

Corpus Linguistic Analysis of the Idiolects of Gollum and Sméagol

Vannessa L. Milom*

Seattle, Washington, USA

*Corresponding author: tess.milom@gmail.com

Received November 27, 2021; Revised January 06, 2022; Accepted January 14, 2022

Abstract The topic of this paper is a linguistic analysis of the speech patterns of Gollum and/or Sméagol from J. R. R. Tolkien's Lord of the Rings books. The goal of this research was to apply forensic linguistic methods and corpus analyses to the speech patterns of the characters in order to determine if Tolkien might have designed the idiolect of Gollum-Sméagol with intention and whether it was crafted with the same professional creativity that he applied to his other fictional languages. The results reveal that Tolkien used orthographic features as a detailed characterization device. The language used by Gollum and Sméagol was thoughtfully curated to subtly convey information regarding their identity, their motivations, and the progress of their character arc. This aspect of Tolkien's character writing may be a unique literary device of his own invention.

Keywords: Tolkien, Lord of the Rings, Gollum, corpus analysis, forensic linguistics, authorship analysis, concordance, keyword analysis

Cite This Article: Vannessa L. Milom, "Corpus Linguistic Analysis of the Idiolects of Gollum and Sméagol." *Journal of Linguistics and Literature*, vol. 5, no. 1 (2022): 1-5. doi: 10.12691/jll-5-1-1.

1. Introduction

The topic that is central to this research concerns the distinct speech patterns of a character (or two) from the literary works of J.R.R. Tolkien. This paper describes a linguistic analysis of the idiolect of Gollum and/or Sméagol. Tolkien was a phenologist and historical linguist with a passion for constructing languages. His collection of Middle-earth writings, which he referred to as his "legendarium", contain several languages with detailed linguistic and historical features. The goal of this research was to determine if Tolkien might have designed the idiolect of the Gollum-Sméagol character with the same level of linguistic detail as he did his other languages and dialects.

The processes and analyses described in this paper resemble those a forensic linguist might perform on a document with multiple authors, where the intent is to determine which author might have supplied which parts. In a typical forensic authorship analysis, authorship attribution would include compiling and referring to corpora comprised of verified language samples from each suspected author. In this study, however, Tolkien is the one actual author, and he portrayed one fictional person who was actually two different characters. As such, the researcher relied on literary cues in order to divide their language samples into authorship corpora for comparison.

This research exemplifies the value of applying quantitative analysis to literature. By analyzing the

speech patterns of certain characters, the results of this research reveal that Tolkien used orthographic features as a detailed characterization device. The ways Gollum/Sméagol use language and the differences and changes in their speech subtly convey information regarding their identity, their motivations, and their character arc. This facet of Tolkien's linguistic creativity seems to have been previously overlooked.

2. Research Materials

J.R.R. Tolkien only lived to publish *The Hobbit* and *The Lord of the Rings* trilogy, but much of his legendarium were compiled, edited, and published posthumously by his son, Christopher Tolkien [1]. The collection of novels, manuscripts, drafts, notes, letters, and linguistic works that compose the legendarium comprise seven books, in a total of twenty volumes [2]. Tolkien was an iconoclast in fantasy writing, world-building, and language construction, so background context may be sourced from any of his works. However, only four volumes from his legendarium will be included in this project's corpus, as that is where we find Gollum.

Gollum, aka Sméagol, has speaking lines in the four books included in the corpus for this project: *The Hobbit* and *The Lord of the Rings* trilogy. Gollum-Sméagol makes his first appearance in the fifth chapter of *The Hobbit*. In *The Lord of the Rings* trilogy, he is in *The Fellowship of the Ring* as a brief second-hand quotation, but he is a central character in *The Two Towers*, and critical to the final conflict in *The Return of the King*.

To include the books in this corpus, the researcher used her personal eBook copies of the books and used a digital book managing program called Calibre [3] to convert all the (.pub) files of the books into text documents (.txt) that AntConc could analyze [4]. The eBooks thusly converted from her personal collection are all professionally digitized and published by HarperCollins Publishers. These texts include forwards, book reviews, indices, published works lists, publisher information, and appendices. The publisher's information released with each eBook file is shared in Appendix B.

All texts of this corpus fit under the register of fiction. The total word count of the four novels, excluding appendices, notes, etc., is approximately 576,499 combined. *The Hobbit* contains 95,356 words; *The Fellowship of the Ring* has 187,790 words; *The Two Towers* has 156,198 words; and *The Return of the King* has 137,155 words. Three distinct corpora were compiled to analyze only the language of Gollum and Sméagol; the joint corpus which contains their lines and literary context contains a total of 9,707 words, the corpus of only Gollum's language is a total 2,822 words, and the corpus of only Sméagol's language totals 3,142 words.

3. Research Discussion

A corpus analysis project was determined to be an efficient method of analyzing Gollum/ Sméagol's idiolect. The corpus that has been compiled to that end is comprised of all of Tolkien's published works which include Gollum and/or Sméagol speaking. This corpus can potentially inform any research about his idiolect; from his vocabulary to his abuse of pronouns. This corpus could be applied to a fair amount of research questions. Some research questions that the researcher considered for this project are these:

1. What rules, if any, does Gollum/Sméagol follow for sibilant additions?
2. Does Gollum's speech shift in correlation to his influence under the One Ring?
3. What rules, if any, define the idiolect of Gollum/Sméagol?

For this project, a single research question was chosen to investigate further: What rules, if any, define the idiolect of Gollum and/or Sméagol? To the study of this research question, three most likely hypotheses were considered:

Hypothesis 1: Gollum/Sméagol's dialect has some rules for its distinct use of pronouns, verb inflections, and adds sibilants, which depend on the environment of his speech.

Hypothesis 2: Gollum and Sméagol speak Sméagol's regional dialect in a way that is corrupted by centuries of isolated, unnatural life and corruption from the One Ring.

Hypothesis 3: Gollum/Sméagol's idiolect is simply an orthographic means of differentiating Gollum as an evil, snake-like creature, as well as illustrating Sméagol's personal struggles with villainous motivations.

This project applies techniques of analyzing linguistic corpora to testing the corpora of Gollum-Sméagol language for evidence that suggests the existence or lack

of linguistic rules within the idiolect of Gollum and/or Sméagol, and further seeks to define those linguistic rules should they be evident. This study examined the linguistic nuances of Gollum-Sméagol's speech to analyze any features Tolkien might have used to differentiate the two characters and whether their idiolect was designed with more detail and intention than previously assumed.

4. Analyses and Results

As barely separate characters, Gollum-Sméagol has a peculiar manner of speaking: He converses and argues with himself, exaggerates and/or adds sibilants, refers to himself with plural pronouns, hisses, produces a "gollum" sound, and famously refers to both himself and the One Ring as "Precious" [2,5,6,7]. For this project, the research relied on corpora analyses to examine those features for any evidence they could reveal about the linguistics of Gollum-Sméagol's idiolect. Some rules were made apparent, and others only strongly implied.

Collective results, analyzed together, support the hypotheses that Gollum-Sméagol's idiolect both follows some linguistic rules, and that it changes to reflect which of the two characters are speaking, and therefor acting, within the books. Differences in the speech of Gollum-Sméagol appear as degrees of use of their idiolect's peculiar features and as contrasting adherence to standard grammar; as indicated by analyses of their independent corpora. Potential rules regarding linguistic features were investigated through analysis of the combined Gollum-Sméagol corpus.

4.1. Keyword Analysis

Keyword analysis compares a reference corpus and a target corpus. By comparing the word use of a smaller, target corpus against a larger, more general corpus, the particularly frequent and therefor characteristic word choices of the target corpus become statistically apparent. For this analysis, the reference corpus is the source text of the four books (*The Hobbit* and the *Lord of the Rings* trilogy), and the target corpora are, in turn, three corpora of Gollum-Sméagol language.

Keyness analyses were performed with a statistical significance set to a 3.84 minimum threshold, as advised by a professor of corpus linguistics. Results were integral to further testing hypotheses. The top ten keywords of all three corpora are charted below in Figure 1.

As Figure 1 is sorted by keyness and not frequency, the order of priority indicates likeliness of preference, not most occurrence, within a target corpus. This means that Figure 1 is organized by the words distinct to each of the three target corpora against the larger corpus that is the entire text of the four books. For example, the word "Sméagol" appears frequently across the joint Gollum-Sméagol corpus in ways that the statistical algorithms used by AntConc identify as significantly distinct from the text of the four books. This can be understood to mean that when Gollum/Sméagol say "Sméagol," they say it in different contexts than it appears elsewhere in the books.

Both			Gollum			Sméagol		
Keyword	Freq	Keyness	Keyword	Freq	Keyness	Keyword	Freq	Keyness
yes	149	764.98	it	240	17.45	i	29	9.55
sméagol	98	568.61	us	64	13.17	the	213	9.25
precious	88	496.39	my	47	9.36	of	52	8.48
nice	66	389.98	curse	11	9.23	master	61	8.29
we	163	293.85	precious	88	8.93	you	30	7.87
it	239	217.94	fish	8	6.71	black	10	7.82
t	93	201.66	fishh	8	6.71	lots	10	7.82
master	61	196.01	hurts	8	6.71	gate	9	7.04
no	103	158.72	sss	8	6.71	there	45	6.86
us	64	133	lost	14	6.54	stair	8	6.26

Figure 1. Top keyness results for Gollum-Sméagol, Gollum, and Sméagol keywords

Key words, on the other hand, refer to relative frequency of word appearance. Words that appear in the target corpora more often than across the reference corpus are identified as being key words. The most frequent words used by Gollum-Sméagol are, expectedly, those he calls himself and others (Sméagol, precious, master) and his first-person plural pronouns. Other words appear as a result of frequent repetition within the text.

Differences in keyness and keywords between the Gollum and Sméagol corpora support the hypothesis that the two characters use language distinctly, albeit subtly; probably to indicate the identity of whichever is dominant. Keyness analysis indicates that Gollum is the character who most frequently uses first-person plural pronouns for himself (we/us), calls other people “it”, hisses, and addresses Precious. Sméagol, in contrast, is the one to most frequently use the first-person singular pronoun (I) and refers to others by the second person pronoun (you). This difference in pronoun use is consistent with narration that describe a shift in behavior indicative of whether Gollum or Sméagol is dominant of their body. Consider the following excerpts as examples of this:

“From that moment a change, which lasted for some time, came over him. He spoke with less hissing and whining, and he spoke to his companions direct, not to his precious self.” (The Two Towers, Book II, Chapter I)

“For one thing, he noted that Gollum used I, and that seemed usually to be a sign, on its rare appearances, that some remnants of old truth and sincerity were for the moment on top.” (The Two Towers, Book II, Chapter III)

Otherwise, Sméagol’s word choices become more nuanced as he speaks on other topics as he guides the Hobbits through Middle-earth. The higher total keyness of Gollum’s keywords within the joint Gollum-Sméagol corpus indicates that Gollum spent more time speaking for the two throughout the books.

4.2. Concordance Analysis

Concordance analyses assemble hundreds of examples of text and present them in a format that is simple to analyze for its context. For this project, the versatile concordance feature of AntConc was used quite frequently. Most of the analyses in this study began with a concordance search to observe particular linguistic features in context. The following concordance analyses were all performed on the joint Gollum-Sméagol corpus in order to examine the linguistic behavior of sibilant

addition/exaggeration used frequently in Gollum-Sméagol’s idiolect.

4.2.1. Sibilant Reduplication

One characteristic feature of Gollum-Sméagol’s language is the redundant addition of word-final sibilants, especially in addition to nouns already containing a plural suffix. To determine if this is a linguistic feature that follows rules, and what those rules might be, it was necessary to identify the feature in context; best achieved with a concordance analysis.

To populate a concordance of nouns that end in at least two sibilants, AntConc was used to search the joint Gollum-Sméagol corpus for occurrences of different spellings of plural suffixes. The search terms used included “*ses”, “*es”, and “*s”. AntConc understands that the asterisk allows for any number of characters to fill that position, so searching for those terms will produce all occurrences of words that end in those letters.

noun + /z/ + /z/	#	noun + /z/	#
pocketses	9	pockets	0
goblinses	6	goblins	0
hobbitses	2	hobbits	38
orcse	2	orcs	13
eyes	1	eyes	10

Figure 2. Samples of plural suffix reduplication

Since many words in standard English end with those letters, it was necessary to account for each example of verbs and correctly pluralized nouns. Excluding those problematic tokens, the concordance lists contained just over a dozen token examples of nonstandard sibilant endings.

To compare phonemic environments, additional concordances were performed for all versions of the affected nouns without extra sibilant suffixes. For example, to analyze surrounding words for potential linguistic influence over occurrences of “hobbitses” verses “hobbits,” concordances for both words were compared and contrasted.

noun + /z/, /z/	#	noun	#
fishes/fisshes	5	fish/fissh	16
mices	1	mice	0
peoples	5	people	2

Figure 3. Samples of plural suffix addition

Some of the most telling results of these sibilant-suffix concordance comparisons are displayed in [Figure 2](#) and [Figure 3](#). Lexical comparisons indicated no significant consistencies or complimentary distributions that might indicate any possible phonemic rules.

4.2.2. Verb Conjugation

Another sibilant-based feature of Gollum-Sméagol's idiolect is the addition of <s> to the end of verbs; making those verbs resemble the standard conjugation of the third person singular, regardless of subject-verb agreement. By analyzing concordances from the joint Gollum-Sméagol corpus, the sibilant additions appeared to occur quite exclusively in certain tenses and with particular pronouns. To test this apparent relation, sibilants were tested as a potential conjugation feature. This angle of approach revealed some patterns that could be potential rules for verb conjugation unique to Gollum/Sméagol's idiolect.

First, recognizing where their idiolect *lacks* "errors" in conjugation was key to identifying potential rules. Gollum/Sméagol tend to conjugate verbs along standard grammar in most circumstances, Sméagol especially. They are both capable of constructing complex verb phrases, including the use of auxiliary and/or modal verbs. Additionally, both repeatedly demonstrate a propensity for reciting from memory and quoting others without applying their peculiar inflections. There are three particular environments where Gollum-Sméagol appear to use non-standard rules to conjugate verbs to a third person singular tense: present-tense verbs following a non-inclusive "we" pronoun, within the subjunctive mood, and a few times following a third-person plural "they" pronoun.

Most sibilant-based conjugation disagreement occur when a third-person plural pronoun "we" is the subject and its verbs are present tense. Concordances for the occurrences of "we" revealed that the simple present tense verbs in those phrases will almost always be conjugated to the third person singular. There are about thirty examples of Gollum/Sméagol pairing the first-person plural pronoun with a present tense third person singular verb.

To summarize, the results of concordance analyses suggest that whether or not verbs receive non-standard inflection seems to be mostly determined by the following conditions:

- Non-Standard Conjugations Present:
 - With self-referent "we" subject pronoun
 - Possibly within subjunctive mood
 - Sometimes with plural "they" pronoun
 - Within complex verb phrases
- Standard Conjugation Present:
 - Past-tense verbs
 - Within quoted speech
 - With most pronouns other than self-referent "we"
 - With most other referents, pronouns, and tenses

It is important to consider that almost all verbs in English are inflected for the third person singular present indicative form by the addition of the <s> suffix. Except for the third person singular simple present tense, such verbs are identical to other person cases within a tense. So, theoretically, in tenses of future, preterite, conditional, etc., Gollum-Sméagol could be conjugating a first person plural pronoun with a third person singular verb, but the

cases are morphologically identical and so appear grammatical. However, phrases involving multiple verbs do not follow these patterns. This may not even be rule of conjugation so much as a preventative environment. If "we" is separated from a main verb by auxiliary verbs, modal auxiliary verbs, or negative adverbs, the conjugation will usually follow standard grammar rules.

A few exceptions to this pattern are found in questions that use "do" in various conjugations, especially where the order of "do" and "we" are reordered to form a question. This environment seems to allow for exceptions to the conjugation of auxiliary verbs; even auxiliary verbs will be conjugated to the third person singular.

Some concordance samples that appeared to be outliers to these other potential conjugation rules turned out to be subjunctive phrases that followed a particular conjugation pattern. Though there are also instances of Gollum/Sméagol using the subjunctive mood without any special conjugation efforts, there are also a few examples of subjunctive mood with standard subjunctive conjugations. Still, this is evidence for the former potential rule, as it explains those few outliers as potentially following some other rule of exception. The following excerpts are examples of Gollum/Sméagol using the subjunctive mood with different levels of standard grammar:

- "If it asks us, and we doesn't answer, then we does what it wants, eh?" (The Hobbit, Chapter V)
- "Ha ha! What does we wish?" (The Two Towers, Book IV, Chapter II)

Finally, the third person plural pronoun "they" also sometimes receives disagreeing verb conjugations. There are eight examples of Gollum/Sméagol pairing the third person plural pronoun "they" with third person singular verbs. However, two of those examples occur in the subjunctive mood and two others are parts of a recitation. Even so, there are over fifty samples of "they" receiving third person plural verbs, observant of standard English grammar. Overall, concordance analysis of this pronoun revealed that this may be more of an inconsistency or error rather than a completely realized grammatical rule.

Outside of these environments, Gollum largely follows standard verbal conjugation rules, Sméagol especially so. When Gollum/Sméagol uses any other pronoun, including an "inclusive we" that refers to others, the person case of their verbs will agree. If the addition of a sibilant to the end of verbs was simply an orthographic device, it could be applied randomly or to verbs regardless of their conjugation. Since there are limited environments where sibilants are reduplicated and other environments where they are not, it is logical to conclude that grammar rules may be a factor. However, given the number and variety of exceptions, those grammar rules may be incomplete.

4.3. Vocabulary

Tolkien used a creative orthographic choice to describe Gollum-Sméagol's speech; specifically, the addition of extra <s> letters to some words, to give the impression of hissed speech. To account for these artistic spellings, all spelling variations were added to the lemma list. However, the AntConc program would treat each variant as a unique token in some types of analysis regardless. As seen in

Figure 1, this had an impact on keyword statistics, but was otherwise mostly harmless.

When word choice is accounted for, vocabulary accounts for a small portion of Gollum-Sméagol's idiolect. In fact, only two words stand out as unique to Gollum-Sméagol, and both are archaic English words: "durst" and "praps." While Gollum-Sméagol uses the modern "perhaps" multiple times in the *Lord of the Rings* trilogy, he only says "praps" in *the Hobbit*, and only twice. Also only seen in *The Hobbit* is the archaic form of dared; "durst," is only used twice as "dursn't," its contraction with *durst* and *not*; Gollum-Sméagol never uses the modern verb in the books. Another word unique to Gollum-Sméagol is "ourselves," peculiar for its unvoiced fricative; Gollum-Sméagol otherwise add voice to any fricatives that receive a sibilant when pluralized (Elves, themselves). Additionally, Gollum-Sméagol use some adjectives that receive an <sy> suffix - a traditionally diminutive morpheme: "creepsy," "tricksy," and "bitsy."

5. Conclusion

Based on the analyses described in this paper, two primary conclusions can be made regarding the idiolects of Gollum and/or Sméagol. Firstly, the characters' idiolect appears to follow some general rules. Secondly, the linguistic mannerisms of the Gollum-Sméagol are distinct and change depending on which of the two are speaking and the context of character development.

The results of keyword analysis and concordance analysis support the conclusion that the idiolect of Gollum and Sméagol observes some rules concerning its unique features of pronoun use and verb inflection. Analyses revealed patterns in non-standard verb inflection; specifically, whether verbs were conjugated contrarily to subject-verb agreement.

Regarding sibilant reduplication, analysis revealed that sibilant reduplication is most likely an orthographic narrative device, possibly to illustrate the cave-dwelling Gollum as a snake-like and surreptitious character. Notably, Gollum was the more likely speaker to reduplicate sibilants, especially those in word-final positions. The most significant factor regarding sibilant reduplication appears to be that plural nouns were most likely to be

injected with extra sibilants. No other consistencies or complimentary distributions were apparent.

Gollum and Sméagol share most of their linguistic features, but there are nuanced differences in language use that can differentiate them as speakers. The most significant features of Gollum's speech are his use of non-standard pronouns and a relevantly higher frequency of sibilant reduplication. In contrast, Sméagol tends to lack sibilant reduplication and correctly uses personal pronouns. In general, the closer Sméagol is to redemption and rehabilitation, the closer he adheres to standard language rules.

In conclusion, there are some apparent rules that define the Gollum-Sméagol idiolect, especially in regard to its unique features of pronoun use, verb inflection, and sibilant reduplication. It is clear that the language used by Gollum and Sméagol was thoughtfully curated to reflect the characters' long isolation from society, social norms, and conversation. As a narrative device, these features and how intensely they are affected serve to illustrate their psychological turmoil, convey their character development, and distinguish the two characters.

References

- [1] Carpenter, Humphrey; Tolkien, Christopher, eds. (1981). *The Letters of J. R. R. Tolkien*. London: George Allen & Unwin.
- [2] Tolkien, J. R. R. (2009b). *The Lord of the Rings: The Fellowship of the Ring*. [Ebook version]. London: HarperCollins.
- [3] Goyal, Kovid (2019). Calibre (Version 3.42.0) [Computer Software]. Centresource Interactive Agency. Available from <https://calibre-ebook.com/about>.
- [4] Anthony, L. (2019). AntConc (Version 3.5.8) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net/software>.
- [5] Tolkien, J. R. R. (2009a). *The Hobbit: Or, there and back again*. [Ebook version]. London: HarperCollins.
- [6] Tolkien, J. R. R. (2009c). *The Lord of the Rings: The Two Towers*. [Ebook version]. London: HarperCollins.
- [7] Tolkien, J. R. R. (2009d). *The Lord of the Rings: Return of the King*. [Ebook version]. London: HarperCollins.
- [8] Carpenter, Humphrey (1977). *Tolkien: A Biography*. New York: Ballantine Books.
- [9] Tolkien, J. R. R. (1968). Interviewed by John Izzard for BBC, 30 March. Available at: <http://www.bbc.co.uk/archive/writers/12237.shtml> (Accessed: May 22, 2019).
- [10] Tolkien, J. R. R. (2012). *The Silmarillion*. [Ebook version]. London: HarperCollins.

