DO NEW WORDS PROPAGATE LIKE MEMES? AN INTERNET USAGE BASED TWO-STAGE MODEL OF THE LIFE CYCLE OF NEOLOGISMS

Menghan Jiang[1], Kathleen Ahrens[2], Xiangying Shen[3], Sophia Yat Mei Lee[2], Chu-Ren Huang[2]

[1]*Shenzhen MSU-BIT University, Shenzhen*

[2]*The Hong Kong Polytechnic University, Hong Kong*

[3]*Southern University of Science and Technology, Shenzhen*

ABSTRACT

Neologisms reflect new ideas or new concepts in our life and play an important role in cultural transmission and the vitality of human language. The explosion of neologisms, especially in the past two decades, can also be ascribed to the popularity and accessibility of digital content and social media. In this paper, we focus on the issue of how neologisms arise by looking at the trajectory of developments in terms of their usage over time, i.e., their life cycle. By studying neologisms *in vivo*, instead of as *fait accompli*, we hope to better understand the nature of neologisms and to enable better prediction and earlier inclusion of neologisms. To achieve this goal, we examine the memetic model for the life cycle of neologisms and compare it with a recently studied epidemic model. We present a longitudinal modeling of the development of neologisms based on internet usage data aggregated from Google Trends, covering the 90 most influential Chinese neologisms from 2008-2016. Our study verifies that the memetic model can describe and predict the life cycle of the neologisms robustly for the early stages (i.e., the ascending stages) of its cycle, but not for its full life cycle, and crucially cannot predict the inflection point. We conclude that two models are needed for word propagation: a memetic model for the initial stages and an epidemic model for the latter stage, particularly the inflection point. This two-stage/two-model approach allows for neologisms to be more easily identified as potentially new words, as it is easier to write a program to automatically filter for emerging terms using a memetic model.

KEYWORDS

Language Modeling **M**emetic Model **E**pidemic Model **N**eologisms

---

Acknowledgments We are grateful to the journal editors and the anonymous reviewers for their detailed reviews and valuable suggestions, which helped to improve the manuscript. This work is supported by 广东省普通高校青年创新人才类项目“基于互联网用户数据的流行语发展周期量化研究”(2022WQNCX073).

Menghan Jiang (蒋梦晗; corresponding author) [menghan.jiang@connect.polyu.hk]; 1 International University Park Road, Dayun New Town, Longgang District, Shenzhen, Guangdong, China; https://orcid.org/0000-0001-5005-147X

Chu-Ren Huang (黄居仁; corresponding author) [churen.huang@polyu.edu.hk]; The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China; https://orcid.org/0000-0002-8526-5520

1. NEOLOGISMS AS MARKERS OF HUMAN LANGUAGE EVOLUTION

Advances in science and technology in the first decade of the 21st century have changed our way of life profoundly. The world evolves, knowledge increases and is redefined, and language expresses this renewal, which has resulted in the emergence of new linguistics units to describe a new concept or a new reality (Baayen and Renouf 1996; Rey 2005; Castellví et al. 2012; Hacken 2020;). The most representative and significant linguistic units of this change are lexical units, as advances of all kinds, especially technological and scientific innovations are expressed through new terms. Thus, neologisms are useful markers of language change over time (Sonnad 2015), and a clear indication of the vitality of a language (Castellví et al. 2012). The contribution to cultural transmission and linguistic vitality and the typically rapid development of neologisms in a relatively short time make them both accessible and intriguing in the study of the quantitative modeling potential of human language evolution (Cavalli-Sforza and Feldman 1981).

Memes, a neologism proposed as the cultural counterparts of genes by Dawkins (1976), provides the conceptual foundation for some of the most influential theories of cultural evolution (Blackmore 2000). More recently, with the rapidly growing popularity and impact of internet memes, not just on social media but in all aspects of our society, the memetic model has been invoked to describe not only the spread of internet memes but also the changes in society (e.g., Davison 2012). Subsequently, neologisms are frequently treated as memes (e.g., Pagel 2009; Tsur and Rappaport 2015), both as a simile (i.e., the linguistic memes) and in association with the most dominant media for the creation and spread of neologisms (i.e., social media and internet). The memetic model has often been adopted to account for (internet) neologisms. On the other hand, it has also been pointed out that a simple memetic model may not be adequate to account for the complex behaviors of neologisms (e.g., Poulshock 2002; Jing-Schmidt and Hsieh 2019; Jiang et al. 2021). However, to the best of our knowledge, no detailed studies of neologisms based on the quantitative memetic model have been implemented or evaluated. The current study aims to address this gap in order the better understand the theoretical strengths and weaknesses of a memetic model for neologisms and the spread of lexical changes.

1.1 Lexicographical Approaches to the Study of Neologisms

Lexicographic considerations have guided previous academic studies on neologisms and concentrated on research issues such as a neologism's life cycle (e.g., Metcalf 2004; Renouf 2007, 2013) or its integration into a language's lexicon (e.g., Baayen and Renouf 1996; Schmid 2008; Kerremans 2015). Neologisms offer a fertile testing ground for the basic lexicographic issue of how to identify a new lexical entry (or the semantic shift (Fišer and Ljubešić 2019)), especially the issues involving rules governing lexical forms of a specific language (Hacken and Koliopoulou 2020), and empirical methodologies for determining the acceptance of a neologism into the lexicon (Hacken 2020; Klosa-Kückelhaus and Wolfer 2020). Such studies usually take the ‘gatekeeper’ perspective. That is, the lexicographer is the gatekeeper at the end of the road of the emergence of neologisms to decide on which ones to officially enter a dictionary.However, one question that arises is whether all emerging neologisms follow the same model, and eventually enter the lexicon of the language via the same criteria. Hence another perspective on neologisms is to examine the life cycles of neologisms. Several empirical studies focused on the variables that optimize a new word’s survival rate, i.e., to decide whether a neologism would vanish or would become a daily lexical entry (Fischer 1998; Metcalf 2004; Renouf 2007; Klosa-Kückelhaus and Wolfer 2020). Altmann et al. (2011, 2013), for example, have found that word frequency and the niche are significant factors influencing the survival rate of a new word. Other studies have focused on investigating the evolution pattern of neologisms. For example, Zhang (2017) adopted internet usage data from Baidu Index to discover the diffusion mode and path of the neologisms 山寨 *shan1zhai4* ‘re-engineered copy’, and found (1) the spread of 山寨 *shan1zhai4* shows the features of periodicity; (2) its diffusion pattern demonstrates the feature of radiating from the centers.

1.2 The Epidemic Model

In addition to these types of observational studies on the spread of neologisms, we argue that it is essential to propose a comprehensive theoretical model that can clearly describe a neologism’s life cycle since a successful model of the life cycle of neologisms will have implications as to how languages change and evolve as well as how new ideas propagate and influence collective human behavior. One such theoretical model was proposed by Jiang et al. (2021). They utilized internet-based data from Google Trends for modeling the propagation and life cycles of neologisms. The results demonstrated that the epidemic model(i.e., the mathematical model to account for the competition between the transmissibility of the virus and the infected populations’ ability to recover) provides a reliable description of the life cycle of neologisms, particularly in terms of identifying the peak point, but leaves open the question as to whether the memetic model may be advantageous for modeling neologisms at a different stage in their life cycles.

1.3 The Memetic Model

The memetic model is another possible model that to model a neologism’s life cycle (Dawkins 1976). The memetic model differs from epidemic models in that it captures the degree of propagation of the memes, instead of the transmissibility of a virus. Neologisms are often described as textual memes (or lexical memes). Similar to cultural memes, such as humorous images, lexical memes propagate quickly. Thus, in this paper, we examine if the life cycle of neologisms may be predicted by a memetic model in addition to an epidemic model (Jiang et al. 2021). Given the close link between the explosion of neologisms and new media and the view that language is a "culturally transmitted replicator" (Pagel 2009), this hypothesis seems to be reasonable (Dawkins 1976; Blackmore 2001; Blackmore 2007), as memetics arise from an evolutionary model of the propagation of ideas and cultural concepts.

Memetics is a model to study the replication, spread, and evolution of memes based on the biological evolutionary theory of Darwin (He 2005; Heylighen and Chielens 2009; Kronfeldner 2014). Memetics relies on the idea of memes. Memes are defined as basic ontological units of culture/language based on gene selectionism that have analogous properties and causal roles to those of genes in biological evolution (Kronfeldner 2014). The dominant memetic model is genetic replication (and vice versa (Dawkins 1976; Blackmore 2001; Blackmore 2007)). In this model, fast-spreading neologisms can be predicted in terms of the rates of replication of the memes, similar to simple cells/genes propagating through replication and copying. This is a memetic model of evolution, where memes are analogous to the genes of a culture. Their life cycles are determined according to each gene's ability to propagate via copying (Blackmore 2001; Blackmore 2007). This hypothesis underlies most studies of memetics prompting the adoption of the computational model of self-replication of genetic information as models for the spread of memes and neologisms (Dawkins 1976; Hull 1982; Dennett 1991; Hull 2000; Jing-Schmidt and Hsieh 2019). The memetic model may also provide insights into language purism and phonological adaptation (Marello 2020; Hacken 2020). That is, neologisms (especially loan words) could be cultural memes that are entering another culture (Ogilvie 2008).

A variety of research has adopted a memetic approach to investigate the mechanisms behind the propagation of internet neologisms (e.g., Marino 2015). According to Jing-Schmidt and Hsieh (2019), the CNKI database (China National Knowledge Infrastructure: a key national research and information publishing institution in China (Wikipedia 2020)) holds more than 1800 journal articles published in Chinese between 2008 and 2016 on the memetics approach to communication and language. The majority of the studies were conducted under a lexicographical paradigm in which scholars examine how the features of biological evolution can be adapted to account for the spread of Chinese neologisms, and how different categories of lexical memes can be developed via the process of replication and transmission (e.g., He 2005, 2008, 2014; Chen and He 2006; Wei 2011; Liu 2014; Li 2019). However, these qualitative models were not created to predict the behavior of neologisms. As such, these studies do not provide empirical data to support the predictive power of model fitness of the memetic model or the fitness of the model.

1.4 Research Questions

In addition, the existing literature has pointed out several reservations concerning utilizing the memetic model for word spread. For example, many studies have challenged the conclusions reached as there has been almost no empirical data to back up the theories that were put forth (Heylighen and Chielens 2009); moreover, memetics has been viewed as merely a conceptual framework lacking in explanatory power (Benitez-Bribiesca 2001; Edmonds 2002; Gil-White 2005; Greenberg 2005). Although at least one empirical study has supported the validity of the memetic model (e.g., Tsur and Rappoport 2015), its data is limited to hashtags and new phrases. A hashtag is by its nature a word meme, while a new phrase may become, but is not yet a neologism. Thus, the developmental trends of lexical neologisms have yet to be investigated with empirically reliable models (Edmonds 2002; Gil-White 2005). In this current study, we utilize the internet-usage data for the memetic model fitting, to empirically verify whether the memetic model can well describe and predict the life cycle or certain stages of the life cycle of neologisms.

Jiang et al. (2021) focused on the epidemic model and showed that even though both the epidemic and memetic models perform well, the epidemic model is more suited to model the full life cycle of internet neologisms. In particular, they showed that the monotonic growth predicted by the memetic model fails to predict the leveling end of the life cycle of successful neologisms. Given their focus on accounting for the explanatory power of the epidemic model, this paper did not explain why the neologisms are often considered to be like memes, nor did it attempt to test a more complex memetic model to address the issues observed. Our initial observation suggests that the early stage (i.e., the fast-spreading stage) of the neologisms is similar to replicating memes described by the memetic model (Heylighen and Chielens 2009). This may be because the self-replicating process of memes (the ascending stage of the neologisms' development) is believed to be a simple exponential growth (Heylighen and Chielens 2009). However, for the epidemic model, if the population of the infected people is much smaller than the total number of susceptible people at the very beginning of the virus epidemic, the rising curve used to indicate the infected population in the SIR model (an epidemic model which simulates the interactions among three populations: Susceptible, Infectious, Recovery) is also believed to be expressed as a simple exponential growth function. Hence, what we are looking at in this paper is if and how the memetic model can describe and predict the life cycle of the neologisms, as well as to what extent the memetic model can capture the point when a neologism's usage saturates and stabilizes, and when the point of inflection occurs.

To answer these questions, we adopt the methodology of Jiang et al. (2021) to study the aggregated data from Google Trends, which provides the popularity of a selected word within a flexible range of timescales. The daily search frequency of a specified neologism reported on Google Trend is renormalized to a range from zero to one hundred, which is adopted as a reflection of each word's popularity. In our current study, we also follow the lexical extraction method in Jiang et al (2021) and selected the same 90 neologisms from the year 2009-2016 in 咬文嚼字 *Yao3Wen1Jiao2Zi4*. By performing the model fitting processes, we aim to ascertain the following:

- Is a memetic model an alternative model for the life cycle of neologisms?
    - If yes, what are the most important features and applications?
    - If not, why not, and which part of the life cycle of the neologisms is best predicted with the memetic model?

In what follows, we describe our studies of modeling the extracted data using a memetic model.

2. DATA AND METHODOLOGY

2.1. Data Source

This study includes 90 neologisms (the same 90 neologisms as reported in Jiang et al. (2021)) published by the 咬文嚼字 *Yao3Wen1Jiao2Zi4* journal, covering the nine years from 2008 to 2016 (Yao 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017). The selection criteria adopted by 咬文嚼字 *Yao3Wen1Jiao2Zi4* is similar to the criteria used by the American Dialect Society in selecting the Words of the Year (Metcalf 2004): *Fashionable:* the words can reflect the characteristics of discourse in each year; *Popular*: the words are familiar to the people who read the press regularly; *Expressive:* the words have an expressive effect, e.g., being creative, humorous, ironic, or euphemistic.

Our definition of neologisms follows the standard definition of neologisms in the literature. This definition has been accepted as a norm, including being adopted in most dictionary definitions of neologism, such as “a new word or expression, or a new meaning of a word” from Oxford Learner’s Dictionary (Oxford University Press 2022). This definition is theoretically well-motivated based on the definition of a word as a unique form-meaning pair, first advocated by De Saussure (2011) and now followed by all linguistic theories. This means a new word can be formed in three different ways: a new form is assigned to an existing meaning, a new meaning is assigned to an existing form or a pair of new meanings and new forms. All the neologisms in our study meet the requirement of this well-motivated definition.

For instance, the existing word 最美 *zui4mei3* has the cited form mapped to the meaning of ‘the most beautiful’. In 2012, a new meaning emerged in the formal announcement from the government as well as in newspapers: 最美 *zui4mei3* ‘exemplary, the paragon of’. This new meaning is typically used to modify a profession or a social role, e.g., 最美护士/老师/科技工作者 *zui4mei3 hu4shi4/lao3shi1/ke1ji4gong1zuo4zhe3* ‘an exemplary nurse/teacher/scientific and technological worker’. This new word has the old form of 最美 *zui4mei3* mapped to the meaning of being exemplary or being the paragon of, and different from the meaning of ‘the most beautiful’. By adopting the above definition and trusted sources of neologisms, it should be clear that neologisms include both words that eventually enter the lexicon (i.e., new words), and words with a short life cycle (i.e., new buzz words). To investigate the life cycle of neologisms, we include both types of words.

We access Google Trends to generate a data pool of these 90 neologisms’ search frequencies as our source data. The popularity of a word is calculated by its search frequency on Google, to model the rapidly rising and decaying pattern of the neologisms.

Note that a central methodological issue we raise is to study neologisms *in vivo* and not as *fait accompli*. Lei et al. (2021) pointed out that the reason why past studies of lexical changes focused on replacement changes is that it is not possible to have reliable documentation for the process of changes before a new word enters the lexicon unless it has a corresponding old form (or meaning) that it replaces. Without the ability to study other types of changes or unsuccessful emerging changes, our knowledge of linguistic changes has been quite limited. The current vitality of neologisms as well as their digitally recorded life cycle trajectory provided a hitherto unavailable opportunity to study lexical changes *in vivo* and to model different factors that contributed to the success of lexical changes.

More specifically, we claim that the search frequency data is crucial to our model-fitting study as it is a reasonable approximation of both how the words spread across the internet and the frequencies of their use (e.g., Burrows and Savage 2014; Madsen 2015). Firstly, note that linguistic usages are loosely classified as 1) active usages: speaking and writing; 2) passive usages: hearing and reading. Frequencies from written corpora have been treated as a reliable approximation of total usage frequency for the following reasons: 1) written corpora are by nature documentation of writing (and occasionally speaking/hearing), which serves as 2) reading materials. This writing/reading duality is essential in the design of balanced corpora as the criteria of the composition of sampled text from the Brown corpus was based on a survey of reading habits (Francis and Kucera 1979). However, it can only be an approximation not just because such corpora typically under-represent speaking/hearing, but also because it cannot reflect the actual frequency of reading usages. Note that modern corpora are typically either unbalanced or balanced according to criteria established by other balanced corpora previously and not by a survey of actual reading habits of the targeted current readers. As such, available current corpora can longer claim to have comprehensive coverage of both active and passive usages of language.

In terms of neologisms, the usage situation can be described as dominated by reading and writing, since speaking/hearing activities involving neologisms are unlikely before they enter the active vocabulary of the general public. The search engine uses provided a very good approximation of usage frequency because they include 1) actual information-seeking searches (by speakers who have acquired the word); 2) dictionary look-up searches, which are most likely the reaction of reading unfamiliar or novel words; i.e., first encounters of a neologism. Hence the total online search frequencies offer comprehensive coverage and a good approximation of the total usage of the neologism. This interpretation of search behaviors of neologisms is corroborated by the fact that it correctly predicts that search engine uses will drop off significantly when neologism enters the lexicon. This is because a successful neologism in the lexicon no longer triggers a dictionary look-up search when read/heard.

2.2. The Rise-Decay Pattern of the Neologisms

As reported by Jiang et al. (2021), we also observed the same tendency that the popularity of most of the neologisms (62 out of 90) have similar patterns: the rise and decay patterns, as shown in Figure 1. It is these 62 neologisms that we are focusing on for model-fitting in this study.

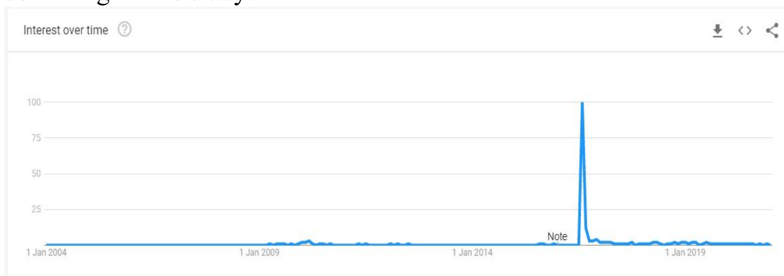

Figure 1: The rise and decay pattern of the neologism 洪荒之力 *hong2huang1zhi1li4* 'with all one's might'

2.3. Data Filtering

We follow the methodology of Jiang et al. (2021) and exclude the other 28 neologisms for model fitting for different reasons. The criteria and examples can be found in Jiang et al. (2021).

1) **Data Scarcity:** Twelve of the 28 neologisms are omitted because of data scarcity.

(1) Seven of these twelve words do not have enough data in Google Trends to adequately assess our model fitting.

(2) Moreover, five words show only rising patterns, indicating that they are still in the emergent stage, and have not fully developed yet.

2) **Neologisms with multiple meanings/functions/interpretations:** Further analysis reveals that for the remaining 16 neologisms, each word form stands for multiple words or has multiple meanings, including established and conventionalized meanings and new interpretations/functions (Thornton and Burdette 2017), similar to add 新 *xin1* 'novel' to 新冠状病毒 *xin1guan4zhuang4bing4du2* 'novel coronavirus.' As the Google Trends data is word form-based and not disambiguated, the neologisms' actual frequency cannot be teased apart. In these cases, the fluctuating pattern is a sum of tendencies of multiple-word meanings and cannot be used for our current study. These data cannot be modeled for neologisms without disambiguation; thus we must leave this question for future studies since automatic disambiguation cannot yet be done reliably.[1]

3. RESULTS

3.1. Model Fitting

The memetic model is a possible model to account for neologisms, but so far there has been a lack of empirical research on this topic (Heylighen and Chielens 2009). Thus, the main purpose of this study is to fill this significant gap and to understand if the memetic model predicts the life cycle of neologisms successfully or not, and why. We used the memetic model proposed by Heylighen and Chielens (2009) to fit the observed data.

In previous studies, scholars believe that the formation of memes requires four stages: assimilation $A(m)$, retention $R(m)$, expression $E(m)$, and transmission $T(m)$. $A(m)$ denotes the proportions of memes encountered by the host that are assimilated. $R(m)$ represents the proportions of these assimilated memes that are memorized. $E(m)$ is the number of times the host expresses a retained meme. $T(m)$ is the number of potential new hosts reached by a copy of the expression. Integrating these four stages is considered a qualitative generalization of the meme formation process. However, a mathematical function that involves four stages (denoting at least four parameters) is complicated. A mathematical equation that can well describe the data evolution is favored to have as few parameters as possible, and a simpler form is beneficial to fitting regression analysis. Therefore, a simplified mathematical function of the memetic model is necessary. For the model proposed by Heylighen and Chielens (2009), the sum of all four stages is characterized by one parameter, i.e, the fitness $F(m)=A(m)R(m)T(m)E(m)$, indicating the rate of growth for the expected number of observed data $N(t)$. Thus, the evolution process of $N(t)$ over time can be expressed by the following differential equation:

$$dN/dt = (F(m) - 1)N \rightarrow N(t) = c\exp(F(m) - 1)$$

The arrow in the above equation indicates that $N(t) \equiv c\exp(F(m) - 1)$ is the solution of the differential equation $dN/dt = (F(m) - 1)N$. This result is a traditional exponential growth if $F(m)>1$, exponential decay if $F(m)<1$, and stability if $F(m)=1$. Thus, in the propagation of a neologism, four parameters are reduced to one, which simplifies the memetic model to a simple exponential function so that the fitting calculation can be performed.

Mathematically, the memetic model varies according to the parameter of fitness, which indicates the self-replicating ability of the meme/word. This model can be reduced to an exponential growth mathematically if $F(m)>1$ (i.e., a meme is a survivor that can spread), or an exponential decay if $F(m)<1$ (i.e., the replication of the meme is not sustainable, and the meme will fade away), and a constant value if $F(m)=1$. The memetic model swaps among different mathematical expressions according to the different values of fitness, indicating the different types of self-replicating modes (growth, decay, and stabilization) of the meme/word. The model further assumes that fitness should be independent of time or remain constant for a long enough time. Hence, mathematically, the memetic model is a semi-quantitative model that for a certain $F(m)$, the *N(t)* is a monotonic function of time. In other words, the model can describe and predict the monotonic change in the popularity of the neologism, e.g., the ascending (early) stages of the neologism's life cycles. Henceforth, we acquired the *P(t)* data from Google Trends and fitted the ascending stages (early stages) of the neologisms' life cycles. To estimate the fitting effects, we adopted the fitting parameter $R^2 \equiv 1 - \sum(f_i - x_i)^2/\sum(\bar{x} - x_i)^2$. In the expression, x_i, f_i are the i^{th} values of the data set and fitting function respectively, and $\bar{x}$ is the average value of the whole data. If $R^2 \to 1$ the fitting function is in good agreement with the actual data. We ran a model-fitting analysis of all the neologisms' ascending stages and obtained a mean value of $R^2=0.9161$. The fitting results and the memetic fitting functions of three neologisms (穿越 *chuan1yue4* 'travel back in time', 最美 *zui4mei3* 'the best', and 低碳 *di1tan4* 'low-carbon') are illustrated in Figure 2 respectively. In the picture, the *P(t)* data obtained from Google Trend are dots. The lines denote the memetic model fitting functions. The vertical axis and the horizontal axis represent the popularity *P(t)* and the prevalent periods *t*, respectively. The *P(t)* is normalized and dimensionless. The *t* is in the unit of 'day' implying the number of days a neologism is popular before its *P(t)* starts to decay.

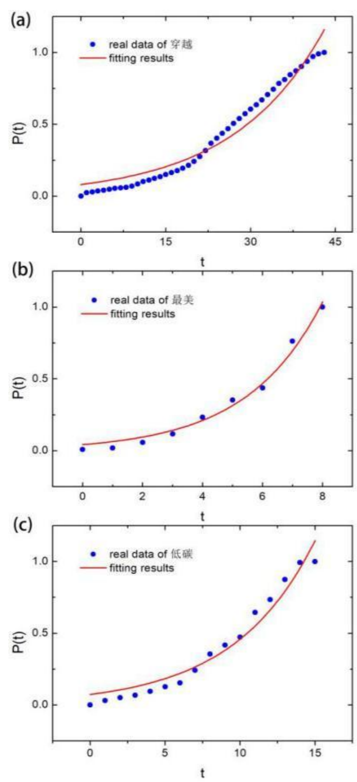

Figure 2. The growth of 穿越 *chuan1yue4* 'travel back in time' (R^2=0.9543), 最美 *zui4mei3* 'the best' (R^2=0.9862), and 低碳 *di1tan4* 'low-carbon' (R^2=0.9616).

However, as we have observed for most neologisms (also stated in Jiang et al. (2021)), their full evolution pattern is not a monotonic growth curve since a neologism, just like any other word, must have an inflection point in its life cycle where the growth of the word plateaus or turns into decline, because otherwise the popularity of a word would increase to infinity which would not be possible (as the popularity must have a maximum value determined by the total population of the language users). In addition, even if a neologism becomes a frequent word and enters a dictionary, its search frequency on the browser would also decay and remain at a stable level. To model the decreasing part of the popularity, we set the parameter fitness of the memetic model to a value less than one, which turns the exponential growth memetic model into an exponentially decaying function. The fitting results show that the descending parts of the neologisms' life cycles could also be well described by the memetic model with a mean value of R^2=0.9019.

In addition, we should also note that for a neologism that contains both rising and decaying patterns (i.e., the neologism with a full life cycle), the memetic model cannot be used for modeling its life cycle because of its monotonicity. To fix this problem, we turned the monotonic memetic model into a piecewise function consisting of several domains with different fitness parameters. We combined both monotonically growing and decaying memetic models, and the complete lifespan of a neologism can be written as:

$$P(t) = \frac{1-\mathrm{sign}(t-\tau)}{2}c_1\exp(F_1-1) + \frac{1-\mathrm{sign}(t-\tau)}{2}c_2\exp(F_2-1)$$

where c_1, c_2 are fitting coefficients; F_1, F_2 are fitness for rising and decaying regime respectively; τ is the time where $P(t)$ reaches its inflection point and sign (...) represents a sign function that holds

$$\begin{cases} \mathrm{sign}(t-\tau) = 1, t-\tau \geq 0 \\ \mathrm{sign}(t-\tau) = -1, t-\tau < 0. \end{cases}$$

We fitted the full life cycle of the neologisms with the fitting functions and obtain a mean value of R^2=*0.8988*. The result is very similar to the fitting effect of the SIR (epidemic) model (R^2=*0.8975*), as shown by Jiang et al. (2021).

The memetic model describes changes, especially with rapid growth. Such a model cannot describe words that are not changing or words that change with decreasing frequencies. This can also be underlined by the design of our study to fit the memetic model in two separate stages: the ascending part and the decaying part. To corroborate this theoretical point, we selected one of the most frequently used words '一 *yi1* 'one'' from 国家现代汉语语料库分词类词频表 'the frequency word list of Chinese National Corpus' for comparison. We ran a model-fitting analysis and obtained the value of R^2=0.617736. R^2 less than 0.7 is generally not considered a strong effect size (i.e., the data pattern cannot be well described by the model) (Moore et al. 2017). The result corroborates the logic that the memetic model for changes is not applicable in describing the life cycle of words not undergoing significant changes.

3.2 Summary of Model Fitting Results

The fitting parameter R^2 indicates that the exponential growth of the neologisms' life cycle best fits the memetic model, followed by the exponential decay, and then the overall evolution pattern. In general, the memetic model can successfully model similar characteristics of exponential growth, exponential decay, and the full life cycle.

However, recall that the simple memetic model can only model either the ascending or descending stage of the neologisms' life cycle due to its monotonic nature. To avoid the monotonicity for modeling the complete path of neologisms' evolution, we combined both monotonically growing and decaying memetic models and turned the monotonic memetic model into a piecewise function. It is worth noting the piecewise function consists of several domains with different fitness parameters. (This means the parameters in different stages needed to be acquired to do the model fitting). The step of combining different domains also turns the memetic model into an *a posteriori* model, as the piecewise fitness parameters were assigned to retrofit known data. The deduction of the piecewise function of the memetic model does not inherently indicate the emergence of an inflection point, and the observed inflection point τ is a given parameter, which exists only for pure mathematical fitting. Since the crucial parameter τ can only be determined when the inflection point exists in the data of the neologism evolution, the modified memetic model itself cannot predict the location of the inflection point (i.e., cannot capture the time when the word usage saturates or stabilizes).

In that sense, we argue that the memetic model is suitable for modeling the neologisms that are still at the emergent/ascending stage of their evolution. However, the full life cycle of neologism cannot be predicted by the memetic model. Suppose the life cycle of a word has proceeded through different stages. In that case, i.e., the growth of the word plateaus or turns into decline, its life cycle can only be described by a modified *a posteriori* piecewise function consisting of several domains with different parameters. By adopting this *a posteriori* piecewise function two weaknesses of the memetic model are addressed:

1) As the piecewise function involves several domains with different fitness parameters, we have to acquire different fitting parameters for different stages (the rising stage, the decaying stage, and the stabilized stage (if any)) for fitting purposes.

2) The step of combining different domains turns the memetic model into an *a posteriori* model, which indicates that the model cannot predict the location of the peak inflection point (the point when a neologism's usage starts to saturate), as the observed inflection point τ has to be determined beforehand for fitting.

4. DISCUSSION

Previous research showed that language data-driven modelling, especially lexical changes in the context of language evolution and change, can be used to support models of genetic and biological evolution (Cavalli-Sforza and Feldman 1981; Cavalli-Sforza and Wang 1986); hence we adopt the memetic model to empirically simulate the evolutionary pattern of the neologisms. We have demonstrated that the ascending stage of the life cycle of the neologisms can be well described and predicted by a memetic model, while the complete propagation path of neologisms cannot, as they can only be *a posterior* described by a modified piecewise function.

The memetic model is a self-adaptive model that simulates the replication and transmission of memes (i.e., the new words). It assumes that memes are selfish and hungry in the sense that their only goal is to maximize their growth and lifespan (Dawkins 1976; Blackmore 2007). The fast-growing of memes is driven by the natural desire for survival of the fittest, and their development patterns mathematically follow exponential growth. This trend has also been reported in a recent study on emergent COVID-19 neologisms (Lei et al. 2021), which demonstrated that pandemic terms quickly entered the stage of diversification and showed a very steep increase. In that sense, the simple self-replicating model with a peak time variable is generally adequate for the memetic model. This explains why the model is suitable to simulate the growing stage of the life cycle of neologisms. In our data set, for example, we have four neologisms that show only ascending trends (i.e., they are still in the emergent stage): e.g., 颜值 *yan2zhi2* ‘appearance-value’, 葛优躺 *ge3you1tang3* ‘slouching on a chair (like Ge You)’, 网红 *wang3hong2* ‘internet celebrity’. These four neologisms were all generated and became popular in 2015 or 2016, hence the evolution time has not been long enough for them to develop a full life cycle pattern. We tested the four words and the results illustrate they all fit the memetic model closely. For example, we obtained a value of R^2=0.9460 for the word 颜值 *yan2zhi2* ‘appearance-value’, as shown in Figure 3. For comparison, we also fitted these words with the SIR epidemic model, and the results show that the words with only ascending trends do not fit the epidemic model very well. For example, we obtained a value of R^2=0.82716 for the word 颜值 *yan2zhi2* ‘appearance-value’ in the fitting of the epidemic model. This indicates that the memetic model is more suitable than the epidemic model for modeling the emergence of neologisms.

The unsatisfactory performance of the epidemic model in modeling the rising part can be explained by the nature of the SIR model. To fit the popularity’s evolution over time one has to solve three differential equations (SIR Model) as given below:

$$\begin{cases} \dfrac{dS}{dt} = -\beta IS \\ \dfrac{dI}{dt} = \beta IS - \alpha I, \\ \dfrac{dR}{dt} = \alpha I \end{cases}$$

where S, I, and R respectively represent the number of susceptible, infected, and recovered. α and β are parameters to characterize the increasing rates of recovery and infectious people. Mathematically, if the data of a neologism is still in development and has not experienced a full cycle from rising to decaying, the fitting results of the epidemic model are unreliable due to the lack of data in giving the explicit solutions of the dI/dt. Therefore, the epidemic model is not suitable for providing a good fitting R^2 to those neologisms that are still in the rising stages.

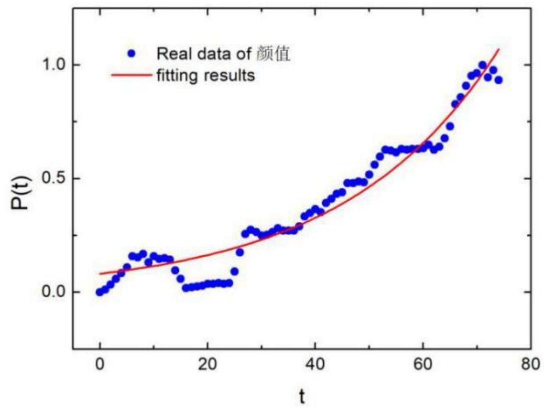

Figure 3. The popularity of 颜值 *yan2zhi2* ‘appearance-value’.

However, when the growth of the word plateaus, saturates, stabilizes, or turns into decline, the memetic model is constrained in terms of peak time and fails to capture the time when the stage changes. The memetic theory assumes that a meme as a replicator is copied whenever information is transmitted from one individual to another via communication or imitation. Due to the desire to survive, every meme tries to produce the largest number of replicas over an extended period (Heylighen and Chielens 2009). The exponential memetic fitness cannot offer a good explanation as to why the fast rates of replications of the selfish memes start to slow down sharply and then drop to a very low value that is typical of a non-frequently used word. In addition, the model cannot predict the location of the inflection point. Based on this, we argue that the memetic model is not an optimal model for modeling the complete life cycle of neologisms.

The changes in developmental stages may imply that there is competition involved in the spreading process of memes. Jiang et al. (2021) argue that the propagation of neologisms can be considered an event as the result of the interaction between sentient (the people who hear/use the neologisms) and non-sentiment actors (the neologism itself), hence the propagation and life cycle of neologisms can be well described by a host-driven SIR epidemic model. For an epidemic model, since it is established according to the competition between the transmissibility of the virus and the infected populations’ ability to recover, it intrinsically implies the existence of an inflection point where the number of recovered people exceeds the number of infected ones. Hence, an epidemic model is an *a priori* model that can provide a good overall description of the neologisms’ evolution patterns observed from the data, as well as a powerful predictive power to locate the inflection point. The memetic model, as a simple self-adaptive model, does not take the action of its sentiment hosts (people who hear/use the neologisms) into direct consideration; hence, it can’t predict the stage changes and the location of the inflection point.

According to Cavalli-Sforza and Feldman (1981), if the diffusion of the neologisms is proceeding over a long period of time, the frequency of the usage of neologisms is always an S-curve. Hence, we could also try to interpret our modeling results with the S-curve frequency development predicted for most complex systems involving cognitive behaviors (Ellis and Freeman 2006; Freeman and Cameron 2008; Fagyal et al. 2010).[2] Note that the S-curve model of complex systems expects a change to start slowly, followed by a period of rapid development (an approximately linear increase), and a plateau with possible residues when the change is completed. The two relatively slow periods of change are the two legs of the S, and the rapid rise (or fall) is often described as the snowball effect (Ogura and Wang 1994; Blythe and Croft 2012). The memetic model can model the sharp rising snowball effect portion of the life cycle.

Given the complex system S-curve interpretation, we can provide further interpretations of the fitting results. The neologisms that are better fitted by the memetic model most likely are still in the ascending part of the S-curve. The ones which are not well fitted by the memetic model have probably reached their maturity. In other words, ignoring the onset part where there is likely not enough data, the memetic model has good predictions for the monotonic growing part but does not perform very well for modeling the maturity of a neologism. One of the possible factors contributing to the discrepancies in the prediction of the memetic model may be "linguistic purism" and/or phonological adaptation issues that arise at the point of a loan word's acceptance to the lexicon of another language.

5. CONCLUSION

In this paper, we obtained internet-based data from Google Trends for the temporally marked popularity of neologisms in Chinese to test whether the memetic model is the optimal model for the life cycle of neologisms. Our study showed that the memetic model is a robust model for the emergence of neologisms but not for their full life cycles, which are better modeled by the epidemic model (Jiang et al. 2021). That is, although the memetic model predicts the early stages of neologisms being coined and gaining popularity, it has inherent limitations in predicting the full life cycle of neologisms, especially the location of the inflection point. As we discussed above, the memetic model displays the replication and transmission of selfish memes. The memes grow quickly to maximize their lifespan, which is driven by the natural desire for survival of the fittest. This fast-growing pattern is mathematically described by the exponential growth found. In that sense, the simple self-replicating model with a peak time variable is generally adequate for the memetic model. This explains why the memetic model is suitable for modeling the emerging stage of neologism's life cycle.

Our modeling results confirmed earlier doubts about the memetic model of neologisms, but also provide empirical evidence to justify the application of the memetic model to describe the fast ascents of neologisms. Since emerging terms are the most relevant and challenging targets of automatic term extraction, this result suggests that the memetic model is a hitherto under-explored model for the automatic discovery of new lexical terms. The memetic model's mathematical form of the neologism's ascending stage is a well-defined exponential growth function. Compared to the epidemic model, more specifically, the SIR model (susceptible, infectious, recovery) which includes three variables and three differential equations, and only has numerical solutions but no analytical solution, the memetic model for a new lexical term is much more straightforward and can more feasibly be implemented in programming for automatically filtering the emerging terms, as it only requires one variable for model fitting (i.e., the replication rate of the meme), and furthermore, has an analytical solution. In addition, our study on linguistic evolution supports the macro-theory that the evolution of human languages mirrors biological evolution (Cavalli-Sforza and Feldman 1981). The evolutionary pattern of language can be clearly described by the theories and models of evolutionary biology. This also has implications for the importance of the independent discovery of basic evolutionary principles using linguistic data.

The study is theoretically grounded in the model of language changes. In particular, it addresses the issues of non-replacement changes with emergent neologisms (Lei et al. 2021). It is an attempt to explore alternative models to the traditional S-curve model, which does not apply to non-replacement changes (Blythe and Croft 2012). The result has implications in both lexicology and social linguistics. In lexicology, we provide an improved model to predict the life cycle of neologisms (Metcalf 2004). In sociolinguistics, our results show that lexical changes, unlike many other highly contagious human behaviors in a highly connected society, are not simply memetic (e.g., Gil-White 2005; Greenberg 2005). The memetic model has been applied to genetic evolution with natural selection as a constraining factor (e.g., Dawkins 1976; Blackmore 2001). The current study reinforces the study of Jiang et al. (2021) to show that speakers as hosts play a central role in language changes and that speakers' decisions reflect social changes (Labov 1966; Li et al. 2020; Lei et al. 2021; Wang et al. 2022). Lastly, given the nature of language as a complex self-adaptive system, other complex system models, in addition to the two models explored in the paper, can also be explored to study neologisms in future research.

NOTES

1. These 16 neologisms can mainly be categorized into the following five classes: (1) The words that did not gain a new meaning, but merely have an old meaning that became more popular because of external incidents; (2) The nouns that acquired new functions, such as addressing terms and/or pronominal functions; (3) The neologisms extended their usages to different conceptual domains and retained their original meaning; (4) The neologisms which borrow existing forms but are less frequently used than the original form; (5) Neologisms involving affixes or bound roots, such as 微 *wei1/wei2* 'micro-', 控 *kong4* 'fervent fan', and 裸 *luo3* 'naked, without strings attached'. More details and examples can be found in Jiang et al. (2021).

2. Note the theoretical foundation of a complex system is that behaviors of such systems are the aggregation of the complex interaction among all members of the systems and of their interaction with a dynamic environment. Hence one signature feature of a complex system is that behaviors of individual members cannot be satisfactorily predicted. And this unpredictability can also be extended to the interaction between members and individual external factors (e.g., Holland 1992, 1996). Given the complex self-adaptive system model of language, external factors are modeled in terms of its aggregated influence on the system and not in terms of the behavior of each individual factor.

REFERENCES

ALTMANN, Eduardo G., Janet B. Pierrehumbert, and Adilson E. Motter. 2011. Niche as a determinant of word fate in online groups. *PLoS ONE* 6(5): e19009.

ALTMANN, Eduardo G., Zakary L. Whichard, and Adilson E. Motter. 2013. Identifying trends in word frequency dynamics. *Journal of Statistical Physics* 151(1-2): 277-288.

BAAYEN, R. Harald, and Antoinette Renouf. 1996. Chronicling the Times: Productive lexical innovations in an English newspaper. *Language* 72(1): 69-96.

BENITEZ-BRIBIESCA, Luis. 2001. Memetics: a dangerous idea. *Interciencia* 26(1): 29-31.

BLACKMORE, Susan. 2000. *The Meme Machine*. Oxford: Oxford University Press.

———. 2001. Evolution and memes: The human brain as a selective imitation device. *Cybernetics & Systems* 32: 225-255.

———. 2007. Those dreaded memes: The advantage of memetics over "symbolic inheritance". *Behavioral and Brain Sciences* 30: 365-366.

BLYTHE, Richard A., and William Croft. 2012. S-curves and the mechanism of propagation in language change. *Language* 88: 269-304.

BURROWS, Roger, and Mike Savage. 2014. After the crisis? Big data and the methodological challenges of empirical sociology. *Big data & society*, 1(1), 2053951714540280.

CASTELLVÍ, M. Teresa Cabré, Rosa Estopà Bagot, and Chelo Vargas Sierra. 2012. Neology in specialized communication. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 18(1): 1-8.

CAVALLI-SFORZA, Luigi Luca, and Marcus W. Feldman. 1981. *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton: Princeton University Press.

CAVALLI-SFORZA, Luigi Luca, and William SY Wang. 1986. Spatial distance and lexical replacement. *Language* 62(1): 38-55.

CHEN, Linxia 陈琳霞, and Zi-ran He 何自然. 2006. Yuyan moyin xianxiang tanxi 语言模因现象探析 (Analysis of memes in language). *Waiyu jiaoxue yu yanjiu: waiguo yuwen shuangyuekan 外语教学与研究: 外国语文双月刊 [Foreign Language Teaching and Research (bimonthly)]* 38(2): 108-114.

DAWKINS, Richard. 1976. *The Selfish Gene*. New York: Oxford University Press.

DAVISON, Patrick. 2012. The language of internet memes. *The Social Media Reader*: 120-134.

DENNETT, Daniel C. 1991. *Consciousness Explained*. New York: Little Brown and Company.

DE SAUSSURE, Ferdinand. 2011. *Course in General Linguistics*. New York: Columbia University Press.

EDMONDS, Bruce. 2002. Three challenges for the survival of memetics. *Journal of Memetics - Evolutionary Models of Information Transmission* 6: 45-50.

ELLIS, Nick C., and Diane Larsen-Freeman. 2006. Language emergence: Implications for applied linguistics—Introduction to the special issue. *Applied Linguistics* 27: 558-589.

FAGYAL, Zsuzsanna, Samarth Swarup, Anna María Escobar, Les Gasser, and Kiran Lakkaraju. 2010. Centers and peripheries: Network roles in language change. *Lingua* 120: 2061-2079.

FIŠER, Darja, and Nikola Ljubešić. 2019. Distributional modelling for semantic shift detection. *International Journal of Lexicography* 32(2): 163-183. https://doi.org/10.1093/ijl/ecy011.

FISCHER, Roswitha. 1998. Reviewed work: Word: Proceedings of an International Symposium, Lund 25-26 August 1995 by Jan Svartvik. *AAA: ArbeitenAusAnglistik Und Amerikanistik* 23(1): 91-94.

FRANCIS, Nelson W., and Henry Kucera. 1979. *Brown Corpus Manual - Manual of Information to Accompany A Standard Corpus of Present-Day Edited American English*. Technical Report. Department of Linguistics. Brown University.

FREEMAN, Diane Larsen, and Lynne Cameron. 2008. Research methodology on language development from a complex systems perspective. *The modern language journal* 92: 200-213.

GIL-WHITE, Francisco J. 2005. Common misunderstandings of memes (and genes): The promise and the limits of the genetic analogy to cultural transmission processes. In *Perspectives on Imitation: From Neuroscience to Social Science*, Vol. 2, eds. by Susan Hurley, and Nick Chater, 317-338. Cambridge MA: MIT Press.

GREENBERG, Mark. 2005. Goals versus memes: Explanation in the theory of cultural evolution. In *Perspectives on Imitation: From Neuroscience to Social Science*, Vol. 2, eds. by Susan Hurley, and Nick Chater, 339-353.Cambridge MA: MIT Press.

HACKEN, Pius ten. 2020. Norms, new words, and empirical reality. *International Journal of Lexicography* 33(2): 135–149. https://doi.org/10.1093/ijl/ecaa005.

HACKEN, Pius ten, and Maria Koliopoulou. 2020. Dictionaries, neologisms, and linguistic purism. *International Journal of Lexicography* 33(2): 127-134.

He, Zi-ran 何自然. 2005. Yuyan zhong de moyin 语言中的模因 (Memes in language). *Yuyan kexue 语言科学 [Linguistic Sciences]* 6: 54–64.

———. 2008. Yuyan moyin jiqi xiuci xiaoying 语言模因及其修辞效应 (Linguistic memes and their rhetoric effects). *Waiyu xuekan 外语学刊 Foreign Language Research* 1(6): 68-73.

———. 2014. Liuxingyu liuxing de moyinln jiedu 流行语流行的模因论解读 (Memetic understanding of language in fashion). *Shandong waiyu jiaoxue 山东外语教学 [Shandong Foreign Languages Journal]* 35(2): 7-13.

HEYLIGHEN, Francis, and Klass Chielens. 2009. Cultural evolution and memetics. In *Encyclopedia of Complexity and Systems Science*, ed. by Robert A. Meyers, 3205-3220. Berlin, Germany: Springer.

Holland, John H. 1992. Complex adaptive systems. *Daedalus* 121(1): 17-30.

———. 1996. *Hidden Order: How Adaptation Builds Complexity*. Addison Wesley Longman Publishing Co., Inc..

HULL, David L. 1982. The naked meme. In *Learning, Development and Culture: Essays in Evolutionary Epistemology*, ed. by Henry C. Plotkin, 273-327. New Jersey: Wiley.

———. 2000. Taking memetics seriously: Memetics will be what we make it. In *Darwinizing Culture: The Status of Memetics as a Science*, ed. by Robert Aunger, 43-67. New York: Oxford University Press.

JIANG, Menghan, Xingying Shen, Kathleen Ahrens, and Chu-Ren Huang. 2021. Neologisms are epidemic: modeling the life cycle of neologisms in China 2008-2016. *PLoS ONE* 16(2): e0245984.

JING-SCHMIDT, Zhuo, and Shu-Kai Hsieh. 2019. Chinese neologisms. In *The Routledge Handbook of Chinese Applied Linguistics*, eds. by Chu-ren Huang, Zhuo Jing-Schmidt, and Barbara Meisterernst, 514-534. Abingdon, UK: Routledge.

KERREMANS, Daphné. 2015. A web of new words. A corpus-based study of the conventionalization process of English neologisms. *English and American Studies in German* 5(1): 8-10.

KLOSA-KÜCKELHAUS, Annette, and Sascha Wolfer. 2020. Considerations on the acceptance of German neologisms from the 1990s. *International Journal of Lexicography* 33(2): 150–167. https://doi.org/10.1093/ijl/ecz033.

KRONFELDNER, Maria. 2014. *Darwinian Creativity and Memetics*. London and New York: Routledge.

LABOV, William. 1966. The effect of social mobility on linguistic behavior. *Sociological Inquiry* 36(2): 186-203.

LEI, Siyu, Ruiying Yang, and Chu-Ren Huang. 2021. Emergent neologism: A study of an emerging meaning with competing forms based on the first six months of COVID-19. *Lingua* 258:103095.

LI, Juan 李娟. 2019. Jiyu moyinlun shijiao xia wangluo liuxingyu de yanjiu - yi jinwunian yaowenjiaozi shida liuxingyu weili 基于模因论视角下网络流行语的研究 -- 以近五年《咬文嚼字》十大流行语为例 (A study on popular network language from the perspective of memetics -- An example of ten buzzwords of Yao Wen Jiao Zi in the past five years). *Heilongjiang jiaoyu xueyuan xuebao 黑龙江教育学院学报 [Journal of Heilongjiang College of Education]* 38(9): 122-125.

LI, LongXing, Chu-Ren Huang, and Vincent Xian Wang. 2020. Lexical competition and change: A corpus-assisted investigation of GAMBLING and GAMING in the past centuries. *SAGE Open* 10(3): 2158244020951272.

LIU, Yi 刘懿. 2014. Liuxingyu moyin de leixing yu tedian - yi jinwunian yaowenjiaozi shida liuxingyu weili 流行语模因的类型与特点 -- 以近五年《咬文嚼字》"十大流行语"为例 (Types and characteristics of buzzword memes: An example of ten buzzwords of Yaowen Jiaozi in past five years). *Guangzhou daxue xuebao 广州大学学报 (社会科学版) [Journal of Guangzhou University (Social Science Edition)]* 13(4): 64-69.

MADSEN, Anders Koed. 2015. Between technical features and analytic capabilities: Charting a relational affordance space for digital social analytics. *Big Data & Society*, 2(1), 2053951714568727.

MARELLO, Carla. 2020. New words and new forms of linguistic purism in the 21st century: the Italian debate. *International Journal of Lexicography* 33(2): 168-186.

MARINO, Gabriele. 2015. Semiotics of spreadability: A systematic approach to Internet memes and virality. *Punctum: International Journal of Semiotics* 1: 43-66.

METCALF, Allan. 2004. *Predicting New Words: The Secrets of their Success*. Boston, New York: Houghton Mifflin Harcourt.

MOORE, David S., William I. Notz, and Michael A. Flinger. 2017. *The basic practice of statistics*. New York, NY: W. H. Freeman.

OGILVIE, Sarah. 2008. Rethinking burchfield and world Englishes. *International Journal of Lexicography* 21(1): 23–59. https://doi.org/10.1093/ijl/ecn005.

OGURA, Mieko, and William S-Y. Wang. 1994. Snowball effect in lexical diffusion: the development of -s in the third person singular present indicative in English. In *English historical linguistics 1994: papers from the 8th International Conference on English Historical Linguistics* (8. ICEHL, Edinburgh, 19–23 September 1994), ed. by Derek Britton, 119-141. Amsterdam: John Benjamins.

OXFORD UNIVERSITY PRESS. 2022. *Oxford Learner's Dictionary*. Retrieved October 28, 2022, from https://www.oxfordlearnersdictionaries.com/

PAGEL, Mark. 2009. Human language as a culturally transmitted replicator. *Nature Reviews Genetics* 10: 405–415.

POULSHOCK, Joseph. 2002. The problem and potential of memetics. *Journal of Psychology and Theology* 30(1): 68-80.

RENOUF, Antoinette. 2007. Tracing lexical productivity and creativity in the British Media. In *Lexical Creativity, Texts and Contexts*, ed. by Judith Munat, 61-89. Amsterdam: John Benjamin.

———. 2013. A finer definition of neology in English: The life-cycle of a word. *Corpus Perspectives on Patterns of Lexis* 57: 177.

REY, Alain. 2005. The concept of neologism and the evolution of terminologies in individual languages, translated by Juan C. Sager. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 11(2): 311-331.

SCHMID, Hans-Jörg. 2008. New words in the mind: Concept-formation and entrenchment of neologisms. *Anglia* 126(1): 1-36.

SONNAD, Nikhil. 2015. How brand-new words are spreading across America. *Quartz*. available July 30, 2015. https://qz.com/465820/how-brand-new-words-are-spreading-across-america/

THORNTON, Brett F., and Shawn C. Burdette. 2017. The neodymium neologism. *Nature Chemistry* 9: 194.

TSUR, Oren, and Ari Rappoport. 2015. Don't let me be# misunderstood: Linguistically motivated algorithm for predicting the popularity of textual memes. In *The 2015 International AAAI Conference on Web and Social Media (ICWSM)*, 426-435.

WANG, Shan, Ruhan Liu, Chu-Ren Huang. 2022. Social changes through the lens of language: A big data study of Chinese modal verbs. *PLoS ONE* 17(1): e0260210.

Wei, Dai-xiang 魏代香. 2011. Xinciyu ciqun de moyin shijiao yanjiu 新词语词群的模因视角研究 (The memetic analysis on Chinese neologism clusters). *Bijie xueyuan xuebao 毕节学院学报 [Journal of Bijie University]* 3 (29): 95-98.

WIKIPEDIA. 2020. CNKI. Latest modification 17 November 2020. https://en.wikipedia.org/wiki/CNKI

YAO WEN JIAO ZI EDITORIAL OFFICE 咬文嚼字编辑部. 2009. 2008 Niandu shida liuxingyu 2008 年度十大流行语 (Top ten neologisms of 2008). *Yao Wen Jiao Zi 咬文嚼字* 2: 4-6.

———. 2010. 2009 Niandu shida liuxingyu 2009 年度十大流行语 (Top ten neologisms of 2009). *Yao Wen Jiao Zi 咬文嚼字* 2: 4-6.

———. 2011. 2010 Niandu shida liuxingyu 2010 年度十大流行语 (Top ten neologisms of 2010). *Yao Wen Jiao Zi 咬文嚼字* 2: 4-6.

———. 2012. 2011 Niandu shida liuxingyu 2011 年度十大流行语 (Top ten neologisms of 2011). *Yao Wen Jiao Zi 咬文嚼字* 2: 4-6.

———. 2013. 2012 Niandu shida liuxingyu 2012 年度十大流行语 (Top ten neologisms of 2012). *Yao Wen Jiao Zi 咬文嚼字* 2: 4-6.

———. 2014. 2013 Niandu shida liuxingyu 2013 年度十大流行语 (Top ten neologisms of 2013). *Yao Wen Jiao Zi 咬文嚼字* 2: 4-6.

———. 2015. 2014 Niandu shida liuxingyu 2014 年度十大流行语 (Top ten neologisms of 2014). *Yao Wen Jiao Zi 咬文嚼字* 2: 4-6.

———. 2016. 2015 Niandu shida liuxingyu 2015 年度十大流行语 (Top ten neologisms of 2015). *Yao Wen Jiao Zi 咬文嚼字* 2: 4-6.

———. 2017. 2016 Niandu shida liuxingyu 2016 年度十大流行语 (Top ten neologisms of 2016). *Yao Wen Jiao Zi 咬文嚼字* 2: 4-6.

ZHANG, Jianhua. 2017. Internet corpora-based diachronic research on the diffusion of Shanzhai through the internet. *Journal of Quantitative Linguistics* 24(2-3): 197-212.

流行语是模因吗？
基于互联网用户数据的流行语发展周期量化模型

蒋梦晗[1], 安可思[2], 沈翔瀛[3]，李逸薇[2]，黄居仁[2]
[1]深圳北理莫斯科大学 [2]香港理工大学 [3]南方科技大学

摘要
流行语反映了我们生活中新思想或新观念的引入，并在文化传播和人类语言的发展中发挥着重要作用。特别是在过去的二十年里，互联网的发展和社交媒体的普及为流行语的爆炸式增长提供了土壤。在本文中，我们通过考察关注流行语如何随时间变化-（即流行语的生命周期）而考察流行语是如何产生以及发展的。通过将流行语比拟为生命体的发展，而不是理解为语言中的既成事实，我们希望能够更好地了解流行语实质的传播模式，从而能够更好地预测流行语的发展。为了实现这一目标，我们基于模因模型研究了流行语的生命周期，并将其与最近研究

的流行病模型进行了比较。我们选取了 2008-2016 年间 90 个最具影响力的汉语流行语，并根据谷歌趋势展现的互联网实际使用情况数据，对新词的发展趋势进行了纵向建模。我们的研究验证了模因模型可以在其周期的早期（即上升阶段）可靠地描述和预测新词的生命周期，但无法预测其发展的拐点以及整个生命周期。我们因此得出结论，模拟词汇的传播需要两个模型：初始阶段的模因模型和后期阶段（尤其是拐点处）的流行病模型。运用这种两阶段/两模型的方法，有助于在发展初期识别流行语为潜在的新词汇。这是因为模因模型在数学形式上更简单，有利于编写适用于流行语发展初期阶段的过滤程序。

关键词
语言模型　**模**因模型　**病**毒模型　**流**行语