

Using Simple NLP Tools to Trace the Globalization of the Art World

Alix Rule

Dept. of Sociology
Columbia University
New York, NY, USA
aer2132@columbia.edu

Zhongyu Wang and Rupayan Basu

Dept. of Computer Science
Columbia University
New York, NY, USA
{zw2259, rb3034}@columbia.edu

Mohamed AlTantawy

Agolo, Inc.
New York, NY, USA
mohamed@agolo.com

Owen Rambow

Center for Computational Learning Systems
Columbia University
New York, NY, USA
rambow@ccls.columbia.edu

Abstract

We introduce a novel task, that of associating relative time with cities in text. We show that the task can be performed using NLP tools and techniques. The task is deployed on a large corpus of data to study a specific phenomenon, namely the temporal dimension of contemporary arts globalization over the first decade of the 21st century.

1 Introduction

This paper reports on the use of fairly simple Natural Language Processing (NLP) technology as a tool for social research. We seek to understand the globalization of contemporary art, relying on releases for art events worldwide over the period of 1999-2012. A first phase of the project exploited Named-Entity Recognition (NER) to extract cities named in the corpus.

In this second phase of the research, we turn our attention to a novel task: the temporal dimension captured in these texts. By identifying the timing of geographically-cited events, we are able to examine how the history and projected future of the art world evolve alongside, and as a result of, its current geographical structure. To determine whether cities named in press releases refer to events that occur in the past, the present, or the future, we need methods for resolving time expressed in text relative to the time of the release of the text. We use the Stanford Temporal TaggerSUTime (Chang and Manning, 2012), as well as rules we have built on top of the Stanford part-of-speech tagger (Toutanova et al., 2003), to identify the temporal referent of each city mentioned. The two systems in combination perform slightly better than either does alone, and at a high

enough level of accuracy to produce meaningful data for network analysis. We start by describing the project in more detail in Section 2, as well as the data we use in Section 3. Section 4 discusses our method for identifying the temporal location of named events. The networks we build using this data afford some preliminary insights into the dynamics of contemporary arts globalization, which we present in Section 5. Our aim here is not, however, to offer definitive conclusions about change in the art world. Rather, we present these analyses as proof of concept to demonstrate how a novel NLP task can help advance a particular agenda in social research.

2 The Goals of the Project

The shifting geography of contemporary art is of interest to social scientists as an instance of cultural globalization (Bourdieu, 1999; Crane, 2008). Scholars and contemporary art practitioners alike have remarked on changes in the geography of globally-significant art activity in recent decades (Vanderlinden and Filipovic, 2005; McAndrew, 2008; Quemin, 2006). Accounts of these dynamics often converge: for example, many have pointed to the role that speculation in the art of the developing world plays in fragmenting the once geographically-concentrated field of contemporary art (e.g., (Lee, 2012; Stallabrass, 2006)). Such claims remain largely a matter of conjecture, however, having never been subjected to systematic empirical scrutiny.

Because of the difficulties of compiling appropriate data, contemporary arts globalization has rarely been studied using quantitative methods. The few quantitative analyses that exist focus on particular cities (Velthuis, 2013), individual artists (Buchholz, forthcoming), or a particular category

of art-related events, for example fairs (Quemin, 2013). Our project relies on NLP techniques to solve the problems of data collection that have limited previous research endeavors. We extract information from a large and growing corpus of press releases for contemporary art events occurring worldwide. An initial phase of the project revealed for the first time an object that has resisted empirical description: the contemporary global art world. The research was able to track changes in the art worlds geographic structure over the period of 2001 until 2012. Named Entity Recognition (NER) was used to identify cities mentioned in the body text of announcements as the location of other significant contemporary art events. Pooling documents over a given time window yielded a dynamic network of cities, enabling the researchers to track the evolution of the art world. The analysis revealed that a distinct central core of cities all in Europe and North America enjoyed a remarkably stable ranking as top art locales over the period; however, as the art world expanded, connection to these capitals became less important as an entry to the art world for locations outside the global north (Rule and Brandt, 2013).

“The city that I believed was my past, is my future, my present; the years I have spent in Europe are an illusion, I always was (and will be) in Buenos Aires” (Borges, 1978). Borges’ remarks capture a paradox about places of significance: important places are places where important things *have happened* and *will happen*—but this is so from the viewpoint of the present, which is always evolving. This insight implies that in order to understand the dynamics shaping the geography of significant art activity, analysts need to disentangle the real time in which attributions of significance to locations occur, from the relative timing of the events that make them significant. The current phase of our research attempts to assess how the geographical evolution of the art world shapes contemporary arts history and its projected future. It does so by considering how events mentioned in the corpus are located in time, relative to the event in question. We build upon the NER-based strategy developed in Phase 1 for identifying cities as the location of important events. Here, however, we distinguish cities as the location of events in the past, present, or future.

[The artist] has exhibited extensively in group shows and biennials across Europe, Asia and the Americas including Manifesta 8, the Third Moscow Biennale and Sharjah Biennial. His work is the subject of a major monograph, Laurent Grasso: The Black-Body Radiation (Les Presses du Reel, 2009). As the 2008 Laureate of the Marcel Duchamp Prize, Grasso presented a special exhibition at the Centre Georges Pompidou (2009). Grasso’s work is currently on view at La Maison Rouge in Paris in an exhibition of the Olbricht Collection; forthcoming are solo exhibitions at the Galerie nationale du Jeu de Paume, Paris in 2012 and at the Musée d’art contemporain de Montréal, Canada in 2013.

Figure 1: A sample press release

3 Data

We draw on corpus of press releases for contemporary art events worldwide, distributed over the premier email digest service for contemporary art. The digest serves a professional clientele, and sends out 3-5 press releases a day. These emails are free to the digest’s 90,000 subscribers (compare to *Artforum*’s circulation of 35,000). Press releases are both written and paid for by the institutions sponsoring the events they announce. The corpus covers the years 1999 to 2012; it contains 10,566 documents and 6362284. Our automated detection, checked by human annotators, identified 1007 unique cities.

We conceive of each press release as an attempt to raise the status of the event in question by mentioning other high-status events to which it is related. Pooling the references of cities across documents thus gives rise to a network, analogous to a citation network. The press releases are rather conventionalized, and name cities almost exclusively as the location of important related events. The format is not just used by this particular digest service, but is common to the art world generally. A sample is shown in Figure 1.

4 Time Identification

Cities associated with art events are identified using an NER-based approach, which performs at an f-measure of 64.5 and which we do not further describe in this paper. We use *two* approaches for temporal resolution to assess whether the cities mentioned are the locations of events that happened in the past, are currently happening, or expected to happen in the future.

The first approach analyzes explicit time ex-

	Accuracy	Past			Current			Future		
		<i>P</i>	<i>R</i>	<i>FI</i>	<i>P</i>	<i>R</i>	<i>FI</i>	<i>P</i>	<i>R</i>	<i>FI</i>
Baseline*	63.8	0	0	0	63.8	100	77.9	0	0	0
Lexical (L)	66.6	79.3	56.6	66.1	75.3	75.5	75.4	21.6	36.5	27.1
Grammaticalized (G)	73.4	70.0	48.8	57.5	74.9	89.0	81.4	64.5	38.5	48.2
L_G	69.2	77.2	73.6	75.4	81.2	71.8	76.2	23.2	42.3	29.9
G_L	74.0	71.0	51.2	59.5	75.5	89.0	81.7	64.5	38.5	48.2
L_G & G_L	78.0	70.5	76.0	73.1	82.2	85.6	83.9	65.5	36.5	46.9

Table 1: Accuracy and precision, recall and F-measure for past, current and future events. *Baseline: Tagging all events as *current*.

pressions in dates, time durations, holidays, etc. The second approach uses verbal tense to resolve grammaticized reference to time. In both approaches: (1) We use distance within a sentence as a heuristic to associate a temporal expression to a city. (2) Cities associated with art events are tagged as *past*, *current* or *future*, where **current** is the default tag. This section describes both approaches and how they can be combined to enhance the performance.

4.1 Explicit Lexical Temporal Expressions

Explicit temporal expressions could be partial or complete dates (*Jan-01*, *March 2nd*, *2014*), specific named days (*Christmas 2001*), deictic time expressions (*last Wednesday*), time durations (*the past five years*), seasons (*next fall*), etc.

We use the Stanford Temporal Tagger, SUTime (Chang and Manning, 2012) to detect and evaluate Temporal expressions relative to the publication date of the press release. Temporal expressions that evaluate to incomplete dates take on the missing date components from the publication date. If the temporal expression resolves to a date before, same as or after the publication date, then the event associated with the city is tagged as past, current, or future event respectively. Cities that are not associated with explicit temporal expressions are given the default tag: *current*.

4.2 Grammaticalized Temporal Expressions

Tense is grammaticalization of location in time (Comrie, 1985). Our tense-based approach has two steps. First, the Stanford part-of-speech tagger (Toutanova et al., 2003) is used to identify the POS tags of all tokens in the documents. Second, we use hand-written rules to identify the temporal interpretation of every event. The rules are mostly based on the POS tags of verbs; we use

the tag sets from the Penn Treebank Project (Marcus et al., 1993). We use only the verbal tags: VB (*Verb, base form*), VBD (*Verb, past tense*), VBG (*Verb, gerund or present participle*), etc. and MD (*Modal*). Here are some examples of the rules; events associated with: (1) VBP or VBZ are tagged as *current*. (2) VBD are tagged as *past*. (3) VB that are preceded by will/MD are tagged as *future*.

4.3 Results and Combining of the Two Approaches

During development, we found that the two time tagging approaches perform well in different situations. For example, the lexical approach has the higher recall (66.1%) for past event as writers often mention the exact date of past events. However, in the case of current events dates are seldom mentioned, the present tense is enough to communicate that the event mentioned in a sentence is ongoing. This accounts for the higher recall of the grammaticized (89%) as against the lexical approach (75.5%) for current events. If we do not assign *current* by default to the failed cases in both the lexical and grammaticized approaches, the recall for current events drops drastically for the lexical approach (75.5% \searrow 0.69%) compared to the grammaticized approaches (89% \searrow 81.5%).

Combining the two approaches improves the performance of the time tagging task. We start with one approach to tag an event and when that approach fails (does not provide an answer), the other approach kicks in. If both approaches fail, the event is tagged by default as *current*. For example, in approach Grammaticalized_{Lexical} (G_L), when the Grammaticalized approach fails to tag an event, the Lexical approach is used. The best combination is achieved by running the Lexical_{Grammaticalized} (L_G) first; if the output tag

is *past*, accept its answer; otherwise run the Grammaticized_{Lexical} system (G_L) and accept its verdict. If G_L has no answer, choose *present*. The accuracy of this approach is 78.0%. For more details, see table 1. Please note that analyses in section 5 rely on the Grammaticized_{Lexical} (G_L) approach.

		2001-04	2005-08	2009-12
# Documents		1063	3352	4687
Past	Nodes	238	600	645
	Edges	6477	34901	51213
	Share	21.5%	25.8%	31.3%
Curr.	Nodes	509	957	995
	Edges	37087	152279	186928
	Share	70.3%	63.8%	49.8%
Fut.	Nodes	158	352	460
	Edges	2047	11270	16851
	Share	8.2%	10.4%	14.0%

Table 2: Properties of networks of past, present, and future art events, for three periods. “Share” refers to the percentage of events in relative time during a given period.

5 Creation and Interpretation of Networks

To analyze the geography of the art world, we generated networks of events in relative time over three four-year periods: 2001-2004; 2005-2008; and 2009-2012. Nodes represent cities named as the location of art events, and edges co-citations from a particular geographical and temporal standpoint. Specifically, we attribute ties between cities of the past, present or future of a given period when they are mentioned together in press releases for events currently occurring in the same city. Ties are weighted to reflect the number of co-mentions. For example, if Basel is noted twice as a future event location by press releases for events in New York, and LA is mentioned by New York press releases three times, the edge between Basel and LA receives an edge weight of 2, reflecting the number of shared mentions.

Basic properties of the nine resulting networks are displayed in table 2. Notice the marked difference in the distribution of events in relative time between 2005-2008 and 2009-2012. Cities mentioned in connection with events in both the future and the past figure as a greater proportion of the

total in the last, as compared to the middle period. Though we consider the possibility that this is an artifact of the data, it seems more likely that the shift reflects the impact of a real world development: namely, the global economic recession, which began in 2008 and had profound effect on funding for contemporary art. In the context of the economic contraction, it seems that both contemporary arts future and its history become more important relative to its present.

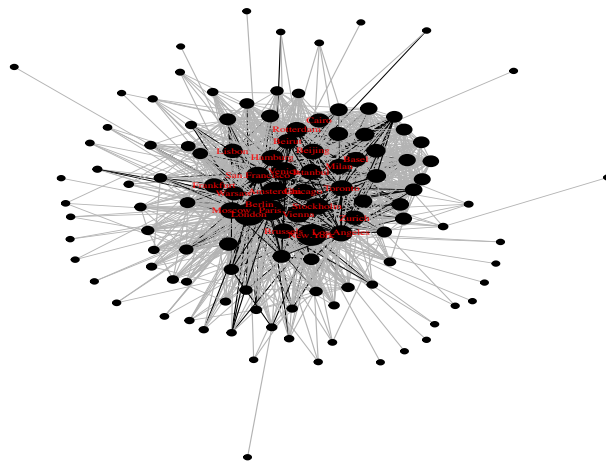


Figure 2: Future of the global contemporary art world, 2009-2012

Figure 2 shows the network of future events arising from 2009-2012. Recall that the graph captures the world’s future, as it is structured by the geography of significant art world activity during this period. Just as in Borges remarks about Buenos Aires, we observe here how important places in the art worlds present shape the map of its future.

This graph suggests an art world that is rather conservative, more so, at least, than accounts of the speculation driving its globalization would imply. The network exhibits a classic core-periphery structure; the cities at its center are the “big” art capitals of the global north: London, New York, Berlin, Paris, Venice. In other words, the common proposition that the “hype” around emerging art markets has de-centered the contemporary art world is not borne out by a close empirical examination from the perspective of the hype-generators themselves. Rather, it would seem that the role of these cities as the prime location for contemporary art in the present enables them to project themselves as the location of significant art activity in the future.

References

- Jorge Luis Borges. 1978. *Obra Poetica 1923-1976*. Emece.
- P. Bourdieu. 1999. The social conditions of the international circulation of ideas. pages 220–228.
- Angel X Chang and Christopher Manning. 2012. Su-time: A library for recognizing and normalizing time expressions. In *LREC*, pages 3735–3740.
- Bernard Comrie. 1985. *Tense*, volume 17. Cambridge University Press.
- Diana Crane. 2008. Globalization and cultural flows/networks. *The Sage handbook of cultural analysis*, pages 359–381.
- Pamela M Lee. 2012. *Forgetting the Art World*. MIT Press.
- Mitchell P Marcus, Mary Ann Marcinkiewicz, and Beatrice Santorini. 1993. Building a large annotated corpus of english: The penn treebank. *Computational linguistics*, 19(2):313–330.
- Clare McAndrew. 2008. *The International Art Market: A Survey of Europe in a Global Context*. European Fine Art Foundation (TEFAF).
- Alain Quemin. 2006. Globalization and mixing in the visual arts an empirical survey of high culture and globalization. *International sociology*, 21(4):522–550.
- Alain Quemin. 2013. International contemporary art fairs in a globalized art market. *European Societies*, 15(2):162–177.
- Alix Rule and Philipp Brandt. 2013. Dynamic art objects: The evolving structural composition of the global art scene. Citation Networks Section of the SUNBELT International Social Networks Research Association Conference.
- Julian Stallabrass. 2006. *Contemporary art: a very short introduction*. Oxford University Press.
- Kristina Toutanova, Dan Klein, Christopher D Manning, and Yoram Singer. 2003. Feature-rich part-of-speech tagging with a cyclic dependency network. In *HLT-NAACL - Volume 1*, pages 173–180.
- Barbara Vanderlinden and Elena Filipovic. 2005. *The Manifesta decade: debates on contemporary art exhibitions and biennials in post-wall Europe*. The MIT Press.
- Mordechai Edgar Velthuis. 2013. The art market in new york, basel and tokyo: Deconstructing the economics of aesthetic consumption. *Art and Money*.