Semantic and Frequency Shifts in Economic Discourse Due to the Great Recession: A Corpus-Based Study

Mahmoud Hosseini Pour Shamsi

Abstract

This research examines the Great Recession's impact on American English public discourse through a frequency and semantic shift analysis of the Corpus of Contemporary American English (COCA). The research used COCA to create eleven yearly sub-corpora (2005–2011) from magazine, newspaper, and academic texts. The frequency analysis of recession-related lemmas showed significant usage increases in 2009 which proved that economic distress received increased linguistic focus. The increase was verified through Pearson's chi-square tests. The research also included both a qualitative collocate analysis and vector-based semantic shift analysis for the terms *economy*, *job*, and *crisis*. The semantic transformation of *economy* proved most significant between 2005 and 2009 before showing partial recovery of pre-crisis meanings in 2011. The term *job* went through a steady transformation while *crisis* shifted from its metaphorical and geopolitical applications to economic interpretations during the peak of the recession. The study reveals how economic disruptions at the national level create changes in both word frequency and meaning within public discourse.

Keywords

Semantic shift, Lemma frequency, Great Recession

1. Introduction

The Great Recession of 2007–2009 stands out as one of the most disruptive economic crises in recent history. While its effects on employment, housing, and financial markets have been extensively studied, much less attention has been given to how it influenced the way people talked about the economy. Language is not just a tool for describing events; it's a window into how we understand and emotionally respond to them.

Previous studies have shown that major events can change the way we use language–affecting everything from metaphorical framing, to lexical frequency distributions, and the collocational behavior of economic terms. But few have looked at these patterns together across different stages of a single crisis.

The present research fills this knowledge gap through its dual examination of economic term occurrence rates and their semantic changes. The Corpus of Contemporary American English

(COCA) served as the data source for creating eleven annual sub-corpora which spanned from 2005 to 2011 across magazines, newspapers and academic publications. The research analyzed six key economic lemmas–*Unemployment*, *recession*, *layoff*, *crisis*, *foreclosure* and *jobless*–for their significant frequency changes while performing vector-based semantic shift assessment and qualitative collocate analysis on the following central terms: *Economy*, *job* and *crisis*. Using this multidimensional approach, the study aims to explain how economic discourse in American English shifted in response to one of the most significant financial events of the 21st century.

2. Data and Experiments

For this study, the Corpus of Contemporary American English (COCA) was used as the primary data source. COCA is a large, balanced corpus of American English that includes multiple genres and time periods which makes it quite appropriate for a diachronic analysis of language usage in relation to socio-economic events.

In order to look into the linguistic changes surrounding the Great Recession, a custom dataset was made by creating eleven sub-corpora, each made-up of texts from the magazine, newspaper, and academic sections of COCA, related to the years 2005 to 2011. Each annual sub-corpus includes 5,000 texts, and the total word count per year is shown in Table 1.

To give a sense of the dataset's linguistic texture, a few examples of target terms in context from the 2009 sub-corpus (peak of the crisis) are presented below:

- Unemployment: "In March, employers shed 663,000 jobs, pushing the nation's unemployment rate to 8.5%."
- Jobless: "More than 4.5 million Americans are currently receiving jobless benefits — more than at any time during the past 26 years."
- Foreclosure: "It's simply unacceptable to have the volume of foreclosures we're having here and have no action from the House at all in the middle of this crisis."

Table 1. Word counts per sub-corpus

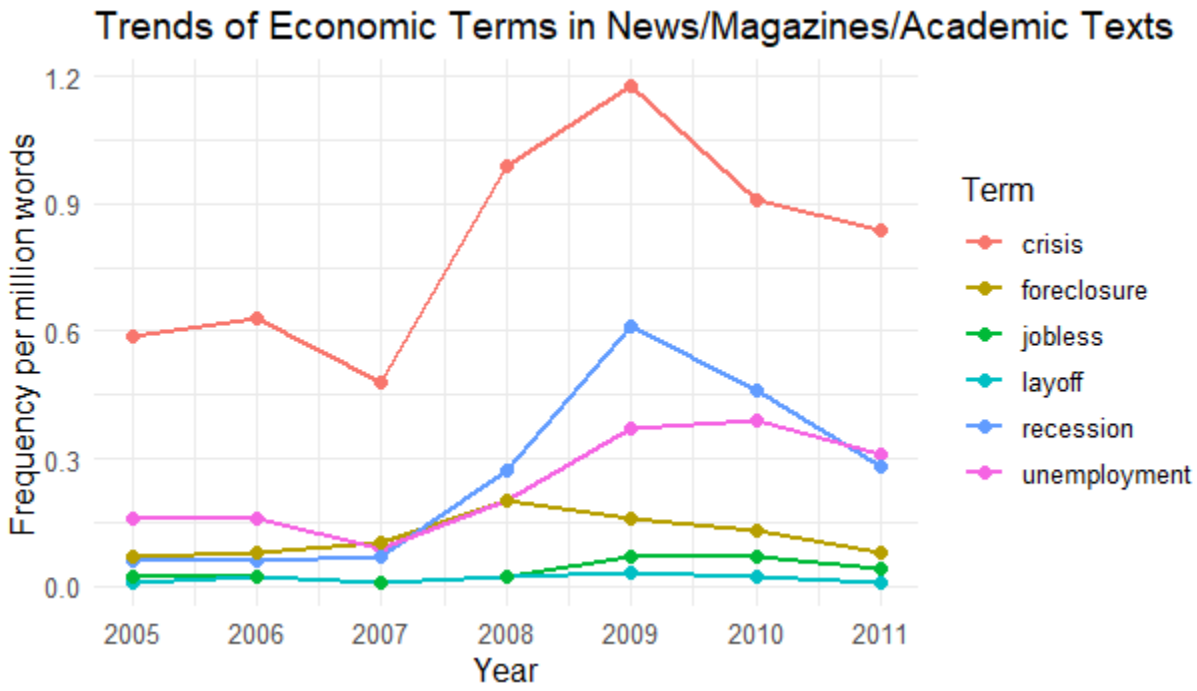
Year	Number of Words
2005	10,637,762
2006	10,226,949
2007	8,941,434
2008	9,348,676
2009	8,795,165
2010	8,309,840
2011	7,305,831

The dataset was examined in two ways: lexical frequency and semantic shift, with a specific focus on the years 2005 (pre-recession), 2009 (crisis peak), and 2011 (early recovery) to record socio-linguistic changes over the course of the economic downturn.

2.1 Frequency Analysis

A frequency-based analysis was conducted on six lemmas that were considered to be central to the discussion of economic distress: U*nemployment*, *recession*, *layoff*, *crisis*, *foreclosure*, and *jobless*. These lemmas were queried using COCA's List function, which returns both raw frequency and normalized frequency per million words. Results across the full seven-year duration are visualized in Figure 1, which shows the substantial frequency increase of terms related to recession in 2009 and their relative decline by 2011.

Figure 1.



Pearson's chi-squared tests were utilized to determine whether these frequency changes were statistically significant. The tests were performed on term frequencies in 2005 vs. 2009 and 2009 vs. 2011. Results are shown in Tables 2 and 3.

Table 2. Comparative Frequencies: 2005 vs. 2009

(Magazine/News/Academic sub-corpora)

Term	**2005 (raw / per mil)**	**2009 (raw / per mil)**	**χ^2**	**p-value**
Recession	55 / 0.06	610 / 0.61	579.65	< 0.001
Crisis	585 / 0.59	1172 / 1.18	326.19	< 0.001
Unemployment	159 / 0.16	372 / 0.37	131.80	< 0.001
Foreclosure	65 / 0.07	159 / 0.16	59.82	< 0.001
Jobless	20 / 0.02	71 / 0.07	39.43	< 0.001
Layoff	13 / 0.01	32 / 0.03	12.14	< 0.001

- **Corpus size**:
    - 2005: 10,637,762 words
    - 2009: 8,795,165 words

As shown, all six terms had statistically significant increases in 2009 which confirms a significant shift in public discourse toward economic struggle. It is notable to mention the tenfold increase in *recession*.

Table 3. Comparative Frequencies: 2009 vs. 2011

(Magazine/News/Academic sub-corpora)

Term	**2009 (raw / per mil)**	**2011 (raw / per mil)**	**χ^2**	**p-value**
Recession	610 / 0.61	281 / 0.28	68.83	< 0.001
Crisis	1172 / 1.18	822 / 0.84	13.87	< 0.001
Unemployment	372 / 0.37	311 / 0.31	0.01	0.933
Foreclosure	159 / 0.16	83 / 0.08	11.98	< 0.001
Jobless	71 / 0.07	36 / 0.04	5.94	0.015
Layoff	32 / 0.03	11 / 0.01	6.80	0.009

- **Corpus size**:
    - 2009: 8,795,165 words
    - 2011: 7,305,831 words

By 2011, most terms experienced a decrease in frequency which can be a sign of gradual return to pre-recession discourse. However, *unemployment* remained notably stable, indicating its persistent prominence even during early recovery phases which is consistent with the delayed rebound of labor markets in post-crisis periods.

2.2 Semantic Shift Analysis

Two complementing analyses were conducted in order to see how the contexts and meanings of economic terms changed due to Great Recession.

2.2.1 Qualitative Collocate analysis

A qualitative assessment was carried out by identifying the top mutual information (MI) collocates of each target word–*economy*, *job*, and *crisis*–in the years 2005, 2009, and 2011. These collocates were extracted from the COCA corpus (Magazine, Newspaper, and Academic sub-corpora). Surface forms (words) were used instead of lemmas to avoid noise from punctuation and the search was conducted using a span of 3 words to the left and 4 to the right (L3-R4), with filtering thresholds of MI > 5 and raw frequency ≥ 5. By looking at the changes in the lexical environment of each word throughout the three periods (pre-crisis, crisis peak, and early recovery), I learned interesting things about how the surrounding words of each term reflected the economic changes.

In 2005, before the recession hit, *economy* was often framed in optimistic and globally oriented language. Words like *booming*, *robust*, and *capitalist* were signs of strength and confidence. Collocates such as *Brazilian* and *global* reflected a strong international orientation and a thriving world economy. During the height of the crisis in 2009, the tone considerably shifted. On the one hand, words like *faltering*, *ailing*, and *worsening* reflected widespread concern and economic pain. On the other hand, hopeful words like *stimulate*, *reviving*, and *recovers* showed efforts to improve the situation. The variety and number of strong collocates this year suggest a lot of public debate and attention focused on the economic situation. In 2011, the language softened to an extent. Words like *recovers* and *stabilizing* pointed to cautious optimism, but terms like *sluggish* and *weak* revealed ongoing concerns. Interestingly, the existence of *global* in the top MI score collocates of this year shows a renewed focus on international economic ties. Due to space constraints, detailed tables including all the top MI score collocates of *economy* in 2005, 2009 and 2011 are provided in Appendix A (see tables A1–A3).

The meaning and associations of the word *job* also changed quite a bit across the same three years. Before the recession, the terms surrounding *job* centered on availability and types of employment. collocates like *openings*, *full-time*, *part-time*, and *applicants* were quite common which is a sign of an opportunity-focused labor market. The presence of *quit* shows voluntary labor mobility, indicating a healthy job market. By 2009, as the recession peaked, the collocates of *job* underwent a clear shift towards dissatisfaction and insecurity. Words like *burnout*, *dissatisfaction*, *insecurity*, and *losses* were reflective of stress and fear while the continued appearance of *full-time* and *part-time* points to the ongoing negotiation of job types. In 2011, during the recovery period after the recession, the lexicon surrounding *job* connoted a mix of cautious optimism and lingering anxiety. Positive terms such as *creation* and *fairs* coexisted with stress-related terms like *stressor*. Importantly, the new presence of *workplace* and *description* hinted at a change in focus toward job structure. The complete tables of top MI-ranked collocates of *job* for 2005, 2009, and 2011 are included in Appendix B (Tables B1–B3).

Back in 2005, *crisis* had a broad and often metaphorical use. It was linked to phrases like *midlife crisis*, or historical and political contexts like the *Cuban* or *missile crisis*. While some economic terms like *fiscal* showed up, they were relatively rare. At the height of the financial crash in 2009, *crisis* became strongly linked to the economy. High-frequency and high-MI collocates such as *financial* and *economic*, and terms like *foreclosure* and *banking* strongly tie *crisis* to the unfolding global economic collapse. By 2011, the semantic profile of *crisis* widened again; retaining some of the economic associations while increasingly reflecting new global developments. Collocates such as *Fukushima* and *Suez* reflect geopolitical and environmental crises, marking a return to more diversified interpretations of the term. The continued presence of *midlife* also reflects the persistent metaphorical use of the term. For a full list of the top MI collocates of *crisis* in 2005, 2009, and 2011, refer to Appendix C (Tables C1–C3).

2.2.2 Vector-Based Semantic Shift Analysis Using Cosine Similarity

To supplement the qualitative analysis with a quantitative measure, a Collocate Vector Space Model method was applied. For each target word–*economy*, *job*, *crisis*–the top 20 collocates with the highest MI score for each year (2005, 2009, 2011) were retrieved using the same statistical criteria and span as the qualitative collocate analysis: Word forms were used as queries and the collocational search was conducted within a span of three words to the left and four to the right (L3–R4), filtering with a minimum raw frequency of 5. Importantly, each keyword was treated independently and a separate collocate union list was created per target word to construct its temporal semantic vector space. This approach is theoretically motivated by the principle that collocational context is highly word-specific; different target words activate different semantic fields, and using global collocate lists across multiple words would obscure those distinctions.

Normalized weighted values were calculated so as to measure the strength of association between each collocate and its target word; Specifically, for each collocate, I computed the

natural logarithm of (1 plus 100 times the per million frequency) of that collocate within the corpus for the given year, rounded to 3 decimal places. This formula for normalization was chosen for two main reasons: Firstly, the logarithm normalized the skewed frequency distribution that typically exists in linguistic data, reducing the impact of extremely frequent collocates. Secondly, the per million frequencies were very small. By multiplying them by 100, they were scaled to a range that is much more appropriate for an effective vectorial comparison. Collocates absent in a particular year were assigned a weight of zero as they had no association during that period. Hence, each target term's semantic profile for each year was represented as a vector of these weighted collocate frequencies.

I computed pairwise cosine similarity scores between the vectors corresponding to 2005, 2009, and 2011. Cosine similarity measures the cosine of the angle between two vectors in a multi-dimensional space, with values closer to 1 showing greater similarity and values closer to 0 indicating greater difference. All vector operations and similarity computations were carried out using the NumPy and Scikit-learn libraries in Python.

As it can be clearly seen in Table 4, *economy* experienced the most significant semantic change. The extremely low cosine similarity score between 2005 and 2009 points to an almost complete reshaping of the words commonly associated with *economy* at the start of the crisis. The higher similarity scores between 2009 and 2011, and between 2005 and 2011, suggest that by the post-crisis period, the term had started to regain some stability or, in a sense, return to earlier patterns. The word *job*, on the other hand, followed a steadier and more gradual path. Its relatively high similarity scores across all three time comparisons prove that even though the term took on some crisis-related associations, its core collocational profile remained mostly unchanged. *Crisis*, meanwhile, underwent a sharp shift from 2005 to 2009, with further changes continuing into 2011. The low similarity between 2005 and 2011 implies that the word's meaning continued to change and did not return to its earlier, more varied or metaphorical usage.

All in all, while all three words changed in response to the Great Recession, *economy* experienced the most dramatic transformation, reflecting how deeply the crisis reshaped public and institutional discourse around economic issues.

Table 4.

Cosine Similarity Scores Indicating Semantic Shifts of *Economy*, *Job*, and *Crisis* Between 2005, 2009, and 2011

Target Term	2005 vs 2009	2009 vs 2011	2005 vs 2011
Economy	0.1475	0.2609	0.5733
Job	0.5672	0.6273	0.5098
Crisis	0.4624	0.6158	0.3844

3. Discussion and Conclusion

The present study, using clear empirical evidence, proves that the Great Recession increased the frequency of key economic lexicon in American English, as well as changing their semantic associations. The data shows statistically significant rises in recession-related lemmas, particularly in 2009, which implies the heightened public attention and media focus during the peak of the crisis.

Semantic analyses clarify that these frequency changes were accompanied by meaningful contextual shifts. The term *economy* underwent the most considerable transformation, moving from globally positive in 2005 to contexts related to crisis and recovery in 2009 and 2011. The cosine similarity scores confirm a sharp semantic shift between 2005 and 2009, followed by partial recovery in 2011. Conversely, *job*, had a relatively stable semantic profile, while, *crisis* transitioned from metaphorical and geopolitical usages to more narrowly economic contexts during the recession, and then diversified again in 2011.

The integration of both qualitative and quantitative methods is considered to be one the strong points of this study. The mutual information-based collocate analysis captures detailed contextual information and the cosine similarity scores provide a replicable, vector-based measure of semantic drift. However, one limitation is the focus on only three COCA genres, which may not include broader public discussions that are expressed in online media, blogs, or spoken interactions.

Future research could extend this framework to other crises–such as the COVID-19 pandemic–to examine whether similar patterns exist. Additionally, utilizing sentiment analysis could offer deeper insights into how economic anxiety is expressed through language.

References

1. Davies, Mark. (2008-) *The Corpus of Contemporary American English (COCA)*. Available online at https://www.english-corpora.org/coca/.
2. Davies, Mark. (2008-) Word frequency data from *The Corpus of Contemporary American English (COCA)*. Data available online at https://www.wordfrequency.info.
3. Davies, Mark. (2008-) Collocates data from *The Corpus of Contemporary American English (COCA)*. Data available online at https://www.collocates.info.
4. R Core Team (2025). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.
5. Wickham, H., François, R., Henry, L., & Müller, K. (2023). *dplyr: A Grammar of Data Manipulation* (Version 1.1.3) [R package]. https://CRAN.R-project.org/package=dplyr
6. Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., & Yutani, H. (2023). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics* (Version 3.4.0) [R package]. https://CRAN.R-project.org/package=ggplot2

7. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). *Scikit-learn: Machine learning in Python*. Journal of Machine Learning Research, 12, 2825–2830.
8. Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... & Oliphant, T. E. (2020). *Array programming with NumPy*. Nature, 585(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Appendix

Appendix A:

Table A1.

Collocates of *economy* in 2005 with MI > 5 and Frequency ≥ 5:

	Words	**Raw Frequency**	**Per Million Frequency**	**MI**
1	Booming	11	0.01	7.39
2	Asante	6	0.01	7.10
3	Capitalist	6	0.01	6.62
4	Robust	8	0.01	6.35
5	Fuel	46	0.05	6.24
6	Global	79	0.08	6.23
7	Natal	5	0.01	6.15
8	Unemployment	6	0.01	5.77
9	Brazilian	5	0.01	5.66
10	Nez	5	0.01	5.44

Table A2.

Collocates of *economy* in 2009 with MI > 5 and Frequency ≥ 5:

	Words	**Raw Frequency**	**Per Million Frequency**	**MI**
1	Recovers	15	0.02	7.94
2	Faltering	7	0.01	7.64
3	Worsening	5	0.01	7.09
4	Jump-Start	6	0.01	7.08
5	Reviving	6	0.01	6.98
6	Stimulate	17	0.02	6.83
7	Lousy	5	0.01	6.59
8	Ailing	5	0.01	6.59
9	Stimulating	7	0.01	6.55
10	Shaky	5	0.01	6.22
11	Revive	7	0.01	6.18

12	Slowing	7	0.01	6.05
13	Improves	7	0.01	5.95
14	Sectors	11	0.01	5.94
15	Stalled	5	0.01	5.82
16	Mpg	5	0.01	5.61
17	Fuel	33	0.03	5.22
18	Battered	5	0.01	5.12
19	Boost	14	0.01	5.11

Table A3.

Collocates of *economy* in 2011 with MI > 5 and Frequency ≥ 5:

	Words	**Raw Frequency**	**Per Million Frequency**	**MI**
1	Recovers	5	0.01	8.76
2	Faltering	5	0.01	7.99
3	Stabilizing	7	0.01	7.28
4	Sluggish	5	0.01	7.17
5	Slowing	6	0.01	6.85
6	Stimulate	9	0.01	6.78
7	Global	81	0.08	5.66
8	Weak	12	0.01	5.26
9	Sectors	5	0.01	5.25

Appendix B:

Table B1.

Collocates of *job* in 2005 with MI > 5 and Frequency ≥ 5:

	Words	**Raw Frequency**	**Per Million Frequency**	**MI**
1	Openings	6	0.01	6.02
2	Applicants	10	0.01	5.82
3	Part-time	15	0.02	5.79
4	Full-time	26	0.03	5.74
5	Quit	18	0.02	5.32

Table B2.

Collocates of *job* in 2009 with MI > 5 and Frequency ≥ 5:

	Words	**Raw Frequency**	**Per Million Frequency**	**MI**
1	Burnout	8	0.01	7.59
2	Seekers	17	0.02	7.54

3	Satisfaction	52	0.05	6.35
4	Dissatisfaction	8	0.01	5.96
5	Full-Time	26	0.03	5.77
6	Quit	17	0.02	5.66
7	Loses	10	0.01	5.57
8	Insecurity	5	0.01	5.38
9	Openings	5	0.01	5.36
10	Part-Time	14	0.01	5.26

Table B3.

Collocates of *job* in 2011 with MI > 5 and Frequency ≥ 5:

	Words	**Raw Frequency**	**Per Million Frequency**	**MI**
1	Fairs	5	0.01	7.41
2	Stressors	16	0.02	7.11
3	Satisfaction	44	0.04	6.75
4	Seekers	5	0.01	6.73
5	Quitting	5	0.01	6.46
6	Quit	19	0.02	6.02
7	Creation	37	0.04	5.84
8	Openings	6	0.01	5.79
9	Full-Time	14	0.01	5.42
10	Workplace	11	0.01	5.30
11	Malware	5	0.01	5.25
12	Description	14	0.01	5.13

Appendix C:

Table C1.

Collocates of *crisis* in 2005 with MI > 5 and Frequency ≥ 5:

	Words	**Raw Frequency**	**Per Million Frequency**	**MI**
1	Midlife	5	0.01	7.81
2	Cuban	5	0.01	7.13
3	Missile	5	0.01	7.05
4	Fiscal	6	0.01	5.91
5	Solve	7	0.01	5.78
6	Facing	8	0.01	5.37

Table C2.

Collocates of *crisis* in 2009 with MI > 5 and Frequency ≥ 5:

	Words	Raw Frequency	Per Million Frequency	MI
1	1997-98	5	0.01	9.58
2	Hotline	5	0.01	8.48
3	Deepening	5	0.01	7.48
4	Hostage	7	0.01	7.09
5	Cuban	6	0.01	6.72
6	Humanitarian	14	0.01	6.66
7	Financial	186	0.19	6.47
8	Foreclosure	10	0.01	6.08
9	Economic	142	0.14	5.87
10	Missile	5	0.01	5.83
11	Fiscal	12	0.01	5.57
12	Darfur	8	0.01	5.57
13	Banking	10	0.01	5.24
14	Current	69	0.07	5.22
15	Worst	17	0.02	5.18
16	Global	57	0.06	5.12
17	Asian	15	0.02	5.10

Table C3.

Collocates of *job* in 2011 with MI > 5 and Frequency ≥ 5:

	Words	Raw Frequency	Per Million Frequency	MI
1	MIDLIFE	13	0.01	9.82
2	HOSTAGE	5	0.01	7.57
3	SUEZ	6	0.01	7.52
4	FINANCIAL	142	0.14	7.11
5	LOOMING	6	0.01	7.04
6	HUMANITARIAN	12	0.01	6.74
7	FUKUSHIMA	8	0.01	6.30
8	DEBT	33	0.03	6.05
9	INQUIRY	7	0.01	5.91
10	DEPARTMENTS	5	0.01	5.27
11	GREEK	7	0.01	5.14
12	GREECE	5	0.01	5.11