

Word Frequency Analysis of American Contemporary Art Reviews from 1960 to 2015

Yanhui Jiang¹, Junhui Gao^{2*}

¹University of California, Berkeley, California, USA

²Wuxi American and European International Study Center, Wuxi, China

Email: *jhgao68@163.com

How to cite this paper: Jiang, Y. H., & Gao, J. H. (2018). Word Frequency Analysis of American Contemporary Art Reviews from 1960 to 2015. *Open Journal of Modern Linguistics*, 8, 87-98.

<https://doi.org/10.4236/ojml.2018.84010>

Received: June 24, 2018

Accepted: July 24, 2018

Published: July 27, 2018

Copyright © 2018 by authors and
Scientific Research Publishing Inc.

This work is licensed under the Creative
Commons Attribution International
License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This research parsed entries of eight contemporary art journals from 1960 to 2015 by Natural Language Processing (NLP) and got 7000 words. These words were then analyzed by three methods: 1) word frequency, 2) word frequency trend analysis based on linear regression, and 3) word frequency variation model analysis based on curve clustering. The data were classified into a five-category set and a ten-category set, along with a ten-word-frequency-variation-model. Based on the above data study, this paper analyzed the resulted trends from political, economical, technological and art perspectives. The conclusion is that hot topics in American contemporary art vary with contemporary social backgrounds.

Keywords

Contemporary Art, Text-Mining, Word Frequency, Trend Analysis, Cluster Analysis

1. Introduction

The term *contemporary art*, is widely referred in all occasions. However, it is almost impossible to give a universal definition to this term because of the close connection with history, geography, cultures, politics, technology and economics, and the huge diversity of themes, forms and media of contemporary art. Therefore, “what is contemporary art?” naturally becomes a question for every art historian and everyone visiting a modern art museum. One difficulty to define contemporary is that for a certain period in history, all the art forms created in that period can be called “contemporary,” so the definition of contemporary for art historians varies from one person to the other. Alexander Alberro (Alberro, 2008), for example, defined contemporary art in a specific period of

time. He sees the years following 1989 as a new historical period for art history, in the context of the collapse of the Soviet Union, globalization and the beginning of the “digital era.” However, according to art theorist Richard Meyer (Meyer, 2016), contemporary art “might include not only newly produced works by living artists but also those times travelers that arrive ‘in our midst’ from earlier moments and historical contexts.” Taking its diversified and real-time response features into account, we tracked the variation of contemporary art reviews in recent fifty-five years’ art history journals by means of word frequency analysis, trend analysis and cluster analysis.

In the constantly evolving machine learning field, the application of deep learning methods in art domain is not unprecedented. For example, Saleh and Elgammal designed a classification algorithm that is capable of distinguishing different art genres from a range of fifteen centuries in a relative high accuracy (Elgammal & Saleh, 2015). Moreover, Passonneau, Yano, Lippincott and Klavans used machine learning to categorize and label artworks based on the corresponding art history text (Passonneau, Yano, Lippincott, & Klavans, 2008). However, the difficulty of contemporary art interpretation comes from not only the abstraction of the work itself, but also the necessity to study the historical background of the work and the artist’s life experience. Today, deep learning in art history is still limited to visual recognition of the work itself and tends to ignore other related information. We hope this research could provide a textual interpretation of the trend in contemporary art based on art history journals by the means of word frequency analysis, trend analysis and cluster analysis.

2. Data and Methods

2.1. Data Source

There is no clear starting point of contemporary art in art history field. However, most scholars choose 1960s as an important period for emergence of contemporary art due to the intensive social-political revolutions (e.g. human rights, feminism and gay right movements) and explosive development of science and technology. The rapidly changing nature of this specific period stimulates the artists’ ideologies and their opinions on contemporary art.

Therefore, this research chose the year 1960 as its starting point, including articles of eight influential art history journals from 1960 to 2015 in every fifth year as our sample data. The journals are *Afterall*, *Art Journal*, *Frieze*, *Grey Room*, *Leonardo*, *October*, *the Art Bulletin* and *The Journal of Aesthetics and Art Criticism*. All of them are leading journals covering an international range of art themes and focus on discussion of art from 20th and 21st centuries. Among these journals, *Art Journal*, *The Art Bulletin* and *The Journal of Aesthetics and Art Criticism* have the broadest coverage of the time period. The number of articles from these three journals included at each time point is summarized in **Table 1**.

Table 1. Summary of amounts of articles from the three journals in each time point.

	Art Journal	The Art Bulletin	The Journal of Aesthetics and Art Criticism
1960	60	49	100
1965	220	85	64
1970	218	108	67
1975	93	124	76
1980	72	125	-
1985	98	60	-
1990	66	50	76
1995	55	73	114
2000	81	54	55
2005	149	43	68
2010	69	67	72
2015	67	72	100
Total	1248	910	792

2.2. Word Frequency Analysis

All articles were saved in TXT format, with articles from the same year under the same directory. There are twelve directories in total with names 1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010 and 2015 respectively. For each directory, a word count was performed on all articles under it. We used Natural Language Toolkit (NLTK) to parse texts and remove stopwords (NLTK 3.3 Documentation, 2018). NLTK is the most used Python library in natural language processing. We only kept words with frequencies larger than zero in all twelve time points and there are more than 8000 of them. After filtered out all punctuation symbols and numbers, about 7000 words were left. Ten of them were chosen randomly as examples in Table 2.

2.3. Word Frequency Trend Analysis

From Table 2, we can see each row can form a time series. We processed each row by linear regression. The resulted slope of the curve shows the trend of frequency variation of a word. To distinguish these trends, we categorized the slopes for all words into five categories (Table 3). Table 3 shows the values of all slopes, i.e. the absolute trends, are very small generally. Then, we calculated the slope-to-mean ratios as the relative trend. Again, we categorized the relative trends into ten categories (Table 4).

2.4. Word Frequency Curve Cluster Analysis

Different from Section 2.3, in which trend cluster analysis was introduced, non-trend cluster analysis is discussed in this section.

Table 2. Word frequencies of ten sample words at twelve time points.

Id	New	however	example	make	life	often	show	still	fact	used
wp1960	0.00576	0.00196	0.00104	0.00117	0.00169	0.00129	0.00075	0.00129	0.00136	0.00103
wp1965	0.00636	0.00184	0.00103	0.00102	0.00182	0.00117	0.00076	0.0012	0.0013	0.00136
wp1970	0.00724	0.0016	0.00132	0.00133	0.00123	0.00105	0.001	0.00135	0.00133	0.00125
wp1975	0.00691	0.00206	0.00151	0.0012	0.00129	0.00114	0.00125	0.00111	0.00137	0.00194
wp1980	0.00576	0.00197	0.00179	0.00137	0.00134	0.00109	0.00104	0.00115	0.00133	0.0013
wp1985	0.00632	0.0019	0.00174	0.0014	0.00223	0.00145	0.0009	0.00152	0.0016	0.00161
wp1990	0.00707	0.00186	0.00174	0.00158	0.00192	0.00129	0.00071	0.00109	0.00127	0.00168
wp1995	0.00594	0.00202	0.00191	0.00151	0.00154	0.00147	0.00092	0.0013	0.00161	0.0013
wp2000	0.00628	0.00214	0.00194	0.00156	0.00207	0.00133	0.0008	0.00136	0.00154	0.00151
wp2005	0.00654	0.00198	0.00199	0.0017	0.00193	0.00144	0.00087	0.00129	0.00152	0.00134
wp2010	0.00676	0.00188	0.00169	0.00148	0.00182	0.00131	0.00081	0.00131	0.00133	0.00136
wp2015	0.00511	0.00251	0.00218	0.0019	0.00164	0.00162	0.00159	0.0014	0.0013	0.00129

Table 3. Absolute trend of word frequency.

Category	Average Slope	Gloss	Word Count
1	$-4.7 \times 10^{**}(-4)$	obvious decrease	3
2	$-9.0 \times 10^{**}(-5)$	decrease	104
0	$-1.4 \times 10^{**}(-7)$	constant	6793
4	$3.5 \times 10^{**}(-5)$	increase	231
3	$1.3 \times 10^{**}(-4)$	obvious increase	27

Table 4. Relative trend of word frequency.

Category	Relative Slope	Gloss	Word Count
1	-0.109810946	obvious decrease	22
4	-0.105143772	obvious decrease	303
9	-0.09516075	obvious decrease	5
3	-0.091128015	obvious decrease	3
8	-0.086133244	obvious decrease	78
5	-0.008250087	constant	5615
7	0.08176863	obvious increase	11
0	0.09351104	obvious increase	928
2	0.09616939	obvious increase	35
6	0.102154761	obvious increase	158

A fifth-degree polynomial was generated to fit the data of each word at the twelve time points. Ignoring the offset terms, we performed cluster analysis on the five polynomial coefficients. The coefficients were categorized into 10 groups. For each group, the number of words it contains was summarized in **Table 5**, and its average variation trend was plotted in **Figure 1**.

Table 5. Ten categories generated by cluster analysis with examples.

Category	Word Count	Example Words
0	5405	biennial, challenged, economy, hybrid, Jewish, linguistic, rationality, sculptural, skepticism, politics
1	17	D, journal, la, e, aesthetics, inches, p, aesthetic, history, Music, see, church, et, le, x, college, de
2	4	Jstor, leonardo, book, use
3	102	abstract, camera, decorative, Madonna, tones, photography
4	85	historical, Indian, renaissance, hegel, color
5	1	Art
6	1014	colonial, documentary, geometry, imagery, logic, machines, Kandinsky, Picasso, Titian, Zen
7	516	Dante, ethical, fashion, men, tapestry, woman
8	1	university
9	13	Artists, science, archive, http, information, utc, technology, mit, one, film, org/terms, hu, fig

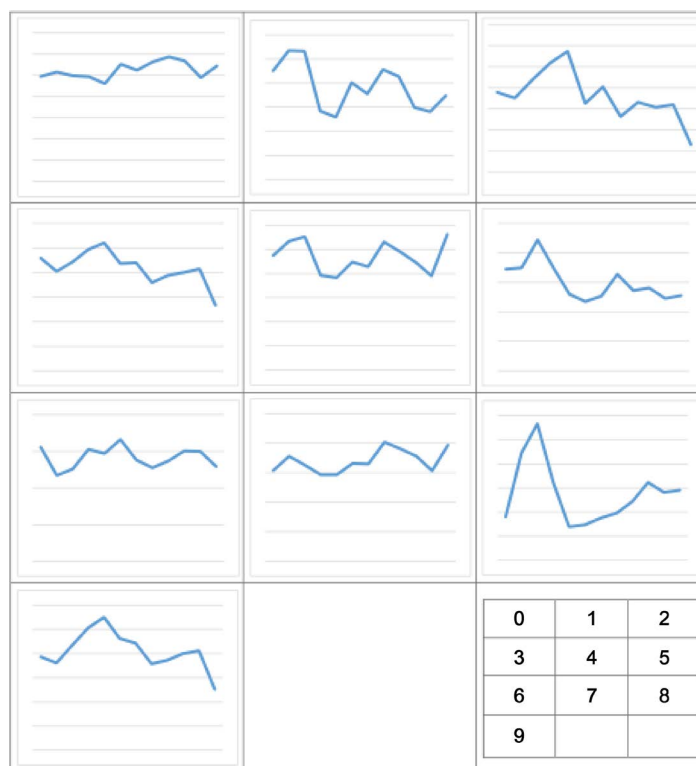


Figure 1. Average trends of the word frequency of ten categories.

3. Results and Analysis

Multiple types of information were obtained by word frequency trend analysis. Sixty words with the most rapidly increasing frequencies were listed in **Table 6**. More tables and graphs were put in Github due space limitation, including the lists of two hundred words with the most rapidly increasing/decreasing frequencies. Part of the codes were also published in Github:

<https://github.com/gaojunhui68/Word-frequency-analysis-python>.

Since it is impossible to answer “what is contemporary art?” by a simple due to its dynamism, we instead tried to describe what has happened in contemporary art field from several aspects which are intimately associated with contemporary art: politics, economics, technology and, in the end, art concept itself. In order to do this, we use the results from our word frequency data to guide our analysis.

3.1. Trend under Political Environment

The result generated by word frequency analysis will be discussed here. As **Figure 2**

Table 6. Sixty words with the most rapidly increasing rates.

1 - 10	11 - 20	21 - 30	31 - 40	41 - 50	51 - 60
hegel	virtual	modernism	spatially	politics	agenda
reconciliation	grounded	solo	impossibility	guitar	audible
copyright	web	jonathan	layered	highlighted	commodity
frieze	track	subjectivity	wine	audiences	screens
multiple	giacometti	addressing	judges	film	tracks
biennale	shot	crew	protests	corporate	collective
temporality	agency	stillness	interview	interviews	regulated
engages	featuring	disciplinary	featured	biennial	conversation
bonds	practices	performers	exposes	interact	collaborators
fictional	close-up	reflective	collaborative	network	narrative

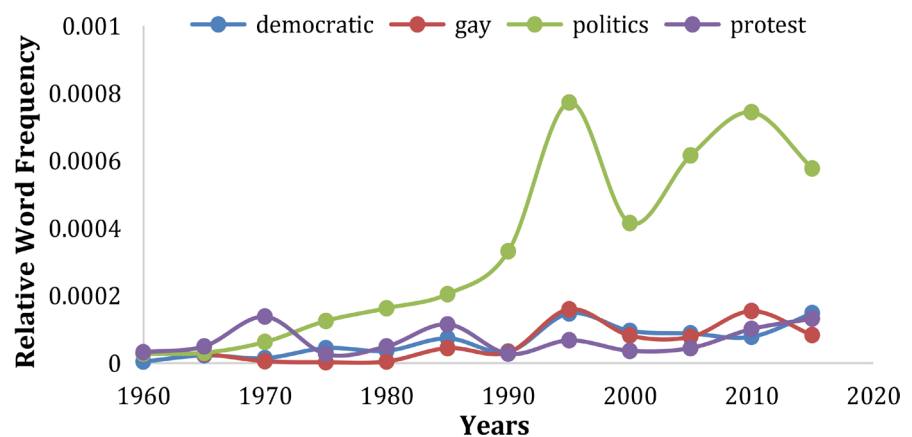


Figure 2. Trend for word frequencies of the four political terms.

shows, from 1960 to 2015, the frequencies of four terms, *politics*, *democratic*, *gay* and *protest* all shows an increasing. From **Table 7**, *politics* is the term with most obvious increasing slope. Its explosively rising frequency at the year 1995 is paralleled with the turbulent political environment at that time, such as World Trade Center Bombing in 1993, Nelson Mandela elected as the president of South Africa in 1994 and Oklahoma City Bombing in 1995.

Responds to contemporary politics is one of the major themes of contemporary art. In the 20th century, especially the series of worldwide civil rights movement starting at the sixties provided the art field with abundant topics. For example, political events like the Cuban Revolution in 1950s, Vietnam War, African-American civil right movements all appears frequently in various artists' works.

Artists made their calls through their artworks. The U.S. validated its global leading position after the end of the Cold War in the twentieth century. Abstract expressionism represented by Jackson Pollock and Greenberg responded to the national call for a representation of a powerful country and freedom of aesthetics (Doss, 2002). The second generation of Appropriation artist Fred Wilson conveyed his resistance to the colonization, slavery and racial inequality problems in Maryland to the public by his exhibition Mining the Museum (Corrin, 2002). Artists did not only respond to the contemporary political environment, but also made efforts to change the environment via the power of art. In 1989, the queer artist Robert Mapplethorpe's photography exhibition in Corcoran Gallery of Art was forced to cancelled by the government of Cincinnati due to "inappropriate homosexual contents". On the night of June 30 of the same year, gay right supporters protested outside Corcoran Gallery of Art, projecting ten works of Mapplethorpe's photography on the walls of the gallery building and raised great public attention (Doss, 2002). Artists' attention to political issues like this is reflected in the change of frequencies of political keywords in contemporary art journals.

3.2. Trend under Technological Environment

Table 5 contains ten categories of words classified by curve cluster analysis in section 2.4. In category 9, the similar frequency variation patterns of science and technology indicates that the influence of technological development on contemporary increased sharply from the year 1970. Some of the influential events include JVS launched videocassette recorders in 1975 and IBM launched first

Table 7. Slope values of economic related words.

Id	Slope	Category
economy	1.56E-05	0
commodity	1.06E-05	0
goods	1.01E-05	0
markets	4.48E-06	0

personal computer in 1981. The newly invented mediums attracted a wide attention from contemporary art field. **Figure 3** shows that, the word frequencies of new art mediums such as *films* and *media* began to surge from 1980. Simultaneously, these new forms of art mediums started to replace some of the traditional medium and ideologies: in **Figure 4**, the frequencies for words associated to relatively “classical period” like *tonal*, *rococo*, *madonna*, *annunciation* have obvious negative trends. We are going to discuss the positive and negative trends below.

One reason for the rising frequency of technology-related words is that photography and public media gained popularity during period of industrialization. Especially the invention of the camera in the 19th century, it not only brought intense discussion on the temporality of art, but also pushed artists to go beyond the traditional *spatial* representation on canvas, and to explore new possibilities of spatiality. At the beginning of the 20th century, the French critic Rivière criticized the that “nothing is more hypocritical than perspective” on the canvas (Rivière, 1912). In the French critic, Olivier Hourcade’s opinion, “illusion is a trick, a lie, and flatness, line and shape are the truth.” (Hourcade, 1912) The influence of camera does not stop at the surface of physical representations but extends to artists’ internal ideology. Because of the incomparable precision of the

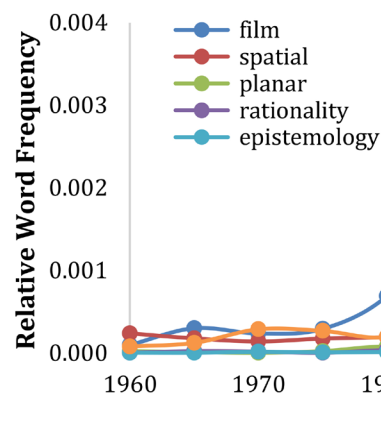


Figure 3. Words with rising frequencies.

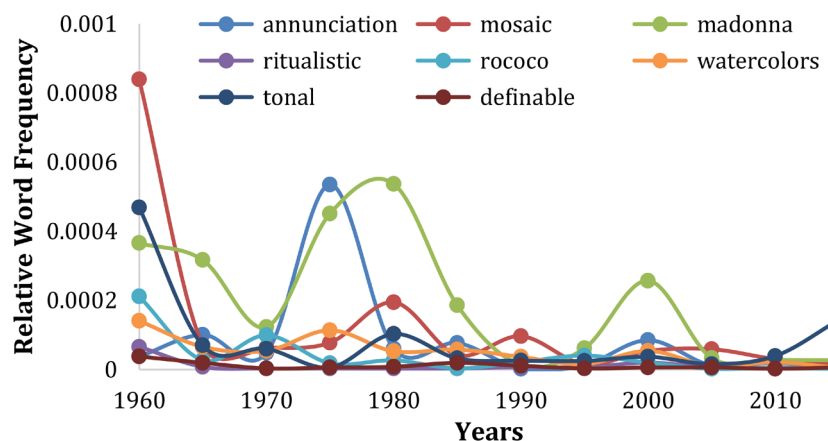


Figure 4. Words with lowering frequencies.

camera in capturing physical forms, artists were forced to switch their focus from forms to more abstract ideas. Conceptual art, for example, emphasizes on recording of concepts more than physical forms. Moreover, the objectivity of the camera, the feature that makes the optical machines free of human *subjectivity*, became a signal of contemporary art in 1960s. For instance, the logic and *rationality* revealed in pop art and minimalism still exerts meaningful influence on today's art society.

Technology development also diluted artists' concentration on some traditional mediums and topics. Science added diversity to art medium inventory, while decentralized pigments from the position of dominant medium. Therefore, the related words like *tones* and *watercolors* show negative change in frequencies. At the same time, the transfer of focus from classic historical and religious paintings to industry related topics causes the decline of frequencies of words like ancient Greeks and Rome, *mosaic* art, *renaissance* and *rococo*. From the 15th to 18th century, cultures, architectures and sculptures from Greeks and Rome were the sacred classic sources of arts at that time. And religion, especially the *ritual* of Christianity stood at the center of the classic resources. Whether an artist could adequately make references to these sources is a direct measurement of the quality of the artist's work. Nevertheless, technological explosion in the twentieth century subverted religion dominant hierarchy. Narratives from the Bible like *Madonna* and the *annunciation* were replaced by objects of industrialization, such as the food cans in Andy Warhol's pop arts, cushions and light bulbs in Robert Rauschenberg's Neo-Dada work *Odalisk*. In a broader historical perspective, since the 19th century, artists began to pay attention to human response towards the painted object, rather than religious and historical references. Then, impressionists in 1870s switched their focus from narratives and actual forms to people's impression of the painted objects, and the cubists such as Picasso made their works further abstracted. Until the end of the twentieth century, the physical forms of objects completely disappeared from the canvas. This trend shows the deviation from traditional narrative painting, and hence the loss of the readability of the artwork itself. Audience can no longer understand the ideas which the artist wants to convey merely from the appearance of the work. This is one of reasons that contemporary art is hard to understand generally: audience should not stare at the work itself but consider the background and history behind the work as well.

Words with increasing frequencies: media, film, spatial, planar, subjectivity, rationally, rationality, epistemology, epistemological.

Words with decreasing frequencies: sarcophagus, basilica, pillars, Vitruvius (Rome architect), mosaic, mosaics, annunciation, triptych, madonna, ritualistic, rococo, watercolors, tones, tonal, definable.

3.3. Trend of Art Concepts

In history of art, meanings of certain concepts evolve along time, but others re-

main relatively constant throughout the history. We are going to discuss some examples from Table 3. The variation of word frequencies of the chosen words is displayed in Figure 5: all chosen words show constant or a rising trend during the fifty-five years, such as modernism.

- *Modernism*: it has been one of the major topics of art historians since 1870s. At the beginning, modernism was defined as an opposition to traditional elitism, stating art should come from ordinary like just as the food cans in Andy Warhol's prints. After the 19th century, modernism became the opposite to consumerism as we discussed in section 3.2. The content of modernism changes with specific historical and social contexts. Therefore, it is a good starting point to understand the nature of contemporary art.
- *Audiences, reflective*: the changing relationship between artworks and audience also marks the difference between contemporary art and traditional ones. The autonomy of the narrative components in traditional form make the artwork relatively independent of audience. Whereas, audience's interaction with the artwork is a necessary part of contemporary art. For example, the representative minimalism work *Die*, 1962 by Tony Smith itself is just a 6' × 6' × 6' cube in the center of a room. The purpose of the work is to induce audience with a sense of surprise or disappointment of its minimalism, and hence emphasize audience's self-consciousness that I am in this room looking at a simple cube. Without audience, this *Die* will lose its meaning. Most of the contemporary artworks in 20th century care about the connections among audience, artworks, and the space of exhibition.
- *Temporality*: the relationship between artwork and time is also a crucial topic in art history. The invention of camera brought us the discussion of temporary and permanence. Temporality is an inseparable factor of contemporary art forms like land art, public art, architecture and sculptures etc.
- *Skepticism, challenge*: Artists' challenge to the past and present exists in all time period. From the 15th century until now, we have the Renaissance artists challenged the dark Middle Age; we saw the great works of impressionists

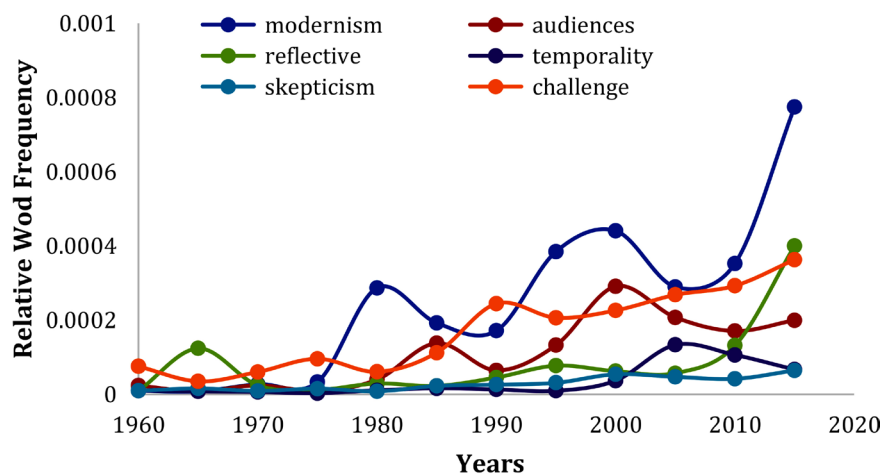


Figure 5. Words with lowering frequencies.

reform the artwork from narrative-orientation to human response to the nature; we have neo-expressionists in 1970s challenged the idea of pure abstraction from expressionists in 1950s.

4. Discussion and Conclusion

We analyzed the articles from eight art history journals from 1960 to 2016 by word frequency analysis, trend analysis and cluster analysis, and found that the word frequency variation of contemporary art terms is closely related to the change of the social contexts in corresponding time period. This linkage suggests that word frequency statistics is a sensitive tool of studying the trend of contemporary art, or even the development of the contemporary society. The data of word frequency, i.e. text, are a reliable medium of history recording. Therefore, text analysis could become our magnifier to explore the development of art and society. Furthermore, compared to human researches, the relative objective machine algorithm reduces possible bias from traditional data analysis methods, and hence provides scholars with suggestions of new perspectives and directions.

This research has several limitations: 1) the studies data include only art journals in English due to authors' limited language ability, and half of the journals are published in United States. The lack of articles from other countries like European, Asia and Africa results in bias towards the situation in the US. 2) The number of art history journals included in the data is relatively small compared to the number of current publishing art journals. Also, the time interval (five years) is also relatively long, so the result might not be able to reflect completely the trend within the five-year interval. 3) The algorithm parsed the texts in units of word. Therefore, terms existing in form of phrases or sentences are ignored by the algorithm, such as names and titles of artworks. 4) Words with similar or the same semantic meanings but different parts of speech were not merged into one unit, which could cause the statistical frequency of certain concept to be lower than its weight in reality.

To conclude: text analysis is a reliable tool to study the trend of arts and the change in society. In future studies, we believe the combination of text analysis with other methods like visual recognition, mechanical translation can bring a broader view of data science from a global perspective.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- Alberro, A. (2008). Periodising Contemporary Art. *Australian and New Zealand Journal of Art*, 9, 66-73. <https://doi.org/10.1080/14434318.2008.11432794>
- Corrin, L. (2002). Mining the Museum: An Installation Confronting History. *Curator: The Museum Journal*, 36, 302-313. <https://doi.org/10.1111/j.2151-6952.1993.tb00804.x>
- Doss, E. (2002). *Twentieth-Century American Art*. Oxford: Oxford University Press.

- Elgammal, B., & Saleh, A. (2015). *Large-Scale Classification of Fine-Art Paintings: Learning the Right Metric on the Right Feature*.
- Hourcade O. (1912). *The Tendency of Contemporary Painting. Included in Fry (1966), Cubism*. New York: McGraw Hill.
- Meyer, R. (2016). *What Was Contemporary Art?* Massachusetts, MA: MIT Press.
- NLTK 3.3 Documentation (2018). *Natural Language Toolkit*. <http://www.nltk.org/>
- Passonneau, R. J., Yano, T., Lippincott, T., & Klavans, J. (2008). Functional Semantic Categories for Art History Text—Human Labeling and Preliminary Machine Learning. In *Proceedings of the 1st International Workshop on Metadata Mining for Image Understanding, MMIU 2008—In Conjunction with VISIGRAPP 2008*, 13-22.
- Rivière J. (1912). Sur les Tendances Actuelles de la Peinture. *Revue d'Europe et d'Amérique*, March 1912, 397.