formulas: a variationist study of *that*. *Linguistics* 47:1-43.