

University of Heidelberg  
English Department  
Proseminar I Literary Studies: Beckett's Dramatic Works  
Instructors: Armen Hesse, Joshua Elwer  
WS 2022/23

## Analysis and Comparison of Beckett's *Waiting for Godot* and *Endgame* using Natural Language Processing (NLP)

Polina Kuznetcova  
qr269@stud.uni-heidelberg.de  
English Studies/Computational Linguistics, 2<sup>nd</sup> Semester  
3680142  
31<sup>st</sup> of March 2023

## Contents

Introduction.....	2
Samuel Beckett.....	2
Summary of the plays.....	2
<i>Waiting for Godot</i> .....	2
<i>Endgame</i> .....	2
Natural Language Processing (NLP).....	3
Text Pre-processing .....	3
<i>Bag-of-words</i> .....	3
<i>Lower casing</i> .....	4
<i>Elimination of punctuation and special characters</i> .....	4
<i>Elimination of stop words</i> .....	4
<i>Lemmatization</i> .....	4
Text analysis .....	4
Analysis.....	5
Concluding remarks .....	14
Supplementary .....	15
References.....	16

# Introduction

## Samuel Beckett

Irish novelist, poet, and playwright Samuel Beckett (13 April 1906 – 22 December 1989) is widely regarded as one of the most influential writers of the 20th century.<sup>1</sup> His name is closely associated with the theater of the absurd, a movement that rejected conventional plot and character development in favor of exploring the human condition in a surrealistic style. Beckett's *Waiting for Godot* (1952) and *Endgame* (1957) are notable examples of this movement.<sup>2</sup> In reading the plays, I realized they share several similarities, including existential themes, minimalist settings, and repetitive structures. My interest in the theatre of absurd and computational linguistics led me to investigate whether certain linguistic patterns in Beckett's use of language would support my claims. To accomplish this, I applied my knowledge of Natural Language Processing (NLP) to conduct the linguistic analysis of the plays and compare the plays.

## Summary of the plays

### *Waiting for Godot*

Vladimir and Estragon, the main characters of this tragicomedy in two acts, wait for someone named Godot on a country road by a tree. In the first act, they encounter Pozzo and his slave Lucky, who add to the surreal nature of the play. Later on, a boy appears and announces that Godot will not come that day, but promises to come the next day. The second act is similar to the first, except that Pozzo is blind and Lucky is speechless. Vladimir and Estragon are still waiting for Godot, and in the evening the envoy of Godot appears again and conveys the same message as in the first act. The play is set in a desolate landscape and consists mainly of meaningless conversations between Didi and Gogo, as Vladimir and Estragon call each other. Throughout the play, Vladimir and Estragon consider leaving or committing suicide several times, but remain onstage without moving or taking any action.<sup>3</sup>

### *Endgame*

The play features four characters: Hamm, an old man who is blind and cannot get up from his chair; Clov, his servant, who takes care of Hamm; Nagg and Nell, Hamm's elderly parents,

---

<sup>1</sup> Linda Ben-Zvi and Angela Moorjani, eds., *Beckett at 100*, (Oxford University Press, 2008), 3.

<sup>2</sup> Ruby Cohn, *A Beckett Canon*, (University of Michigan Press, 2001), 21, 382.

<sup>3</sup> Samuel Beckett, *Waiting for Godot: A Tragicomedy in Two Acts*, 1st edition (Grove Press, 2011).

who live in adjacent trash bins. Throughout the play, Hamm tries to find an end to his suffering by asking Clov to kill him or to get him a painkiller. Hamm and Clov are dependent on each other, but the same time, while Hamm often orders Clov around and insists on his presence, Clov threatens to leave and longs for freedom. The characters are engaged in a series of circular conversations often about the nature of their existence, their memories, and the meaning of life. In the final of the play, Clov prepares to leave Hamm, but doesn't.<sup>4</sup>

### Natural Language Processing (NLP)

In the most basic sense, NLP is a subfield of artificial intelligence and computational linguistics.<sup>5</sup> NLP techniques allow researchers to analyze and understand text at a deeper level than traditional close reading methods, leading to more elaborate perspectives. This method offers an objective analysis of the text without the influence of personal bias or interpretation. It can also identify patterns and trends in literature, such as recurring themes and motifs, which can provide insights into the text. Considering Beckett's repetitive and minimalistic use of language, computational analysis of his works appears especially effective.

### Text Pre-processing

To make the text analysis more efficient and accurate, the data needs to be cleaned and standardized. Normally, in order to pre-process the text for analysis, the following steps are applied to the data:<sup>6</sup>

#### *Tokenization*

Tokenization is a fundamental step in NLP. It is a way of separating a text into smaller units (words or sentences) called tokens. A vocabulary is then created using the following tokens.<sup>7</sup> In this term paper, word tokenization was conducted.

#### *Bag-of-words*

The bag-of-words model is a representation of text in NLP. By this model, a text (such as a sentence or document) is represented as a collection of words, disregarding word order and grammar while maintaining the number of words.<sup>8</sup>

---

<sup>4</sup> Samuel Beckett, *Endgame and Act Without Words* (Grove Press, 2009).

<sup>5</sup> Daniel Jurafsky and James H. Martin, *Speech and Language Processing*, 2nd ed. (Prentice Hall, 2008), 10.

<sup>6</sup> Peter Jackson and Isabelle Moulinier, *Natural Language Processing for Online Applications: Text Retrieval, Extraction and Categorization*, (John Benjamins Pub, 2002), 9-13.

<sup>7</sup> Peter Jackson and Isabelle Moulinier, 10.

### *Lower casing*

Words "new" and "New" carry the same information, but are recognized as two different words. To avoid that, all words are transformed into lowercase.<sup>9</sup>

### *Elimination of punctuation and special characters*

Punctuation marks are not words and can introduce unnecessary noise to the data. Removing punctuation helps to standardize the data, making it more consistent and easier to compare across different sources.<sup>10</sup>

### *Elimination of stop words*

Words that can be found in every text and do not carry much relevant information, such as "the" and "of", are removed.<sup>11</sup>

### *Lemmatization*

After a morphological analysis (Part-of-Speech Tagging) inflected forms of a word are grouped together ("caring" -> "care"). The goal of lemmatization is to reduce inflectional forms and bring derivationally related forms of a word to the same base form.<sup>12</sup>

## **Text analysis**

### *n-grams*

An n-gram is a sequence of n items (letters, words etc.) from a sample of text or speech.<sup>13</sup> In this paper, bigrams are of interest. Bigrams are pairs of consecutive words in a text. Analyzing them can help to identify frequent word combinations and collocations, which words tend to co-occur with each other, and how their order and frequency of occurrence affects the overall meaning of a text.<sup>14</sup>

### *Verb-object analysis*

The verb in a sentence typically expresses the action or state the subject is performing or experiencing, and the object usually receives the verb's action. By analyzing the verb-object

---

<sup>8</sup> Daniel Jurafsky and James H. Martin, 642.

<sup>9</sup> Peter Jackson and Isabelle Moulinier, 9-10.

<sup>10</sup> Peter Jackson and Isabelle Moulinier, 9-10.

<sup>11</sup> Peter Jackson and Isabelle Moulinier, 9.

<sup>12</sup> Peter Jackson and Isabelle Moulinier, 11-12.

<sup>13</sup> Christopher D. Manning and Hinrich Schütze, *Foundations of Statistical Natural Language Processing*, (MIT Press, 1999), 192.

<sup>14</sup> Christopher D. Manning and Hinrich Schütze, 193.



Estragon, Hamm, and Clov, followed by Pozzo, Lucky, Nagg, and Nell. Unexpectedly, there is a striking similarity in number of dialogues between Hamm and Vladimir, and Clov and Estragon respectively (Table 1).

Name	HAMM	VLADIMIR	CLOV	ESTRAGON
Number of dialogues	349	346	332	331

Table 1. Number of dialogues by Hamm, Vladimir, Clov, and Estragon

After the data was tokenized and the noise (punctuation marks and stop words) was removed, the vocabulary of *Waiting for Godot* was equal to 1859 words, whereas that of the *Endgame* was 1846 words. This relationship primitively shows that the extent of vocabulary used in the two plays was similar, and performing a computational comparison between the two is relevant. To improve the efficacy of lemmatization and stop word detection, parts of speech were assigned to the tokens before removing stop-words and lemmatizing. After performing noise reduction, lemmatization was conducted. This removal of the excess words led to the bag of words containing 1425 different words for *Waiting for Godot*, and 1461 words for *Endgame*. Finally, words consisting of 1 or 2 letters, and all the words with frequency lower than 2 were removed, giving a total of 388 words in *Waiting for Godot* and 370 words in *Endgame*. This approach shows that the count of word types is similar between the two plays. The whole process is summarized in the Table 2 below.

Step	Number of tokens in <i>Waiting for Godot</i>	Number of tokens in <i>Endgame</i>
Tokenization including punctuation	2105	2092
Stop-words removal	1859	1846
Lemmatization	1425	1461
Punctuation removal	1405	1447
1-2 letter tokens removal	1383	1432
Removal of words with frequency lower than 2	388	370

Table 2. Vocabulary count after each text pre-processing step

Table 2 demonstrates that there is no big difference in the vocabulary count between both plays after every step of text pre-processing. It means that the plays have an approximately same amount of punctuation, stop-words, and 1-2 letter tokens.

Even though the final numbers of word types we end up with are of the same range, it is important to note that plays do not share a lot of common words. In total, both plays contain 563 word types, 195 of which are shared (a mere 34.6%), denoting that the majority of words used in the plays were different, which is dramatically illustrated in Figure 2.





vocalizations and pauses conjoin”.<sup>17</sup> Beckett's characters often have difficulty expressing themselves, and their silence might depict their inability to convey their thoughts and feelings. As a result, silence often illustrates the futility of language and the limitations of human communication. Further, Vladimir and Estragon's long periods of silence in *Waiting for Godot* create a sense of anticipation for the arrival of the titular character. Beckett also makes use of pauses and ellipses in *Endgame* to create a sense of emptiness and loneliness. Beckett understands the basic structure of conveying information, and provides a contrast to the dialogues themselves by adding silence. Just like how more black in a picture makes the white pop out, “speech induces confusion by being antagonistic, discordant, and fragmented, while the pauses become an exploration of creative multiplicity, an element of stability which resists indeterminacy and contradiction.”<sup>18</sup>

	<i>Waiting for Godot</i>	<i>Endgame</i>
<i>pause</i>	53	379
<i>silence</i>	56	6

Table 3. Frequency of *pause* and *silence* in *Waiting for Godot* and *Endgame*

While analyzing the most frequent words in a text can give us a general idea of the topics discussed, it may not reveal how those words are used in relation to each other. To define frequently occurring word combinations and recurring phrases, the bag of words was split into bigrams. Analyzing bigrams provides more context and insight into word relationships. The most frequent bigrams are represented in Figure 4.

---

<sup>17</sup> Pia Brinzeu, “‘Without a Word’: Pauses in Beckett’s Drama,” *Samuel Beckett Today / Aujourd’hui* 2 (1993), 229.

<sup>18</sup> Pia Brinzeu, 231.



Figure 4. Most frequent bigrams in *Waiting for Godot* and *Endgame*

A lot of bigrams have their names in their, which happens because each line in a play starts with a name. This shows “a limitation” of such an assay performed on plays, and is something that one must keep in mind. As it follows from the Figure 4, the most frequent combination of words is *pause hamm*. This bigram most possibly happens when Hamm speaks after a pause some other character makes. A bit more rare is a situation when a pause is followed by Clov's line (bigram *pause clon*). Since Hamm and Clov have approximately the same number of dialogues, it means that they make pauses equally often, showing that Beckett gives almost equal weightage to both his characters in the play. Similarly, in *Waiting for Godot* silence is broken by Vladimir and Estragon equally frequently (bigrams *silence vladimir*, *silence estragon* have approximately same frequency.) Next on the list are bigrams *vladimir estragon*, *clon hamm*, *estreron vladimir*. This can be misinterpreted by someone who has not read the play as one characters calling each other by their full name in a beginning of the dialogue (i.e. VLADIMIR: Estragon...), hence it shows another shortcoming of the approach used. One should remember that the analyzed data has stop-words and words consisting of 1 or 2 letters removed. As Vladimir and Estragon did not refer to each other by their full name, it means that often their dialogues were extremely short and consisted of a stop-word or short words, such as:

VLADIMIR: Ah yes, the two thieves. Do you remember the story?

ESTRAGON: No.

[...]



bigrams that is consistently repeated is *yes sir*. The Boy in *Waiting for Godot* said it 14 times, which again points at Beckett's repetitive use of language.

Among the most frequent bigrams, there is a curious bigram *reason unknown*. It can be easily overlooked when reading *Waiting for Godot*, as it is a part of Lucky's fragmented speech. He repeats the phrase *for reasons unknown* ten times in his dialogues, which underscores the sense of confusion and uncertainty in the play.

The majority of the most frequent bigrams represented in Figure 5 (*put down, sit down, get down, take off, look like, move chair* etc) contain a verb. To gain a better understanding of the actions described, verb-object analysis was conducted.

### Verb-Object Analysis

Further analysis was conducted on the most frequently occurring predicate-argument structures in each play, particularly on the verb-object relations. Verbs together with their arguments play a crucial role in shaping the meaning of a phrase.<sup>19</sup> Identifying the most commonly used verbs and their associated nouns or pronouns can provide a better understanding of the actions and relationships described in the text. A function was written to identify presence of an object argument in the sentence with the verb. Only characters' dialogues and no stage directions were considered for the analysis. The results are presented in the Table 4 and Table 5.

Word	Count	Has an object	Has no object	% has an object
Know	43	31	12	72%
Say	42	38	4	90%
Let	33	33	0	100%
Come	31	5	26	16%
Tell	30	26	4	87%
Think	30	15	15	50%
Wait	28	7	21	25%
Look	17	13	4	76%
Put	15	15	0	100%

Table 4. Most frequent verbs in *Waiting for Godot*

Verbs *let* and *put* require at least one object to complete their meaning, thus in 100% of cases they have it, and the rest of the given predicates can occur without one. It is significant that in contrast to all verbs having an object in more than half the cases, *come* and *wait* only have an object in 16% and 25%, respectively. Even if the destination of *come* or the

<sup>19</sup> Daniel Jurafsky and James H. Martin, 339, 506-507.

object of *wait* are implied by the context, in most of the cases Beckett decides not to name them. Regardless of whether Beckett eliminates the objects intentionally or not, the idea of waiting, simply waiting for nothing or no one aligns with the overall idea of the play – endless, or even aimless waiting. Moreover, it goes together with the speculation that the name Godot comes from the Irish word "go deo" meaning "forever," as "waiting forever" is what the main characters of the play seem to be doing.

Remarkably, the bigram *put down* (Figure 5) appears in the play 10 times, which is a mere 67% of this verb's count, and in other 4 cases *put* is complemented by the adverb *on*. For further analysis, the window around the verb *put* was expanded, and one can see that the sentences it appears in repeat themselves.

*ESTRAGON: Why doesn't he **put down his bags**? – this question by Estragon appears 4 times in the entire play, which again points at the circularity of the play.*

*ESTRAGON: Ask him what? VLADIMIR: Why he doesn't **put down his bags**.*

*POZZO: You want to know why he doesn't **put down his bags**, as you call them.*

*ESTRAGON: **Bags**. Why? Always hold. Never **put down**.*

*VLADIMIR: Since he has **put down his bags**...*

*VLADIMIR: Anyway he has **put them [bags] down**.*

*ESTRAGON: And why has he **put them [bags] down**?*

Similarly, the predicate *put on* appears in the exact same context. When talking about the hat, Pozzo and Vladimir repeat an entirely identical phrase thrice in the same dialogue:

*POZZO: You must **put it on his head**.*

*ESTRAGON: (to Pozzo). Tell him to take it.*

*POZZO: It's better to **put it on his head**.*

*VLADIMIR: I'll **put it on his head**.*

As demonstrated by the examples above, the characters in the play often repeat themselves word for word or with slight variations. It creates an impression that the characters do not listen to each other, as their conversations reoccur multiple times and do not bring them anywhere. Thereby, Beckett deconstructs communication to show the futility of language, and it reaches its peak in Lucky's fragmented and disconnected speech:

*LUCKY: Given the existence as uttered forth in the public works of Puncher and Wattmann of a personal God quaquaquaqu with white beard quaquaquaqu*

*outside time without extension who from the heights of divine apathia divine athambia divine aphasia loves us dearly...*

Additionally, the phrases *come on* and *let's go* are repeated in the play 9 and 6 times, respectively. The phrase *let's go* is used *to express a wish, request, or suggestion to depart*,<sup>20</sup> and *come on* is used in the sense of *encouraging someone to do something such as make a greater effort or stop being sad*.<sup>21</sup> It is a curious observation, as the characters do not depart or make any effort neither after repeating *come on* thrice in this dialogue,

VLADIMIR: *Are you mad? We must take cover. **Come on.***

ESTRAGON: *Pity we haven't got a bit of rope.*

VLADIMIR: ***Come on.** It's cold.*

ESTRAGON: *Remind me to bring a bit of rope tomorrow.*

VLADIMIR: *Yes. **Come on.***

nor after expressly agreeing to leave in the closing dialogue of the play:

ESTRAGON: *Well, shall we go?*

VLADIMIR: *Yes, **let's go.***

Similarly, in *Endgame*, Clov threatens to leave Hamm 11 times. However, he remains motionless even after covering Hamm's face with a handkerchief at the end of the play. The characters' inability to escape creates an atmosphere of entrapment and despair. Overall, the use of the verb *leave* establishes relations between characters in *Endgame*. Clov constantly threatens Hamm to leave, and Hamm does not let him go. Both characters rely heavily on each other. Clov depends on Hamm for food and shelter, and Hamm needs Clov as his eyes and legs. Their dependency on each other is illustrated in the lines below:

CLOV: (to Hamm) [...] *I'll leave you, I have things to do.*

HAMM: (to Clov) *I can't leave you.*

CLOV: (to Hamm) *So you all want me to leave you.*

HAMM: (to Clov) *You won't be able to leave me.*

Verbs *get*, *look*, *give*, and *take* are used by Hamm in an imperative mood as commands to Clov. It is illustrated in Table 5 that in the majority of cases, these verbs have an object, which can be explained by the aim to make the commands precise and clear.

---

<sup>20</sup> "Definition of LET'S GO," accessed March 31, 2023, <https://www.merriam-webster.com/dictionary/let's%20go>.

<sup>21</sup> "COME ON (Phrasal Verb) Definition and Synonyms," accessed March 31, 2023, [https://www.macmillandictionary.com/dictionary/british/come-on\\_1](https://www.macmillandictionary.com/dictionary/british/come-on_1).

The verb *come* has an object only in 25% of cases, which is close to 16% in *Waiting for Godot*. All in all, 6 out of 9 most frequent verbs in *Waiting for Godot* are among the most frequent verbs in *Endgame*. This means that the majority of phrases in both plays have similar predicates referring to the same actions.

Word	Count	Has an object	Has no object	% has an object
get	49	33	16	67%
know	34	16	18	47%
say	34	34	0	100%
come	32	8	24	25%
look	30	17	13	57%
leave	26	26	0	100%
give	22	22	0	100%
think	19	17	2	89%
hear	18	16	2	89%
feel	17	17	0	100%
tell	16	16	0	100%
take	15	14	1	93%

Table 5. Most frequent verbs in *Endgame*

## Concluding remarks

Overall, the analysis demonstrates that there are certain linguistic patterns in *Endgame* and *Waiting for Godot*. The plays share many similarities at every level of text processing. There is remarkable similarity in the number of dialogues between Hamm and Vladimir, and Clov and Estragon (349 and 346, 332 and 331 dialogues respectively). The count of vocabulary used in both plays lies in the same range (388 word types for *Waiting for Godot* and 370 for *Endgame*), but only 195 word types are shared, meaning that the plays are worth comparing. Furthermore, both *Waiting for Godot* and *Endgame* share an abundance of pauses and silences, which Beckett uses as a powerful tool to show the futility and limits of language in both plays.

It was proven with examples that Beckett's language use in *Waiting for Godot* and *Endgame* is repetitive and circular. His characters often repeat the same questions or requests to each other, and they do not get a response, as communication in both plays is deconstructed. Finally, waiting in *Waiting for Godot* is indeed *go deo* (meaning *forever* in

Irish<sup>22</sup>), as the verb *wait* is complemented by an object only in 25% of cases. Overall, *Waiting for Godot* and *Endgame* share most of the frequently used verbs, meaning that the majority of actions in the plays are similar.

NLP approach made it possible to carry out a representative quantitative analysis of Beckett's works. However, as computational linguistics is an evolving field, NLP analysis should be approached critically. This study showed the common problems and limitations one can encounter when analysing a play using conventional NLP methods. The limitations were mainly connected with pauses and names, as they created noise in the data. It is good practice to remove stage directions and the characters' names from the analysed data.

Further research involving NLP techniques in analysis of Beckett's works could explore semantic roles in predicate-argument structures, semantic or syntactical parsing, or sentiment analysis of the plays or separate acts and passages.

## Supplementary

In order to deepen my understanding of neural networks and deep learning, as a creative side project, I trained deep learning long short-term memory (LSTM) network to generate text based solely off of *Endgame* and *Waiting for Godot*. For more details please refer to the document attached.

---

<sup>22</sup> "Foclóir Gaeilge–Béarla (Ó Dónaill): Deo," accessed March 30, 2023, <https://www.teanglann.ie/en/fgb/deo>.



## References

- Beckett, Samuel, trans. *Endgame and Act Without Words*. New York: Grove Press, 2009.
- Beckett, Samuel. *Waiting for Godot: A Tragicomedy in Two Acts*. 1st edition. New York: Grove Press, 2011.
- Ben-Zvi, Linda and Angela Moorjani, eds., *Beckett at 100*. New York: Oxford University Press, 2008.
- Brinzeu, Pia. “‘Without a Word’: Pauses in Beckett’s Drama.” *Samuel Beckett Today / Aujourd’hui* 2 (1993): 229–34.
- Chowdhary, K. R. “Natural Language Processing.” *Fundamentals of Artificial Intelligence*, edited by K.R. Chowdhary, 603–49. New Delhi: Springer India, 2020.
- Cohn, Ruby. *A Beckett Canon*. Ann Arbor: University of Michigan Press, 2001.
- “COME ON (Phrasal Verb) Definition and Synonyms | Macmillan Dictionary.” Accessed March 31, 2023. [https://www.macmillandictionary.com/dictionary/british/come-on\\_1](https://www.macmillandictionary.com/dictionary/british/come-on_1).
- “Foclóir Gaeilge–Béarla (Ó Dónaill): Deo.” Accessed March 30, 2023. <https://www.teanglann.ie/en/fgb/deo>.
- Jackson, Peter, and Isabelle Moulinier. *Natural Language Processing for Online Applications: Text Retrieval, Extraction and Categorization*. Amsterdam: John Benjamins Pub, 2002.
- Jurafsky, Daniel, and James H. Martin. *Speech and Language Processing*. 2nd ed. Prentice Hall, 2008.
- “Let’s Go Definition & Meaning - Merriam-Webster.” Accessed March 31, 2023. <https://www.merriam-webster.com/dictionary/let's%20go>.

Manning, Christopher D, and Hinrich Schiitze. *Foundations of Statistical Natural Language Processing*. 2nd ed. Cambridge: MIT Press, 1999.

“MATLAB - MathWorks.” Accessed March 31, 2023.

<https://de.mathworks.com/products/matlab.html>.

NobelPrize.org. “The Nobel Prize in Literature 1969.” Accessed March 31, 2023.

<https://www.nobelprize.org/prizes/literature/1969/beckett/biographical/>.

Song, Jaeyeol, Jin-Kook Lee, Jungsik Choi, and Inhan Kim. “Deep Learning-Based Extraction of Predicate-Argument Structure (PAS) in Building Design Rule Sentences.” *Journal of Computational Design and Engineering* 7, no. 5 (October 13, 2020): 563–76.

<https://doi.org/10.1093/jcde/qwaa046>.

Tomlin, Russell S. *Basic Word Order: Functional Principles*. Croom Helm Linguistics

Series. London: Croom Helm, 1986. <http://bvbr.bib->

[bvbr.bib-bvb.de:8991/F?func=service&doc\\_library=BVB01&local\\_base=BVB01&doc\\_number=000349070&line\\_number=0001&func\\_code=DB\\_RECORDS&service\\_type=MEDIA](http://bvbr.bib-bvb.de:8991/F?func=service&doc_library=BVB01&local_base=BVB01&doc_number=000349070&line_number=0001&func_code=DB_RECORDS&service_type=MEDIA).