"Let's Go Gank Ourselves a Paris Hilton": A Textual Analysis of the Dialogue of *Supernatural* (the First 10 Years)

Liorah Golomb

It may seem counterintuitive to examine an audiovisual medium through its dialogue alone, but some linguists have been doing just that. Using a freeware concordance toolkit and other computer programs, I analyzed the dialogue of the first 10 seasons of the American television program *Supernatural* (2005-present) through several facets in an attempt to demonstrate originality, variety, and how language contributes to character individuation.[1] Using fan-made transcripts, I created three main corpora for the study: the SPN Corpus, which includes all dialogue from the first 218 episodes, and the DEAN and SAM Corpora, comprised of all of the dialogue of each of the two main characters. The $SUBTLEX_{US}$ corpus of American film and television dialogue was used as a reference corpus. A sampling of Supernatural episodes was also compared to that of other contemporary genre shows.

Background: Why *Supernatural*

> "*Supernatural* has everything. Life. Death. Resurrection. Redemption. But above all, family." (Calliope, *Supernatural* episode 10.05, "Fan Fiction")

To give the barest of summaries, *Supernatural* is an American television show about two brothers, Dean and Sam Winchester (played by Jensen Ackles and Jared Padalecki, respectively), who were raised as hunters of supernatural beings. Created by Eric Kripke, *Supernatural* has

[1] Generally, episodes referenced in this paper are identified only by season and episode number within the season, so 5.22 refers to the twenty-second episode of the fifth season. *Supernatural* is also be referred to as *SPN*.

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/317082

been running since 2005, first on the WB Network and then on its successor, the CW. It is now airing its thirteenth season.

When the series begins Sam is three years into an ivy league education and is reluctant to return to hunting, whereas Dean left high school early and never stopped hunting. The brothers travel the country in a classic Chevy Impala in pursuit of ghosts, demons, and other monstrous beings. *Supernatural* sometimes has an unusual take on monsters: ghosts, vampires, and werewolves can be good, angels are often as bad as demons, and humans can commit some of the worst acts of inhumanity. Over the years the show has dealt with issues such as the ethics of doing bad things to reach good outcomes, what constitutes a "monster," the importance of loyalty and family, free will, the nature of evil, and the existence of God.

Supernatural encompasses several genres, sometimes within a single episode. It could be classified as horror, having many truly terrifying moments; as detective procedural, as Sam and Dean take on cases where they need to identify, find, and defeat the monsters; as buddy comedy, with the focus on Sam and Dean's very different personalities; as drama, as overarching plots unfold over multiple episodes or even seasons; as fantasy, set in a universe where mythical creatures, pagan gods, and witches exist; and as metadrama—in some episodes the show calls attention to itself either directly or through an invented series of novels and the fanworks and activities inspired by them.

What This Study Does and Does Not Do

There has been a good deal of close-reading analysis of *Supernatural*, with the aired episodes being the text. In addition to numerous journal articles, at least four collections of essays have been published. *TV Goes to Hell: An Unofficial Road Map of Supernatural* (Abbott), *Supernatural and Philosophy: Metaphysics and Monsters … for Idjits* (Foresman), *Supernatural* (Zubernis and Larsen), and *Supernatural, Humanity, and the Soul: On the Highway to Hell and*

Back (George and Hansen) all discuss the show and its relationship to issues such as class, gender and sexuality, free will, theology, morality, evil, and the existence of the soul. Scholars have viewed *Supernatural* through just about every lens, including existentialism (Silva), neoliberalism (Blake), postindustrialism (Robinson), and postmodernism (Hansen).

This study takes a different approach. It is concerned with only one element of *Supernatural*: its dialogue. In a traditional thematic analysis of a television show, dialogue is only one of many elements delivering meaning to the viewer. But for this project, visual elements such as set design, lighting, camera angles, makeup, special effects, or props are not taken into account, nor are gestures, actions, weighty pauses, fight scenes, significant glances, or other actor or directorial choices. Even the element that is being examined is not placed in any context; we don't know why, when, or to whom dialogue is spoken. Rather, quantitative evidence in the form of, essentially, word lists serves as the raw data from which conclusions are drawn.

The motivation for subjecting the dialogue of *Supernatural* to textual analysis was simple: I believed that the writing of the show was superior to that of many other so-called genre shows, and that there ought to be a way to support that objectively. I believed that the characters in the show each had his or her own way of speaking, and that those ways remained largely consistent despite the fact that the show has used a large pool of writers. I also believed that *Supernatural*'s dialogue was more clever and original than that of other network shows targeting the same audience. Another deciding factor in analyzing only dialogue is that the unsaid is not easily measured, whereas computer-driven corpus analysis tools can provide word counts and frequencies, identify parts of speech, and create concordances of a written text or texts.

In recent scholarship, corpus linguists have used computer-generated corpora to analyze television dialogue. Corpus linguistics, according to Douglas Biber, is "a research approach that

facilitates empirical descriptions of language use. … [R]esearch carried out on a corpus has the goal of describing the patterns of language use in the target textual domain" (Biber 15). Biber goes on to enumerate the characteristics of corpus analysis:

"1. it is empirical, analyzing the actual patterns of language use in natural texts;

"2. it utilizes a large and principled collection of natural texts, known as a 'corpus,' as the basis for analysis;

"3. it makes extensive use of computers for analysis, using both automatic and interactive techniques;

"4. it depends on both quantitative and qualitative analytical techniques." (16)

Monika Bednarek is among the most prominent scholars in the area of analyzing television dialogue using "corpus stylistics," defined as "the application of corpus linguistic methodology to the analysis of diegetic fictional texts" ("Stability" 186). Her book-length study of the dialogue of *Gilmore Girls*, an American television program, looks at a number of different factors, including individual character's speech, the functions of words and short word clusters, and whether certain clusters or n-grams (strings of any given number of words) are over- or under-represented in comparison to natural language (*Language of Fictional Television*). In a further study of *Gilmore Girls* Bednarek examines words and word clusters from corpora of (a) the dialogue of all the episodes of *Gilmore Girls*, (b) the dialogue of a sampling of episodes from ten shows of different genres (including five episodes from the first season of *Supernatural*), and (c) various corpora of spoken, unscripted language ("Language"). Bednarek did not find that, in terms of repetition and usage frequency, *Gilmore Girls*'s dialogue was significantly different from that of her multi-genre sampling of fictional television or from natural speech. However, her research did suggest that "the language of a specific fictional television series will include aspects both unique to its character as a particular popular cultural artifact and, presumably,

aspects particular to the genre it belongs to" (*Language of Fictional Television* 72), and she continued to explore this avenue in later work ("Get Us the Hell out of Here").

Bednarek also finds computer-constructed concordances useful in character analysis. She returns to her *Gilmore Girls* corpus to determine whether the character Lorelei's speech changes over the course of the series and according to whom she is speaking ("Stability"). Analyzing the character of Sheldon from the American comedy series *The Big Bang Theory* ("Constructing 'Nerdiness'"), Bednarek is able, among other things, to categorize Sheldon's speech into various characteristics attributed to "nerds," such as his belief in his own intellectual superiority and his struggle with social skills (208-209). She concludes, "a corpus-based analysis is highly useful in demonstrating repeated patterns of behavior and may give us an impetus to further investigate particular features in more detail" (215).

In his study comparing the dialogue of the sitcom *Friends* to natural speech (Quaglio, *Television Dialogue*), Paulo Quaglio uses corpus analysis in the field of conversation studies. Using fan-made transcripts for his *Friends* corpus and the American English conversation portion of the *Longman Grammar Corpus* as a reference, Quaglio tested for language functions such as vague, emotional, and informal speech. His goal in doing so was to test the extent to which television dialogue resembles natural conversation, or indeed whether it is meant to.

Other computer-assisted scripted film and television corpus studies have been done not to learn about the programs per se, but rather to gauge their usefulness in language learning or translation studies. See, e.g., "Vocabulary Demands of Television Programs" (Webb and Rodgers), "Narrow Viewing: The Vocabulary in Related Television Programs" (Rodgers and Webb), and "The Language of Film: Corpora and Statistics in the Search for Authenticity. *Notting Hill* (1998)—A Case Study" (Taylor).

How the Corpora Were Created

Because some limits had to be set, the corpora created for this project include only dialogue from the first 10 seasons of *Supernatural*. Also, the only specific characters' speech examined is that of Sam and Dean. But there are a number of recurring characters in *Supernatural*, each with a distinctive voice. The tools and methods I have used to analyze the speech of the two principal characters can be used on any of the characters, should anyone wish to do so. The characters Bobby, Castiel, Crowley, Meg, Ruby, or Charlie would all be interesting choices.

The data for this project consists primarily of the SPN Corpus and its subcorpora, the SAM Corpus and the DEAN Corpus. The SPN Corpus contains all of the dialogue of every character for every episode of seasons 1 through 10, as collected from fan-made transcripts of the aired episodes. This is the same method used by Bednarek to create the *Gilmore Girls* corpus (*Language of Fictional Television*) and by Quaglio to create the *Friends* corpus ("Television Dialogue and Natural Conversation"). The SPN Corpus includes the speakers' identities, either by name or description (PSYCHIATRIST, OLDER MAN) as well as certain information: episode name and number, writer(s), director, and original air date. Actual scripts used in creating the show are tightly held by the copyright holder, Warner Brothers, and were not available for use. The fan-made transcripts are housed on the "Supernatural Wiki," an encyclopedic site with over 3,000 articles related to *Supernatural*. Because fans are human it is possible that there are mistakes in the transcripts; however, the wiki format helps ensure that the transcripts are accurate. And because the author is human, it is likely that there were some

mistakes made in creating the corpora from the transcripts. However, I have taken great care to avoid mistakes. Each episode is saved as a separate plain-text file—218 files in all.[2]

The SAM and DEAN corpora were derived from the SPN Corpus and consist of all dialogue spoken by Sam or Dean, individually. There are 218 text files in each corpus. Determining what to include was more challenging than it would seem, because there are a number of instances in which Sam and Dean are not, or not entirely, Sam and Dean. The brothers have been impersonated, hallucinated, thrown into an alternate reality, possessed, body-swapped, cursed, or temporarily altered in a way that affects their personalities. How should these not-Sams and not-Deans be treated for the purpose of data creation? In some cases the decision to include dialogue or not was clear. For example, a creature taking the form of Dean is clearly not Dean; a body-swapped Sam is Sam in another body, and the character who looks like Sam is not. Other cases were much less cut-and-dried. Both Sam and Dean have been possessed, poisoned, or cursed by supernatural forces which have caused them to express anger and resentment that they later pass off as being due to the supernatural force. But the history of the brothers' relationship, along with subtle visual cues, make the viewer suspect that the sentiments do in fact belong to Sam and Dean. The urge to express them can be blamed on the monster, but not the words themselves. In these circumstances I attributed the dialogue to Sam or Dean.

There are also four smaller corpora used in this project consisting of dialogue from a sampling of episodes from contemporary supernatural genre shows *American Horror Story*, *iZombie*, *The Vampire Diaries*, and *The Walking Dead*. These were created using subtitle files

[2] A noble attempt to write a program to automate corpus creation was made by Frank Branch, a graduate student at the University of Washington. Unfortunately, there were too many variables to produce the kind of accuracy this project required.

collected from tvsubtitles.net. These files contain dialogue but no speakers. A sampling of nine episodes of *Supernatural* dialogue was made from the SPN Corpus.

Tools

The corpora for this project were processed using AntConc 3.4.4w (Anthony, *AntConc*), a freeware suite of corpus analysis tools that can produce a word list, show words in context, show frequency, produce word clusters, and compare one corpus to another and TagAnt 1.2.0 (Anthony, *TagAnt*), a freeware part-of-speech tagger.

Research Questions and Findings

Many corpus linguistic studies are concerned with words and collocates that appear frequently in any given corpus. Such information may be useful in understanding how language is learned, for example. This study, however, is attempting to show the breadth and variety of language in *Supernatural* ("SPN") and therefore is more concerned with the total list of words and, in particular, words that are relatively uncommon in American television. To determine the size of the vocabulary in *Supernatural* I created a word list from every episode in the first 10 seasons of the show. To determine the degree of uncommonness, I compared the SPN corpus against $SUBTLEX_{US}$ (Brysbaert), a freely-available corpus of 74,286 words created from subtitle files from over 8,300 American films and television programs, mostly from 1990-2007. $SUBTLEX_{US}$ is an appropriate reference corpus for this project as it uses contemporary American English from the same medium that this study is concerned with. I also compared dialogue from a sampling of episodes from SPN and certain contemporary genre shows.

How Large is the Vocabulary of SPN?

Using the Word List tool of AntConc we find that the total number of distinct words used in the 218 episodes of the SPN Corpus is 20,003, the SAM Corpus is 8,281, and the DEAN Corpus is 10,129. The difference between the brothers' vocabulary size is interesting because Sam, with

his Stanford pre-law education and penchant for researching, is considered the "smart" brother, yet Dean, proud possessor of a GED, has a vocabulary 20% larger than Sam's. There are several instances of Sam being described as more intelligent than Dean: by the demon Casey in 3.04, "Sin City" (DEAN: "I got somebody coming for me and, uh … he did pay attention in class." CASEY: "Oh right—Sam. Everyone says he's the brains of the outfit."); by Lilith in 4.18, "The Monster at the End of this Book" ([to SAM] "You were always the smart one"); by Crowley in 10.22, "The Prisoner" ([to SAM]: "I thought you were the smart one"); and, poignantly, by Dean himself : "I'm a grunt, Sam. You're not. You've always been the brains of this operation." (8.14, "Trial and Error"). But if vocabulary size is an indication of intelligence, Dean is brighter than he gives himself credit for.

How does the Language of Supernatural *Compare to SUBTLEX$_{US}$?*

As noted earlier, SUBTLEX$_{US}$ contains 74,286 words collected from subtitle files from U.S. movies and television aired in or before 2007. The Supernatural corpus contains 20,003 words. Obviously, there are going to be many words—54,283 to be exact—in SUBTLEX$_{US}$ that do not occur in *Supernatural*. The more interesting question is, how many words are in *Supernatural*'s vocabulary that are not in SUBTLEX$_{US}$? Using AntConc, the SUBTLEX$_{US}$ Corpus served as a stop word list against the SPN Corpus, producing a list of 3,386 words unique to *Supernatural*. Many of these are foreign words, largely Latin, which is used frequently in the show in exorcism and other rituals, and "Enochian," the language of angels as allegedly revealed to John Dee and Edward Kelley in the late 16th century. Others are proper names. A few are typos, and some are variant spellings or relaxed pronunciation spellings (e.g., gunna, havta), or the 16 variations of sighing or screaming, from aahh to aaaaaaaaaaaaaaaaah. The SUBTLEX$_{US}$ Corpus also contains these elements, including people and place names and words like "gonna." It would require

human examination to pick out all the typos or variant spellings, foreign words, and onomatopoeias.

Short of doing a comprehensive elimination of all proper nouns, foreign words, etc., it is interesting to visually scan the list of words that appear in the SPN Corpus and not in the $SUBTLEX_{US}$ Corpus, and take note of some of them. Leaving out words meant to convey a sound, partial words that indicate the speaker was interrupted, and obvious typographical errors, these words can be divided into categories:

- common words that happen to not be in $SUBTLEX_{US}$ (e.g. eyeshadow, whiniest);
- proper nouns, including species of monsters;
- words from a foreign or invented language;
- words that refer to popular culture or current events;
- slang, whether current or outdated;
- words referring to social media or technology which were nonexistent or uncommon before $SUBTLEX_{US}$'s cutoff year of 2007 (sexting, snapchat);
- *Supernatural* neologisms.

This last category would include words that seem unique to *Supernatural*. Though I can't guarantee that they originated with the show, I have checked them against the Oxford English Dictionary online edition and the Merriam-Webster online edition, and they are either not included or they have a different meaning. In one case a word has been repurposed: the verb "to gank," meaning "to kill" in *Supernatural*, is used in gaming but does not primarily refer to killing.

"Gank" and its various forms is one of the most common and most interesting words in the *Supernatural* lexicon. It is spoken 62 times in the SPN Corpus, and with increasing frequency after its first appearance. Three other hunters and one demon use "gank," while one

civilian is perplexed by it, suggesting that the term as used to mean "kill" is unique to the world of hunters and monsters. Notably, Dean uses "gank" and its variants considerably more than Sam—41 times to Sam's eleven. This, along with Dean's use of other slang terms for "to kill" (e.g. smoke, croak, drop, ice, eighty-six, waste, ventilate) helps place Dean solidly in the hunter community. It may also reinforce a connection between hunting and the criminal underworld. Indeed, in most other contexts Sam, Dean, and hunters generally would be considered criminals. Breaking and entering, car theft, impersonating federal agents, committing credit card and computer fraud—not to mention crossing state lines with large caches of unlicensed weapons and grave desecration—are common activities for the hunters we meet in *Supernatural*.

Supernatural *Neologisms*

One hypothesis of this study is that the language of *Supernatural* is creative. Another is that the characters are written in such a way that makes them distinct. Figure 1 supports both notions. It lists *Supernatural* neologisms or new usages in context, the speaker, the episode in which the term was first used, and the episode's writer or writers.

***WORD* IN CONTEXT**	**FIRST USE SPEAKER**	**FIRST USE EPISODE**	**WRITER(S)**
And if you *affirmate* me, I'm gonna punch you in the face.	Dean	7.07	Ben Acker & Ben Blacker
Enough *Ally McBealing*.	Dean	7.04	Adam Glass
Hey! *Assbutt!*	Castiel	6.20	Ben Edlund
Well, it's gotta be better than this G-rated *assfest.*	Non-recurring character	4.07	Julie Siege

Of all the lame-ass things you've ever said, that's got to be the lame-*assiest*	Dean	8.15	Brad Bucker & Eugenie Ross-Leming
It was the *bendiest* weekend of my life.	Dean	3.02	Sera Gamble
This all sounds like sad times at *Bitchmont High*	Dean	9.15	Adam Glass
The guy Molotov-*cocktailed* himself	Dean	10.15	Jenny Klein
[Some sort of a twisted cosplay?] *Cosreal.*	Non-recurring character	9.15	Jenny Klein
I'm the *daringest* devil you've ever met, love.	Crowley	8.21	Ben Edlund
But he rejected her because she was already *dehymenated*?	Dean	6.12	Adam Glass
I told you, it's *eekish*.	Charlie	10.18	Robbie Thompson
See if any co-ed *ganked* herself there.	Dean	2.15	John Shiban
You *ganky* putrescent *skanger*.	Crowley	9.06	Robert Berens
Right now, I'm the *goodest* guy you got.	Crowley	9.10	Andrew Dabb
You can't tell me this joint doesn't give you the *heebs* and/or *jeebs*.	Dean	8.08	Andrew Dabb
I just had a 12-inch *herpe* crawl out of my ear.	Dean	6.16	Brett Matthews

There was nothing *hexy* found on him.	Dean	10.16	Brad Bucker & Eugenie Ross-Leming
You're *junkless* down there, right?	Dean	4.10	Eric Kripke
Said she wouldn't get *kabobed* if she brought you back.	Dean	9.03	Brad Bucker & Eugenie Ross-Leming
I got a gander at Dick's big plan, right before he *Lincolned* me.	Bobby	7.20	Robbie Thompson
You know, a little *manburger* helper.	Dean	4.04	Cathryn Humphris
See, so not just a robot, more of a … *Mandroid.*	Non-recurring character	2.12	Ben Edlund
I used to skipper this *meatboat* for awhile.	(Soulless) Sam	6.22	Eric Kripke
Dean Winchester's behind you, *meatsack*.	Meg	6.10	Brett Matthews
Makes my *meatsuit* all dewy.	Meg	6.10	Brett Matthews
Did you just *Molotov* my brother with holy fire?	Lucifer	5.22	Eric Kripke & Eric Gerwitz
Anything more and our *mookie* pals here may just throw you out.	Crowley	8.02	Andrew Dabb & Daniel Lofflin
So, what are we looking for? An *octovamp*? A *vamptopus*?	Dean	7.14	Andrew Dabb & Daniel Lofflin

Most of those websites wouldn't know a ghost if it bit 'em in the *persqueeter*.	Dean	1.17	Trey Callaway
We're gonna drive out and make a lotta noise a long way from here, keep the *safeboat* safe for you.	Demon pretending to be Sam	8.21	Ben Edlund
Did you and *Samateur* hour really think I wouldn't have a back-up plan?	Metatron	10.18	Robbie Thompson
Meaning, where is the *Samulet*?[3]	Non-recurring character	10.05	Robbie Thompson
She has a thing for *smutton* chops.	Non-recurring character	8.17	Robbie Thompson
They still making *spleenburgers*?	Dean	7.03	Andrew Dabb & Daniel Lofflin
I just got *thraped*.	Dean	5.11	Andrew Dabb & Daniel Lofflin
Vampire pirates? That's what you guys are? *Vampirates*.	Dean	8.05	Ben Edlund
10 ccs of *Vamptonite*.	Dean	7.22	Andrew Dabb & Daniel Lofflin

[3] This term originated with the *Supernatural* fandom, and refers to an amulet Sam gave to Dean when they were children.

Must've *Windexed* your brain.	Ash	5.16	Andrew Dabb & Daniel Lofflin

Table 1, *Supernatural* Neologisms

It will be apparent, and no surprise to even casual viewers of *Supernatural*, that Dean gets some of the most colorful lines. In the table above, fully half of the coinages are spoken by Dean. The writers are particularly fond of giving Dean portmanteaus: vampirates, manburger, thraped (analyzed by a therapist without his consent). The only time Sam gets to play with language in the same way is when he hallucinates a version of himself. Indeed, in episode 8.21 the character Kevin deduces that a Sam imposter isn't Sam *because* he speaks too much like Dean.

Another point of interest is that the coinages are of a piece even though 19 different writers, working under three different showrunners, have created them. Most are either portmanteaus or derivations, including of proper nouns (affirmate, eekish, Ally McBealing). Part of what keeps the character Dean consistent and recognizable, even when looked at in isolation from visual and aural elements of *Supernatural*, is his vocabulary. This is understood and utilized by many writers, across many seasons. Sam's more straightforward, less colorful use of language is consistent with his characterization as the more serious, studious brother.

How does Sam's Vocabulary Compare to Dean's?

I compared words in each brother's vocabulary that were unique in comparison to the other brother by creating a word list of the DEAN Corpus using the SAM Corpus as a stop word list, and then doing it again in reverse. In the ten seasons under examination here, Dean has used 4,576 words that Sam has not used, and Sam has used 2,727 words that Dean has not used. A few of the words in Sam's list are Latin, since the tasks of exorcism and spell-casting often fall

to him. Most, however, are ordinary English words. A handful of Sam's unique words would be considered current slang or popular culture references: bromancing, buttload, cosplay, craphole, Disneyland, douche, gnarly, Hogwarts, [Dr.] Huxtable, juggalos, photobombing. By contrast, Dean's unique vocabulary features terms such as agro, anyhoo, assload, babydaddy, barfing, bazillionaire, bejeezus, Benjamin (for $100 bill), bestie, Bieber, Bitchmont, Bitcoin, biz, boneyard, brainiac, bro, bubkes, buzzkill, and so on, through the alphabet. In both the number of unique words and the variety, Dean's vocabulary proves to be more varied and more peppered with slang than Sam's. Writers for *Supernatural* seem to be well-aware of the ability of individual words to differentiate one character from another.

What's in a Word?

Some quantitative comparisons of words in both brothers' vocabulary also help distinguish Sam from Dean. Sam, who is portrayed as preferring research to slaying, uses forms of "to hunt" 146 times to Dean's 192. Dean, who for the most part exhibits great enthusiasm for the bloody aspect of his job, uses "lore" just 16 times to Sam's thirty-five. In keeping with his persona, Dean uses "chick" or "chicks" a whopping 51 times to Sam's 10. Excepting conditions such as being under a spell or losing his soul, it would be out of character for Sam to use any potentially reductive term for women. Dean uses "babe" or "hot" in referring to attractive women; Sam never uses "babe" and uses "hot" in that sense twice—one of which is quoting Dean. Dean uses "laid" to express having sex seven of the eight times he uses the term. Sam does so once. Dean's vocabulary is of a piece with his nature. Being slightly outdated, Dean's use of slang also connects him to an older generation, significant in the first few seasons especially, when Dean is often accused of being merely a copy of his father John.

Supernatural fans know that "awesome" is one of Dean's favorite words and would not be surprised to learn that he uses it 89 times in 218 episodes. What might be more surprising is

how rarely Sam says "awesome": only eight times. The writers affiliate words like "chick" and "awesome" with Dean not just by giving them to him, but also by *not* giving them to Sam.

How Often do They Say …?

In the episode "Changing Channels" (5.08), Sam and Dean are forced to be characters in a variety of genre shows, including a sitcom. Dean's sitcom catchphrase is "son of a bitch"–-an acknowledgement of how often Dean speaks those words. Indeed, if Dean can be considered to have a catchphrase, "son of a bitch" would be it.[4] Out of the 201 times the phrase is used in the show, 118 are spoken by Dean. It is a phrase he uses multiple times per season; sometimes multiple times per episode. Out of all of the four-word clusters spoken by Dean, "son of a bitch" is second only to "I don't know" (counted as four words because AntConc treats the character after the apostrophe as a separate word). Of course, the show airs in prime time on a network channel, not on HBO or Showtime. If the writers were less restricted in Dean's vocabulary, his expletives might be more colorful. A hint of this is evident in the episode "Ghostfacers" (3.13), where Sam and Dean run into a crew of amateur ghost hunters hoping to find fame in the reality tv market. Dean does use what we assume to be coarse language, but it is bleeped out and his mouth covered by an image of a skull.

And what about Sam? His most-used four-word cluster is also "I don't know," and it leads the next highly used 4-gram, "What do you think," 229 to 49. But if Sam can be said to have a catchphrase it would be "get this," an utterance he has used when he's made a discovery

[4] For a video fanwork highlighting Dean's use of "son of a bitch" from seasons 1-9, see "Son of a B****! – Dean Winchester" by Shawniegore, posted to YouTube Dec. 30, 2014. https://youtu.be/Rsm9zeu_Idw. Accessed Feb. 19, 2018.

pertinent to a monster-hunting case. Still, even though fans recognize "get this" as a Sam-ism, he has only used the phrase in this context 28 times in the first 218 episodes.[5]

How does Vocabulary Size Correlate to Talkativeness?

First, to determine whether and how Sam and Dean's vocabulary changes over the course of ten years, I created lists of the words spoken by each brother, by season. Figure 1 shows the number of words spoken by each brother, season by season.

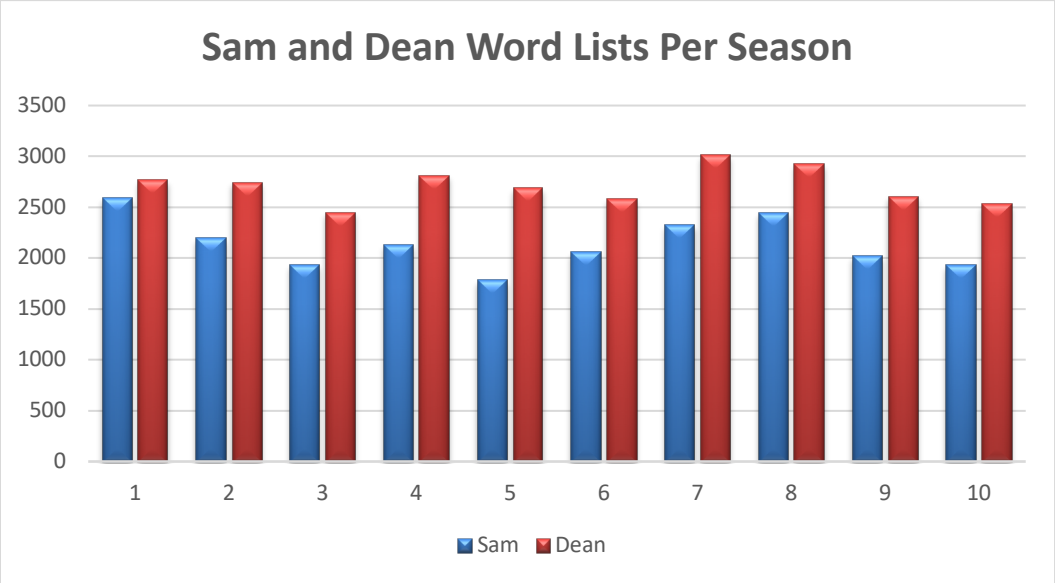


Figure 1, Sam and Dean Word Lists Per Season

In the first season of *Supernatural* Sam and Dean have roughly the same sized vocabulary. The writers are only beginning to find the characters' "voices" at this point. In all the

[5] For a video fanwork highlighting Sam's use of "get this" from seasons 1-9, see "So Get This" by Shawniegore, posted to YouTube Oct. 1, 2014. https://youtu.be/hY6ew_ytMrc. Accessed Feb. 19, 2018.

subsequent seasons, there is a marked difference between the two brothers, with Dean's vocabulary being consistently larger. Season 5 is particularly interesting here; Dean uses over 900 distinct words more than does Sam.

As to whether a connection exists between size of vocabulary and loquacity, a sampling of dialogue size from Season 5 suggests that there is. Using every third episode from that season, I did a word count of all of the speeches of Sam and Dean, individually. Figure 2 shows how much dialogue Sam and Dean had in the Season 5 sampling.

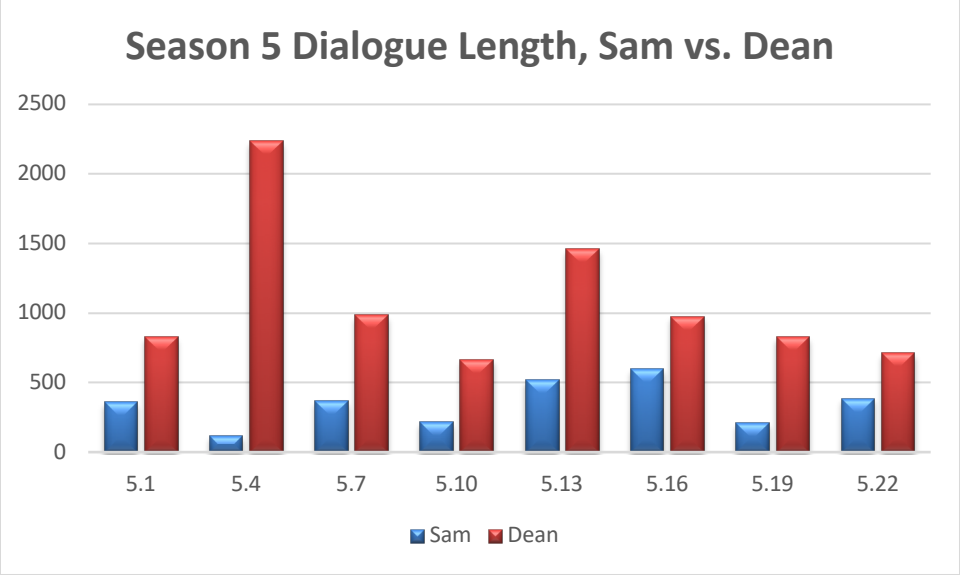


Figure 2, Season 5 Dialogue Length, Sam vs. Dean

The extreme difference in Episode 5.4, "The End," is because Sam is barely in that episode while Dean encounters a future version of himself and therefore has about twice as much dialogue as usual. That is not true of the other sampled episodes, though, and the difference between how much each brother speaks is still quite dramatic. So, not only does Dean use a much larger vocabulary than Sam, it appears he also does much more talking.

Whose Vocabulary is More Modified?

The SPN Corpus was tagged for parts of speech using the freeware program TagAnt. The corpus contains 2,256 regular, comparative, and superlative adjectives and 540 regular, comparative, and superlative adverbs. The results were not cleaned up and are therefore subject to error. However, a scroll through the results indicates that any degree of error would be slight. On the contrary, TagAnt's ability to recognize modifiers is impressive. It identified several uncommon words and hyphenated adjectives (douche-y, grand-coven-approved, NC17, Xeroxed).

I combined the list of adjectives and adverbs and used it as a limiter in AntConc to test how many were used by Sam and Dean, individually. Properly speaking, AntConc counts word types, not words. Usually the number of word types is virtually the same as the number of words. However, AntConc treats both sides of a hyphenated term as a word type, which does affect the count. The SAM and DEAN corpora were both subject to the same rule, so no effort to correct for hyphenated modifiers was made. The number of combined modifiers found in the DEAN Corpus as compared to the SPN Corpus is 1,563. The number found in the SAM Corpus as compared to the SPN Corpus is 1,397. Once again, the *Supernatural* writers seem to favor Dean when it comes to creative dialogue.

How does the Language of Supernatural *Compare to that of Other Horror Shows?*

Using subtitle files from tvsubtitles.net, I compared a sampling of nine *Supernatural* episodes (the SPN SELECTED corpus) to samplings from contemporary horror genre shows, specifically, nine episodes each of *The Vampire Diaries* (TVD) (CW Network, 2009-2017), *American Horror Story* (AHS) (FX Network, 2011-present), and *The Walking Dead* (TWD) (AMC Network, 2010-present), and *iZombie* (CW Network, 2015-present),[6] Horror genre shows were chosen in

[6] See Appendix for a list of the episodes used in the sampling.

an attempt to compare like to like. All were episodic, hour-long, weekly shows featuring at least some of the same supernatural beings as are found in *SPN*.

Samples were selected from the second, middle, and penultimate episodes of the first, middle, and latest seasons of the shows, as of July 2016. Vocabulary size was measured by a simple word count of each of the sampled corpora. As seen in Figure 3, "Total Individual Words," *iZombie* had the highest word count, with *Supernatural* using only about 750 words more than *The Walking Dead*.

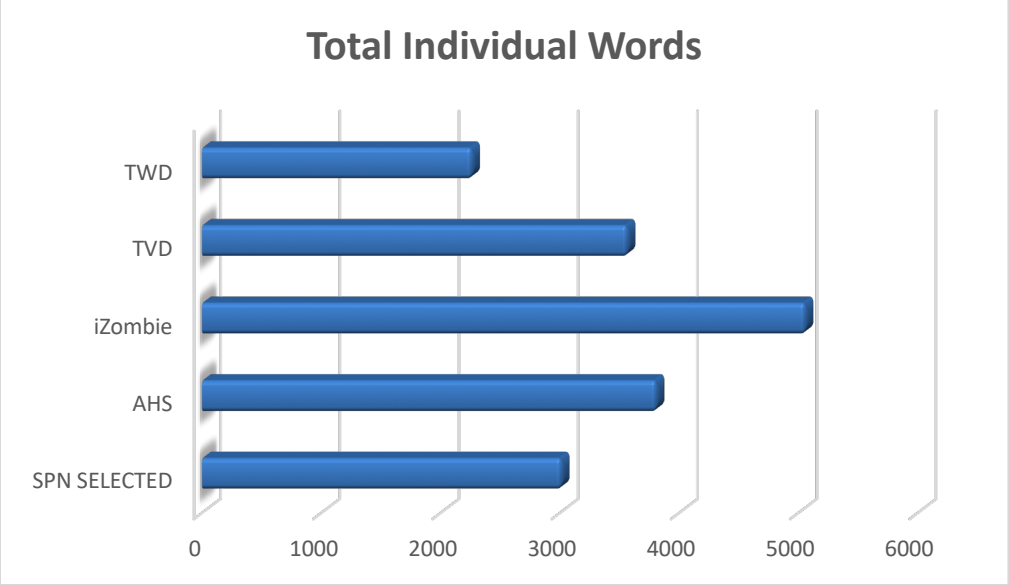


Figure 3, Total Individual Words

To determine the uniqueness of vocabulary for each show, words used only once in each corpus ("Single Use Words") were tallied and expressed as a percentage of the total number of words in each sample. For example, *American Horror Story* used 3,859 different words in the nine episodes sampled, and of those, 2,100 were used only once: 54%. The higher the percentage of single-use words, the more varied the vocabulary. AntConc was used to compile the word lists. The lists were created without setting any conditions, so there will be an undetermined but

miniscule margin of error as AntConc treats contractions as two separate words and makes no accommodation for variant spellings. See Figure 4, "Percentage of Single Use Words for Each Show."

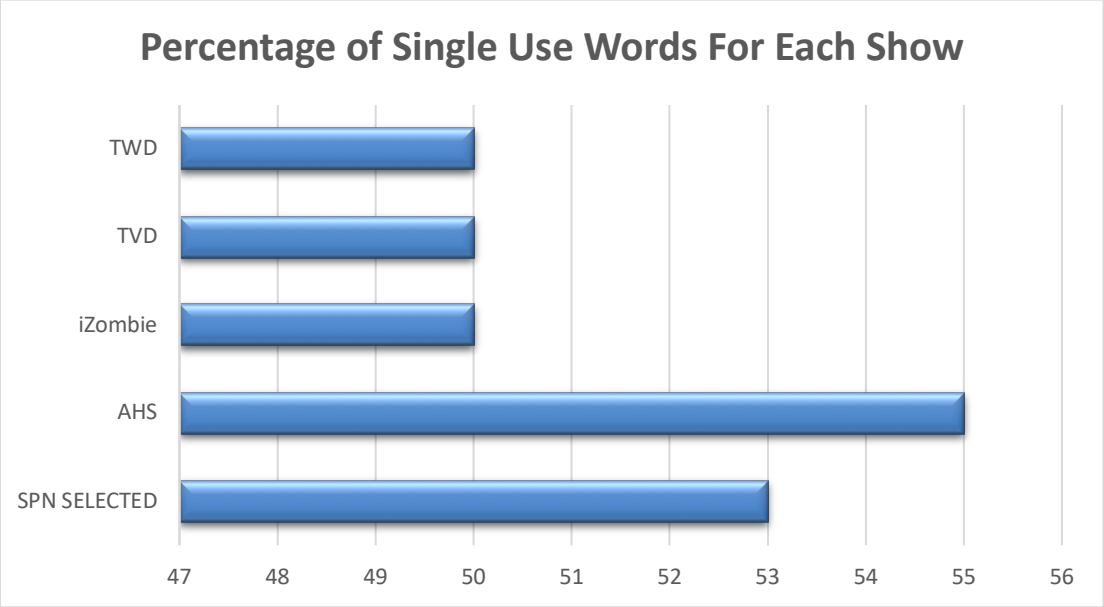


Figure 4, Percentage of Single Use Words for Each Show

So in this respect, *Supernatural* is modestly exceptional. The percentage of unique words is a little lower than *American Horror Story* and a little higher than *iZombie*, *The Vampire Diaries,* and *The Walking Dead*. All of the shows ranged between 50 and 54 percent in terms of words used only once. It may turn out that these percentages are about average for scripted television overall. A quick test of the dialogue of the pilot episode of *Gilmore Girls*, one of the most banter-filled, fast-talking shows in recent television history, had only 48 percent unique words.

In comparing the SPN SELECTED corpus against the corpora of the same sampling of episodes used above, *Supernatural* did not always win on uniqueness of vocabulary, as determined by quantity. This was tested using the Word List feature of AntConc. The full word

list of each of the sampled horror programs was used as a stop list against the full word list of the SPN SELECTED corpus. Then the SPN CORPUS was used as a stop list against the full word list of each of the compared shows. The words that remained in each case are unique. The results are shown in Figure 5, "Unique Vocabulary, Shows vs. SPN SELECTED." I only compared shows to the sampling from *Supernatural*, not to each other.

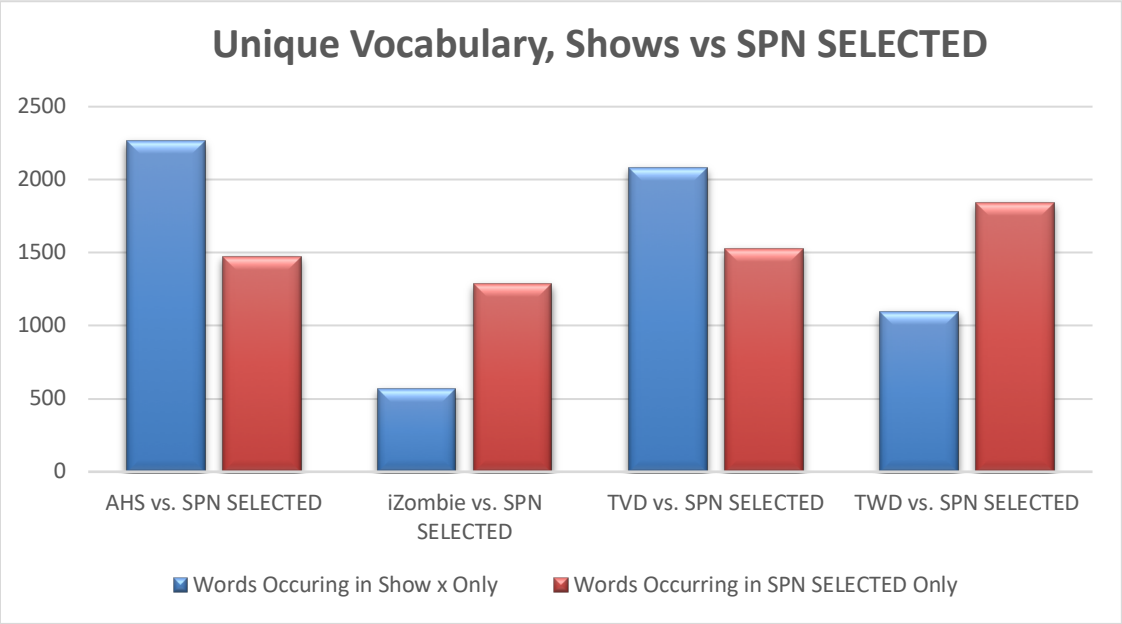


Figure 5, Unique Vocabulary, Shows vs. SPN SELECTED

Judging by numbers alone, *Supernatural*'s vocabulary is less unique than that of *American Horror Story* and *The Vampire Diaries* but more than *iZombie* and *The Walking Dead*. It should be noted that the word lists used are from a sampling of episodes, and words that don't appear on a list may in fact be used in the larger corpus. For example, the word "zombies" appears in two out of the four compared shows' corpora and in the $SUBTLEX_{US}$ corpus, but it does not appear in the SPN SELECTED corpus. However, it appears 18 times in the full SPN corpus.

The *iZombie* results in the test above are interesting. That show's sampling of episodes had the largest vocabulary of any of the shows, by far. Yet when it came to comparing unique vocabulary against *Supernatural*, *iZombie* fell short by more than half.

Conclusion

My thesis that *Supernatural*'s dialogue could be shown quantitatively to be varied and original was not definitively borne out from the examination of the sampled episodes of four of its contemporary horror genre shows. Other experiments were more conclusive. I was able to demonstrate that Dean's vocabulary is more colorful than Sam's. When the human eye was applied to the data, it was easy to show that *Supernatural*'s writers excel in inventing neologisms. Dialogue stayed true to the characters despite the number of writers creating it.

It was surprising to find that Dean had so much more dialogue than Sam. I offer a couple of unscientific observations as a fan who has seen every episode more than once: Dean seems to assert his role as the older brother by taking the lead more often when he and Sam interview witnesses or meet other hunters. He may also use talking to mask fear and other emotions; when in mortal danger, Dean certainly seems to engage with the bad guy much more than does Sam. I suspect the writers also find it more fun to write Dean than Sam.

Though a vocabulary analysis may not adequately support *Supernatural*'s (in my opinion) excellence, the SPN, SAM, and DEAN corpora can be useful for exploring other aspects of the show's writing. These include:

- The speech of other major characters such as Bobby, Castiel, Meg, Ruby, or Crowley;
- Female vs. male speech;
- Comparison of different writers' dialogue;
- How fans write certain characters' dialogue as compared to canon.

Overall, corpus analysis as a method of understanding a television series or other audiovisual medium provides an interesting perspective and may prove quite useful as one tool for understanding a multimodal text. Word or phrase frequency analyses, for example, can point a researcher directly to scenes that can then be examined within the context of the nonverbal elements. Tagged parts of speech can be studied more thoroughly to see how writers develop a character. Developers are writing new programs for this field, and perhaps we can look forward to tools that can detect intention or find words with similar meaning. It will be much more challenging to automate the creation of corpora such as I've used here. But if one only needs the words and not the speakers and is willing to put up with the odd errors common to captioned television, fan-produced transcripts or subtitle files are easily available on the internet for just about any scripted program.

Liorah Golomb is an Associate Professor and Humanities Librarian at the University of Oklahoma. This is an unpublished paper ©2018. If the reader is interested in using this paper in whole or in part, or in the corpora created for this project, please contact the author at lgolomb@ou.edu. The title of this article refers to a line spoken by Dean in episode 5.05, "Fallen Idols."

APPENDIX: Horror Show Episodes Used in Sampling

American Horror Story	1.02 Home Invasion 1.06 Piggy Piggy 1.11 Birth 3.02 Boy Parts 3.07 The Dead 3.12 Go to Hell 5.02 Chutes and Ladders 5.07 Flicker 5.11 Battle Royale
iZombie	1.02 Brother Can You Spare a Brain 1.07 Maternity Liv 1.12 Dead Rat Live Rate Brown Rat White Rat 2.02 Zombie Bro 2.10 Method Head 2.18 Dead Beat 3.02 Zombie Knows Best 3.07 Dirt Nap Time 3.12 Looking for Mr. Goodbrain part 1
Supernatural	1.02 Wendigo 1.11 Scarecrow 1.21 Salvation 5.2 Good God, Y"All 5.11 Sam, Interrupted

	5.21 Two Minutes to Midnight 10.2 Reichenbach 10.12 About a Boy 10.22 The Prisoner
The Vampire Diaries	1.02 The Night of the Comet 1.11 Bloodlines 1.21 Isobel 4.2 Memorial 4.11 Catch Me if You Can 4.22 The Walking Dead 7.02 Never Let Me Go 7.11 Things We Lost in the Fire 7.21 Requiem for a Dream
The Walking Dead	1.02 Guts 1.03 Tell it to the Frogs 1.05 Wildfire 3.02 Sick 3.08 Made to Suffer 3.15 This Sorrowful Life 6.02 JSS 6.08 Start to Finish 6.15 East

Works Cited

Abbott, Stacey, editor. *TV Goes to Hell: An Unofficial Road Map of Supernatural*. ECW Press, 2011.

Anthony, Laurence. AntConc. 3.4.1m, Waseda University, 2014, http://www.laurenceanthony.net/software/antconc.

---. *TagAnt*. 1.2.0, Waseda University, 2015, http://www.laurenceanthony.net/.

Biber, Douglas. "Corpus Linguistics and the Study of Literature: Back to the Future?" *Scientific Study of Literature*, vol. 1, no. 1, May 2011, pp. 15–23.

Blake, Linnie. "All Hell Breaks Loose: Supernatural, Gothic Neoliberalism and the American Self." *Horror Studies*, vol. 6, no. 2, Oct. 2015, pp. 225–238.

Brysbaert, Marc. *SUBTLEXus Sources*. 29 Mar. 2016.

Foresman, Galen A. *Supernatural and Philosophy: Metaphysics and Monsters-- for Idjits*. 2013.

George, Susan A., and Regina Hansen. *Supernatural, Humanity, and the Soul: On the Highway to Hell and Back*. Palgrave Macmillan, 2014.

Hansen, Regina M. "Deconstructing the Apocalypse? Supernatural's Postmodern Appropriation of Angelic Hierarchies." *Supernatural, Humanity, and the Soul: On the Highway to Hell and Back*, edited by Susan A. George and Regina M. Hansen, Palgrave Macmillan US, 2014, pp. 13–26, http://dx.doi.org/10.1057/9781137412560_2.

Quaglio, Paulo. "Television Dialogue and Natural Conversation: Linguistic Similarities and Functional Differences." *Corpora and Discourse: The Challenges of Different Settings*, John Benjamins Pub. Co, 2008, pp. 189–210.

---. *Television Dialogue: The Sitcom Friends vs. Natural Conversation*. John Benjamins Pub. Co., 2009.

Robinson, Gregory J. "Metal and Rust: Postindustrial White Masculinity and Supernatural's Classic Rock Canon." *Supernatural, Humanity, and the Soul*, edited by Susan A. George

and Regina M. Hansen, Palgrave Macmillan US, 2014, pp. 197–210. *link.springer.com*, doi:10.1057/9781137412560_15.

Rodgers, Michael P. H., and Stuart Webb. "Narrow Viewing: The Vocabulary in Related Television Programs." *TESOL Quarterly*, vol. 45, no. 4, Dec. 2011, pp. 689–717.

Silva, Alyssa. "Dean Winchester: An Existentialist Hero?/Dean Winchester: ¿Un Héroe Existencialista?" *Sesión No Numerada: Revista de Letras y Ficción Audiovisual*, vol. 2, 2012, pp. 67–83.

Taylor, Christopher John. "The Language of Film: Corpora and Statistics in the Search for Authenticity. Notting Hill (1998)--A Case Study." *Miscelánea: A Journal of English and American Studies*, vol. 30, 2004, pp. 71–85.

Webb, Stuart, and Michael P. H. Rodgers. "Vocabulary Demands of Television Programs." *Language Learning*, vol. 59, no. 2, June 2009, pp. 335–366.

Zubernis, Lynn S., and Katherine Larsen, editors. *Supernatural*. 2014.