New Wine in Old Bottles

Ideological Transformation and Rhetorical Creation of Market in China's People's Daily, 1946 - 2003

THE UNIVERSITY OF CHICAGO

Shilin Jia & Linzhuo Li

Introduction

How could a new ideological regime be established upon lasting old ideas, institutions, and culture that tend to have strong inertia to persist? For example, one of the most surprising transformations in the 20th century was China's embrace of market economy under the leadership of a communist party. How could it have happened?

Market is never created ex nihilo. The "liberalization" of market is itself an engineering project in ideology (Polanyi, 1944). To make it happen in the communist China, there was not just some changes in economic relation of production on the ground but also a huge ideological transformation in words to resolve the incompatibility between "free market" and orthodox Marxism. In this study, we attempt to answer how this transformation could have happened by applying computational content analysis to the full text of the communist party's mouthpiece, *the People's Daily*, from 1946 to 2003.

In sociology, there has been a long debate about whether ideology and culture have any independent explanatory power in historical processes. Operated as an authoritarian regime's mouthpiece, the *People's Daily* perhaps is the least expected place to find cultural autonomy. Nevertheless, we found that during the course of China's economic transformation, the party state's official rhetoric has progressed in a very linear and smooth fashion. Newness always comes out of old repertoires, and controversial concepts like 'market economy' only became stabilized by being attached to a existing stable rhetorical subspace. Our findings echo the Weberian idea that ideology and culture should be viewed as a semi-autonomous social sphere that interacts with other social processes with its own logic.

Data and Methods

Full texts from the very first issue published in 1946 to the last issue in 2003 were used for analyses. After applying word segmentation and stopword removal to the text, we applied various exploratory techniques to detect changing patterns in word frequencies and word embedding spaces in the newspaper articles. Many of the analyses are computationally intensive and made only possible by running on the Midway high performance cluster of the University of Chicago.

Word frequencies

The frequency of a word or phrase w_i in a month (or year)'s corpus C_t can be calculated and normalized as $P_t(w_i) = \frac{\#(w_i \in C_t)}{|C_t|}$. A time series $\mathbf{d}$ can be constructed with:

$$d_t = D_{\mathrm{KL}}(P_t \| Q_t), \tag{1}$$

where P_t is the probability distribution of words in the tth month, Q_t is a prior reference distribution, and D_{KL} is the Kullback-Leibler(KL) divergence. To find out common trends in word frequencies, we employed correspondence analyses (CA) (Benzêcri, 1973), which given an input matrix $\mathbf{X}$, applies a singular value decomposition to $\mathbf{D_r^{-1/2} X D_c^{-1/2}}$ where $\mathbf{D_r}$ and $\mathbf{D_c}$ are diagonal matrices whose diagonal entries are the row sums and column sums of $\mathbf{X}$. In our analyses, the input matrix $\mathbf{X} = [x_{ij}]$ is a term frequency-inverse document frequency (tf-idf) matrix where x_{ij} is the weighted frequency of ith word in the jth month.

Word embeddings

Every year's words were embedded into a 400-dimensional vector space based on their context words. Specifically, we used the skip-gram model implemented in the Python package *gensim*. Given a sequence of training words $w_1, \ldots, w_T$, the objective of the skip-gram model is to find d-dimensional vector representations of words to maximize the average log probability,

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{j=-k}^{k} \log p(w_{t+j}|w_t) \tag{2}$$

where k is a window size (Mikolov et al, 2013). The probability of correctly predicting the word w_i given the word w_j is defined as

$$p(w_i|w_j) = \frac{\exp(u_{w_i}^{\top} v_{w_j})}{\sum_{l=1}^{\mathcal{V}} \exp(u_l^{\top} v_{w_j})}, \tag{3}$$

where $v_{w_j}, u_{w_j} \in \mathbb{R}^d$ are the input and output vectors of the word w_j, and $\mathcal{V}$ is the set of all unique words in the corpus. Huffman Trees were used to speed up computation, and optimal solutions were found through stochastic gradient descent. Equation (2) and (3) allow extrapolating the probability of co-occurrence of a set of words given a year's learned vector space by utilizing both the input and output representation of the words (Taddy, 2015).

Procrustes analyses (PA) were applied separately to align the learned word embedding spaces (Hamilton et al, 2016). Given every year's row-normalized word embedding matrix $\mathbf{W}^{(t)} =\in \mathbb{R}^{|\mathcal{V}_t|\times d}$, the rows of which are the input vectors of all words in that year, the optimal rotation and reflection matrix $\mathbf{R}^{(t)} \in \mathbf{R}^{d\times d}$ is found such that

$$\mathbf{R}^{(t)} = \underset{\mathbf{Q}^{\top}\mathbf{Q}=\mathbf{I}}{\arg\min} \left\| \mathbf{W'}^{(t+1)} - \mathbf{W'}^{(t)}\mathbf{Q} \right\|_F, \tag{4}$$

where $\mathbf{W'}^{(t)}, \mathbf{W'}^{(t+1)} \in \mathbb{R}^{|\mathcal{C}_{t,t+1}|\times d}$ are sub-matrices of $\mathbf{W}^{(t)}$ and $\mathbf{W}^{(t+1)}$, indexed by the common vocabulary set $\mathcal{C}_{t,t+1} = \mathcal{V}_t \cap \mathcal{V}_{t+1}$. After rotation, words in different years can be aligned to the same vector space. PA allowed us finding the *stability* of every word from year t to $t+1$ as the cosine dissimilarity between the vector representations of the word in t and $t+1$. An alternative measure is the substitution rate of its 100 nearest neighbors, $s_t = 1 - \frac{|\mathcal{N}_t \cap \mathcal{N}_{t+1}|}{|\mathcal{N}_t \cup \mathcal{N}_{t+1}|}$, where $\mathcal{N}_t$ is the neibours set at time t (Barabasi, 2007). We used both.

Findings

Path dependency in the frequency domain

We found that, first, there was persistent path dependency in the party state's ideology, especially in the economic domain. Heatmaps based on year-to-year KL divergences are shown in Figure 1. Domain-specific words are the words that are most similar to the word "经济(economy)" (and "政治(politics)") in the word-embedding spaces.

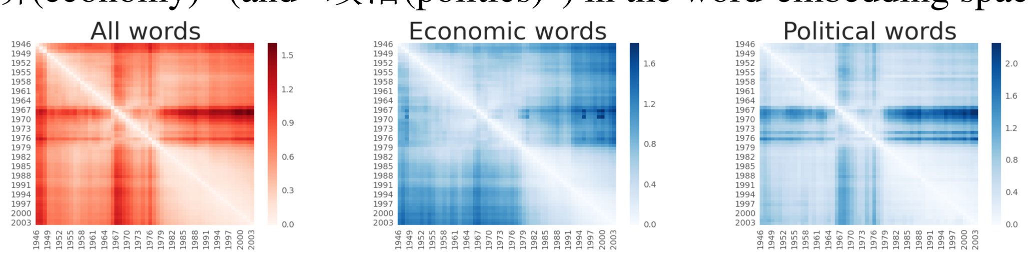

Figure 1: Heatmaps based on year-to-year KL divergences

After the Cultural Revolution, the party state's official rhetoric, in the grand scheme, had moved in a very smooth and linear fashion in almost all of the time. Time series constructed with two choices of Q_t are plotted in Figure 2.

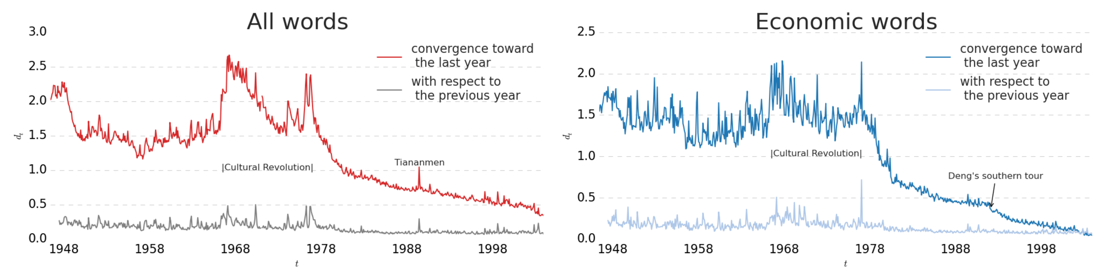

Figure 2: Monthly frequency change measured in KL divergence. The time series on the top were constructed with Q_t equal to the average probability distribution of words in the final year, 2003. The ones at the bottom were constructed with Q_t equal to the average probability distribution of words over the 12 months preceding the month t. It turns out that when $\mathbf{d}$ is constructed in the second way, it can be fitted well with an AR(1) model with 1 breakpoint (Bai and Perron, 1998). The breakpoint is found to be around 1979, which is roughly the beginning of China's economic transformation.

New wine in old bottles

Second, we found that, the transformation was initiated in the late 1970s by first utilizing some existing 1950s repertoires. In both the word frequency domain (as shown in the CA biplot in Figure 3) and the word embedding spaces (as shown in Figure 4), after extreme radicalization during the Cultural Revolution, the party state's official ideology first gradually shifted back to the level of 1950s (in a recovering period) before evolving to its current form. This "recovering" process has not been well noticed in previous studies.

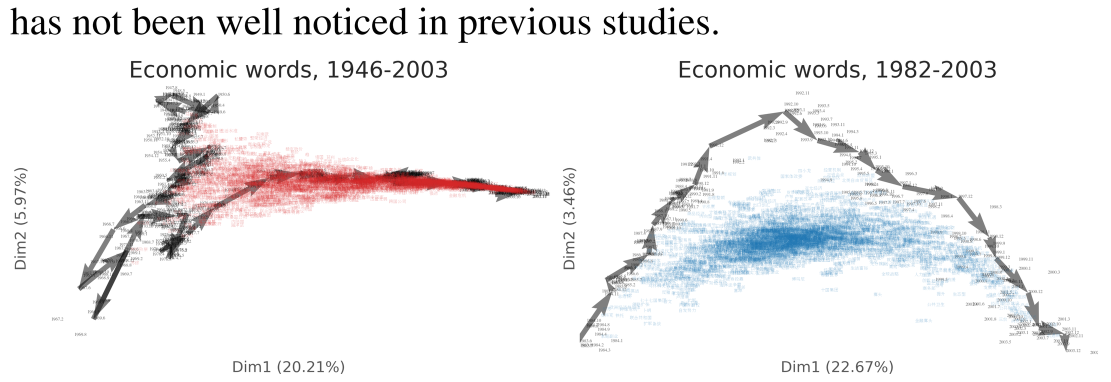

Figure 3: CA biplots in row principal components, dim 1 vs. dim 2. On the left, the first principal dimension captures an ideological shift from the leftist point during the Cultural Revolution toward the latest time point in the reform period. The second dimension captures how the party state's economic ideology became radicalised during the Cultural Revolution and then first reverted back to its 1950s level before embarking on its journal toward reform and opening-up. A second biplot of economic words from 1982 to 2003 is shown on the right. The principle dimension still captures a smooth and linear progression, but it is worth noticing that the second principle dimension correctly captures a sudden critical change in the year 1992. In 1992, although officially in retirement, being dissatisfied about the party state's economic conservatism after the Tiananmen crackdown, Deng Xiaoping made a famous southern tour of China that finally pushed the official leaders to start China's market reform.

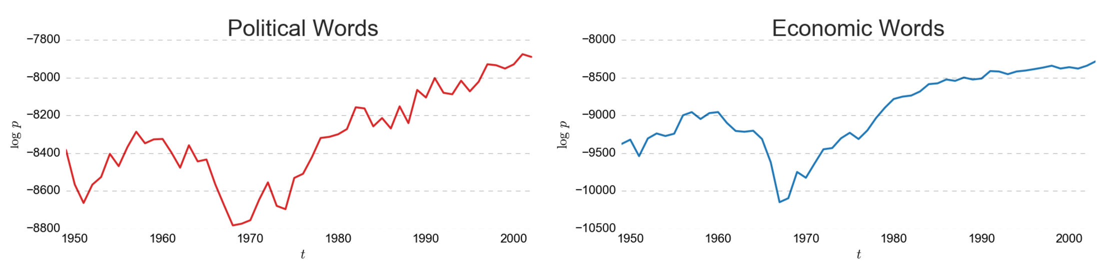

Figure 4: Log-likelihoods of co-occurrences of the most recent (reformist) year's political and economic words given each year's corpus.

Stabilization in sudden changes

A rhetorical field of course cannot be completely autonomous. But we found that even when a sudden change happened like in the year 1992, newly introduced concepts still need to get attached to existing stable repertoires to get themselves stabilized.

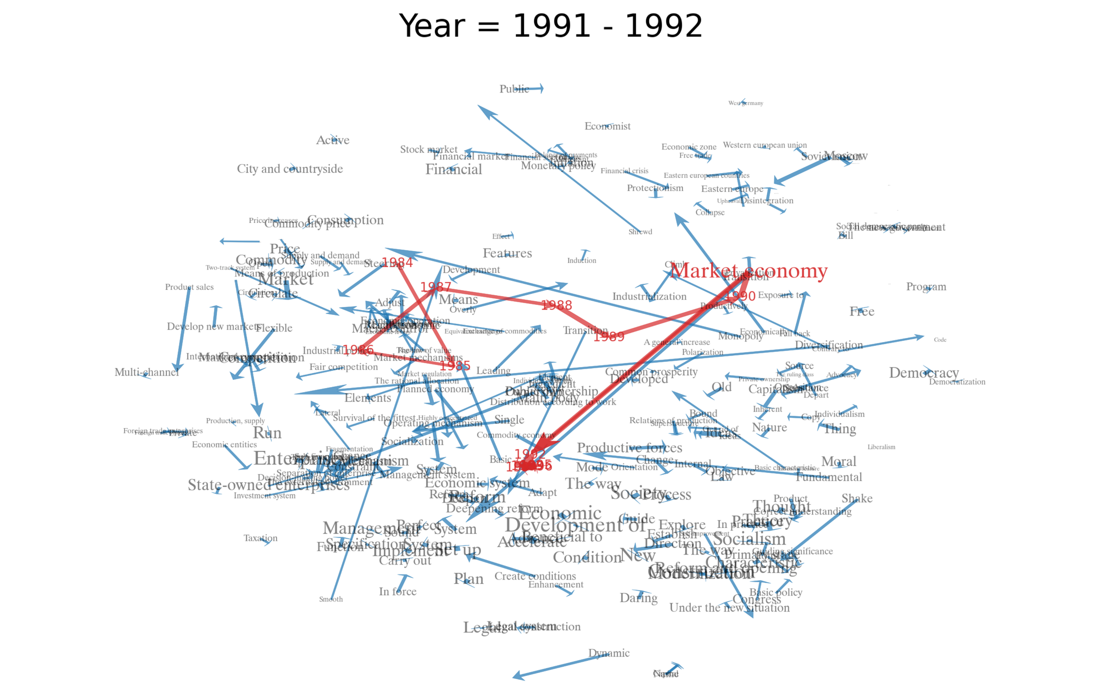

Figure 5: The vector space around the word "市场经济(market economy)" from year 1991 to 1992, visualized in t-SNE (van der Maaten and Hinton, 2008)

As visualized in Figure 5, the word "市场经济(market economy)" never really settled down until 1992. It only became stabilized by being attached to a stable subspace of words. The heatmap in Figure 6 corroborates the fact that its new neighbors in 1992 were already very stable prior to the year of change in comparison with its old neighbors. Instances of sudden stabilization are rare, but "market economy" is not a single case. We found all instances of word that became suddenly stabilized during the course of history, and as shown on the right of Figure 6, their sudden stabilization all took place in a similar manner. The new neighbours they were attached to after their stabilization were mostly already very stable before the sudden changes. The pattern is clearer in the reform period than it is in the earlier years. Results based on two different measures of stability are all similar.

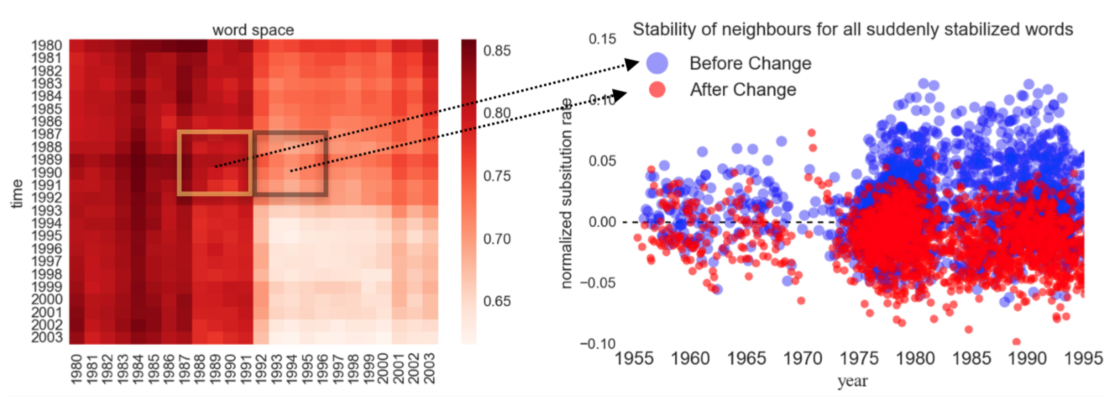

Figure 6: (Left) Heatmap of year-by-year neighborhood stability of the word "market economy". The vertical axes is the time line. The columns on the horizontal axes are indexed by the word' yearly neighbors. A cell (i, j) is colored according to the averaged stability measures of the year i's 100 nearest neighbors in year j. (Right) Plots of the yearly stabilities of the pre-change and post-change neighbours of all the words that experienced a sudden stabilization during the course of history. For each word, the stability measures of its yearly neighbors in each of the 5 years before and after the year of critical change (values of the 25 cells in the black square in Figure 5(left) were first demeaned and then plotted in red and blue respectively.