网络环境下
英语规则化趋势研究

张 权 著

TOWARDS REGULARIZATION OF ENGLISH THROUGH INTERNET

语言服务就是利用语言（包括文字）、语言知识、语言艺术、语言技术、语言标准、语言数据、语言产品等等所有语言的所有衍生品，来满足政府、社会及家庭、个人的需求。　　——李宇明

网络环境下
英语规则化趋势研究

张权 著

TOWARDS REGULARIZATION OF ENGLISH THROUGH INTERNET

网络环境下英语规则化趋势研究
WANGLUO HUANJING XIA YINGYU GUIZEHUA QUSHI YANJIU
著　者：张　权
....

前 言

现代英语有三类语法范畴使用两种语法形态标记：绝大多数规则形态标记与少数不规则形态标记。具体体现为动词的过去时或过去分词形式、名词的复数形式、形容词或副词的比较级形式。

对于规则与不规则形态的习得机制存在两种假说：单一机制记忆理论（Rumelhart & McClelland）和词汇—规则理论（Stephen Pinker）。本研究以后者为理论框架，通过修正的"Wug"不规则习得试验，进一步检测将英语作为外语学习的成年中国学习者的语言能力并对右手中心词规则和系统规则化现象进行了网络化检测。

全球化网络加快了语言发展的步伐，催生了越来越多的国际性的英语使用和使用者。英语规则化发展趋势同样率先在因特网上出现印证。本研究通过"谷歌"（google）等网络检索引擎，收集不规则形态被规则化的各类网络数据并对其规则化的比率进行对比分析与讨论。

本研究主要引出以下发现：

（1）修正的"Wug"试验结果表明：①将英语作为外语的成年中国学习者生造的规则形态多于不规则形态。这为全球性英语使用者在因特网上将不规则形态大量规则化提供了试验性依据。②规则化呈现向名词尤其是形容词或副词延伸的趋势。

（2）在高频率不规则形态中，形容词或副词的规则化比率远大于名词和动词。这在一定意义上对 Zipf 的基本假设及 Stephen Pinker 的词汇—规则理论提出挑战。因此，规则化趋势研究应当转向诸如 good 和 bad 这样高频率形容词的规则化和过度规则化现象。网络字典将其列为词条也表明形容词和副词有可能是优先被全球性英语使用者高度规则化的语言形态。

（3）符合右手中心词规则的某些复合词或短语连同其中的右手中心词规则一并趋于规则化。

（4）大部分符合系统规则化规则的复合词或短语享有很高的规则化比率，即规则形态大大超过不规则形态。

（5）绝大部分规则化使用实例（笔者保留引文中的拼写原貌，未加改动）并非表明其使用者的英语水平低下。不规则形态的规则化并不一定意味着其使用者缺乏英语不规则形态的相关知识。

总而言之，网络势必提供一个英语越来越规则化的全球性平台。目前的不规则形态的拼写形式将来有可能用规则加以清理，从而最大限度凸显其优点，扬弃其弊端。也许有可能只需要几个原则即可满足包括英语本族语使用者在内的所有使用者的需求。借助网络的全球性传播，英语规则化将是英语进化的最终趋势。

著名语言学家乔姆斯基曾经提出充分研究体现为三个方面：充分观察、充分描写与充分解释，同时认为充分解释是一切研究的终极目标。但语言的趋势研究则意味着充分解释依然不能成为研究的终极目标，充分研究还应包括充分评价与充分预测等验证和趋势性研究。本研究正是试图开辟此类研究之先河。其特别意义在于：通过其规则化趋势研究评估英语的发展现状与趋势，以期更好地指导中国英语教学。

本研究得到上海外国语大学戴炜栋教授及业内专家的精心指导与帮助。南京理工大学李娟老师为本研究做了大量数据采集与分析工作。我的亲友给予了无微不至的关爱。笔者不胜感激。

张　权
2016 年春于南京

Introduction

Many languages like English are inflectional in morphing their grammatical categories. English, as well as some other languages, however, does not use only one unified morphological means to mark its important grammatical fields. In modern English, there are three grammatical categories which are morphologically marked by two different forms of inflection, one of which is the majority of the open-class, entirely predictable regulars while the other is the minority of the closed-class, totally unpredictable irregulars. The three grammatical categories are: ① about 180 past-tense or past-participle verbs, as in *do-did-done*, *know-knew-known*, which are not formed in accordance with the usual patterns of English regular past-tense and past-participle verbs, as in *work-worked-worked*; ② plural nouns, which hold a handful of irregulars, as in *man-men*, *child-children*, which violate the regular plural endings as in *student-students*; and ③ comparatives and superlatives of adjectives as well as adverbs, which keep the few irregulars, as in *good-better-best*, *bad-worse-worst*, which are against the regular endings, as in *small-smaller-smallest*, *hard-harder-hardest*.

Linguistics concerns with regularity and irregularity that has long covered these questions:

(1) Why does English have two different morphological forms of inflection instead of one since it takes more time and energy to learn them respectively? Where do regulars and irregulars come from respectively? Are languages like English preferably regular or irregular?

(2) What is the nature of regularity and irregularity? By what criteria do we tell regulars from irregulars?

(3) Since regulars and irregulars coexist in certain grammatical fields, what might be the interaction between the two? Do the two always keep constant or changing? Are they mutually affected? Or does one effect upon the other? If they are mutually affected, is there much evidence for new memberships of regulars from irregulars as well

as of irregulars from regulars? If not so, which effects upon which? Is the irregular circle being shrunk or vice verse? Can we suppose that the irregular circle is being shrunk by the analogy or generational rules of the heavy majority of regulars?

(4) If regularization is the tendency of irregularity, are irregulars regularized randomly or do they follow certain order? For what reason (s) do they tend to be regularized and what evidence can support the event? For what reason (s) do some irregulars tend to be more easily regularized while some strongly resist regularization? Why do irregulars have to be regularized in certain structures or context while they have to keep irregular in some other structures or situations?

(5) What might be a possible mental process that guides the acquisition of regulars and irregulars? Do the two share the same model or are they processed differently? Are there any ways of testing the hypothetical model?

This book tends to discuss some major questions listed above and attempts to make contributions to the following aspects:

(1) Since globalization of English has given birth to more and more international uses and users of English, the author assumes that regulars are more preferable to irregulars regardless of the user's English proficiency. International users of English might not be so poor at mastery of some most familiar and highly frequently used irregular forms as to carelessly make such mistakes as in *take-taked*. Global Englishes seem to more conform with regulars than irregulars since global users vary with native users in their acquisition of regulars and irregulars.

(2) Although much evidence from English studies has shown that the regularized tendency of English irregular morphology has gradually resulted in the transition of English irregular forms from the irregular to the regular, the author still hypothesizes that much more evidence from the Internet, which is becoming a more and more important platform where there tends to be more and more global users of English, will modify or challenge the general hypothesis for the remnants of English irregular forms in modern English proposed by Zipf (1935): among the words characterized with the high frequency of occurrence, the percentage of the old one is much higher than that of the newly coined one. The latter is characterized with the low frequency of occurrence. The higher frequency the word possesses, the more chances it has to secure in the given lexicon, the more resistant it is to alterations. Consequently, the lower frequency of occurrence the word has, the more probable is its loss from the lexicon.

The author finally tends to conclude that the fast global spread of English,

especially by means of Internet will quicken its steps of language evolution. Some linguists (Pagel, Atkinson & Meade, 2007), once tried to track, in the coming centuries, the shift in the regularization of irregular verbs like *sink*, *sting*, *swim*, *wring and spring*, which unlike most English verbs are made past tense by changing the "i" to "u" instead of adding the usual – ed at the end. So one day in the future we might begin to say that *yesterday we sinked*, *stinged*, *swimmed*, *wringed*, or *springed*, instead of *sunk*, *stung*, *swum*, *wrung* or *sprung*. If we trace the course of other irregular verbs from hundreds of years ago that have already been regularized, we can begin to estimate how long it will be before these remaining irregular verbs like *sinked*, *springed* and the rest become the accepted way to speak. At past rates it will take another 300 – 2,000 years (39,000 years for the ultimate verbs, "be" and "have"). But some other linguists (Kelly, 2007) postulated that this timetable is too long. Many historical linguists postulate an original proto-language that fragmented as human beings underwent the diaspora from our point of origin. Sociologists use this theory to track human migration and cultural development. Evolutionary linguists are reviving their field thanks to advances in technology that allow them to model the adaptation of languages. Written language is now changing more rapidly than spoken language, thanks to the Internet and its close relationship with globalization. The English language is morphing as concepts from other languages and cultures are absorbed and spread through online media.

Contents

Chapter One　English Irregular Forms

1. 1　Three Major Grammatical Categories of English Irregular Forms

Irregularity is called so in contrast with regularity. If all linguistic forms are regular morphologically and phonologically, there are by no means irregulars by the name. But it is not otherwise true that all linguistic forms are irregular. Irregularity does not exist in every grammatical category. In modern English, irregularity lies in three grammatical fields: irregular markings of plural nouns, past-tense and past-participle verbs, and comparatives and superlatives of adjectives and adverbs, among which they are all in the minority.

1. 1. 1　Irregular Plurals of Nouns

The majority of English countable nouns are regular and predictable in the spelling of the plural form by adding -s to the end of the singular form or -es to those singulars that end in a sibilant sound (/s/, /z/, /ts/, /dz/). However, there are several nouns which are irregular in their spellings. These nouns are exceptions when it comes to making them plural. When irregular nouns become plural, their spellings change in different ways or they may not change at all from singular forms. Some nouns that end in -f or -fe are changed to -ves in the plural, as in *calf-calves*, *half-halves*. Some nouns change the vowel sound in becoming plural as in *foot-feet*, *goose-geese*, *louse-lice*, *man-men*, *mouse-mice*, *tooth-teeth*, *woman-women*. Some Old English plurals are still in use, as in *child-children*, *ox-oxen*. Some nouns ending in -o take -s as the plural, while others take -es: *auto-autos*, *echo-echoes*. Some nouns do not change at all: *deer-deer*, *offspring-offspring*. Some nouns retain foreign plurals: *criterion-criteria*. Other irregular plurals retain from different languages: *tempo-tempi* (Italian), *schema-schemata* (Greek). More can be found in Appendix I.

1.1.2　Irregular Past-Tense and Past-Participle Verbs

The English language has a large number of irregular verbs. In the great majority of these, the past participle and/or past tense is not formed according to the usual patterns of English regular verbs. Other parts of the verb—such as the present 3rd person singular -s or -es, and present participle -ing—may still be formed regularly.

Among the exceptions are the verb to *be* and certain defective verbs, which cannot be conjugated into certain tenses.

Most English irregular verbs are native, originating in Old English (an exception being "catch" from Old North French "cachier"). They also tend to be the most commonly used verbs. The ten most commonly used verbs in English are all irregular.

In general, English contains about 180 "irregular" verbs that form their past tense in idiosyncratic ways, such as *ring-rang*, *sing-sang*, *go-went*, and *think-thought*. In contrast with the regulars, the irregulars are unpredictable. The past tense of *sink* is *sank*, but the past tense of *slink* is not *slank* but *slunk*; the past tense of *think* is neither *thank* nor *thunk* but *thought*, and the past tense of *blink* is neither *blank* nor *blunk* nor *blought* but regular *blinked*. Also in contrast to the regulars, "irregular verbs define a closed class: there are about 180 of them in present-day English, and there have been no recent new ones" (Pinker, 1999: 5). And they have a corresponding advantage compared with the regulars: there are no phonologically unwieldy forms such as *edited*; all irregulars are monosyllables (or prefixed monosyllables such as *become* and *overtake*) that follow that canonical sound pattern for simple English words. For instance, the past tense of T-forms include: *burnt*, *clapt*, *crept*, *dealt*, *dreamt*, *dwelt*, *felt*, *leant*, *leapt*, *learnt*, *meant*, *spelt*, *smelt*, *spilt*, *spoilt*, *stript*, *vext*. T-forms can be divided into two categories: those with a vowel change and those without a vowel change. T-forms with a vowel change include: *crept*, *dealt*, *dreamt*, *felt*, *leapt*, *meant*. The T-forms with a vowel change are still very common in modern English. In fact, *crept*, *dealt*, *felt* and *meant* are the only accepted forms. In the case of *dreamt* and *leapt*, although *dreamt* and *leapt* are still quite common and acceptable in both written and spoken English, the regular forms *dreamed* and *leaped* seem to be more popular in modern usage. T-forms without a vowel change include: *burnt*, *clapt*, *dwelt*, *leant*, *learnt*, *spelt*, *smelt*, *spilt*, *spoilt*, *stript*, *vext*. The T-forms without a vowel change are slowly disappearing from the language. *Dwelt* is the only form in this category which is more frequently used than the regular -ed form. *Burnt*, *leant* and *learnt* are still

relatively common in spoken English and fairly common in written English. *Spelt*, *smelt*, *spilt* and *spoilt* are quickly disappearing. *Stript*, *clapt* and *vext* are rarely used in contemporary English.

1.1.3 Irregular Comparatives and Superlatives of Adjectives and Adverbs

The regular way to make comparative or superlative adjectives is to add -*er* or -*est* to the end of bare adjectives or adverbs, or to use the peripheral form of *more* or *most* before. A small number of adjectives, however, are irregular and some of these can be regular or irregular. The most important ones are listed in Table 1.1:

Table 1.1   Irregular comparatives and superlatives of adjectives and adverbs

Serial No.	Adjective	Comparative	Superlative
1	bad	worse	worst
2	far	farther	farthest
3	far	further	furthest
4	good	better	best
5	many	more	most
6	much	more	most
7	little	less (littler)	least (littlest)
8	old	elder (older)	eldest (oldest)
9	well (healthy)	better	best

1.2 Etymology of Irregular Forms

Etymology of English irregular forms tends to answer the question where do irregulars come from. As a language, English is derived from the Anglo-Saxon, a West Germanic variety, although its current vocabulary includes words from many languages. The Anglo-Saxon roots can be seen in the similarity of numbers in English and German, particularly, *seven/sieben*, *eight/acht*, *nine/neun* and *ten/zehn*. Pronouns are also cognate: *I/ich*; *thou/du*; *we/wir*; *she/sie*. However, language change has eroded many grammatical elements, such as the noun case system, which is greatly simplified in Modern English; and certain elements of vocabulary, much of which is borrowed from French. Though more than half of the words in English either come from the French

language or have a French cognate, most of the common words used are still of Germanic origin①.

When the Normans conquered England in 1066, they brought their Norman language with them. During the Anglo-Norman period, which united insular and continental territories, the ruling class spoke Anglo-Norman, while the peasants spoke the English of the time. Anglo-Norman was the conduit for the introduction of French into England, aided by the circulation of Langue d'oïl literature from France. This led to many paired words of French and English origin. For example, *beef* is cognate with modern French *bœuf*, meaning *cow*; *veal* with *veau*, meaning *calf*; *pork* with *porc*, meaning *pig*; and *poultry* with *poulet*, meaning *chicken*. In this situation, the foodstuff has the Norman name, and the animal the Anglo-Saxon name, since it was the Norman rulers who ate meat (meat was an expensive commodity and could rarely be afforded by the Anglo-Saxons), and the Anglo-Saxons who farmed the animals.

English words of more than two syllables are likely to come from French, often with modified terminations. For example, the French words for *syllable*, *modified*, *terminations* and *example* are *syllabe*, *modifié*, *terminaisons* and *exemple*. In many cases, the English form of the word is more conservative (that is, has changed less) than its French form.

English has proven accommodating to words from many languages. Scientific terminology relies heavily on words of Latin and Greek origin. Spanish has contributed many words, particularly in the southwestern United States. Examples include *buckaroo* from *vaquero* or "*cowboy*", *alligator* from *el lagarto* or "*the lizard*", and *rodeo*. *Cuddle*, *eerie* and *greed* come from Scots; *honcho*, *sushi* and *tsunami* from Japanese; *dim sum*, *gung ho*, *kowtow*, *kumquat*, *ketchup*, and *typhoon* from Cantonese Chinese; *behemoth*, *hallelujah*, *Satan*, *jubilee* and *rabbi* from Hebrew; *taiga*, *sable* and *sputnik* from Russian; *cornea*, *algorithm*, *cotton*, *hazard*, *muslin*, *jar*, *sofa* and *mosque* from Arabic; *kampong* and *amok* from Malay; and *boondocks* from the Tagalog word *bundok*.

Accordingly, English irregular forms are derived from several sources of languages and their irregular forms have been long preserved during the evolution of language.

1.2.1 Old English

Old English contains about twice as many irregular verbs as Modern English,

---

① For an example of the etymology of an English irregular verb of Germanic origin, see the etymology of the word *go* in section 1.2.1.

including now obsolete forms such as *cleave-clove*, *crow-crew*, *abide-abode*, *childe-chid*, and *geld-gelt*. Bybee (1985) examined the current frequencies of the surviving descendants of the irregulars in Old English and found that it was the low-frequency verbs that were converted to regular forms over the centuries.

Today one can actually feel the psychological cause of this historical change by considering past tense forms that are low in frequency. Most low-frequency irregulars sound stilted or strange, such as *smite-smote*, *slay-slew*, *bid-bade*, *spell-spelt*, and *tread-trod* (in American English) , and one can predict that they will eventually go the way of *chid* and *crew*. In some cases, a form is familiar enough to block the regular version, but not quite familiar enough to sound natural, and speakers are left with no clear choice for that slot in the conjugational paradigm. For example, many speakers report that neither of the past participle forms for *stride*, *strided* and *stridden*, sounds quite right. In contrast, low-frequency regular past tense forms always sound perfectly natural (or at least no more unnatural than the stems themselves). No one has trouble with the preterite or participle of *abate-abated*, *abrogate-abrogated*, and so on.

Most irregular verbs exist as remnants of historical conjugation systems. What is an exception actually followed a set, normal rule long ago? When that rule fell into disuse, some verbs kept the old conjugation. An example of this is the word *kept*, which before the Great Vowel Shift fell into a class of words where the vowel in keep (then pronounced kehp) was shortened in the past tense. Similar words, such as *peep*, that arose after the Vowel Shift, use the regular-ed suffix. Groups of irregular verbs include: The remaining strong verbs, which display the vowel shift called ablaut and sometimes have a past participle in -en or -n: e. g. , *ride/rode/ridden*. This verb group was inherited from the parent Proto-Germanic language, and ultimately from the Proto-Indo-European language, and was originally an entirely regular system. In Old English and in modern German, it is still more or less regular, but in modern English, the system of strong verb classes has almost entirely collapsed.

Weak verbs that have been subjected to sound changes over the course of the history of English that has rendered them irregular acquired a long vowel in the present stem, but kept a short vowel in the preterite and past participle, e. g. , *hear/heard/heard*.

Weak verbs that show the vowel shift are sometimes called "Rückumlaut" in the present tense, e. g. , *think/thought*.

Weak verbs that end in a final -t or -d that made the addition of the weak suffix -ed seem redundant, e. g. , *cost/cost/cost*.

There is a handful of surviving preterite present verbs. These can be distinguished from the rest because their third person simple present singular (the *he*, *she*, or *it* form) does not take a final -s. These are the remnants of what was once a fairly large Indo-European class of verbs that were conjugated in the preterite or perfect tense with present tense meaning. All of the surviving verbs of this class are modal verbs, that is, a class of auxiliary verbs or quasi-auxiliaries, e. g. , *can/could/could*.

Verbs that contain suppletive forms form one or more of their tenses from an entirely different root. *Be* is one of these, as in *go/went/gone* (where *went* is originally from the verb to wend). By the 15th century in southern England, *wende* (wend) had become synonymous with *go*, but its infinitive and present tense forms had ceased to be in frequent use. With a waning, morphing preterite tense (yode), *go* was ripe to receive a new preterite—the preterite of *wende*, the familiar *went*. (Online Etymology Dictionary)

Other verbs have been changed due to ease of pronunciation so that it is shorter or more closely corresponds to how it is spelt.

Some weak verbs have been the subject of contractions, e. g. , *have/had/had*.

There are fewer strong verbs and irregular verbs in Modern English than there are in Old English. Slowly over time the number of irregular verbs is decreasing. The force of analogy tends to reduce the number of irregular verbs over time. This fact explains the reason that irregular verbs tend to be the most commonly used ones; verbs that are more rarely heard are more likely to switch to being regular. For instance, a verb like *ablate* was once irregular, but today *ablated* is the standard usage. Today, irregular and standard forms often coexist, a sign that the irregular form may be on the wane. For instance, seeing *spelled* instead of *spelt* or *strived* instead of *strove* is common.

On the other hand, contraction and sound changes can increase their number. Most of the strong verbs were regular, in that they fell into a conventional plan of conjugation in Old English; there are so few of them left in contemporary English that they seem irregular to us.

1. 2. 2　Borrowings

Loanwords are words adopted by the speakers of one language from a different language (the source language). A loanword can also be called a borrowing. The abstract noun borrowing refers to the process of speakers adopting words from a source language into their native language.

English has many loanwords, due to England coming in contacts with numerous