

Marjorie McAtee

Dr. Brian Ballentine

English 606

3 May 2016

I Can Haz Capshun?: Analyzing Language Choices in LOLcat Memes

Abstract

LOLspeak is a digital dialect characterized by systematic differences in phonology, morphology, syntax, semantics, and lexicography; it is primarily used online, often to caption LOLcats, or humorous pictures of cats and other animals. Despite linguistic analysis of the features of LOLspeak and research into the role LOLcats play in forming group identity among those who create, share, and consume them, no scholar has yet investigated why some LOLcats are written in LOLspeak and why some are written in standard English. Through an analysis of 201 LOLcats with captions in both LOLspeak and standard English, this paper seeks to understand the difference between these two types of LOLcats, arguing that language choices in LOLcats reflect how the cat characters are portrayed, and may serve to give clues as to the group identity of their creators.



Figure 1: Happy Cat, the first LOLcat meme ("Happy Cat")

Introduction

Even if you've never created a LOLcat, you're sure to have seen even shared one. LOLcats are cute and humorous pictures of cats accompanied by captions or speech bubbles. Internet legend has it that the first LOLcat appeared on 4chan as part of the NSFW web community's Saturday celebrations. That first LOLcat was Happy Cat, a perky, gray British Shorthair named Frank who appeared in a Russian

cat food ad in the early 2000s (“Happy Cat”). The photo of Frank may have appeared on the Web as early as 2003, but its first verified appearance as a meme with the caption “I Can Has Cheezburger?” (see fig. 1) took place in 2007, when Happy Cat became the first meme posted to the website ICanHazCheezburger.com (IHC) (“Happy Cat”). With that, a meme genre was born; IHC would soon become well-known as a forum for the creation and sharing of LOLcats, and the meme’s popularity would allow cats to dominate the Internet in short order.

But LOLcats are more than just a fun meme. They are also the subject of serious scholarly study. Linguists like Lauren Gawne, Jill Vaughan, and Jordan Lefler have examined the linguistic features of LOLspeak (LOL), the poorly-spelled, grammatically inventive version of English often used to represent the speech of the cats in the memes. Gawne, Vaughan, and Lefler believe that the linguistic features of LOLspeak play a significant role in constructing the identities of the kitty characters in LOLcats; Kate Miltner believes that LOLspeak is a tool for forming group identity around the creation and sharing of LOLcat memes. However, not every LOLcat speaks LOL; some LOLcat captions are written in standard English (SE), and this paper seeks to understand why. My study of 201 LOLcat memes – 100 written in LOLspeak and 101 written in SE – has found that LOLspeak LOLcats are most likely to feature cat characters that are cute, cuddly, loving, and harmlessly mischievous, while SE LOLcats are most likely to feature cat characters that are sarcastic, selfish, derisive, rebellious, and even violent.

Literature Review

Since my methods for this study are firmly rooted in sociolinguistics and corpus analysis, I will begin my literature review with an overview of how this type of study gets done. Corpus analysis is one of many forms of computational analysis that Stephen Ramsey discusses in *Reading Machines*; while literature and digital humanities scholars may use corpus analysis software to perform Ramsay’s distant

reading and algorithmic criticism, linguistics use it to make claims about language use based on word usages, frequencies, and contextual patterns prevalent in a large corpus of linguistic data (Baker 109).

Corpus analysis typically begins with the collection of word frequency data, which can help the researcher identify important keywords within the corpus (Baker 108-109). A next step is to look for collocates of those keywords, in order to gain insight into their meaning and usage within the corpus; examining concordance lines, which put the keywords in context, can help researchers spot linguistic patterns (Baker 109). Corpus analysis software provides linguists with keyword-in-context displays, concordance lines, and frequency data, all of which Ramsay mentions as useful for computational text analysis (74).

A linguist interested in studying LOLspeak must build his or her own corpus, since there are no linguistic corpora available for the study of this dialect. In a paper presented at the 42nd Australian Linguistic Society Conference in 2011, linguists Lauren Gawne and Jill Vaughan based their study of the linguistic features of LOLspeak on Martin Grondin's *LOLcat Bible*. Gawne and Vaughan argue that LOLspeak is a digital dialect based on complex language play, and that it follows identifiable rules (102-104, 110-118).

Jordan Lefler based her analysis of the linguistic qualities of LOLspeak on a close reading of ninety-five LOLcat memes posted to ICHC in 2007 and 2008 (10). Lefler uses Grounded Theory methodology to organize her data into a spreadsheet, categorizing features of the dataset such as number, date, text, frames, contextual settings, characters present + action, and more (10-11). Lefler extracted data from each meme and form a thesis based on what the data represent (11). She agrees with Gawne and Vaughan that LOLspeak is a dialect of English (1). She identifies LOLspeak as a form of what Charles Ferguson calls "baby talk" or "foreigner talk," a simplified form of English used by those who are not proficient in the language (Lefler 17-18, Ferguson 117-118). Features of Ferguson's "baby

talk” include absence of copula, omission of definite articles and prepositions, failure to conjugate verbs, or reliance on a single verb construction, notably third person singular (119). Lefler’s research into the sociolinguistic significance of LOLspeak reveals that, while primarily an Internet dialect, many users “admit to using LOLspeak orally as well as graphically” (50).

Both Jordan Lefler and Kate Miltner agree that LOLspeak serves as a means of forming group identity. Lefler believes that LOLspeak users “want to show solidarity with others who use and are interested in LOLspeak” (50). Miltner calls LOLcats – and memes in general – a form of “vernacular communication” of which humor is a generic convention. She argues that, like slangs or accent humor, the LOLspeak in LOLcat memes is a form of identification humor; it “create[s] a code that is understandable only within a group context.” Miltner studied a focus group of LOL cat users aged 21-72 years, 86 percent of whom were under the age of 30, sourced from Twitter, Tumblr, Facebook, Craigslist, Reddit, and The Cheez Town Cryer, a forum for users of ICHC. She divided the users into three groups – Cheezfrenz, MemeGeeks, and Casual Users – seeking to understand how each group engages with and creates meaning from LOLcat memes.

Methods

Like Lefler, I chose to approach this study by examining the data and working out a theory through the research process (10). I collected 201 LOLcat memes – 100 memes written in LOLspeak and 101 written in SE – from ICHC, the birthplace of the genre. Like Lefler, I decided to consider a LOLcat caption “written in LOL” if it contains at least three words written in LOL or demonstrates a syntactic formula, snowclone, or grammatical construction unique to LOL. Unlike Lefler, I decided to collect only LOLcat memes which have at least one cat as a character. In an effort to only include humorous or entertaining memes in my study, I collected memes with at least 100 upvotes. I saved each LOLcat to two separate folders in my Microsoft OneDrive cloud storage, labeling the JPEG files 001.LOL through

100.LOL and 001.SE through 101.SE. I numbered them according to the order in which I collected them, and collected them in the order in which they were posted on the site, so that I collected the first 100 LOLcat memes written in LOLspeak with more than 100 upvotes, and the first 101 LOLcat memes written in SE with more than 100 upvotes.

Because I was working with a dataset that is very similar to Lefler's, I followed her example and extracted relevant data in a close reading of each individual meme. I created two Excel spreadsheets,



Figure 2: 016.LOL (top) compared to 017.SE (bottom)

one for the LOL memes and one for the SE memes, with the goal of analyzing each meme for relevant features. In the LOL spreadsheet, I created the following columns: number; text; setting; characters and action; narrator/speaker; dialogue/monologue/interior monologue; recurring characters present; themes; pop culture references; memetic intertextuality; same picture as; missing/added parts of speech; lexical items; subject/verb agreement variants; spelling variants; nonstandard capitalization; nonstandard punctuation; snowclones; and syntactic formulas. In the SE spreadsheet, I created the following columns: number; text; setting; characters and action; narrator/speaker; dialogue/monologue/interior monologue; recurring characters present; themes; pop culture references; memetic intertextuality; same picture as; spelling variants;

nonstandard capitalization; nonstandard punctuation; snowclones. Since linguistics features of LOLspeak weren't a factor in my analysis of the SE memes, I didn't include all of those categories that relate to the linguistic features of LOL, although I did keep the category "spelling variants" since some of the SE memes did contain fewer than three words in LOL. I used many of the same categories in my

spreadsheets that Lefler uses in her own data collection and management documents, but I added and altered categories as dictated by my research question and available data.

Once I had collected my data sample of 201 LOLcats and created the two spreadsheets, it was time to extract all the data I could from each LOLcat. I performed a close reading of each individual meme, transcribing the caption text, describing the setting in a few words (often just *indoors*, *outdoors*, or *?*) and describing the characters and action occurring in each meme. In most cases, the characters were simply *cat*, *kitten*, *dog*, or *human*, but some memes contained recognizable recurring characters like Grumpy Cat, Hamilton the Hipster Cat, or Dr. Tinitycat. I then identified whether the caption in the meme could be attributed to the character itself (speaker) or an unidentified narrator, or both. Since images often appeared more than once throughout the dataset, I recorded which memes used the same image under the category *Same Picture As*. Few, if any, of the memes contained information relevant to every category on the spreadsheet.

Once I had extracted all of the relevant data from the memes, I turned to digital humanities methods to perform a distant reading of the entire dataset. Through this distant reading, I hoped to gain

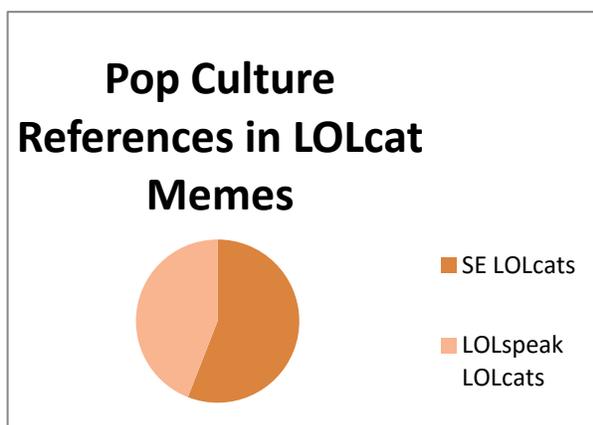


Figure 3: Ratio of pop culture references in 101 SE LOLcats vs. 100 LOLspeak LOLcats

insights into the differences between LOLspeak LOLcats and SE LOLcats. I could already see some differences between the two datasets (see fig. 3), but I wondered if distant reading techniques might help me make generalizations about that difference.

In my analysis, I treated the 201 memes as two datasets – Lolspeak memes and SE memes – in order to compare them to each other. I stripped the caption text for each dataset into two plain text

files; in this way, I created two small corpora for analysis. I also stripped the themes information from each dataset into two plain text files. Then, I used Voyant Tools to perform linguistic corpus analyses of the caption text for both datasets, in order to obtain word frequency data, concordance information, and collocate information for both the LOL and the SE caption text. Next, I used the TAPoRware Comparator to compare themes that appear in both datasets, as well as those that appear in only one of the datasets. I identified 41 themes that appeared in the 201 memes studied; because I used multi-word descriptions to characterize these themes, and the TAPoRware Comparator compares the occurrence of single words within two texts, I had to piece the results together based on the occurrence of unique words in each theme description, and I used frequency data provided by the TAPoRware Comparator to determine which themes were most common in the entire set of 201 memes and in each dataset individually. Finally, I counted the number pop culture references in each dataset to determine if there was a significant difference in the use of pop culture references in LOLspeak LOLcats as compared to SE LOLcats.

Results

While pop culture references were used in slightly more of the SE LOLcats, I didn't think the difference was significant; 33 pop culture references appeared in the SE LOLcats, compared to 26 in the LOLspeak LOLcats (see fig. 3). As Baker recommends (108), I began my analysis of two corpora of caption text by generating frequency lists. Some of the most frequent words in the LOLspeak corpus were *teh*, *dis*, *ai*, *da*, *iz*, *dat*, and *wif*, while some of the most frequent words in the SE corpus were *I'm*, *cat*, *just*, *it's*, *know* and *human* (see Appendix, Tables 1 and 2).

Of the 41 themes I identified across the 201 LOLcats, several were found in both the LOLspeak and the SE datasets (see Appendix, Table 3). The most common theme in both datasets was *cats love food*, occurring in 10.45 percent of all the memes. The most common theme among the LOLspeak

LOLcats was *cats demand attention*, which occurred in three percent of the memes; *cats love paper* and *forever home* were the second and third-most common, each accounting for two percent of the LOLspeak LOLcat memes. The theme *cats are assholes* appeared in 14.85 percent of the SE LOLcats; the next most common theme appearing in that dataset was *cat-related puns*, which occurred in 2.9 percent of the memes. *Cats are violent* occurred in 1.98 percent of the SE LOLcat memes

The results of this study seem to corroborate the results of Gawne and Vaughan's, Lefler's, and Miltner's research. Gawne and Vaughan argue that LOLspeak is a digital dialect with features that affect language at all five levels: phonology, lexicography, morphology, syntax, and semantics (102-104). Gawne and Vaughan identified rules used to create LOLspeak, including orthography based on English phonetics (110), features reminiscent of language acquisition in childhood (111), the over-extension of the regular past-tense suffix *-ed* to irregular verbs, as can be seen in *eated* (114), the omission of basic grammatical elements (117), and reliance on snowclones or phrasal templates (for example, *I made you a X, but I eated it*) (118). I found all of these assertions held true in my analysis of the 100 LOLspeak LOLcat memes.

Lefler pinpoints a number of unique linguistic features of LOLspeak, including the use of a wide range of spelling and copula variants (22-28, 34), unique syntactic formulas, snowclones (39-42), and referential or identification humor (36); I was able to identify all of these features in the LOLspeak memes I studied. Miltner finds that some users – namely, those who share, create, and consume LOLcat memes – are deeply emotionally invested in the genre, using LOLcats as a “form of emotional expression, particularly as ‘reactions’ to situations they (or their family or friends) were going through” and as a means of sharing difficult or negative emotions with online friends or social media contacts. These emotionally invested users – a group Miltner dubs “Cheezfrenz” – said they “shared and created LOLcats with someone else’s enjoyment in mind” – purely for entertainment and to nurture relationship

connections, not for recognition. As a result, members of the linguistic community of practice that has evolved to use LOLspeak – Miltner’s Cheezfrenz – associate LOLcats with emotionality, earnestness, kindness, gentleness, caring, and the provision of emotional support within the context of a community. Of the other groups in Miltner’s study, MemeGeeks were significantly less engaged with the LOLcats, professing a nostalgic interest in them and a respect for their place in the Internet’s cultural canon, but claiming to connect more deeply with computer culture and memes in general. Casual Users were the least engaged, saying they appreciated LOLcats for their humor and cuteness, but engaged with them primarily as a means of staving off boredom at work. Miltner’s findings seem to support my argument that LOLspeak LOLcats and SE LOLcats portray their cat characters in fundamentally different lights.

Discussion

A comparison of word frequency data for both the SE and LOLspeak memes begins to reveal the difference between the two types of LOLcat memes. Some of the most common words in the LOLspeak memes are *teh/da* (the), *dis* (this), *dat* (that). These words direct attention and action at an object; their frequency may be an indication that the creators of LOLspeak LOLcat memes construct characters that are outward-directed, focused on the world around them and, as words like *wif* (with), *yu* (you), and *haz* (has) might indicate, relationships and possession.

Forms of *I* are, however, fairly frequent in both corpora, indicating the creators of both SE and LOLspeak LOLcats construct characters with strong identities. But words like *dog*, *cat*, and *human* appear in the SE caption corpus, suggesting that the creators of SE LOLcat memes construct cat characters as more distant and aloof, prone to using common nouns as brusque forms of address rather than more intimate pronouns.

The differences between the SE and LOLspeak memes really become apparent when the unique themes in each dataset are compared. Most of the themes identified only in LOLspeak LOLcats are

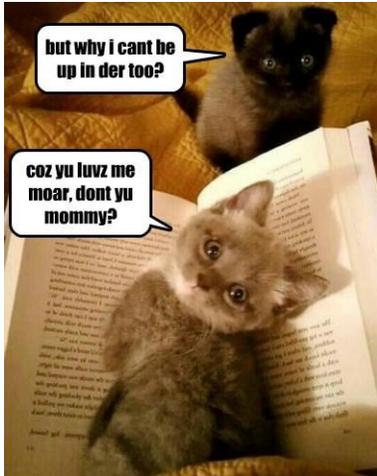


Figure 4: Meme 069.LOL, an example of the theme *cats demand attention*

innocuous references to known aspects of cat behavior – these memes deal with cats’ delight in boxes and scraps of paper, their propensity to beg their human owners for food, and their nocturnal rambunctiousness (see Appendix, Table 4). The most common theme, *cats demand attention*, was applied to memes in which interrupt human tasks in order to demand affection (see fig.4). These themes range from humorous to heartwarming, and suggest that creators of LOLspeak LOLcats construct cat characters as loving, affectionate, cuddly, and cute. As Lefler maintains, the LOLspeak itself works to create that identity; copula absence, reliance on the third person singular construction (as seen in the caption “I can has cheezburger?”), and failure to conjugate verbs all point to LOLspeak as a simplified form of English that creates its speakers –



Figure 5: Meme 052.SE, an example of the theme *cats are assholes*

in these memes, ostensibly the cats themselves – as possessing a childlike command of the language (17-18). Gawne and Vaughan establish that people who engage with LOLcats have a preconceived idea of how the cat characters’ voices must sound – high-pitched and childish, “with strong vowel distinctions” (111). According to Miltner, LOLspeak is a form of identification humor, “an in-joke for those who understand the context of its origins” on the Web, and virtuosity in its performance is seen as a mark of status within the social group that appreciates it. I would go so far as to call this group a linguistic community of practice – a community in which a group of people “who interact around a shared goal” (Dodsworth 270). In the case of LOLcat memes, that shared goal is exchanging emotional support and nurturing relationships.

Themes appearing in the SE caption corpus say something entirely different about the creators of SE LOLcat memes, as the themes of this dataset are far more negative (see Appendix, Table 5). The theme *cats are assholes* (see fig. 5) is not only the most common theme in the SE dataset, but the second most common theme in the entire corpus of 201 memes; it occurs in 15 of the memes, while the theme *cats love food/tuna* occurs in 21 of the memes. Other themes found only in the SE meme corpus include *cats are rebellious*, *cats are superior*, and *cats see humans as their servants*. All of these themes suggest that SE LOLcat creators tend to construct cat characters as selfish, derisive, unpleasant, and impossible to control. Some of these themes also refer to well-known aspects of cat behavior, including *cats like to hide under things* and *catching the red dot*, but others wholly anthropomorphize the animals, constructing cat characters that look forward to Fridays and hate Mondays. Of course, many of the themes evident across both datasets also anthropomorphize the animals, particularly those that construct cat characters as participating in human-like relationships. The common themes construct cat characters as gluttonous (*cats love food*, *cats are fat and lazy*) or concern themselves with aspects of cat behavior, constructing cats as placing themselves on a pedestal they may not deserve. In general, however, SE LOLcat creators are relying on a more mainstream conception of what cats are like to achieve an entirely different purpose – getting a laugh, and not much more.

Conclusion

Ultimately, these results seem to confirm Miltner's findings, that LOLcats written in LOLspeak are created by members of a linguistic community of practice who use both the memes and the dialect as a way to forge group identity and make meaning as emotional beings. Miltner states that these users tend to be women who share an "affinity for cats" and consider proficiency in LOL a mark of status within the community. I think it follows that LOLcat creators who choose LOLspeak for their captions probably like cats, are interested in portraying cat characters in a positive light, and may use the memes

as emotional maintenance tools in relationships with other members of their community of practice. By contrast, those who choose SE for their captions may be users who don't like cats, who see them in a more negative light, or who simply don't belong to the Cheezfrenz community of practice. LOLspeak itself is a form of identification humor, and LOLcat creators who choose not to use it are those who don't get the joke.

Works Cited

- Baker, Paul. "Corpus Linguistics in Sociolinguistics." *Research Methods in Sociolinguistics: A Practical Guide*. Ed. Janet Holmes and Kirk Hazen. Malden, MA: Wiley Blackwell, 2014. 107-118. Print.
- Dodsworth, Robin. "Speech Communities, Social Networks, and Communities of Practice." *A Practical Guide*. Ed. Janet Holmes and Kirk Hazen. Malden, MA: Wiley Blackwell, 2014. 262-275. Print.
- Ferguson, Charles A., and Huebner, Thom. "Absence of Copula and the Notion of Simplicity: A Study of Normal Speech, Baby Talk, Foreigner Talk and Pidgins." *Sociolinguistic Perspectives*. Cary, US: Oxford University Press (US), 1996. ProQuest ebrary. 115-123. Web. 15 April 2016.
- Gawne, Lauren and Jill Vaughan. "I Can Has Language Play: The Construction of Language and Identity in LOLspeak." Proceedings of the 42nd Australian Linguistic Society Conference, 2-4 December 2011. Canberra: Australian National University, 2011. 97-122. PDF file.
- "Grumpy Cat." *Know Your Meme*. Cheezburger, Inc. n.d. Web. 30 April 2016.
- "Hamilton the Hipster Cat." *Public Figure Fan Page*. Facebook. 20 October 2012. Web. 30 April 2016.
- "Happy Cat." *Know Your Meme*. Cheezburger, Inc. n.d. Web. 28 April 2016.
- Lefler, Jordan. *I Can Haz Thesis?: A Linguistic Analysis of LOLspeak*. MA Thesis. Louisiana State University, 2011. PDF file.
- Miltner, Kate. "'There's no place for lulz on LOLCats': The role of genre, gender, and group identity in the interpretation and enjoyment of an Internet meme." *First Monday: A Peer-Reviewed Journal on the Internet* 6.8 (4 Aug. 2014): n. pag. Web. 18 March 2016.
- Ramsay, Stephen. *Reading Machines: Toward an Algorithmic Criticism*. Chicago: University of Illinois Press, 2011. Print.
- Rockwell, Geoffrey, Lian Yan, Andrew Macdonald and Matt Patey. *TAPoRware Comparator*. TAPoR Project. McMaster University. 2005. Web. 17 April 2016.
- Sinclair, Stefan, Geoffrey Rockwell et al. *Voyant Tools*. 2012. Web application. 17 April 2016

Appendix – Tables

Term	Count	Trend
i'm	13	0.011384
cat	7	0.00613
just	7	0.00613
it's	6	0.005254
know	6	0.005254
human	5	0.004378
you're	5	0.004378
can't	4	0.003503
come	4	0.003503
dog	4	0.003503

Term	Count	Trend
teh	27	0.023196
dis	17	0.014605
ai	11	0.00945
da	11	0.00945
iz	11	0.00945
dat	10	0.008591
wif	9	0.007732
yu	9	0.007732
haz	8	0.006873
ur	6	0.005155

Tables 1 and 2: Word frequency data for SE and LOLspeak caption corpora

Theme	Count	Trend
Cats love food/tuna	21	0.1045
Enmity between cats and dogs	12	0.0597
Cats are in charge	8	0.0398
Cats are children and humans are parents	5	0.0249
Parent-child relationships	5	0.0249
Cats are fat and lazy	4	0.0199
Sibling relations	3	0.0149
Cats are hunters	4	0.0199

Table 3: Themes in both SE and LOLspeak memes

Theme	Count	Trend
Cats demand attention	3	0.03
Cats love paper	2	0.02
Forever home	2	0.02
Cats hate the vet	1	0.01
Special = mentally disabled	1	0.01
Children's aversion to the opposite sex	1	0.01
The problem of cat fur	1	0.01
Cats have litter preferences	1	0.01
Cats more active at night	1	0.01
Cats get hairballs	1	0.01
Christmas/Catmas	1	0.01
Cats love boxes	1	0.01
Cats beg for food	1	0.01

Table 4: Themes in LOLspeak memes only

Theme	Count	Trend
Cats are assholes	15	0.1485
Cat-related puns	3	0.029
Cats are violent	2	0.0198
Cats like warm things	1	0.0099
Cats like to climb into refrigerators	1	0.0099
Looking forward to the weekend	1	0.0099
Cats hate water	1	0.0099
Cats are superior	1	0.0099
Mondays suck	1	0.0099
Cats love string	1	0.0099
Cats see humans as their servants	1	0.0099
Catching the red dot	1	0.0099
Cats are rebellious	1	0.0099
Introversion	1	0.0099
Grandparent-Grandchild relations	1	0.0099
Humans can't hunt	1	0.0099
Cats like to hide under things	1	0.0099
Cats steal chairs	1	0.0099
Aphorisms	1	0.0099

Table 5: Themes in SE memes only