

Doing Things with Words: an update

Liorah Golomb, Humanities Librarian
The University of Oklahoma





The project: a computer-assisted analysis of the dialogue of the American television series *Supernatural* (The CW, 2005-present)



The state of the project: Not Done



What's Taking So Damn Long?



Numerous false starts in creating the corpora



Lack of any knowledge about corpus linguistics



Finding an appropriate (free) reference corpus



Kept changing my research questions



They kept making new episodes



It's a complicated show!



The Good News



Real corpus linguistics scholars have done similar work



I've pretty much nailed down the questions



Decisions have been made regarding content



Found a great free [reference corpus](#)



Writing has begun



It's summer!



And so I asked myself ...



How large is the SPN vocabulary?



How many rare ($<n$) adjectives, adverbs, nouns, verbs are used?



What are the most and least frequent n -grams (3? 4? 5?)



How has the language changed over the seasons?



And I also asked myself things that required other corpora, such as:



How does Sam's vocabulary compare to Dean's?



How does the language of SPN compare to that of related shows (a sampling; use maybe 5 eps from American Horror Story, TVD, TWD, iZombie)?



What are some of the synonyms, e.g., how many ways to express "to kill"?



How does the vocabulary of SPN compare to American English as spoken in films and tv (SUBTLEXus), i.e., rarest words in SPN cf SUBTLEXus?

JESS
We were supposed to be there like fifteen minutes ago.

JESS walks off.

JESS

Sam!

MUSIC

I'm a human cannonball

JESS

You coming or what?

Starring JARED PADALECKI

A YOUNG MAN pokes his head around the corner, this is SAM. He's wearing

SAM

Do I have to?

JESS

Yes!

MUSIC

I'm gonna fly high

JESS

It'll be fun.

SAM comes into the room.

JESS

And where's your costume?

MUSIC

I'm gonna fall fall fall

SAM laughs and ducks his head.

JENSEN ACKLES

SAM

You know how I feel about Halloween.

PARTY

INT. BAR – NIGHT

Fan-made transcript

29

00:05:12,156 --> 00:05:14,261

We were supposed to be
there like 15 minutes ago.

30

00:05:15,282 --> 00:05:17,041

Sam, you coming or what?

31

00:05:17,633 --> 00:05:21,322

Do I have to? Yes. It'll be fun.

32

00:05:22,306 --> 00:05:24,051

And where's your costume?

33

00:05:26,690 --> 00:05:29,059

You know how I feel about Halloween.

Tvsubtitles.net

JESS

Sam!

Get a move on, would you?

We were supposed to be there like fifteen minutes ago.

Sam!

You coming or what?

SAM

Do I have to?

JESS

Yes!

It'll be fun.

And where's your costume?

SAM

You know how I feel about Halloween.

SPN Corpus

Bloody Mary manifests as Sam's reflection
Sam is enraged under the influence of the ghost of
Sanford Ellicot

SAM IS POSSESSED BY MEG

For most of ep, Sam is an AU version

SAM IS INFECTED BY A SIREN

Zachariah sends Sam and Dean into an AU

Sam is hallucinating as a result of withdrawing from demon blood. He
hallucinates Dean and also a younger version of himself.

Body swap, Sam for teenaged Gary

Lucifer in Sam

SOULLESS!SAM

LEVIATHAN "SAM"

Sam influenced by love potion

Demon pretending to be Sam

Sam's in a coma and action is taking place in his mind

Ezekiel in Sam

Ezekiel in Sam

Ezekiel in Sam/Sam possessed by Wicked Witch

Ezekiel in Sam

Ezekiel/Gadreel in Sam

Gadreel in Sam

Shapeshifter takes Dean's form

"Dean" phones Sam and tells him he's in the basement

Dean is placed in an AU by a Djinn

Dean encounters a dream version of himself under the
influence of African dream root

DEAN IS INFECTED BY A SIREN

Zachariah sends Sam and Dean into an AU

Sam is hallucinating as a result of withdrawing from demon blood

Fake voicemail to Sam

Future!Dean

Sam hallucinates Dean

LEVIATHAN "DEAN"

Ghost-possessed Dean

Demon pretending to be Dean

Sam's in a coma and action is taking place in his mind

Dean possessed by Wicked Witch

[AntConc](#) set of concordance tools: create a word list with frequency, find KWIC, link back to file, finds collocates, can compare one corpus to another.



Word Types: 19796 Word Tokens: 844458 Search Hits: 0

Rank	Freq	Word
1	1	aaaaaaaaaaaaaaaaah
2	1	aaaaaaaaaaaaahh
3	1	aaaaaaaaaaaaahh
4	1	aaaaaaaaaaaaahh
5	1	aaaaaaaah
6	1	aaaaaaaahh
7	1	aaaaah
8	1	aaaaahhh
9	1	aaaaahhhh
10	1	aaaaargh
11	1	aaaagh
12	1	aaargh
13	1	aagh
14	1	aaah
15	1	aarp
16	1	aba
17	1	abacus
18	1	abaddons
19	1	abandonment
20	1	abandons
21	1	abbinanti
22	1	abby
23	1	abducting
24	1	abe
25	1	abramage
26	1	abercrombie
27	1	abetting
28	1	abhorrent
29	1	abi
30	1	abiding
31	1	abite
32	1	abode

SCREAM ENABLED!!

Word Types: 19783 Word Tokens: 844445 Search Hits: 0

Rank	Freq	Word
1	1	aaah
2	1	aarp
3	1	aba
4	1	abacus
5	1	abaddons
6	1	abandonment
7	1	abandons
8	1	abbinanti
9	1	abby
10	1	abducting
11	1	abe
12	1	abramage
13	1	abercrombie
14	1	abetting
15	1	abhorrent
16	1	abi
17	1	abiding
18	1	abite
19	1	abode
20	1	abortions
21	1	aboveboard
22	1	abracadabra
23	1	abrading
24	1	abramelin
25	1	abrasions
26	1	abrigo
27	1	abroad
28	1	abrupt
29	1	absently
30	1	absinthe
31	1	absolve
32	1	absorb

SCREAM DISABLED

All words sorted by # of occurrences, 1-10

	A	B	C	D	E	F	G	
1	#words to	words occurring 10 times	words occurring 9 times	words occurring 8 times	words occurring 7 times	words occurring 6 times	words occurring 5 times	words o
2	19796	#245	#314	#341	#414	#494	#665	#1029
3		abilities	ab	abducted	abduction	aaah	abductions	acciden
4		ability	abandon	abraham	adoption	accidentally	abomination	acknow
5		ackles	advantage	actions	adversarii	ace	absent	acres
6		admire	agreement	actor	aldo	active	absolute	addictic
7		adopted	ahem	adrenaline	alien	acts	accent	adding
8		ajaw	aid	adults	amulet	additive	accepted	adios
9		amazon	aim	alcohol	angle	adult	accurate	adjurar
10		ambulance	alan	aliases	arrangement	affairs	aces	adjust
11		appeared	aliens	altar	asian	african	acted	advise
12		apr	alike	alter	assignment	aged	actors	agony
13		arms	allergic	angelic	assured	airplane	adjacent	aidan
14		attendant	anyhow	answered	autopsy	alto	adventures	airport
15		attorney	appears	anthony	awol	amateur	afghanistan	allan
16		banging	appointment	antique	babysit	amazons	afterlife	allies
17		barb	armageddon	archives	backwards	amnesia	ahold	alligator
18		beers	asap	arguing	balance	amongst	albert	allman
19		beloved	ashamed	armed	barista	amount	album	alongsic
20		bigfoot	associate	arrogant	bash	andover	alcoholic	amate
21		bills	assure	arsenal	bay	ankle	alibi	amateu
22		binding	attached	asks	beast	app	allows	ambyr
23		blend	attacker	assigned	begins	apparition	ames	amigo
24		blows	authority	assistance	bid	apples	annual	andersc

Corpus Files

101 pilot.txt
 102 wendigo.txt
 103 dead in the wa
 104 phantom travel
 105 bloody mary.tx
 106 skin.txt
 107 hook man.txt
 108 bugs.txt
 109 home.txt
 110 asylum.txt
 111 scarecrow.txt
 112 faith.txt
 113 route 666.txt
 114 nightmare.txt
 115 the benders.tx
 116 shadow.txt
 117 hell house.txt
 118 something wick
 119 provenance.txt
 120 dead man's blc
 121 salvation.txt
 122 devil's trap.txt
 201 in my time of d
 202 everybody love
 203 bloodlust.txt
 204 children shoulc
 205 simon said.txt
 206 no exit.txt
 207 the usual suspe

Total No.

214

Files Processed

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Concordance Hits 73

Hit	KWIC	File
1	Yeah, for now. DEAN How we gonna waste her? SAM I don't know, she	119 provena
2	Winchester in over a year. You're wasting your time Even if I did know	121 salvatio
3	ever cared about. DEAN Sam I wanna waste it. I do. Okay? But it's	121 salvatio
4	be furious. JOHN: What? DEAN: That I wasted a bullet. He wouldn't be proud	122 devil's t
5	, too, huh? Oh wait, I forgot. I wasted 'em. SAM: Dean! No! DEAN: Dad! Dad,	122 devil's t
6	I've got fifty one hours to waste. Maybe tonight we should, uh... No, you	202 everybo
7	't really care. We find 'em, we waste 'em. SAM Why? DEAN What part of '	203 bloodlu:
8	right? GORDON You're right. I'm wasting my time here. This bitch will never	203 bloodlu:
9	know, but uh... ANGELA Haven't we wasted enough time on him already? Come sit	204 children
10	always wanted? DEAN We can't just waste it with a head shot? SAM Dude.	204 children
11	doctor was squeaky clean, why would Andy waste him? DEAN If it is Andy. SAM	205 simon s
12	life. SAM Bottom line, last night, he wasted somebody. DEAN Yeah, but he's not	205 simon s
13	've ever tussled with wanted to be wasted, so why the hell would Claire lead	207 the usua
14	mean? I mean, I'm not gonna waste an innocent man. I wouldn't! SAM	209 croatoar
15	're not careful you will have to waste me one day, Dean. DEAN I never	210 hunted.t
16	know what I mean. DEAN You were wasted. SAM But you weren't. And you	211 playthin
17	I could push you far enough to waste Sam. Should've known you wouldn't	214 born un

Search Term Words Case Regex

wast*

Advanced

Search Window Size

50

Start

Stop

Sort

Kwic Sort

 Level 1 1R Level 2 2R Level 3 3R

Clone Results

Corpus Files

101 pilot.txt
 102 wendigo.txt
 103 dead in the wa
 104 phantom travel
 105 bloody mary.tx
 106 skin.txt
 107 hook man.txt
 108 bugs.txt
 109 home.txt
 110 asylum.txt
 111 scarecrow.txt
 112 faith.txt
 113 route 666.txt
 114 nightmare.txt
 115 the benders.txt
 116 shadow.txt
 117 hell house.txt
 118 something wick
 119 provenance.txt
 120 dead man's blk
 121 salvation.txt
 122 devil's trap.txt
 201 in my time of d
 202 everybody love
 203 bloodlust.txt
 204 children shoulc
 205 simon said.txt
 206 no exit.txt
 207 the usual suspe

Total No.
214

Files Processed

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Total No. of Collocate Types: 125

Total No. of Collocate Tokens: 292

Freq	Freq(L)	Freq(R)	Stat	Collocate
21	2	19	8.23808	time
16	14	2	4.52757	to
14	12	2	3.14783	you
11	11	0	5.27181	re
11	4	7	3.99741	a
10	10	0	4.59362	we
10	1	9	3.40135	the
8	1	7	3.26904	sam
7	5	2	2.31381	i
6	0	6	4.88361	your
6	0	6	5.03485	my
6	2	4	3.09858	it
5	5	0	3.28267	t
5	1	4	3.66248	of
5	5	0	4.48013	not
5	0	5	2.26880	dean
4	4	0	8.29898	wanna

Search Term Words Case Regex

wast*

Advanced

Start

Stop

Sort

Sort by Invert Order

Sort by Freq

Window Span Same

From... 2L To... 2R

Min. Collocate Frequency

1

Clone Results

Corpus Files

- 101 pilot.txt
- 102 wendigo.txt
- 103 dead in the wa
- 104 phantom travel
- 105 bloody mary.tx
- 106 skin.txt
- 107 hook man.txt
- 108 bugs.txt
- 109 home.txt
- 110 asylum.txt
- 111 scarecrow.txt
- 112 faith.txt
- 113 route 666.txt
- 114 nightmare.txt
- 115 the benders.txt
- 116 shadow.txt
- 117 hell house.txt
- 118 something wick
- 119 provenance.txt
- 120 dead man's blk
- 121 salvation.txt
- 122 devil's trap.txt
- 201 in my time of d
- 202 everybody love
- 203 bloodlust.txt
- 204 children shoulc
- 205 simon said.txt
- 206 no exit.txt
- 207 the usual suspe
- 208 crossroad blue
- 209 croatoan.txt
- 210 hunted.txt
- 211 playthings.txt
- 212 nightshifter.txt
- 213 houses of the f
- 214 born under a b
- 215 tall tales.txt
- 216 roadkill.txt
- 217 heart.txt
- 218 hollywood bab
- 219 folsom prison l
- 220 what is and wh

Total No.
214
Files Processed

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Total No. of N-Gram Types 540185 **Total No. of N-Gram Tokens** 844030

Rank	Freq	Range	N-gram
13006	5	5	bobby sam s
13007	5	5	bobby then you
13008	5	5	bobby they re
13009	5	5	bobby what dean
13010	5	4	bobby where s
13011	5	5	bobby you boys
13012	5	5	body was found
13013	5	5	bone of a
13014	5	4	bottom of the
13015	5	5	boy i m
13016	5	5	boy sam i
13017	5	5	boy you re
13018	5	3	break the final
13019	5	5	breaking and entering
13020	5	3	brian come on
13021	5	3	brian it s
13022	5	3	bridge when we
13023	5	5	bring it on
13024	5	5	bring them back
13025	5	5	brother dean yeah
13026	5	5	brother sam you
13027	5	5	brother you re
13028	5	4	brought me here
13029	5	5	buried in the
13030	5	5	but dean but

Words Case Regex N-Grams **N-Gram Size**
 Min. Max.

Search Term Invert Order **Search Term Position**

 On Left On Right

Clone Results

Corpus Files

101 pilot.txt
 102 wendigo.txt
 103 dead in the wa
 104 phantom travel
 105 bloody mary.tx
 106 skin.txt
 107 hook man.txt
 108 bugs.txt
 109 home.txt
 110 asylum.txt
 111 scarecrow.txt
 112 faith.txt
 113 route 666.txt
 114 nightmare.txt
 115 the benders.txt
 116 shadow.txt
 117 hell house.txt
 118 something wick
 119 provenance.txt
 120 dead man's blc
 121 salvation.txt
 122 devil's trap.txt
 201 in my time of d
 202 everybody love
 203 bloodlust.txt
 204 children shoulc
 205 simon said.txt
 206 no exit.txt
 207 the usual suspe
 208 crossroad blue
 209 croatoan.txt
 210 hunted.txt
 211 playthings.txt
 212 nightshifter.txt
 213 houses of the f
 214 born under a b
 215 tall tales.txt
 216 roadkill.txt
 217 heart.txt
 218 hollywood bab
 219 folsom prison l
 220 what is and wh

Total No.

214

Files Processed

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Concordance Hits 460

Hit	KWIC	File
1	. Or not. Get off of me. SAM What the hell are you doing here? DEAN	101 pilot.txt
2	, I was looking for a beer. SAM What the hell are you doing here? DEAN	101 pilot.txt
3	' right here till you tell me exactly what the hell that means. SAM Hi. Are	101 pilot.txt
4	together. JOSEPH Well, I don't know what the hell kinda story you're working	101 pilot.txt
5	in the middle of a job? Dean, what the hell is going on? DEAN Sam?	101 pilot.txt
6	my GPS and my satellite phone. HALEY What the hell is going on? SAM It'	102 wendigo
7	drowned? SAM Yep. In the sink. DEAN What the hell? So you're right, this	103 dead in
8	come out. JAKE I don't know what the hell you're talking about. DEAN	103 dead in
9	Sweeney thirty-five years ago. That's what the hell I'm talking about. ANDREA	103 dead in
10	got to stretch my legs. MAX Hey, what the hell are you doing?! SAM Morning,	104 phantor
11	to come, but, you know, I thought, what the hell, I have to try. SAM:	106 skin.txt
12	! Rich, no! RICH: No, just wait here. What the hell? LORI: Rich, let's go!	107 hook m.
13	. DEAN: Thanks. You ok? SAM: Yeah. DEAN: What the hell happened? SAM: Hook Man. DE	107 hook m.
14	again, if it wasn't some disease, what the hell was it? DEAN That's	108 bugs.txt
15	family? DEAN Wait. You hear it? LARRY What the hell? DEAN All right, it's	108 bugs.txt
16	it, right? DEAN: I guess so, yeah. What the hell are you talking' about? SAM:	109 home.tx
17	? DEAN Dude! You were in there forever. What the hell were you talking about? SAM	110 asylum:
18	's no other way out. GAVIN So what the hell are we gonna do? SAM	110 asylum:
19	me in that same direction? HARLEY: Sure. What the hell? DEAN: Dude, you're fugly.	111 scarecr
20	car. Go! Go! Go! Go! STEVE: What—what the hell was that? DEAN: Don't	111 scarecr
21	. Alright...just wanted you to know. SAM What the hell are you doing here? DEAN	112 faith.txt
22	How? DEAN You know how. SAM Wait, what the hell are you talking about Dean,	112 faith.txt
23	a truck? A big black truck? RON What the hell you talking about? You even	113 route 66
24	having these dreams, I don't know what the hell is happening Dean. What. DEAN	114 nightma
25	'll be here—waiting for you. SAM: What the hell—? SAM: Dude, I gotta talk	116 shadow
26	you do. CRAIG I don't know what the hell to think, man. You guys,	117 hell hou
27	to take a picture of this. SAM What the hell would I do that for?	117 hell hou

Search Term Words Case Regex

Search Window Size

what the hell

Advanced

50

Start

Stop

Sort

Kwic Sort

 Level 1 1R Level 2 2R Level 3 3R

Clone Results

TagAnt: assigns part-of-speech tags to each word in the corpus.



TagAnt 1.2.0

File Help

Input Text Clear Results Files: 22 100%

101 pilot_tagged.txt
 102 wendigo_tagged.txt
 103 dead in the water_tagged.txt
 104 phantom traveler_tagged.txt
 105 bloody mary_tagged.txt
 106 skin_tagged.txt
 107 hook man_tagged.txt
 108 bugs_tagged.txt
 109 home_tagged.txt
 110 asylum_tagged.txt
 111 scarecrow_tagged.txt
 112 faith_tagged.txt
 113 route 666_tagged.txt
 114 nightmare_tagged.txt
 115 the benders_tagged.txt
 116 shadow_tagged.txt
 117 hell house_tagged.txt
 118 something wicked_tagged.txt
 119 provenance_tagged.txt
 120 dead man's blood_tagged.txt
 121 salvation_tagged.txt
 122 devil's trap_tagged.txt

Input Files: 22 Load Clear

101 pilot.txt
 102 wendigo.txt
 103 dead in the water.txt
 104 phantom traveler.txt
 105 bloody mary.txt
 106 skin.txt
 107 hook man.txt
 108 bugs.txt
 109 home.txt
 110 asylum.txt
 111 scarecrow.txt
 112 faith.txt

Horizontal Vertical Clear

Language English Start Stop

101 pilot_tagged - Notepad

File Edit Format View Help

1.01	CD	@card@Pilot	NN	pilot			
Written	VVN	writeby	IN	by:	:	:Eric	NP
EricKripke		NP	Kripke				
Directed	VVN	directby	IN	by:	:	:David	NP
DavidNutter		NP	Nutter				
Air	NP	AirDate	NP	Date:	:	:13	CD
@card@Sep	NN	Sep2005	CD	@card@			
MARY	NP	Mary					
Come	VVN	comeon	IN	on,	,	,let	VV
let's	VBZ	besay	VV	saygood	JJ	goodnight	NN
nightto	TO	toyour	PP\$	yourbrother		NN	
brother.	SENT	.					
DEAN	NN	dean					
'	"	'Night	NP	Night,	,	,Sam	NP
Sam.	SENT	.					
MARY	NP	Mary					
Good	JJ	goodnight	NN	night,	,	,love	NN
love.	SENT	.					
JOHN	NN	john					
Hey	UH	hey,	,	,Dean	NP	Dean.	
SENT	.	.					
DEAN	NN	dean					
Daddy	NN	daddy!	SENT	!			
JOHN	NN	john					
Hey	UH	hey,	,	,buddy	NN	buddy.	
SENT	.	.					
So	RB	sowhat	WP	whatdo	VVP	doyou	PP
youthink	VVP	think?	SENT	?You	PP	youthink	VVP
thinkSammy	NP	Sammy's	VBZ	beready	JJ		
readyto	TO	totoss	VV	tossaround	RP		
arounda	DT	afootball	NN	football	RB	yet?	
SENT	?	.					

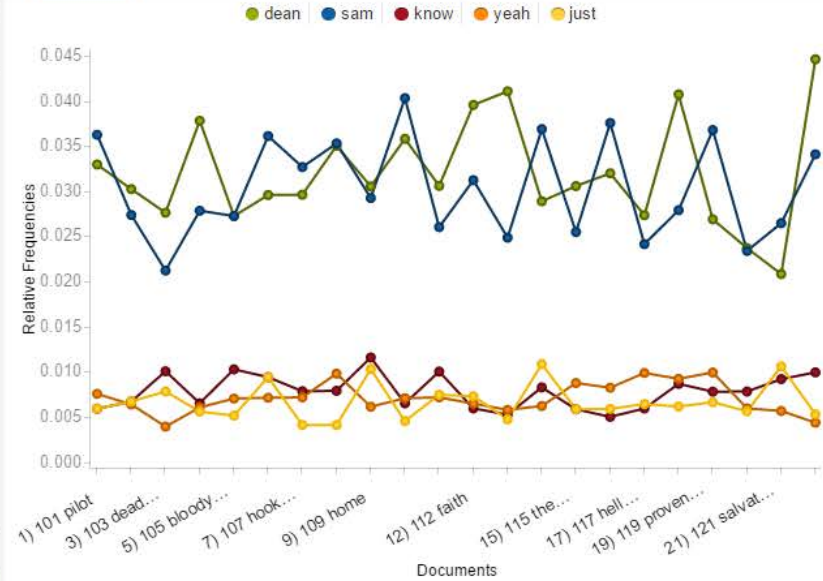
[Voyant Tools](#): does much of the same thing as AntConc, presenting the information visually.





101 pilot

1.01 Pilot
 Written by: Eric Kripke
 Directed by: David Nutter
 Air Date: 13 Sep 2005
 MARY
 Come on, let's say good night to your brother.
 DEAN
 'Night, Sam.
 MARY
 Good night, love.
 JOHN
 Hey, Dean.
 DEAN
 Daddy!
 JOHN
 Hey, buddy.
 So what do you think? You think Sammy's ready to toss around a football yet?
 DEAN
 No, Daddy.



Scale Terms

This corpus has 22 documents with 78,043 total words and 5,337 unique word forms. Created now.

- Document Length:**
- Longest: 104 phantom traveler (4313); 101 pilot (4249); 119 provenance (4243); 117 hell house (4062); 114 nightmare (3879)
 - Shortest: 115 the benders (2749); 121 salvation (2836); 107 hook man (2939); 109 home (3113); 102 wendigo (3144)

- Vocabulary Density:**
- Highest: 102 wendigo (0.262); 115 the benders (0.259); 107 hook man (0.258); 120 dead man's blood (0.250); 118 something wicked (0.240)
 - Lowest: 114 nightmare (0.205); 119 provenance (0.211); 105 bloody mary (0.214); 104 phantom traveler (0.215); 112 faith (0.216)

Most frequent words in the corpus: dean (2506); sam (2381); know (605); yeah (555); just (515)

- Most frequent words in the corpus:
- 101 pilot: constance (21), troy (13), jaffe (12), joseph (15), welch (11).
 - 102 wendigo: haley (56), roy (34), tommy (9), ranger (8), ben (10).
 - 103 dead in the water: andrea (52), lucas (37), jake (33), carlton (19), lake (23).

Items

Document	Left	Term	Right
1) 101 p...	good night to your brother.	dean	'Night, Sam. MARY Good night
1) 101 p...	Good night, love. JOHN Hey,	dean	. DEAN Daddy! JOHN Hey, buddy
1) 101 p...	night, love. JOHN Hey, Dean.	dean	Daddy! JOHN Hey, buddy. So
1) 101 p...	toss around a football yet?	dean	No, Daddy. JOHN No. MARY
1) 101 p...	Sammy. You okay? No! Mary!	dean	Daddy! JOHN Take your brother
1) 101 p...	and don't look back! Now,	dean	, go! JOHN Mary! JOHN No
1) 101 p...	go! JOHN Mary! JOHN No!	dean	It's okay, Sammy. JOHN I
1) 101 p...	you? JESS Crash and burn.	dean	Whoa, easy, tiger. SAM Dean
1) 101 p...	DEAN Whoa, easy, tiger. SAM	dean	? SAM You scared the crap
1) 101 p...	the crap out of mel	dean	That's 'cause you're out of

2,506 context expand Scale

SUPERNATURAL

Thanks for coming! I'm @LiorahLibrarian and lgolomb@ou.edu.