

## Dataset

- Source: Project Gutenberg eBook of *The Brothers Karamazov*
- Preprocessing: Boilerplate stripped, whitespace normalized, chapter headers removed
- Total sentences: 19,561
- Middle 10,000 sentences extracted for analysis

## Part 1: SentenceTransformer Embeddings

### Embedding

- Model: `SentenceTransformer(all-mpnet-base-v2)`
- Sentences embedded: 10,000

### Clustering (k = 142)

- Method: KMeans with elbow method using 30 randomly sampled values of k in the range (2, 500)
- Optimal k: 142
- Largest cluster size: 157, smallest: 13

### Sentence Counts per Cluster (Sorted and Normalized)

#### Raw counts:

[157, 146, 139, 134, 129, 127, 124, 124, 121, 120, 120, 118, 110, 109, 108, 107, 107, 103, 99, 98, 95, 95, 95, 94, 94, 93, 92, 92, 90, 89, 88, 88, 87, 85, 84, 81, 80, 80, 80, 80, 79, 78, 78, 78, 77, 77, 77, 77, 77, 76, 76, 75, 75, 74, 74, 73, 73, 73, 72, 71, 71, 71, 70, 69, 69, 69, 67, 67, 67, 67, 67, 67, 67, 66, 66, 66, 65, 65, 65, 65, 65, 63, 63, 63, 62, 62, 62, 61, 61, 60, 59, 59, 59, 58, 58, 58, 57, 57, 56, 56, 55, 55, 55, 55, 54, 53, 53, 53, 52, 51, 49, 48, 47, 45, 45, 45, 44, 44, 43, 43, 43, 42, 42, 42, 41, 39, 38, 37, 36, 35, 34, 34, 34, 34, 33, 30, 29, 27, 23, 19, 13]

#### Normalized counts:

[1.000, 0.930, 0.885, 0.854, 0.822, 0.809, 0.790, 0.790, 0.771, 0.764, 0.764, 0.752, 0.701, 0.694, 0.688, 0.682, 0.682, 0.656, 0.631, 0.624, 0.605, 0.605, 0.605, 0.599, 0.599, 0.592, 0.586, 0.586, 0.573, 0.567, 0.567, 0.561, 0.561, 0.554, 0.541, 0.535, 0.516, 0.510, 0.510, 0.510, 0.510, 0.503, 0.497, 0.497, 0.497, 0.490, 0.490, 0.490, 0.490, 0.490, 0.484, 0.484, 0.478, 0.478, 0.471, 0.471, 0.465, 0.465, 0.465, 0.459, 0.452, 0.452, 0.452, 0.446, 0.439, 0.439, 0.439, 0.427, 0.427, 0.427, 0.427, 0.427, 0.427, 0.420, 0.420, 0.420, 0.414, 0.414, 0.414, 0.414, 0.414, 0.401, 0.401, 0.401, 0.395, 0.395, 0.395, 0.389, 0.389, 0.382, 0.376, 0.376, 0.376, 0.369, 0.369, 0.369, 0.363, 0.363, 0.357, 0.357, 0.350, 0.350, 0.350, 0.344, 0.338, 0.338, 0.338, 0.331, 0.325, 0.312, 0.306, 0.299, 0.287, 0.287, 0.287, 0.280, 0.280, 0.274, 0.274, 0.274, 0.268, 0.268, 0.268, 0.261, 0.248, 0.242, 0.236, 0.229, 0.223, 0.217, 0.217, 0.217, 0.217, 0.210, 0.191, 0.185, 0.172, 0.146, 0.121, 0.083]

### Print Sentences in Clusters

A cluster with 49 sentences showed strong thematic coherence centered around the concept of *father*.

#### Sample Sentences from Cluster ID 135 (size = 49):

- “So your boy is a good boy, he loves his father, and he attacked me as the brother of your assailant.... Now I understand it,” he repeated thoughtfully.
- “Some idiot comes in, and you put us to shame!” cried the girl by the window, suddenly turning to her father with a disdainful and contemptuous air.

- “Oh, father, father!” the hunchback girl, who had till then been silent on her chair, said suddenly, and she hid her eyes in her handkerchief.
- “Father, father, how can you—with him!”
- “Let him alone!” cried the boy, sitting up in his bed and gazing at his father with glowing eyes.
- “Father,” he cried, “father!” He caught hold of me, hugged me, tried to pull me away, crying to my assailant, “Let go, let go, it’s my father, forgive him!”
- An ordinary boy, a weak son, would have submitted, have felt ashamed of his father, sir, but he stood up for his father against them all.
- “Father,” said he, “father!” “Well?” said I. I saw his eyes flashing.
- “Father,” he said, “anyway don’t forgive it.”
- “Father,” he kept crying, “dear father, how he insulted you!” And I sobbed too.
- “My brothers are destroying themselves,” he went on, “my father, too.”
- At last the child cannot scream, it gasps, “Daddy!”
- A father corrects his child.
- “He will raise your child,” the crowd shouts to the weeping mother.
- “If it is Thou, raise my child!” she cries, holding out her hands to Him.
- What have I to do with you?” was on the tip of his tongue, but to his profound astonishment he heard himself say, “Is my father still asleep, or has he waked?”
- His father heard this announcement with no sign of surprise, and forgot in an unmannerly way to show regret at losing him.
- You must excuse me.” “Come, you might oblige your father.”
- “Don’t be too hard on me!” the father called for the last time.
- Father Iosif, who had witnessed the scene yesterday, exchanged glances with Father Païssy.
- Let him gather round him once a week, some hour in the evening, if only the children at first—the fathers will hear of it and they too will begin to come.
- Of both of them Father Païssy felt for some reason suddenly suspicious—though, indeed, he might well have felt the same about others.
- Father Païssy stood over him for a little.
- Father Iosif walked away, grieving the more since he had put forward his own opinion with little confidence as though scarcely believing in it himself.
- “What have you come for, worthy Father?”
- “Go, Father, and do not trouble the flock!” he repeated impressively.
- “Go away, Father, I command you!” Father Païssy thundered in conclusion.
- “Have you, too, fallen into temptation?” cried Father Païssy.
- “You will come back again!” murmured Father Païssy, looking after him with sorrowful surprise.
- Critical Moment Father Païssy, of course, was not wrong when he decided that his “dear boy” would come back again.
- Father Païssy raised his eyes for an instant from his book, but looked away again at once, seeing that something strange was happening to the boy.
- All the family trembled before the father.
- Only, Father, I don’t know what we’re to do with you.
- His father was standing near the window, apparently lost in thought.

- “He must have killed his old father!” she cried, clasping her hands.
- My father was alarmed and, for the first time, he saw me then, cried out, and sprang back from the window.
- “Why, he’s very likely his son, you know—his natural son.”
- “It was the devil murdered father, it was through the devil that you found it out so soon.”
- “Not from my father, don’t be uneasy.”
- “His father will go out of his mind or hang himself.”
- “My father would be awfully cross.”
- “His father was terribly concerned about him.”
- “His father saw it.”
- The sick boy was holding his hand and calling for his father.
- “Father, father!”
- “Ah, father!”
- “I shall expect you.... Father, father!” His voice broke.
- “Children ought to honor their parents, and yet it would be better if it were he...”
- “I didn’t, of course, expect him to come often, knowing what a lot he has to do as it is...”

Most of the sentences in this cluster contain the word “father” explicitly, suggesting strong lexical cohesion. However, a few sentences, while not directly mentioning the word, convey the concept of fatherhood or paternal relationships. For example:

- “Why, he’s very likely his son, you know—his natural son.”
- “He will raise your child,” the crowd shouts to the weeping mother.”
- ““If it is Thou, raise my child!” she cries, holding out her hands to Him.”
- “Children ought to honor their parents, and yet it would be better if it were he, as you’d have nothing to cry over then, for he did it when he was unconscious...”
- “I didn’t, of course, expect him to come often, knowing what a lot he has to do as it is, vous comprenez, cette affaire et la mort terrible de votre papa.”

This demonstrates that the SentenceTransformer model not only clusters based on direct lexical overlap, but also captures semantic relationships and synonymous references, including multilingual cues (e.g., “papa”) and abstract notions of parenthood.

## Histogram

- Normalized cluster sizes ( $\text{cluster\_sizes} / \max(\text{cluster\_sizes})$ )
- Bin setup: 20 equally spaced bins between 0 and 1 (width = 0.05)
- Histogram: counts how many cluster sizes fall into each bin
- Peak around normalized size 0.35–0.45

### Histogram Bins and Counts:

Bins: [0.00, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 1.00]

Counts per bin: [0, 1, 2, 3, 10, 13, 8, 20, 22, 21, 8, 11, 5, 5, 1, 6, 2, 2, 1, 1]

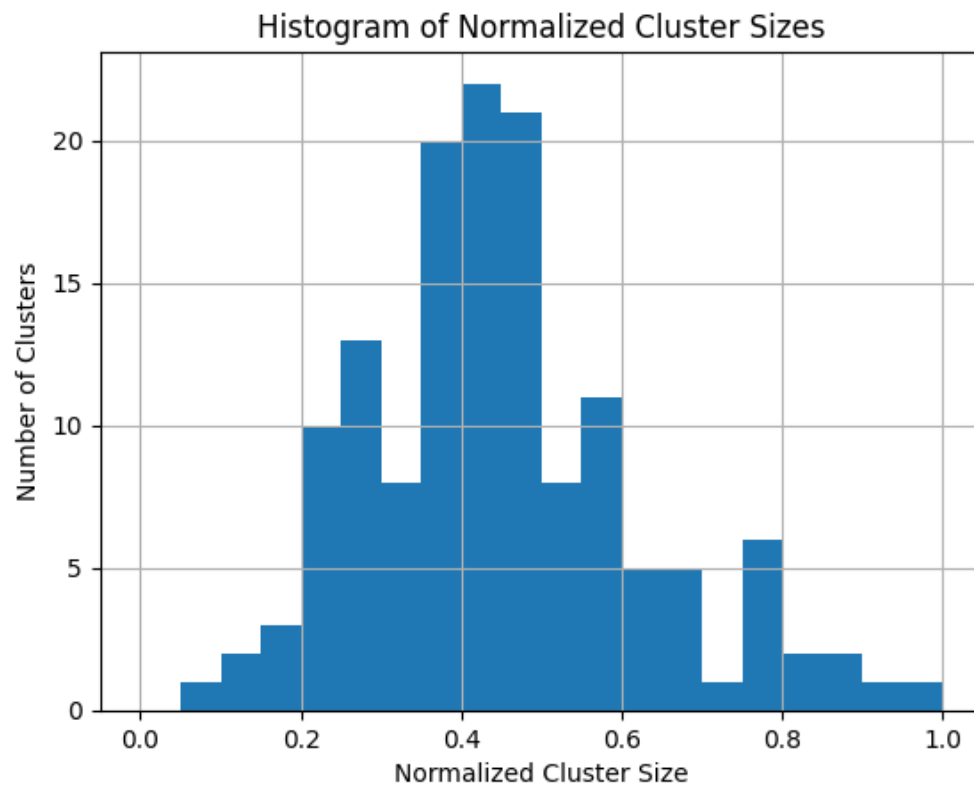


Figure 1: Histogram of Normalized Cluster Sizes (SentenceTransformer(all-mpnet-base-v2, Kmeans (k = 142)))

### Log-Log Plot

- Log-log slope: **-0.0929**

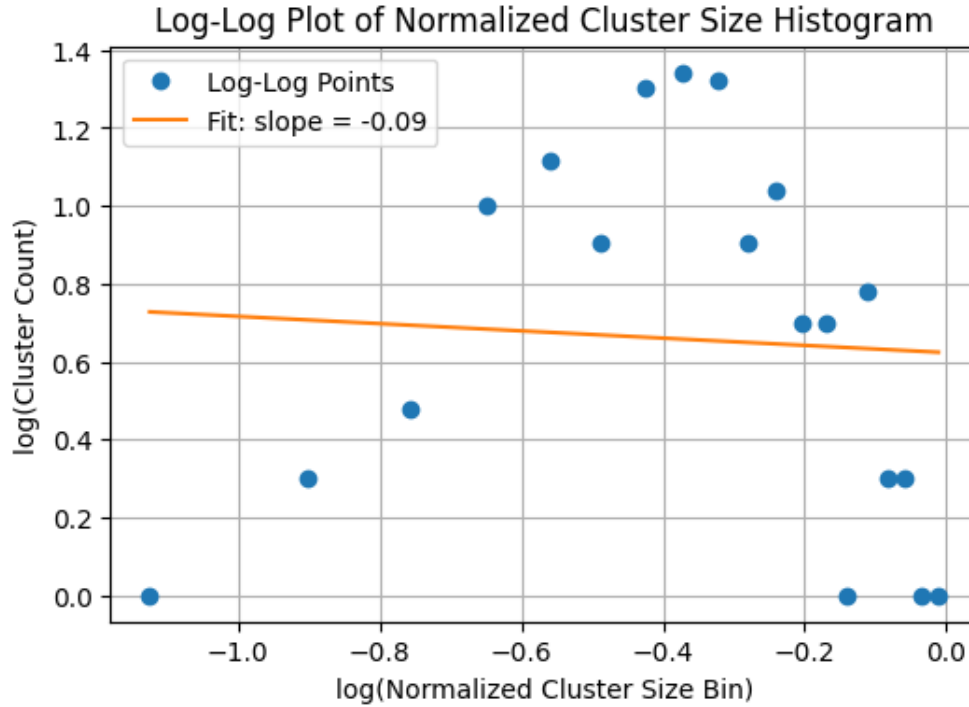


Figure 2: Log-Log Plot of Cluster Sizes (SentenceTransformer(all-mpnet-base-v2, Kmeans (k = 142)))

### Clustering (k = 516)

- Method: KMeans with elbow method using 50 randomly sampled values of k in the range (300, 800)
- Optimal k: 516
- Largest cluster size: 66, smallest: 3

### Sentence Counts per Cluster (Sorted and Normalized)

#### Raw counts:

[66, 66, 56, 52, 50, 50, 49, 48, 46, 46, 46, 46, 46, 45, 45, 45, 43, 41, 41, 41, 41, 40, 39, 38, 38, 38, 38, 37, 37, 37, 36, 36, 36, 36, 36, 36, 36, 35, 35, 35, 35, 34, 34, 34, 34, 34, 34, 34, 34, 33,...3]

#### Normalized counts:

[1.000, 1.000, 0.848, 0.788, 0.758, 0.758, 0.742, 0.727, 0.697, 0.697, 0.697, 0.697, 0.697, 0.682, 0.682, 0.682, 0.652, 0.621, 0.621, 0.621, 0.621, 0.606, 0.591, 0.576, 0.576, 0.576, 0.576, 0.561, 0.561, 0.561, 0.545, 0.545, 0.545, 0.545, 0.545, 0.545, 0.545, 0.530, 0.530, 0.530, 0.530, 0.515, 0.515, 0.515, 0.515, 0.515, 0.515, 0.515, 0.500, 0.500,...,0.045]

### Print Sentences in Clusters

#### Cluster ID: 43 (size = 49)

- “But Mitya persisted and sent his name up again.”
- “Mitya was sitting on a little chair at the entrance, awaiting his fate with nervous impatience.”

- “Mitya was well dressed, in a frock-coat, buttoned up, with a round hat and black gloves in his hands, just as he had been three days before at the elder’s, at the family meeting with his father and brothers.”
- “The old man waited for him, standing dignified and unbending, and Mitya felt at once that he had looked him through and through as he advanced.”
- “Mitya was greatly impressed, too, with Samsonov’s immensely swollen face.”
- “Mitya was somewhat surprised for a moment, and explained that that was what Samsonov had called him.”
- “She was just then expecting the ‘message,’ and was much relieved that Mitya had not been to see her that day or the day before.”
- “Mitya grew thoughtful.”
- “Mitya resolved to sacrifice an hour to it: ‘In an hour I shall know everything, I shall settle everything, and then, then, first of all to Samsonov’s.’”
- “As he worked out this new idea, Mitya was enchanted with it, but so it always was with him in all his undertakings, in all his sudden decisions.”
- “Mitya was stupefied.”
- “It was the old servant who waited on Samsonov, whom Mitya had particularly noticed the day before.”
- “He’s wild with impatience. ... Mitya slipped back at once, and fell to gazing in at the window again.”
- “Mitya watched him eagerly.”
- “Mitya put out his hand and began feeling it.”
- “What am I doing it for? thought Mitya, suddenly pulling himself together.”
- “Mitya looked at his hands again.”
- “Mitya, in desperate haste, scarcely soaped his hands (they were trembling, and Pyotr Ilyitch remembered it afterwards).”
- “He seemed to exercise more and more sway over Mitya, as time went on.”
- “You see I’ll ... Mitya went on with the same confiding ingenuousness, drying his face and hands on the towel, and putting on his coat.”
- “Ech, they won’t be half an hour before us, let alone an hour. Though Mitya bustled about seeing after things, he gave his orders strangely, as it were disconnectedly, and inconsecutively.”
- “Although Mitya was excited and expansive, yet he was melancholy, too.”
- “He’ll remember it afterwards, Mitya remarked.”
- “Quicker! Mitya cried again, nervously.”
- “But this all seemed to Mitya perfectly right and not to be questioned.”
- “Something seemed for the first time to flash upon Mitya’s mind.”
- “Mitya began to feel anxious.”
- “Mitya, too, returned, apologizing for having kept them waiting.”
- “You’ll be glad to hear it. The little pan was taken aback and looked apprehensively at Mitya.”
- “Take it and go your way. The Pole gazed open-eyed at Mitya, with a searching look.”
- “Let him keep it as a consolation. Bravo, Mitya!”
- “Do you remember, Mitya, do you remember how we made friends here last time! Mitya himself was almost delirious, feeling that his happiness was at hand.”

- “Mitya, dumb with rapture, gazed into her eyes, at her face, at her smile, and suddenly clasped her tightly in his arms and kissed her passionately.”
- “Mitya turned, and saw that some one had, in fact, parted the curtains and seemed to be watching them.”
- “Mitya (he remembered it afterwards) became suddenly intensely interested in his big rings, one with an amethyst, and another with a transparent bright yellow stone, of great brilliance.”
- “Allow me, gentlemen, allow me one minute more, interposed Mitya, putting his elbows on the table and covering his face with his hands.”
- “Mitya, hearing her cry, trembled, jumped up, and with a yell rushed impetuously to meet her, not knowing what he was doing.”
- “Mitya was impatiently anxious not to omit the slightest detail.”
- “Little by little Mitya began to grow surly.”
- “Mitya was surprised.”
- “Next came the account of Mitya’s sudden determination to ‘step aside’ and make way for their happiness.”
- “He’s come to look at me dressed up, thought Mitya.”
- “Mitya stared at it with open eyes.”
- “But Pan Vrublevsky confirmed the statement, and Mitya, after thinking for a moment admitted, frowning, that it must have been as the Poles stated, that he had been excited at the time, and might indeed have said so.”
- “This circumstance had been the one piece of evidence which, insignificant as it was, had hitherto told, to some extent, in Mitya’s favor.”
- “Mitya was calm, and even looked more cheerful, but only for a moment.”
- “All stared at Mitya.”
- “I’ve taken a chill, thought Mitya, twitching his shoulders.”
- “And now that day Mitya had taken it into his head to be fearfully jealous.”

**Analysis:** This cluster tightly revolves around the character Mitya (Dmitri Fyodorovitch), capturing his actions, internal monologues, and emotional states. The repetition of his name across all sentences strongly signals lexical cohesion, while the consistent narrative perspective and emotional themes support semantic similarity.

#### Cluster ID: 111 (size = 40)

- “At that thought he was completely comforted.”
- “It made him feel stronger.”
- “At the same time there was a sort of grin on his lips.”
- “The papa was glad that the birch was covered with twigs.”
- “He was carried away as he talked, and spoke with excitement; when he had finished, he suddenly smiled.”
- “But a change had passed over him, too.”
- “And all of a sudden he felt very happy.”
- “For two hours afterwards he felt almost happy, and sat drinking brandy.”
- “A marvelous change passed over him, his spirit seemed transformed.”
- “And the lad was delighted that the bear had walked away without hurting the saint, and that Christ was with him too.”

- “On the contrary, his friends, as I observed already, seeing him that night apparently so cheerful and talkative, were convinced that there was at least a temporary change for the better in his condition.”
- “He was not shaken by it, but awaited what would come next without fear, watching with penetration and insight for the outcome of the general excitement.”
- “But even this thought of the smell of corruption, which had seemed to him so awful and humiliating a few hours before, no longer made him feel miserable or indignant.”
- “He was trembling with delight.”
- “After eating, his spirits and his heart grew lighter.”
- “But a load had been lifted from his heart, anyway.”
- “At first fortune seemed to smile upon him.”
- “His first horror was over, but evidently some new fixed determination had taken possession of him.”
- “He suddenly stood up, smiling dreamily.”
- “The game cheered him.”
- “But that moment flew by like a spark.”
- “He caught his breath.”
- “He seemed to have become suddenly gentle and subdued.”
- “He seemed to have forgotten everything, and was looking round at every one with a childlike smile of delight.”
- “In his mood of doglike submissiveness all feeling of rivalry had died away.”
- “The fresh air revived him.”
- “His face beamed.”
- “He seemed completely transformed in a moment.”
- “He recovered himself instantly.”
- “He grew more good-humored at every word.”
- “But her face reassured him at once.”
- “Where?” he exclaimed opening his eyes, and sitting up on the chest, as though he had revived from a swoon, smiling brightly.”
- “I’ve had a good dream, gentlemen,” he said in a strange voice, with a new light, as of joy, in his face.”
- “Next day he had a slight attack of nervous fever, but he was in high spirits and well pleased with himself.”
- “At last he could go out.”
- “All this was very becoming to him, and he looked quite handsome.”
- “He hurried, because he was happy himself.”
- “The justice of the peace smiled and was vexed with himself at once for having smiled.”
- “He had by now completely recovered his dignity, though he was still a little uneasy.”
- “The news evidently impressed him.”

*Sentences expressing personal transformation, emotional relief, and calm joy.*

**Analysis:** This cluster captures internal psychological transitions such as comfort, happiness, or spiritual peace. The sentences are semantically unified through positive emotional outcomes and reflection, indicating clustering based on tone and affect rather than content keywords.



### Cluster ID: 230 (size = 36)

- “You see, I wanted to ask you, my dear Alexey Fyodorovitch, to go to him, to find some excuse to go to them—I mean to that captain—oh, goodness, how badly I explain it!—and delicately, carefully, as only you know how to manage to give him this assistance, these two hundred roubles.”
- “Listen, Alexey Fyodorovitch, at a moment like this you must listen, for you can’t understand what these two hundred roubles mean to me now.”
- “And now I can get a servant with this money, you understand, Alexey Fyodorovitch, I can get medicines for the dear creatures, I can send my student to Petersburg, I can buy beef, I can feed them properly.”
- “Katerina Ivanovna will send you as much more as you need, and you know, I have money too, take what you want, as you would from a brother, from a friend, you can give it back later....”
- “Mamma has just told me all about the two hundred roubles, Alexey Fyodorovitch, and your taking them to that poor officer ...”
- “He might have killed him yesterday on account of Grushenka ... but he won’t steal. He is in very great need of money now—the greatest need.”
- “That is just how it is with people who, like Dmitri, have never had anything to do with money ... and have no notion how money is obtained.”
- “You have saved a man from a violent death ... My eternal gratitude—I will give you more, infinitely more than three thousand!”
- “I only need that fatal three thousand ... I can give security for that sum with infinite gratitude ... I have promised to save you, and I will save you.”
- “Enough! as Turgenev says. ... “It is yours, Dmitri Fyodorovitch,” Madame Hohlakov cut in at once.”
- “The money is as good as in your pocket, not three thousand, but three million ...”
- “The depreciation of the rouble keeps me awake at night, Dmitri Fyodorovitch ...”
- “I shall perhaps set off ... to the gold-mines ... that three thousand you so generously ... if you could today ...”
- “For the last time I entreat you, tell me, can I have the sum you promised me to-day, if not, when may I come for it?”
- “As soon as the sun rises to-morrow ... ask her whether she did stump up that three thousand or not.”
- “In spite of the thousands of roubles he had saved, Trifon Borissovitch was very fond of emptying the pockets of a drunken guest ...”
- “You gave the peasants champagne to drink the other day ... and had picked up a hundred-rouble note under the table.”
- “No, I’m not going to play any more,” observed Kalganov ...”
- “Here, I stake ten roubles, the knave leads.” “And I put a rouble on the queen ...”
- “But he won’t come back ... “make them give you back the money you lost.”
- “At five o’clock this afternoon Dmitri Fyodorovitch borrowed ten roubles from me ...”
- “When I asked him where he had got so much money, he answered that he had just received it from you ...”
- “If you would write three lines ... stating that you did not give Dmitri Fyodorovitch money ...”
- “She scribbled ... I’ve never in my life lent ... three thousand roubles today.”

- “By the bed, behind the screen, they picked up ... A present of three thousand roubles ... but it had been torn open and was empty.”
- “He stole fifteen hundred ... and then, without even hiding the money, ... went off to the girls.”
- “Can you not, at least, tell us what sum you had ... “Perhaps I did.””
- “You stated just now in your evidence that you spent three hundred roubles ...”
- “We’ve done what we could in your interest ... “What is the stone in your ring?” Mitya interrupted suddenly ...”
- “That’s where I got that money yesterday.... “This is almost miraculous,” murmured Nikolay Parfenovitch.”
- “Yes, what you said just now ... up to the last hour you were still contemplating going to Katerina Ivanovna ...”
- “Surely I didn’t declare so positively that I’d brought three thousand?” “You did say so, Dmitri Fyodorovitch.”
- “The old man at once confessed ... he had borrowed “ten roubles in my poverty” ... ready to pay it back.”
- “Had he noticed how much money Dmitri Fyodorovitch held in his hand ... he declared that there was twenty thousand.”
- “And did you never, once, hear ... that Dmitri Fyodorovitch had saved half that sum for his own use?” “No, I never heard that,” answered Grushenka.”
- “He’s been brought from Petersburg for three thousand roubles, they say ... she paid two thousand for the doctor from Moscow herself.”

*Discussions about money, financial needs, loans, and obligations.*

**Analysis:** This cluster is unified around the theme of money and debt. Named entities like “Dmitri Fyodorovitch” and “Madame Hohlakov” recur in multiple dialogues, often with references to “three thousand roubles.” The model likely grouped these based on consistent mention of amounts, finance-related verbs, and character interactions about resources.

### Cluster ID: 121 (size = 36)

- “The three windows, which consisted each of four tiny greenish mildewy panes, gave little light, and were close shut, so that the room was not very light and rather stuffy.”
- “In The Open Air — “The air is fresh, but in my apartment it is not so in any sense of the word.”
- “There was a wind blowing and no sun, and a feeling of autumn; twilight was coming on.”
- “The guards lead their prisoner to the close, gloomy vaulted prison in the ancient palace of the Holy Inquisition and shut Him in it.”
- “The day passes and is followed by the dark, burning, ‘breathless’ night of Seville.”
- “The air is ‘fragrant with laurel and lemon.’ In the pitch darkness the iron door of the prison is suddenly opened and the Grand Inquisitor himself comes in with a light in his hand.”
- “The wind had risen again as on the previous evening, and the ancient pines murmured gloomily about him when he entered the hermitage copse.”
- “It was already beginning to get dark, the room was lighted up by the lamps and the candles before the ikons.”
- “The old nurse would come in and say, “Let me light the lamp before the holy image, my dear.” And once he would not have allowed it and would have blown it out.”

- “It was a bright evening, the sun was setting, and the whole room was lighted up.”
- “It was a fine day, and I remember to-day, as though I saw it now, how the incense rose from the censer and softly floated upwards and, overhead in the cupola, mingled in rising waves with the sunlight that streamed in at the little window.”
- “He made his way in the dark to her bedroom, where a light was burning.”
- “And before dawn the news of the death reached the town.”
- “And, behold, soon after midday there were signs of something, at first only observed in silence by those who came in and out and were evidently each afraid to communicate the thought in his mind.”
- “It was quite dark when Rakitin and Alyosha entered her rooms, yet they were not lighted up.”
- “The fresh, motionless, still night enfolded the earth.”
- “It was a very large and gloomy old house of two stories, with a lodge and outhouses.”
- “They went into that room and lighted a tallow candle.”
- “The candle burnt dimly, a cricket chirped; it became insufferably close in the overheated room.”
- “At last he realized that the room was full of charcoal fumes from the stove, and that he might die of suffocation.”
- “The candle guttered and was about to go out.”
- “Yes, the old man’s bedroom is lighted up.”
- “The light of the window fell on the front part of the bush.”
- “The old man’s profile that he loathed so, his pendent Adam’s apple, his hooked nose, his lips that smiled in greedy expectation, were all brightly lighted up by the slanting lamplight falling on the left from the room.”
- “Through the pale darkness of the night loomed a solid black mass of buildings, flung down, as it were, in the vast plain.”
- “The village of Mokroe numbered two thousand inhabitants, but at that hour all were asleep, and only here and there a few lights still twinkled.”
- “But next morning they would get up at dawn, as usual, sweep out the rooms with a birch-broom, empty the slops, and clean up after lodgers.”
- “There was a lighted candle on a small deal table in the corner.”
- “A fearful and terrible light!”
- “But it was dark there, and she could only hear the invalid beginning to gasp and struggle.”
- “But still— “Oh, to keep off dogs.... Oh, because it was dark....””
- “The lights had been extinguished long ago.”
- “Not far off was a village, he could see the black huts, and half the huts were burnt down, there were only the charred beams sticking up.”
- “It was a pitch-dark night without a moon.”
- “Two red lights gleamed out of the darkness; the monster roared as it approached.”
- “We lighted some in a pomatum pot, it burned splendidly, it all burnt away leaving only a tiny ash.”

*Descriptions of darkness, lighting, atmosphere, and environmental setting.*

**Analysis:** The sentences are mostly descriptive and focus on setting, time of day, or ambient lighting. The model’s embeddings must have captured the consistent semantic field involving space, visual cues, and time (e.g., “dark,” “candle,” “lighted up,” “night”).

## Histogram

- Normalized cluster sizes ( $\text{cluster\_sizes} / \max(\text{cluster\_sizes})$ )
- Bin setup: 20 equally spaced bins between 0 and 1 (width = 0.05)
- Histogram: counts how many cluster sizes fall into each bin

### Histogram Bins and Counts:

Bins: [0.00, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 1.00]

Counts per bin: [2, 21, 33, 109, 75, 61, 77, 38, 31, 13, 26, 8, 5, 9, 2, 3, 1, 0, 0, 2]

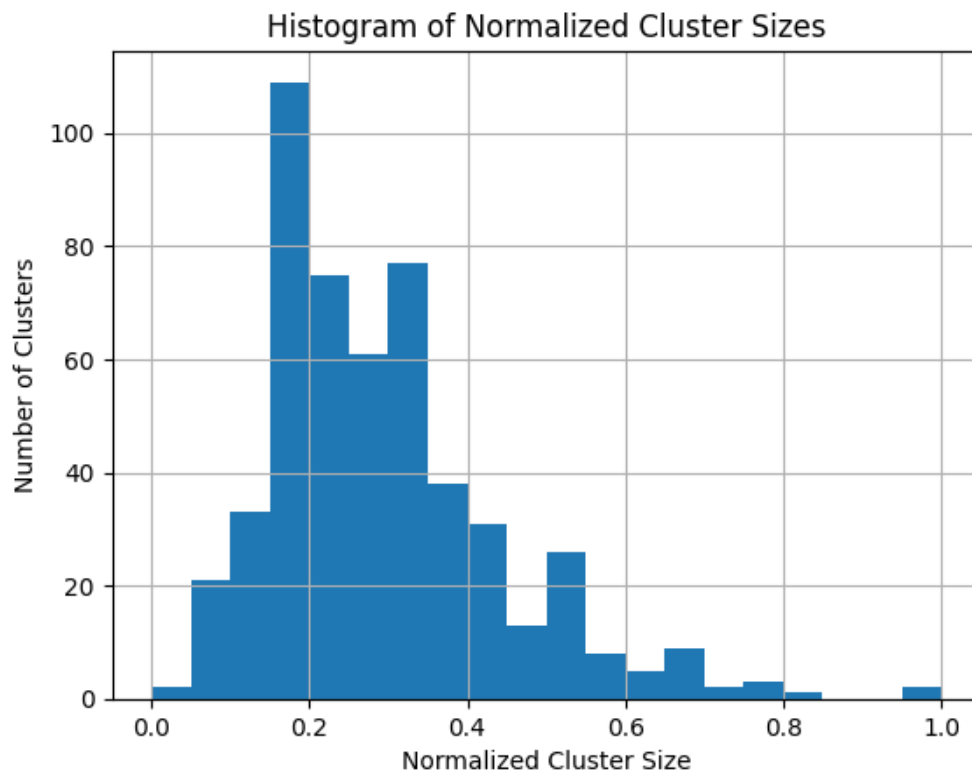


Figure 3: Histogram of Normalized Cluster Sizes (SentenceTransformer(all-mpnet-base-v2), Kmeans (k = 516))

## Log-Log Plot

- Log-log slope: **-0.43**

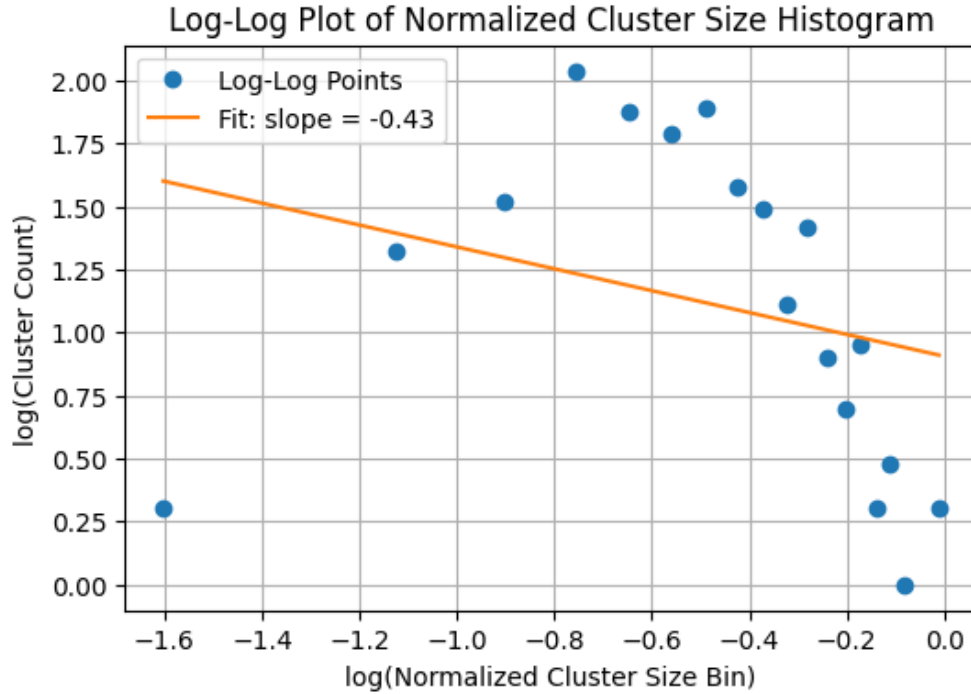


Figure 4: Log-Log Plot of Cluster Sizes (SentenceTransformer(all-mpnet-base-v2), Kmeans ( $k = 516$ ))

## Part 2: Hidden State Clustering Based on Next-Token Prediction (GPT-2)

1. **Model Setup:** GPT-2 was loaded with hidden state outputs enabled, using the `transformers` library.
2. **Sentence Processing:** Each of the 10,000 sentences was tokenized and processed in batches. The hidden state of the final token in each sentence (after padding) was extracted from the final layer.
3. **Next-Token Prediction:** For each sentence, the most probable next token was computed from the model logits.
4. **Clustering:** Sentences were grouped into clusters based on their predicted next-token ID.

### Embedding

- Model: `gpt2`
- Representation: Last token's hidden state

## Token-wise Cluster Sizes

Table 1: Cluster Sizes Based on Predicted Next Token (GPT-2)

| Token ID | Token               | Cluster Size |
|----------|---------------------|--------------|
| 198      | \n                  | 5748         |
| 447      | (Unicode character) | 1412         |
| 314      | I                   | 1119         |
| 679      | He                  | 635          |
| 366      | "                   | 291          |
| 564      | (Unicode character) | 265          |
| 383      | The                 | 117          |
| 1375     | She                 | 105          |
| 1        | "                   | 46           |
| 887      | But                 | 37           |
| 775      | We                  | 36           |
| 1119     | They                | 31           |
| 921      | You                 | 28           |
| 843      | And                 | 21           |
| 632      | It                  | 21           |
| 40       | I                   | 16           |
| 960      | --                  | 12           |
| 290      | and                 | 7            |
| 4362     | Because             | 5            |
| 326      | that                | 3            |
| 1867     | What                | 3            |
| 4162     | Why                 | 3            |
| 357      | (                   | 2            |
| 1550     | On                  | 2            |
| 318      | is                  | 2            |
| 7214     | Take                | 2            |
| 5338     | Who                 | 2            |
| 3966     | Oh                  | 2            |
| 1992     | President           | 2            |
| 1301     | Trump               | 2            |
| 3244     | Then                | 2            |
| 37653    | Reload              | 1            |
| 5660     | Run                 | 1            |
| 34048    | Thou                | 1            |
| 1114     | For                 | 1            |
| 392      | and                 | 1            |
| 47977    | Samson              | 1            |
| 673      | she                 | 1            |
| 286      | of                  | 1            |
| 345      | you                 | 1            |
| 2094     | Don                 | 1            |
| 14026    | Tell                | 1            |
| 1526     | Mar                 | 1            |
| 4360     | but                 | 1            |
| 339      | he                  | 1            |
| 262      | the                 | 1            |
| 428      | this                | 1            |
| 1902     | Gr                  | 1            |
| 785      | com                 | 1            |
| 828      | ),                  | 1            |
| 2644     | ...                 | 1            |
| 12911    | Money               | 1            |

## Thematic Cluster Example

### Predicted Token Cluster: 'But' (Token ID: 887, Size = 37)

- “I know that I ought not to say this, that it would be more dignified on my part simply to leave you...”
- “In 1812 there was a great invasion of Russia by Napoleon...”
- “Some driveling consumptive moralists—and poets especially—often call that thirst for life base.”
- “I understand solidarity in sin among men.”
- “It is true there were many miracles in those days.”
- “That is true.”
- “We have taken the sword of Cæsar, and in taking it, of course, have rejected Thee...”
- “Oh, ages are yet to come of the confusion of free thought...”
- “Who knows, there may have been some such even among the Roman Popes.”
- “I know myself that Samsonov, her merchant, was laughing with her about it...”
- “But whether this was only the conversation that took place then...”
- “But let them wonder, for soon all will be explained.””
- “It is true, alas, it is true, that there are many sluggards, gluttons, profligates and insolent beggars among monks.”
- “Of course, I don’t deny that there is sin in the peasants too.”
- “That is why the philosophers say that we cannot apprehend the reality of things on earth.”
- “If I had been righteous myself, perhaps there would have been no criminal standing before me.”
- “Even though I would gladly give my life for others, it can never be...”
- “And, behold, soon after midday there were signs of something...”
- “Of course there had been, in former times, saints in the monastery...”
- “For it would have been discreditable to insist on burdening with the common regulations so great an ascetic...”
- “It is true that being had for so long been accepted by him as his ideal...”
- “There is happiness for her, too.... She has gone to the feast...”
- “Oh, the shame of it!””
- “It may be wondered why, if he felt so certain, he had not gone to her at first...”
- “People used to tell, laughing at Mitya, how he had given champagne to grimy-handed peasants...”
- “At first he meant to marry her himself.”
- “He may be right.”
- “At first I thought he was ashamed to talk before his great big Pole.”
- “I understand, of course, in what character I am sitting before you.”
- “Well, gentlemen, you see, that Samsonov to whom I went then ...”
- “I understand, of course, that I have come to the vital point.”
- “But at first when I began to give my evidence, it was all still far away and misty...”
- “I understand that that is necessary.”
- “He could have made them stand side by side and played soldiers with them...”

- “I am not opposed to Christ, if you like.”
- “Well, perhaps I am talking nonsense, I agree.”
- “Well, you see, a man may be sitting perfectly sane and suddenly have an aberration.”

*Sentences that offer qualification, contrast, hesitation, or narrative redirection—often involving philosophical reflections, conditionality, or rhetorical clarification.*

**Analysis:** This cluster demonstrates how the model captures nuanced discourse patterns and argumentative pivots. Many of these sentences contain implicit or explicit contradictions, hypotheticals, or reflections, often opening with “but,” “if,” or “perhaps.” The sentences appear thematically disjointed but structurally related, often functioning as exceptions, clarifications, or internal reasoning turns within narrative or dialogue.

### **Predicted Token Cluster: ‘You’ (Token ID: 921, Size = 28)**

- “You have done no harm.”
- “Oh, yes; you are not screaming.”
- “You are not lonely.”
- “She entreats you to accept her help.... You have both been insulted by the same man.”
- “I am absurd and small, but you, you!”
- “You are a steadfast person, Alexey.”
- “Do you understand that, friend and brother, you pious and humble novice?”
- “I am making you suffer, Alyosha, you are not yourself.”
- “Do you know, Alyosha—don’t laugh!”
- “Faith to the End Remember particularly that you cannot be a judge of any one.”
- “You are working for the whole, you are acting for the future.”
- “You are talking some nonsense, you want to say something nasty.”
- “She wounded you, but don’t be angry.”
- “You see, you cannot, in any case, lose over it.”
- “No, you’ve contracted for the job and turned out a scamp.”
- “Look at your hands, Dmitri Fyodorovitch.”
- “But it doesn’t matter if you don’t.”
- “For you were frightened; you couldn’t speak.”
- “No, you didn’t steal it.”
- “That’s just what you’re doing now.”
- “You may not trust a criminal or a man on trial tortured by your questions...”
- “But you don’t understand that.”
- “He spoke to you, he did not speak to us, you stupid.”
- “Come, you want obedience and mysticism.”
- “You really are not like every one else...”
- “You are a prophet.”
- “I tell you you are silly, Alyosha.”
- “You are not fit to be a husband.”

*Second-person direct address in emotional, advisory, or reflective dialogue—often involving appeals, reprimands, or supportive reassurance.*

**Analysis:** The model clustered direct interactions centered around the pronoun “you,” capturing personal tone and interpersonal dynamics. The sentences convey concern, confrontation, or compassion and often involve Alyosha, reflecting emotional intensity or moral questioning. The shared syntactic pattern and emotional engagement around the second person pronoun likely led to this semantic grouping.



### Predicted Token Cluster: 'They' (Token ID: 1119, Size = 31)

- “They are children and stupid.”
- “They had sent for Herzenstube; they had sent for the aunts.”
- “Pine-trees are not like people, Alexey Fyodorovitch, they don’t change quickly.”
- “They don’t matter.”
- “The Turks, of course, have gone into it, but they are foreigners.”
- “I have specimens from home that are even better than the Turks.”
- “They are not sentimentalists there.”
- “They will cast down temples and drench the earth with blood.”
- “They went out, but stopped when they reached the entrance of the restaurant.”
- “The Maslovs have it all their own way.”
- “They reached Volovya station.”
- “The monks ... had all been his devoted friends for many years.”
- “Friends came to see us.”
- “We found the people of the town hospitable, rich and fond of entertainments.”
- “They were well-to-do people of influence and position.”
- “In that they see freedom.”
- “They live only for mutual envy, for luxury and ostentation.”
- “It is different with the upper classes.”
- “They are not servile ... and not revengeful and not envious.”
- “He and his wife earned their living as costermongers in the market-place.”
- “And those who came were far from being all of the humbler classes.”
- “There they are among the Turks, they have forgotten everything.”
- “There they are bringing new wine.”
- “They went into the hut.”
- “They are not men who do anything real.”
- “They must appreciate you, they must understand you.”
- “You see what our peasants are, Mavriky Mavrikyevitch.”
- “Our fellows began going there of themselves.”
- “You know, they spring from the growing artistic instincts of the young.”
- “The classical languages, too ... they are simply madness, nothing more.”
- “So they thought of Greek and Latin.”

*Group references in descriptive or reflective commentary focusing on collective actions, social dynamics, or generalized human behavior.*

**Analysis:** This cluster includes third person plural references to groups (“they”), capturing patterns of communal behavior, societal observations, or character group dynamics. The model likely identified structural cues (use of “they”) and thematic similarity in reflecting on social class, cultural attitudes, or relational judgments across different contexts. The sentences form a shared semantic role as commentaries on external others.

## Histogram

- Normalized cluster sizes ( $\text{cluster\_sizes} / \max(\text{cluster\_sizes})$ )
- Bin setup: 20 equally spaced bins between 0 and 1 (width = 0.05)
- Histogram: counts how many cluster sizes fall into each bin

### Histogram Bins and Counts:

Bins: [0.00, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 1.00]

Counts per bin: [47, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1]

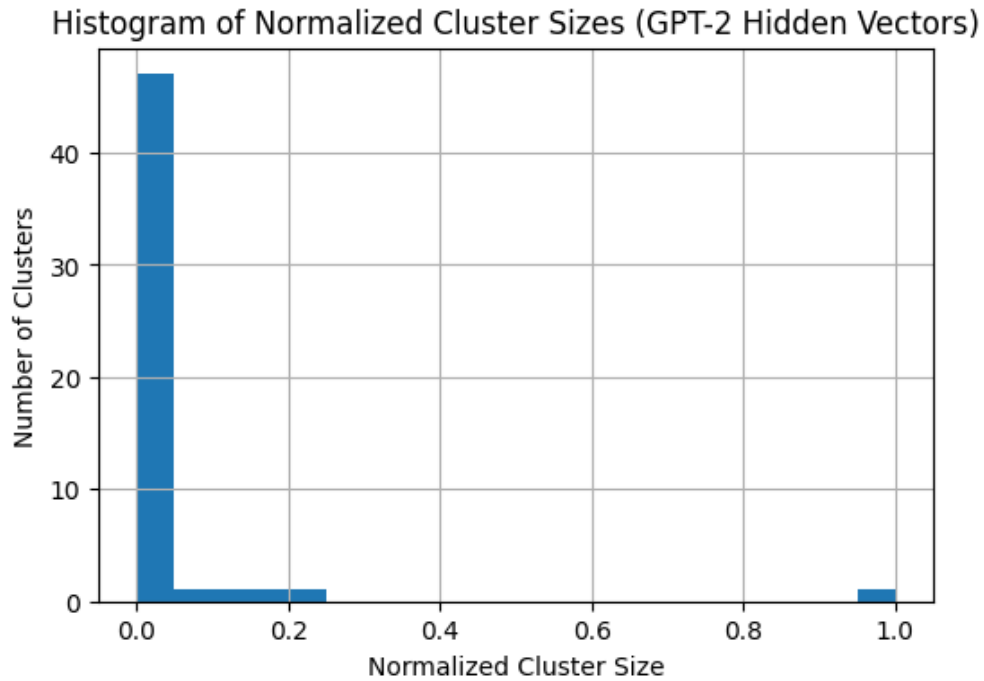


Figure 5: Histogram of Normalized Cluster Sizes (GPT2, next token prediction based clustering)

## Log-Log Plot

- Log-log slope: **-0.43**

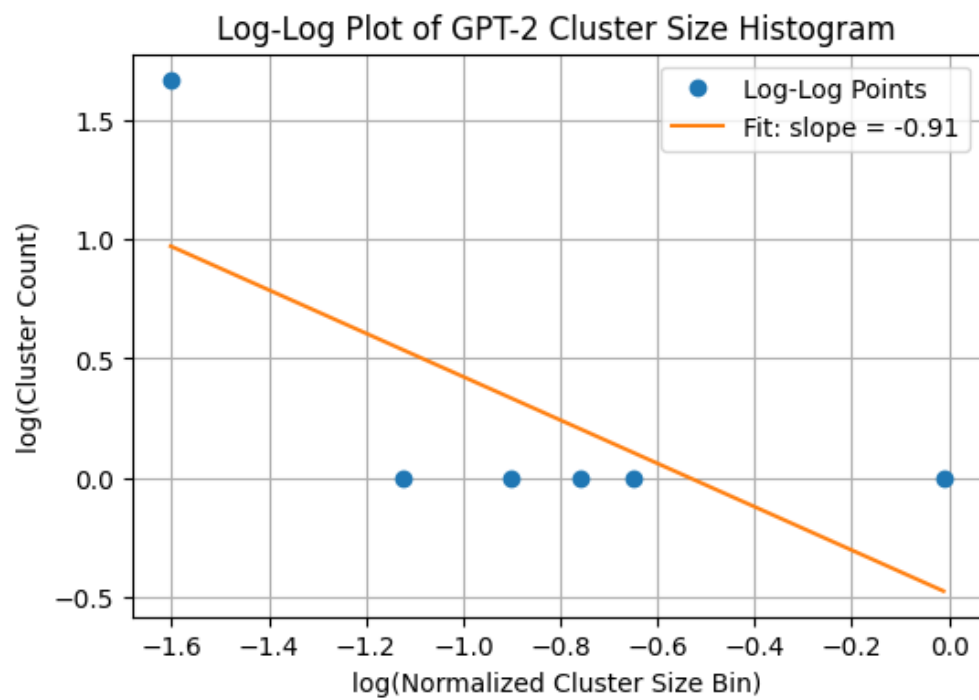


Figure 6: Log-Log Plot of Cluster Sizes (GPT2, next token prediction based clustering)