Emerging Chinese-English Hybridized Internet Neologisms: a Big Data Study Based on Baidu Index

Huayong Li

Foreign Language School, Sichuan University of Arts and Science, Sichuan Province, People’s Republic of China
E-mail: lihuayong2008@126.com

Received: January 16, 2020   Accepted: February 20, 2020   Online Published: March 7, 2020
doi:10.11114/ijecs.v3i1.4688

Abstract

This article studies the emerging neologisms in the Chinese cyber context: 笑叫, 打住, 话语, which are termed Chinese-English hybridized internet neologisms (CEHINs). CEHINs share the features of Chinese words and English words. Based on the research tool called Baidu Index, it is found: (1) the structure and pronunciation of CEHINs (笑叫, 打住, 话语) is different from the ‘pure’ Chinese or English internet neologisms; (2) the diffusion of them is Chinese netizen-driven, not the meme-driven; (3) the top users of them are usually aged between 30-39, while the ‘pure’ Chinese internet neologisms are much more frequently used by netizens aged 20-29; and (4) the popularization of internet, fast cultural communications in cyber times, and the further pervasiveness of English as a Lingua Franca in China are the reasons for the emergence and popularity of them in China.

Keywords: Neologism, English word, Chinese character, cyber context

1. Introduction

With the popularization of the internet, and the pervasiveness of English in China, a new type of internet neologism is emerging in recent years: such as 笑叫 (pinyin: da call, ‘cheer for’), 打住 (pinyin: hold zhu, ‘control’), 话语 (pinyin: word ge, ‘amazing’, ‘my brother’). These neologisms are formed by a Chinese character plus an English word, or an English word plus a Chinese character. Hence, the author terms them Chinese-English hybridized internet neologism (CEHIN). They are frequently and widely used by Chinese netizens: 笑叫 is the top one annual Chinese internet neologism in 2017, 打住 is the top one annual Chinese internet neologism in 2011, and 话语 is the ninth top annual Chinese internet neologism in 2016, according to the report from official weibo of the journal 语文教子. Through the lens of internet neologisms we can observe the behaviour of netizens, the thought change of the netizens, the internet culture change and even the Zeitgeist of an era so that the internet neologisms have attracted the eyes of the scholars (Roigmarín, 2016; Zhou & Shu, 2017; Zhang, 2017; Xu & Tian, 2017).

However, the scholars have not paid much attention to the emerging CEHINs. One reason may be that they are emerging and used in recent years: 笑叫 (initially in May, 2017), 话语 (initially in August, 2016), 打住 (initially in September, 2011). These neologisms are worth being studied in details for the following reasons: (1) they show the blending of English words and Chinese characters in cyber times in China; (2) their structure is different from the typical ‘pure’ Chinese or English internet neologisms that are completely formed by Chinese character or English letters; (3) they show the creativity and ways of thinking of Chinese netizens; and (4) they are frequently and widely used by Chinese netizens.

There is a trend toward internet neologism research with the empirically reliable methods, such as corpus linguistic method, big data or quantitative linguistic method. The data-based method is helpful to reveal the authentic using status and the intrinsic characteristics of the neologisms, discovering what could not be noticed previously or what could not be studied by the traditional methods. Consequently, the author conducts the research to 笑叫, 打住, 话语, based on the big data retrieval tool called Baidu Index.

2. Research Questions

Based on the Baidu Index, the author tries to answer the following three questions related to the CEHINs 笑叫, 打住, 话语: (1) what are the features of them? (2) why are they emerging? and (3) what are the prospects for them?
3. Research Methodology

3.1 Research Tool

The author is based on Baidu Index in this research. It is developed by Baidu, Inc., the second biggest search engine company in the world. Baidu search engine holds 76.05% market shares in China’s search engine market for 0.751 billions netizens in 2017, according to The 40th China Statistical Report on Internet Development.6 Baidu Index is regarded as one of the most important analysis tools in the internet times and even the data times7 so that it can relatively much more accurately reveal or represent the authentic behaviours and needs of Chinese netizens. Baidu Index, which is based on the tremendous big data of Baidu search engine, not only analyse the ‘search heat’ degrees of the keywords but also mine the public opinion information, the market demands, the user features, and so on.8 It is updated everyday and provides the data and analysis on the pc net and mobile net. For this research, Baidu Index can quickly and automatically provide the following retrieval information related to the three neologisms ʃʃ call, hold ʃʃ, word ʃʃ: the users, the behaviors of netizens, the diachronic and synchronic frequency, the gender and age of netizens, the distributional areas, and so on, which can substantially answer the previous research questions. Therefore, Baidu Index can be viewed as a tremendous big data based research tool and market demand analysis tool.

There are advantages of big data based method: it can reflect some intrinsic features of languages, one of which is the probability of language (Bod, et al., 2003); it can also reveal more language facts, usage patterns, the diachronic trend, the nature of language and so on, which are impossible to find under the small linguistic data, or introspection data. Consequently, based on Baidu Index, the paper tries to observe and analyse the origin, the diachronic and synchronic features of the neologisms ʃʃ call, hold ʃʃ, word ʃʃ, which are helpful to study the features of the emerging CEHINs.

3.2 Object of Research

The research object is the three CEHINs ʃʃ call, hold ʃʃ, word ʃʃ, focusing on the features, reasons of emergence, and the prospects of them. The reasons for selecting the three neologisms are: (1) they are respectively listed as one of the annual top ten Chinese internet neologisms in 2017 (ʃʃ call), 2016 (word ʃʃ), 2011 (hold ʃʃ), which is released by the official weibo of the Chinese journal called yaowen jiaozi; (2) they are typical CEHINs; (3) they are frequently and widely used by Chinese netizens during cyber communications; and (4) they emerged in different years, which is helpful for the diachronic and synchronic study to CEHINs.

3.3 Procedures

The author mainly conducts this research based on Baidu Index in the following steps:

(1) Analyse the structure and semantic features of the three neologisms ʃʃ call, hold ʃʃ, word ʃʃ. In order to study the structure and semantic features of CEHINs, the author analyses the components of the three neologisms based on the word frequency from the List of Lexicon of Common Words Contemporary Chinese (draft) (the Lexicon of Common Words Contemporary Chinese Project Team, 2008) and the corpus created by Mark Davies.9

(2) Analyse the using features of CEHINs. The author retrieves the neologism ʃʃ call on Baidu Index, focusing on initial searching status, user’s distribution in cities, user’s age and gender distribution, which can be automatically and objectively obtained in the ‘search trend’ function and the ‘user’s portrait’ function of Baidu Index9. In order to observe the life stage of CEHINs, the author also retrieves hold ʃʃ, word ʃʃ in Baidu Index. Based on the above retrieval results, the using features of CEHINs can be summarized and explained.

(3) Analyse the reasons for the emergence of CEHINs. Based on the retrieval results from Baidu Index and other websites, the author analyses the reasons why the CEHINs are emerging and popular.

(4) Analyse the prospects for CEHINs.

4. Features of CEHINs

4.1 Structure and Semantic Features of CEHINs

The emerging CEHINs, such as ʃʃ call, hold ʃʃ, word ʃʃ, 太(out), 他(low), take on the four structure features. (1) The component words of them are single syllable. Regardless of English words or Chinese characters in the neologisms, they are all single syllable, such as ʃʃ /dá/, or call /kɔːl/ in the neologism of ʃʃ call. (2) The component words of them is high frequency words. For example, Chinese characters ʃʃ (409)10, ʃʃ (250), 太 (237), 他 (40) are within the top 500 out of 56008 Chinese words, according to the List of Lexicon of Common Words Contemporary Chinese (draft) (the Lexicon of Common Words Contemporary Chinese Project Team, 2008). Obviously, the English words, call, hold, out, low are within the top 300 high frequency words: call (122)10, hold (213), word (244). (3) The number of letters of English word or strokes of the Chinese character in ʃʃ call, hold ʃʃ, word ʃʃ is less: letters of the English word are no more than four, and the strokes of the Chinese character are no more than 9. The less letters or strokes in the hybridized internet neologisms conform to the economy principle of language, which is convenient for the use and diffusion of the
neologisms in Chinese cyberspace communication context. (4) They show the features of English and Chinese in pronunciation and meanings, containing two different kinds of phonetic sounds: for example, in the neologism \textit{ff call}, the first part \textit{ff} is pronounced /dǎ/ in pinyin of Chinese phonetic system while the second part \textit{call} is pronounced /kɔːl/ in English phonetic system; the meaning of \textit{ff call} is together produced by the components of the Chinese character \textit{ff} and the English word \textit{call}.

The meaning of Chinese-English hybridized internet neologisms gives the audience or users a sense of humor, fashion, or positive feeling when they are used in cyberspace communications in China. The three neologisms possess positive connotation and usually are used to express or construct positive semantic prosody in the cyberspace communications. Hence, they are frequently and widely used in Chinese netizens, which can be showed in the top ten annual Chinese internet neologisms in the passing decade: \textit{ff call} is top one in 2017; \textit{hold 住} is top one in 2011; \textit{word 哥} is the ninth top annual Chinese internet neologism in 2016.

4.2 Using Features of CEHINs

Firstly, they usually spread from ‘the bottom to the up’ that means they are initially used in the informal registers, and then in the much formal registers, even enter the quite formal register such as the mainstream media. They can truly reflect the conditions and the needs of majority common individuals: the netizens in China. The CEHINs are all initially used by Chinese netizens, especially in their cyber communications. With the increase of using frequency, they may spread out to much more formal registers, such as newspapers and TV programs, classrooms, the journals, and even the Chinese government websites. There is some evidence for this claim: (1) in CCTV (www.cctiv.com.cn) and China daily (www.people.com.cn), a great number of concordances on the three neologisms can be retrieved; (2) it can be seen that the neologisms \textit{ff call}, \textit{hold 住}, \textit{word 哥}, are also used in the classrooms by the teachers or students, and in the speeches of the much formal meetings; (3) the three neologisms have appeared in the China journals; and (4) in the quite official and formal website The Central People’s Government of the People’s Republic of China (www.gov.cn), some of the concordances on the three neologisms can also be retrieved, which suggests that these neologisms are accepted by the upper class or the elites.

Secondly, the initial diffusion of them is not based on the traditional so-called meme-driven. For example, when the hybridized neologism \textit{ff call} was coined in the internet, it was not diffused by ‘meme’. Because, initially, the Chinese netizens want to know the meaning and implicature of \textit{ff call}, it is impossible for them to copy, simulate and use it directly. The memetic theoretical explanation to neologisms considers the neologism to be diffused by meme at the whole life cycle, including at the initial stage. The author retrieves the \textit{ff call} in Baidu Index, focusing on the ‘needs of searching users’ of the first month. The retrieval steps are: (1) put the neologism \textit{ff call} into the Baidu Index ‘search box’ and search; and (2) in searching result page, click the ‘demand net’ button in the menu, and then Figure 1 is automatically generated. According to Figure1 it can be found that Baidu Index retrieval results of the hybridized internet neologism \textit{ff call} are against the memetic theoretical explanation to initial diffusion of neologisms: \textit{ff call} is initially searched for the meaning and the Chinese netizens use it, which means that it is not directly used by the netizens at the initial stage, while it is supposed to be used directly according to the memetic theory. From Figure 1, it can be seen that within the first month, almost all the user’s searching of \textit{ff call} is related to the meaning of \textit{ff call}, the action of \textit{ff call}, the course of \textit{ff call}, which are just to know the meaning and implicature of \textit{ff call}. After getting the background information of \textit{ff call}, Chinese netizens gradually use it in their cyber communications. So the author states that the initial diffusion of the Chinese-English internet hybridized internet neologisms \textit{ff call}, \textit{hold 住}, or \textit{word 哥}, are netizen’s need-driven, not the meme-driven.

![Figure 1. The first month searching for \textit{ff call}](image_url)
Thirdly, the using features of them are netizen-driven. Liu & Lin (2018) argue that language is a complex adaptive system driven by human. The neologism is a sub-system of the complex adaptive system. Hence, the author may deduce that the using features of internet neologism are netizen-driven. The author retrieves the distribution of 打call in Baidu Index: (1) put the neologism 打call into the Baidu Index ‘search box’ and search; (2) in searching result page, click the ‘user’s portrait’ button in the menu; (3) and then click the ‘city’ button in the ‘district distribution’ section; and (4) after that, Figure 2 can be automatically generated. It is found that Baidu Index trend of CEHIN 打call shows the obvious netizen-driven characteristics: the more Chinese netizens are in a district, the more frequencies of the word 打call in this district; there are several ‘using centres’ for the neologism of 打call. The top-three ‘using centres’ for the neologism of 打call are Beijing, Shanghai, Guangzhou, where according to the China Statistical Report on Internet Development, the popularization rate of netizens are the top three in China.

Fourthly, there is region limitation for the popularity of Chinese-English hybridized internet neologisms. Based on the Baidu Index retrieval results of the three neologisms, the author finds the CEHINs which are popularly used in Chinese mainland are not used in Hong Kong, Taiwan, and Macao. Though the internet can eliminate the barricades of space and geography, and the communications in the internet are quite convenient and easy, the Chinese-English hybridized internet neologisms 打call, hold住, word 哥 are not used by the netizens in Hongkong, Taiwan, and Macao. From Figure 4 it can be found that the users of the ‘pure’ Chinese internet neologism 图样图森破 are different: the top one group users are aged between 20-29 years old, which account for 37%; the proportion of male users of 图样图森破 is obviously larger than that of female users; surpassing 56%, which means that there is palpable gender difference in using the ‘pure’ Chinese internet neologism 图样图森破. More details of age difference in using the Hybridized and the ‘pure’ neologisms can be seen in Figure 3 and Figure 4.
4.3 Life Stage of CEHINs

The CEHINs usually reaches the peak within short time because of certain events, and goes down quickly, and then keeps stable. Scholars have studied the life stage of neologisms. Based on the structural, social-pragmatic, and cognitive factors, Schmid (2008) put forward three stages: creation, consolidation, and establishing. Quantitative research on neologisms has also made great progress in statistical understanding of neologism life-stage: originated, evolved, die out (Zhuo & Shu, 2018). Quantitative research views the word frequency as a leading factor in explaining life-stage and the prediction force of whether a neologism may survive or not. The frequency is significant to observe and understand the evolution features of CEHINs. The author retrieves the diachronic process of the neologism *打call*: (1) put the neologism *打call* into the ‘search box’ of Baidu Index; (2) click the button ‘all’ in the ‘whole trend’ section; and (3) Figure 5 can be automatically generated. In the same steps, Figure 6 and Figure 7 can be automatically generated. Figure 5 shows the origin, increasing, peak and decreasing of *打call*. There are no more than two months for *打call* to reach its peak. *打call* initially appeared in April, 2017 so the time span of *打call* is limited and is inadequate to observe or describe the life stage of CEHINs. The author retrieves another neologism *hold住* which appeared in August, 2011, see Figure 6 which shows that the neologism *hold住* was suddenly popular among the Chinese netizens, in August, 2011 because of the Taiwan TV program called 大学生了没 (Pinyin: ‘da xuesheng le mei’) on 9th August, 2011. Within a week (from 21th August, 2011 to 28th August, 2011), the frequency of neologism hold住 reached its peak (40,000 hits) and then dramatically decreased, and after 11th November, 2011, it kept stable. From 2012 to 2017, the frequency of hold住 is relatively stable, 251 times per week. Similarly, there is also a peak for the neologism *word哥*, and then goes down quickly, see Figure 7.

Based on Figure 6 and Figure 7, it can be seen that the diachronic tendency of hold住 and word哥 is quite similar. They are continually used by the Chinese netizens and gradually diffuse in other registers, such newspaper, TV program, and so on. Diachronically speaking, the CEHINs *打call, hold住, word哥* are welcomed by the Chinese netizens because of theirs structure features and humorous semantic, which are elaborated in previous ‘structure and semantic features of CEHINs’ section.
5. Reasons for Emergence of CEHINs

The emerging CEHINs in recent years are resulted from the popularization of internet, the frequent cultural communications in the internet, and the further blending of English and Chinese in cyber times.

5.1 Popularization of Internet

The popularization of internet in China is an important reason for the emergence of Chinese-English hybridized internet neologisms. There are 0.751 billion netizens in 2017, and the popularization rate of internet is 54.3%\(^{\text{xiii}}\). The ‘mobile phone’ netizens in China account for 96.3%. The fast development of internet in China makes the work, learning and communications much more convenient, which also promotes the creativity to meet the diverse individual needs of Chinese netizens. The CEHINs initially emerged in 2011, and the number is gradually increasing, with the increase of Chinese netizens. Before 2011, there are no CEHINs used in China mainland cyber world. The author can say that the CEHINs *call*, *hold*，*word* are one of the products of popularization of internet in China.

5.2 Cultural Communications

The CEHINs are one of the products of cultural communications. According to related materials, the original concept of
the hybridized neologisms \textit{call}, \textit{hold} are respectively from Japan and Taiwan. The concept \textit{call} initially appeared in Japan\textsuperscript{xiv} and was introduced into Chinese cyber world, being popular among Chinese netizens: one of the top ten neologisms in 2017. The origin of \textit{hold} was from Taiwan in 2011 but was not hot and popular in Taiwan cyber worlds\textsuperscript{xv}. It was borrowed by Chinese netizens and was quickly popular in the Chinese netizens, being the top one internet neologism of 2011\textsuperscript{xvi}. It can be seen that the neologisms \textit{call}, \textit{hold} are the result of netizens’ communications, especially the cultural communications. If the Chinese netizens have not accepted Japanese-related ‘cheer for’ culture, the hybridized neologism \textit{call} would have not been popular in Chinese netizens. If the Chinese netizens have not watched the Taiwan TV program called ‘大学生了沒’ on 9th August, 2011 and enjoyed it, the hybridized neologism \textit{hold} would have not been the top one internet neologism in 2011, which continues to be used now.

5.3 Blending of English and Chinese

The emerging CEHINs show a special blending pattern of English and Chinese. The hybridized neologism, such as \textit{call}, usually consists of part of English word and part of Chinese character: there are two kinds of spelling systems (e.g. \textit{call} is Chinese spelling system and call is English system.) and pronunciation systems (\textit{call} is pronounced according to Chinese pinyin, and call is pronounced according to the English pronunciation system.) within a Chinese-English hybridized internet neologism, which represents the deep interactive assimilation of Chinese and English in the lexical level in cyber times. The CEHINs are the result of interactive assimilation of English and Chinese, which is promoted by internet and the creativity of Chinese netizens. The rising popularization rate of English in China is another important factor resulting in the emergence of CEHINs. Over a decade, the English level and English popularization rate in China are rising, so that Chinese netizens who can speak or use English are increasing. It is naturally that the English-peaking Chinese netizens coin and tend to use the CEHINs during the cyber communications in China. Consequently, there are more and more CEHINs, which are popular in Chinese netizens.

It is impossible produce the CEHINs in China before 2000: one reason is that the proportion of Chinese who can speak or read English is small; another reason is that the popularization rate of internet is low, which limits the creativity of netizens and the convenience of communications.

In a word, the CEHINs not only show the characteristics of Chinese words but also show the characteristics of English words. They represent a new blending and assimilation of the two different languages in cyber times, which is driven by the Chinese netizens. They also enrich the vocabulary of Chinese English, which is regarded as a ‘developing variety of English’ (Xu, 2010, Xu et al, 2017).

5.4 Prospects for CEHINs

Though CEHINs are emerging in recent years, the number of the neologisms is increasing and they are welcomed by Chinese netizens. Besides the three CEHINs, there are CEHINs ‘fashionable’, ‘low character’, ‘out’ (‘unfashionable’), ‘out’ (‘behind the times’), and so on, which are used in Chinese cyberspace. One of the features of CEHINs is that most of them are within the annual top ten Chinese internet neologisms and they are welcomed by Chinese Netizens. For example, the hybridized internet neologism \textit{hold} is the top one Chinese internet neologism\textsuperscript{xviii} in 2011, and \textit{call} is top one Chinese internet neologism\textsuperscript{xviii} in 2017, and \textit{word}, is the ninth top annual Chinese internet neologism\textsuperscript{xviii} in 2016. The life span of CEHINs is much longer than the other type of Chinese internet neologisms. For instance, Xu and Tian (2017) stated that over time, most of the English lettered words in Chinese have died out. It is true that the majority of the Chinese internet neologisms could not stand the test of time, abandoned by the Chinese netizens and disappeared suddenly. However, the trend of the three neologisms \textit{call}, \textit{hold}, \textit{word}, which are retrieved on Baidu Index, indicates that the CEHINs might stand the test of time and continue to be popular in the netizens, because of the special structures, special semantic features and the special communication effects. Seen from Figure 6, the hybridized neologism \textit{hold} has been used more than seven years and will be continually used by Chinese netizens.

The further rising of the popularization rate of the Internet, the development of the globalization in economies and cultures, and the pervasiveness of English as a Lingua Franca in China, will accelerate the produce and use of the CEHINs. The CEHINs show and reveal the new trend of the blending of Chinese and English, the blending of cultures, and the blending of ways of thought. It is valuable to study them. The author believes that the CEHINs which stand the test of time and expand the using registers will catch the attention of scholars and official institutions. The English-Chinese lexicographers might consider the status of CEHINs, for the high using frequency and a great many users, the structure and semantic features, and the impact on the net language in China. However, whether or not the CEHINs can enter the dictionaries is an issue that time can give the answer to.
6. Conclusion

The emerging CEHINs are one of products of the blending of Chinese and English on the lexical level in cyber times. They share the features of Chinese words and English words so that once they are coined, they are welcomed by the Chinese netizens in net communications, being the annual top neologisms in China. Based on Baidu Index, the author finds: (1) the structure and pronunciation of CEHINs (打call, hold住, word哥) is different from the ‘pure’ Chinese or English internet neologisms; (2) the diffusion of them is Chinese netizen-driven, not the meme-driven; (3) the top users of them are usually aged between 30-39, while the ‘pure’ Chinese internet neologisms are much more frequently used by netizens aged 20-29; and (4) the popularization of internet, fast cultural communications in cyber times, and the further pervasiveness of English as a Lingua Franca in China are the reasons for the emergence and popularity of them in China. The number of CEHINs are increasing and they will have more influence on the Chinese net language, even Chinese.

Acknowledgement

This article is supported by Education Management Section in Sichuan University of Arts and Science [Grant No.: 2020JY013].

References


Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).

Note:

See http://blog.sina.com.cn/s/blog_4c0596f60102x0mx.html
See http://baidu.com/
See http://www.cnnic.net.cn/hlwzjy/hlwxzbg/hlwjbg/201708/P20170807351923262153.pdf
See https://baike.baidu.com/item/%E7%99%BE%E5%BA%A6%E6%8C%87%E6%95%B0/1 06226 ?fr=aladdin
See http://index.baidu.com/Helper/?tpl=help&word=
See http://www.wordandphrase.info/frequencyList.asp
See https://baike.baidu.com/item/%E7%99%BE%E5%BA%A6%E6%8C%87%E6%95%B0/1
The number in the brackets are the ranks of the frequency of Chinese characters in the *List of Lexicon of Common Words Contemporary Chinese (draft)*.

*See* [https://www.wordandphrase.info/frequencyList.asp](https://www.wordandphrase.info/frequencyList.asp)

The number in the brackets are the ranks of the frequency of English words according to the word frequency list of 60,000 words.

*See* [mall.cnki.net/magazine/articlesearchnew.aspx?keys=call](mall.cnki.net/magazine/articlesearchnew.aspx?keys=call)

*See* [http://www.cnnic.net.cn/hlwzdyj/hlwz/bj/201701/P020170122064311?fr=aladdin](http://www.cnnic.net.cn/hlwzdyj/hlwz/bj/201701/P020170122064311?fr=aladdin)

*See* [https://baike.baidu.com/item/%E6%89%93call/22064311?fr=aladdin](https://baike.baidu.com/item/%E6%89%93call/22064311?fr=aladdin)


*See* [http://tech.sina.com.cn/roll/2017-12-20/doc-ifypxmsq8368945.shtml](http://tech.sina.com.cn/roll/2017-12-20/doc-ifypxmsq8368945.shtml)