

Study on Linguistic Features of Chinglish in Web Text

Bo Chen, Chen Lyu and Ziqing Ji
Guangdong University of Foreign Studies
No.2, Baiyundadaobei Road, Baiyun District,
Guangzhou, China

cb9928@gmail.com, lvchen1989@gmail.com, 1163539719@qq.com

Received July 2017; revised March 2018

ABSTRACT. *Chinglish in Web Text is one of burgeoning language phenomena. This paper analyzes the linguistic features of Chinglish in Web Text from three aspects: phonetic, vocabulary and grammar. The phonetic characteristics are mostly reflected on the replacements and combinations of syllables. The word formations include Affixation method, blending method, compound method, homophonic method and embedded method. The grammatical characteristics include POS transfer, mainly with simple sentence and exclamation sentence, and five types of translation methods. The study is helpful to natural language processing.*

Keywords: Chinglish in web text, linguistic characteristics, phonetic, vocabulary, grammar

1. Introduction. Chinglish is a kind of Chinese-style English with Chinese vocabulary, Chinese grammar and Chinese expression of habit. Chinglish in Web Text (CWT), the combination of the network and Chinglish created by netizens and widely disseminated on the network, is a new lingual phenomenon of Chinglish. For example: “*gelivable*”(awesome) and “*ungelivable*” created from Chinese word “给力”(pinyin: *geili*), “*zhuangbility*”(pretentious, ostentatious) and the new synonym “*drunbility*” created from Chinese word “装逼”(pinyin: *zhuangbi*). As a polysemous or polymorphic phenomenon that possesses a variety of forms including characters, words, phrases, clauses, sentences [1-4], CWT has brought difficulties for Natural Language Processing (NLP).

Currently in linguistic fields, systematic research on CWT is not abundant. Wang studied this phenomenon in microblogging text, and applied it to machine translation system [5].

This paper systematically studies CWT based on the web text in the past three years,

constructs an open small-scale CWT corpus, and analyzes its linguistic features for machine translation, sentiment analysis and other related natural language processing areas.

2. Phonetic characteristics of CWT. The phonetic characteristics of CWT are mostly reflected on the replacements and combinations of syllables. On the basis of existing words, new CWT are created by replacing and combining syllables.

2.1. The Replacement of English Words' Syllables. The replacement of English words' syllables often refers to replacing the original tone with a homophone. There are mainly three types of replacements: the replacement of English syllables, the replacement of syllables in CWT and the replacement of Chinese syllables.

2.1.1. The Replacement of English Syllables.

(1) *together*+*gay* = *togayther*

The pronunciation of “*gay*” and the “*ge*” in “*together*” is similar. By replacing “*ge*” with “*gay*”, a new CWT word “*togayther*” is created.

(2) *sex*+*secretary* = *sexretary*

The pronunciation of “*sex*” and the “*sec*” in “*secretary*” is similar. By replacing “*sec*” with “*sex*”, a new CWT word “*sexretary*” is created.

(3) *shit*+*citizen* = *shitizen*

The pronunciation of “*shit*” and the “*cit*” in “*citizen*” is similar. By replacing “*cit*” with “*shit*”, a new CWT word “*shitizen*” is created. “*Shitizen*” often means the residents who doesn't possess much power, the people who are mediocre.

(4) *stupid*+*pig* = *stupig*

In (1) - (4), the pronunciation of “*pig*” and the “*pid*” in “*stupid*” is similar. By replacing “*pid*” with “*pig*”, a new CWT word “*stupig*” is created. “*Stupig*” has a negative meaning which means the people who are stupid and obese.

2.1.2. The Replacement of Syllables in Chinglish Words.

(5) *drunb*(the homophonic of *zhaungbi*)+*ity*(*suffix*)=*drunbility*(*pretentious*)

Based on the creation of the CWT word “*zhuangbility*”, netizens replaced “*zhuangbi*” with “*drunb*” by utilizing the similarity in their pronunciation and created a new CWT word “*drunbility*” that has the same meaning as the Chinese word “装逼”.

(6) *newb*(the homophonic of *niubi*)+*ity*(*suffix*)=*newbility*(*great*)

Based on the creation of the CWB word “*niubility*”, netizens replaced “*niubi*” with “*newb*” by utilizing the similarity in their pronunciation and created a new CWT word “*newbility*” that has the same meaning as the Chinese word “牛逼”.

2.1.3. The Replacement of Syllables in Chinese Words.

(7) **Z+turn** = Z-turn(*zhe teng*, Chinese pinyin of 折腾, *turn from side to side; toss about*)

The two pinyin of the Chinese word “折腾” are “*zhe*” and “*teng*”, which are similar to the pronunciation of “*Z*” and “*turn*”. So netizens decided to replace “折腾” with “*Z-turn*”.

Besides, netizens also use “no Z-turn” to express the intention of “不折腾” put forward by State leaders in China.

(8) *Don*'t+train = don'train (*dong che*, Chinese pinyin of 动车, bullet train)

The two syllables of the Chinese word “动车” are “dong” and “che”, which are similar to the pronunciation of “don't” and “train”. So netizens decided to replace “动车” with “don'train”. As “don't” means no in English, “don'train” also has a meaning that expresses the citizens in satisfaction of the expensive train tickets.

2.2. The combination of syllables. The homophonic replacement is to replace a part of a syllable with a similar pronunciation syllable. Unlike the replacement of homophonic syllables, the combination of syllables is a combination of two completely different syllables. There are two categories, as shown in (5)-(13).

2.2.1. English syllables+English syllables.

(9) *Vegetable* + *steal* = vegeteal (*to steal Vegetable*)

(10) *Smile* + *silence* = smilence (*to smile with silence*)

(11) free + damn = freedamn (*Chinese netizens' ridicule on freedom*)

(12) *Egg* + *ache* = eggache (*people do unreasonable things when they are too bored*)

(13) *Spring* + *brother* = sprother (*a female singer named Yuchun Li, entitled “brother Chun”.*)

This category combines the partial syllables of two English words to form a new CWT.

2.2.2. Chinese syllables+English syllables.

(14) zhuangbi (*Chinese pinyin of 装逼*) + ity (*suffix*) = zhuangbility (*pretentious*)

(15) niubi (*Chinese pinyin of 牛逼*) + ity (*suffix*) = niubility (*great*)

(16) jiujiie (*Chinese pinyin of 纠结*) + ity (*suffix*) = jiujielity (*Tangled*)

(17) geli (*Chinese pinyin of 给力*) + able (*suffix*) = gelivable (*awesome*)

(18) un (*prefix*) + ding (*Chinese pinyin of 顶*) + able (*suffix*) = undingable (*Unreliable*)

This category combines syllable of Chinese word and English affixes to form a new CWT.

3. The Lexical Features of CWT. The vocabulary of CWT contains words and phrases. They are mainly from some buzzwords and current social events in China, such as: “suihide” (*from “躲猫猫”, a social security incident*), “vegeteal” (*from the QQ game “stealing vegetables”*); “stupig” (*from the existing Chinese word “stupid pig”*), “a knife in the smile” (*from the existing Chinese phrase “笑里藏刀 (with murderous intent behind one's smiles)”*) [8]. The lexical features of CWT can be analyzed by word formation [9].

The word formation method of CWT follows the standard English word formation method, such as: affixation, blending and compound method. Netizens have created some new word formation methods, such as the embedded method and homophonic method. These word formation methods also have Chinese characteristics.

3.1. **Affixation method.** By adding the affixes to the root of the word, the affixation forms a new word. It contains the prefix method and the suffix method (*except for the very few infixes in English*). The affixation of CWT also contains the prefix method and the suffix method, and the prefixes in Chinglish words are mostly “un”, such as:

- (19) un + geli + able = ungelivable
 (prefix) (the Pinyin deformation of “给力”) (suffix) (not so good as you think)
- (20) un + ding + able = undingable
 (prefix) (pinyin of “顶”) (suffix) (unbearable)

(19) - (20) are direct transliteration of Chinese vocabulary. A new Chinglish word is created by combining the Chinese pinyin and the English affix.

3.2. **Blending method.** The blending method operates on two original words, chooses the former part or the latter part, and then merges them into a new word. There are three categories, as shown in Table 2:

TABLE 2. THE CLASSIFICATION OF BLENDING METHOD OF CWT

CLASSIFICATION	WORD COMPOSITION
keeping the former part of the first word + the latter part of the second word	Sprother= Spr (spring)+ other (brother)
	Chinsumer= Chin (China)+ sumer (consumer)
	vegeteal= veget (vegetable)+ eal (steal)
keeping the original form of the first word + deleting the former part of the second word	smilence= smile (smile)+ nice (silence)
	shitizen= shit (shit)+ izen (citizen)
	sexretary= sex (sex)+ retary (secretary)
keeping the original form of the second word + deleting the former part of the first word	divoice= di (divorce)+ voice (voice)
	animale= ani (animal)+ male (male)
	suihide= sui (suice)+ hide (hide)

3.3. **Compound method.** The compound method merges two or more words into a new word in a certain order. It is a common word formation method in English, such as:

- (21) free+damn=freedamn(*freedom*)
 (22) egg + ache =eggache(蛋疼, *an existing Chinese word*)

“Freedamn” is composed of the English words “free” and “damn”. The pronunciation of “freedamn” is very similar to the English word “freedom”. “Freedamn” reflects netizens’ ridicule on freedom.

“Eggache” is composed of the English words “egg” and “ache”, which means people do unreasonable things when they are too bored.

3.4. **Homophonic method.**

The harmony method is a new word formation method created by netizens. A part of the word is replaced by a similar pronunciation unit, such as:

- (23) newb + ity = newbilty
 (The homophonic of Pinyin niubi) (great)
- (24) drunb + ity = drunbilty
 (The homophonic of Pinyin zhuangbi) (pretentious, ostentatious)

In (23) -(24), after creating the Chinglish word “*niubility*” and “*zhaungbility*”, netizens replace “*niubi*”, “*zhuangbi*” with their homonyms “*newb*”, “*drunb*” in order to make the words more normative, respectively.

(25) Z + turn = Z-turn(*tossing, zhe teng*)

In (25), the Pinyin of “折腾” is “*zhe teng*”. On one hand, the pronunciation of the letter “Z” is similar with the Chinese pinyin “*zhe*”, and the pronunciation of the word “*turn*” is similar with the Chinese pinyin “*teng*”. On the other hand, “Z” and “*turn*” let people easily think of the winding road. So the Chinglish word “*Z-turn*” vividly expresses the meaning of the word “折腾”.

3.5. Embedded method. The embedded method is also a new word formation method created by netizens. Unlike the compound method, this method embeds a complete word into another complete word, such as:

(26) together + gay = togayther (*get married*)

(27) property + poor = propoorty (*real estate*)

A new Chinglish word “*togayther*” is composed by embedding the word “*gay*” into the word “*together*”, and “*togayther*” becomes an alternative English translation of “终成眷属, *get married*”. It reveals the increasing tolerance of Chinese young people towards homosexuality.

Another new Chinglish word “*propoorty*” is created by embedding the word “*poor*” into the word “*property*”. It indicates that people feel “*poor*” when confronting the mounting house prices.

From these different methods, the word formation methods of CWT not only have some traditional English methods, but also possess their own specific methods, which reflect the abundant imagination and creativity of netizens.

4. The Grammatical Features of CWT. CWT includes vocabulary and sentences. The vocabulary is self-created by netizens and mostly follows the rules of English word formation methods. However, most of the sentences of CWT directly ignore English grammar, such as word order, tense, etc. There are three grammatical characteristics of CWT.

4.1. POS Transfer.

Part-of-speech (POS) transfer is one of characteristics of CWT, such as: from noun to verb, or from verb to noun.

(28) Let’s **niubility**!

(29) If you **tiger** me, I will **moutain** you !

(30) You give me **stop** !

In (28) - (29), *niubility*, *tiger*, *moutain* are nouns, transferring to verbs as the predicate. In (30), *stop* is a verb, transferring to a noun as the object.

4.2. **Mainly with Simple Sentence and Exclamation Sentence.** From syntactical structure, the sentences are short and conform to the principle of linguistic economy. Many sentences are directly translated from Chinese common syntactical structure or idiomatic phrase. Simple sentences are much more than complex sentence, such as (31). Exclamation sentences are more than 60% in whole corpus, such as (32).

(31) five talk, four beauty, three lover

五讲、 四美、 三热爱

Wu jiang simei san reai

five stresses, four points of Beauty and three aspects of love

(32) Release your horse and come!

放 马 过来!

Fang ma guolai

bring it on!

4.3. **Types of Translation Methods.** The translation methods of the sentences of CWT mostly have five types, as Table 3.

TABLE 3. THE 5 TYPES OF TRANSLATION METHODS OF CWT

TYPE	EXAMPLES
<i>Literally word for word translation, following Chinese word order</i>	<i>People moutain people sea (huge crowds of people)</i>
<i>Translating homonymously</i>	<i>Even game win, even so whole. (<u>no</u> pain, no gain.)</i>
<i>Very few to follow English grammar</i>	<i>Many people think they are full of niubility (great, awesome), and like to play zhuangbility(pretentious), which only reflect their shability(stupid).</i>
<i>Combination of English grammar and literally word for word translation</i>	<i>I will give you some colour to see see. (I will teach you a lesson)</i>
<i>Combination of Chinese pinyin and English</i>	<i>no zuo no die. (If you do not do the stupid thing, you will not die)</i>

5. **Conclusion.** This paper systematically analyzes the phonetic, lexical and grammatical characteristics of the phenomena. The study can be beneficial for natural language processing, such as machine translation, sentiment analysis and information extraction.

Acknowledgment. The work is funded by the following projects: Humanities and Social Science Foundation of Ministry of Education of China No.16YJCZH004, Science and Technology Project of Guangzhou No.201704030002 and Major Projects of Chinese National Social Science Foundation No.11&ZD189.

REFERENCES

- [1] J.W.Chen, Study on buzzwords, *Guangxi University* ,2008. (in Chinese)
- [2] Z.Y.Ma, On the Characteristics and causes of CWT, *Journal of Chongqing Technology and Business University* , vol.28, no.5, pp.115-120,2011. (in Chinese)
- [3] J.J.Xue, On CWT from the perspective of memetics, *Journal of Yangtze University*, vol.37, no.7, pp. 108-110,2014. (in Chinese)
- [4] K.Zhang, Study on microblog language characteristics. *Shannxi Normal University* ,2012. (in Chinese)
- [5] W.Ling, L.Marujo, C.Dyer, et al.,Mining parallel corpora from sina weibo and twitter, *Computational Linguistics*, vol.42, no.2, pp. 1-37,2016.
- [6] B.Chen, D.H.Ji and C.Lyu, Building a Chinese semantic resource based on feature structure, *International Journal of Computer Processing of Languages*, vol.24, no.1, pp.95-101,2013.
- [7] B.Chen, C.Lyu, X.M.Wei and D.H.Ji,Chinese Semantic Parsing Based on Feature Structure with Recursive Directed Graph, *Wuhan University Journal of Natural Sciences*, vol.20, no.4, pp. 318-322 ,2015.
- [8] X.G.Yang and X.Q.Liu, On the Characteristics of Emerging Chinglish, *Journal of Nanjing Forestry University*, vol.11, no.3, pp. 49-52,2011. (in Chinese)
- [9] China internet network information center, *The 39th China Statistics Report on Internet Development* ,2017. (in Chinese)