Towards a Model for the Prediction of Chinese Novel Verbs*

Paul You-Jun Chang[a] and Kathleen Ahrens[a, b]

[a] Graduate Institute of Linguistics, National Taiwan University, Taipei, Taiwan R. O. C.
r95142005@ntu.edu.tw
[b] Hong Kong Baptist University, Kowloon Tong, Hong Kong
ahrens@hkbu.edu.hk

Abstract. As previous word adoption models, though proposing potential factors that influence the survival of neologisms, receive little empirical examination, this corpus-based study compares the performance of two such models by providing clear operational criteria for each factor in the models and, consequently, proposes a hybrid model that improves the previous results. We focus on seventy-seven Chinese novel verbs that appeared about ten years ago, defining their survival/failure in the real world, and examine the accurate prediction ratio of the two models. Both models display an overall accuracy of about 60 percent. However, as certain factors, e.g., unobtrusiveness, appear to be invalid predictors for the Chinese data, we attempt to improve the results by deleting inappropriate factors and by adjusting the weightings. As the overall accuracy was improved to about 70 percent, we suggest that this study would shed light on the potential factors that influence the survival of Chinese novel verbs.

Keywords: Neologism, Word Adoption Model, Unobtrusiveness.

1. Introduction

One interesting aspect about the human cognition is its ability to create. And creativity in language has doubtlessly inspired numerous studies. Among others, one phenomenon that commonly occurs would be the emergence of neologisms. For one thing, all words, together with their senses, that exist in our current vocabulary were once new (Klein and Murphy, 2001) and must have undergone a certain developmental process to finally remain in the vocabulary. For another, we still see this process going on in everyday life as can be seen from the coinage of new words, such as *blog* or *Y2K*. Previous studies on neologisms, therefore, mostly focused on the collection of new words (e.g., Algeo and Algeo, 1991), the analysis of their inherent features (Hsu, 1999; Rey, 1995), or their relationship with the society (Hsu, 1999). Metcalf (2002; cf. Chang, 2008) and Kjellmer (2000), among the few, nevertheless attempted to observe the factors that influence the adoption of neologisms in a language and, respectively, proposed a scoring system to assess the possibility for novel words to survive or to fail.[1] Interestingly, these two models are similar regarding certain factors as important, e.g., morphological productivity,

* We appreciate comments from two anonymous reviewers and Siaw-Fong Chung, as well as comments from the audience in CLDC-2 for a previous, smaller-scale paper. We also appreciate Professor Janice Fon for her comments on the statistics and Yu-You Chen, Yao-Zhu Zhang, and Yi-Wei Lin for help in collecting the frequencies. The remaining errors are our own.

[1] Part of this paper (mainly 2.1 and parts of Sections 3 and 4, about Metcalf's model,) were presented at The Second Conference on Language, Discourse, and Cognition (CLDC-2). As an extension, this study contrasts Metcalf's model with Kjellmer's and proposes a hybrid model for improvement of the results.

and yet are contrastive in their evaluation of linguistic gaps and borrowings. Following their studies, Sabino (2005) used the two models to evaluate the possibility of the word *gameday* to survive and found that both models yield a roughly 0.7 probability. While this interesting study displayed the possibility to test Metcalf's and Kjellmer's models empirically, the fact that the study looked at only one word, and that *gameday* itself was a new word in development, leaves the accuracy of prediction, i.e., the performance of the two models, unaddressed. Moreover, although Chang (2008) empirically examined Metcalf's model, no such study was done for Kjellmer's model or for the comparison of the two models. Therefore, for a more adequate evaluation, comparison, and improvement of the performance of both of the models, an empirical study with a set of clear criteria for each factor in the two models would seem necessary.

This study, therefore, attempts to examine the models proposed by Metcalf (2002) and Kjellmer (2000) by focusing on Chinese novel verbs that appeared around ten years ago. In doing so, we propose, for each factor in the models, a set of criteria for the scoring of the neologisms by using corpus tools. We ask the following questions: (1) Can we predict the survival of Chinese novel verbs from 1996 to 2006 according to Metcalf's and Kjellmer's models? (2) If yes, which provides better prediction? (3) If no, how can we better predict the survival of these words by modifying their conditions? And, as these two models contrast in their notion of linguistic gaps and borrowings, we hypothesize that (1) in Chinese, unlike the situation observed in English, borrowings and translated words take no disadvantage, while gap-filling is beneficial, if not essential, for the survival of novel verbs. We demonstrate, by setting up clear criteria, that although the two models perform similarly in predicting successful words, Metcalf's model performs better at predicting failure words. While the results suggest room for improvement of the models, we further analyze the factors, deleting inappropriate ones or adjusting the weightings; consequently, we improved the overall accuracy to about 70 percent.

2. Word Adoption Models

2.1. Metcalf's Fudge Scale

Metcalf (2002) proposed the FUDGE scale to measure the probability of a word's survival based on his observation of English neologisms. The crucial factors, as Metcalf stated, involve (1) frequency of the words; (2) unobtrusiveness, i.e., a successful word should not be exotic or too cleverly coined; (3) diversity of users and situations, i.e., the range of their usage; (4) generation of other forms and meanings, namely the productivity of the word; and (5) endurance of the concept, related to the concept's reference to a historical event. The method of assessing the probability is to rank the new words from level 0 to level 2 in each factor and sum up the total scores in the end. The higher the scores are, (7, as proposed,) the more likely the new words are to survive.

Metcalf downplays cleverness, exoticness, and linguistic gaps. He states that obtrusive new words cannot last long, and that words such as *skycap*, *scofflaw*, *agnostic*, etc., caught on because they do not look exotic, or they do root from a certain source language, which English has opened to all along. In addition, Metcalf claims that lexical gaps do not necessarily provoke new words. For example, although the Hebrew word *hesed* implies God's mercy, tenderness, and love, it was inadequately translated as *mercy* in English, thus losing certain original meanings. Metcalf states that gaps such as these do not seem likely to be filled in the near future; communication can be sustained without a proper word for everything. As will be noticed later, this model differs from Kjellmer's model in terms of humor, gap filling, and exoticness.

2.2. Kjellmer's Model

Kjellmer (2000) presents thirteen conditions to assess potential words, which can be divided into five categories: semantic, phonological, morphological, and graphematic conditions, and others, such as prestige.

Firstly, semantic conditions include the existence of a semantic parallel, namely a pre-existing semantic pattern in the language (e.g., *–able* for adjectives meaning "*capable of being V-ed*") as well as semantic transparency, the straightforwardness of the new word's meaning. Secondly, phonological parallels, i.e., well-established sound combinations in the language, and ease of pronunciation are considered influential factors. Thirdly, a word is more likely to survive if it has a morphological parallel, filling up a gap in a pre-existing morphological pattern in the language; e.g., the potential word *pensivity* would fill up a gap where the suffix pattern *–ive* would turn into its nominal form *–ivity.* In addition to this, Kjellmer states that a new word is also likely to survive if it does not fill up a morphological gap and yet follows the general morphological rules in the language. Besides, Kjellmer states that highly productive affixes (e.g., *-ness*) may facilitate the survival of new words. Furthermore, the more compatible the etymological roots of the new word and its affixes are, the more likely the word will be accepted. In terms of graphemes, a word following the graphematic customs in the language and a word whose spelling agrees with the pronunciation would possibly succeed. Finally, words that carry a prestigious, exotic, or humorous connotation, together with words that are concise, would also possibly survive.[2]

In the following sections, we therefore seek to evaluate the performance of the two models.

3. Methodology

3.1. Real World Cases: Evaluating the Words' Survival in UDN Database

3.1.1. Defining Neologisms and Narrowing Down the Scope

In this study, we gather our data from the collection of neologisms from July 1996 to December 1997 in Taiwan (National Languages Committee, 1998). Containing 5,711 new words, this collection selects new words according to their appearance in the Revised Dictionary of Mandarin Chinese based on a day-by-day examination of major newspapers. For our current study, we narrowed down our scope to the category of fashion words (cf. Hsu, 1999) and focused only on non-sense-neologism verbs (cf. Rey, 1995; Hsu, 1999), since we could not obtain accurate frequencies for sense-neologisms from our current corpus, and verbs and nouns may behave differently (Ahrens, 1999). Excluding inappropriate words, such as English acronyms, monosyllabic words, and foul language, we obtained seventy-seven verbs in the end.[3]

3.1.2. Normalizing the Frequencies and Defining Their Actual Success

As researchers working on LIVAC (Linguistic Variations in Chinese Speech Communities) synchronic corpus (http://www.livac.org/) and Fischer (1998) have conducted similar studies, we would currently collect the year-by-year frequencies of these neologisms in the UDN (a major newspaper in Taiwan) database[4] from 1996 to 2006, since newspapers might capture the emergence and fading away of new words. These frequencies were further normalized to the frequencies per 10,000 characters. Thus, we obtained the normalized ratios of the words.

We further set up criteria to define the words' actual survival or failure by looking into the words' normalized ratios in 2006. As the words' appearance in print would indicate a wide usage, we set the threshold at a low level, in this case 0.3 (about 10 tokens) and 3 (about 100 tokens). Namely, a word having a normalized ratio less than or equal to 0.3 in 2006 (e.g., 哈草, *ha1 cao3*, 'to smoke,' normalized ratio=0.11) would be counted as a failure, and a word having a normalized ratio greater than 3 (e.g., 抓包 *zhua1 bao1,* '(to be) caught doing something,' normalized ratio=4.24) would be counted as a survival. If a word has a normalized ratio greater than 0.3 but less than or equal to 3 (e.g., 哈啦, *ha1 la1*, 'to chat,' normalized ratio=2.67), we would further look at its slope of the normalized ratios throughout the years to observe its

---

[2] Kjellmer actually discusses three more factors: semantic needs, prompting of media, and fashion; however, since no actual scores are assigned to these factors, they are currently not included in the analysis. Please visit http://graftedlife.googlepages.com for details about the factors and scores in Metcalf's and Kjellmer's models (appendices I and II) as well as for all the other appendices of this paper.

[3] For the list of verbs and related data, please refer to appendix III.

[4] For a description of the corpus, please refer to appendix IV.

developmental tendency, since we cannot be sure whether such words would really survive; in this case, only words with a slope of less than -0.06 would be counted as failures. All of these criteria were set up through our observation and comparison in the data. The results of our definition are summarized as follows:

Table 1. Survival and Failure Thresholds

Failure		Survival	
NR≦0.3	0.3<NR ≦3		NR>3
	Slope ≦-0.06	Slope>-0.06	
23	1	29	24

3.2. Prediction of Metcalf's and Kjellmer's Models

3.2.1. Operational Definitions for Metcalf's Model

The five conditions in Metcalf's model will now be discussed in sequence. Firstly, for frequency, the method and rationale would be similar to what we did in the previous section for defining the words' actual survival or failure, except in this case we look at the normalized ratio in the year 1996 in order to simulate the prediction process. We score a word as zero if the normalized ratio is less than or equal to 0.3; as one if the normalized ratio is greater than 0.3 but less than or equal to 3; and as two if the normalized ratio is greater than 3.

Secondly, as unobtrusiveness, in fact, explains for various levels of factors, including phonology, morphology, semantics, and foreign borrowings, its operational criteria would also involve different levels of information in the hope that the scores would be close to the original model's prediction. As a score zero in the model denotes a conspicuous word, exotic, strange-looking, or cleverly invented, we would assign to it words that are borrowings, violate morphological/phonological rules, or lack a clear form-meaning relationship. To score one, noticeable obtrusiveness, we assign words whose meaning can only be (indirectly) inferable from the word form in Chinese, perhaps through certain metonymical/metaphorical links. And to score two, we assign words with a transparent relationship between form and meaning.

Thirdly, since diversity involves the variety of users and situations, we define a diversity score of zero for specialized terms, as Metcalf defined, and for words that never occurred in the database, since Metcalf states that score zero words would only appear in specialized documents and not newspapers; score one for words that need further explanations in the 1996 UDN data[5]; and score two for the remaining words.

Fourthly, generation of other forms and meanings in Metcalf's model refers to the word's ability for POS alternation, inflectional changes, or the word's ability to extend to other forms and meanings. For clear criteria, we would regard a word's productivity as reflected in the variety of the collocates they have. In this case, we look into the Chinese Word Sketch (http://wordsketch.ling.sinica.edu.tw/). If Chinese Word Sketch shows that a word can collocate with more than one type of words, perhaps not belonging to the same domain with its dominant meaning, than we consider the word able to be applied to various types of situations, thus having a greater productivity. Practically, and also through comparison, we set up the criteria that words having more than ten collocates would be scored as two; those with less than ten collocates but having more than three Word Sketch functions would be considered moderately productive and scored as one; and those with less than ten collocates and having less than or equal to three Word Sketch functions would be scored as zero.

[5] In our observation, we would also consider words that need a pair of quotation marks with their usage to be words that need further explanations.

Fifthly, for endurance of concept, we would simply follow Metcalf's definitions that score zero for nonce word forms, one for words pertaining to historical events, and two for words with long-enduring concepts.

3.2.2. Operational Definitions for Kjellmer's Model

The thirteen conditions will be listed briefly with the operational definitions we use to score the words. (1) S1: Since having semantic parallels means filling up semantic gaps, we look into the Chinese Word Sketch to look for the words' near synonyms; if there are no competing synonyms, then we consider the word filling up a semantic gap.[6] In this case, we set minimum similarity at zero, since neologisms usually have low frequencies.[7] (2) S2: For transparency to the layman, we adopt identical operational definitions as in Metcalf's model, i.e., the meanings of transparent words should not be specialized and must be clearly inferable from the form. (3) Ph1: To fill up phonological gaps, two operational criteria should be met. Firstly, a word should have parallel phonological patterns but have no homophones. Secondly, we compare the neologisms (with novel pronunciation) to a non-word list. Only legal non-words, i.e., possible pronunciation (Lin, 1999) that observe the Chinese phonological constraints can be regarded as filling up a phonological gap. (4) Ph2: For ease of pronunciation, as Kjellmer notes its overlapping with the previous condition, we would still compare the word to a non-word list; the criterion would require it not to be an illegal non-word in Chinese, i.e., its pronunciation is allowed in Chinese (whether it is currently existing or not). (5) M1: For morphological parallels, we check the words' morphemes in Souwenjiezi (搜文解字, http://words.sinica.edu.tw/) to look for patterns of a certain morpheme and decide if a word fills up a morphological gap in a certain pattern. (6) M2: For following the morphological principles, as also noted by Kjellmer for its overlapping with the previous condition, here the words do not have to fit into a certain pattern but have to observe the general morphological rules. (7) M3: For productive affixes, we check Souwenjiezi for words formed with a certain affix; if more than ten words are formed with the same morpheme, we consider the morpheme productive. For example, 打 *da3*, 'hit' is a productive morpheme as suggested in the database. (8) M4: To measure affix compatibility (i.e., the affixes deriving from the same root with the stem), we attempt to provide an operational definition that the words should not have morphemes of mixed origins, since in Chinese, often a Taiwanese morpheme combines with a Mandarin affix, e.g., 大 *da4*, 'big,' to form a word. (9) G1: For the filling of graphematic gaps, although rare in Chinese, we check the words in Soucixunzi (搜詞尋字), utilizing its radical searching function to decide whether a word fills up a graphematic gap. (10) G2: For agreement between spelling and pronunciation, we check if the word's pronunciation deviates from the usual case, i.e., the word should not be a homograph that has inconsistent pronunciations or needs to be pronounced in its original language, e.g., Taiwanese. (11) O1: We follow the original model, which recognizes the advantage of borrowings/foreign prestigious words. (12) O2: For conciseness, we examine whether a more concise word exists and can convey the same idea as the new word does. (13) O3: For humorous meanings, we follow basically two criteria proposed by previous scholars, namely incongruity (a discrepancy between the expression and the situation) and surprising meanings, the decision of which would be based on our intuition.

3.2.3. Success Threshold at 0.5

In order to compare the results of the two models, we transfer the scores on each scale into proportions; for example, score zero on Kjellmer's scale ranging from -16 to 14 would be

[6] Note, however, that sometimes the CWS lacks enough data for a thesaurus; then we make a decision on whether a synonym exists based on our world knowledge.

[7] Therefore, if under this setting the near synonym shown in the thesaurus is not likely to be a real synonym of the neologism, we look further into Word Sketch Difference to decide; if the two words do not have shared patterns, or only share certain common patterns, e.g., verbs like *become*, we consider it a sign for a lack of synonyms.

transferred as [0-(-16)]/30=0.53. In this work we adopt a threshold of 0.5 in defining success. Three reasons are provided for this threshold: (1) Metcalf, in his work, merely stated that words scored as 7 or higher are *more likely* to survive; however, in our observation this would lead to a bias where only a few words may be predicted as successful, while most words are predicted as failures in this model. (2) Kjellmer did not mention a threshold for success in his model, but score zero in his model would mean a neutral effect, the probability for which, as shown before, is 0.53. (3) As might be seen later, in a modified model, we find the lowest score for a word to survive would be 0.47 (and the score next to which would be 0.53, since there are only a fixed set of scores on this kind of scaling). Therefore, for consistency and for later comparison, we deem it proper to set the threshold at 0.5, which would be neither suspiciously high nor low.

4. Results and Discussion

4.1. Correct Prediction Ratio and Evaluation of the Two Models

The correct prediction ratio of the two models is summarized in Table 2.[8] It is shown that both models, especially Metcalf's model, perform better in predicting failures than survivals, while overall and in survival prediction the two generally do not differ much, with Metcalf's model slightly better than Kjellmer's model. However, as one might consider, the correctness ratio in both models can still be low.

Table 2. Predictions of the Two Models and Their Correct Prediction Ratio

	Real Case	Metcalf		Kjellmer	
Survival	53	29	54.72%	29	54.72%
Failure	24	20	83.33%	17	70.83%
Overall	77	49	63.64%	46	59.74%

We, therefore, propose that certain modification might be done to improve the results after the evaluation of the two models. In our observation, the advantage of Metcalf's model lies in its simplicity and his consideration of endurance of concept, which could provide explanations for certain data, while Kjellmer's model distinguishes the factors across different linguistic levels and takes conciseness into consideration. However, problems emerge when: (1) previous models, being designed for English, may not appropriately account for the Chinese data, where there might be a trend of accepting borrowed words (e.g., 變身 *bian4 shen1*, 'to transfigure,' a borrowing from Japanese), as contrary to our operational criteria for Metcalf's model (cf. Hsu, 1999); (2) vague definitions render establishing scoring criteria difficult, e.g., three levels of frequency in the FUDGE scale are vaguely defined from usages in family/among friends from 1,000~10,000 users to 'widely used'; (3) overlappings may lead to biased results, as in Kjellmer's M1-M2 (filling up a morphological gap and following the morphological rule) and Ph1-Ph2 (filling up a phonological gap and following the phonological rule); and (4) most importantly, in our observation, the scores in certain factors, such as G1or Ph2, appear almost the same across the seventy-seven items, which might indicate that these factors currently do not qualify as good predictors.[9] We, therefore, propose to examine the distribution of scores within each factor so as to determine the potentially important factors for our data. In addition, we would like to conduct statistical tests on Metcalf's/Kjellmer's models to see whether certain factors, as we hypothesized, would have an effect on the words' frequencies in 2006.

4.2. ANOVAs for Verification of Factors

Assigning scores to a word would, in effect, mean to categorize them; e.g., a word can obtain score one for generation and score two for endurance of concept, which would distinguish it

[8] For predicted scores of the two models, please see note 2 in appendix III for further information.
[9] Except for endurance of concept, since the items chosen mostly pertain to long-enduring concepts.

from other words with different scores. We, therefore, can draw distribution tables to see how many words fall in a certain category, conduct ANOVAs with factors in the models as independent variables, and log normalized ratios in 2006 as the dependent variables. Due to the limits of our number of tokens, we currently select factors in interest for analysis, examining whether Metcalf's/Kjellmer's concepts are correct. If Metcalf/Kjellmer are right in their hypotheses, that the factors they propose are influential for the words' survival, then we probably can detect certain main effects in ANOVAs that support their ideas. In this study, we select unobtrusiveness and generation from Metcalf's model and S1 (semantic gaps), M1 (morphological gaps), and O1 (prestige) from Kjellmer's model for analyses. These factors either reflect the conflicting concepts of the previous models (gaps and prestige), or represent certain general concepts that both models deem as important (productivity).[10]

The distribution tables of the factors in the two models, as given below, suggest three kinds of possible ANOVAs that we can conduct: (1) a one-way ANOVA for G0, G1, and G2 when all the words are under U0; (2) a one-way ANOVA for U0, U1, and U2 when all the words are under G0; (3) a 2×2 ANOVA between U (0 and 2), and G (0 and 2); and (4) a 2×2×2 ANOVA among S1, M1, and O1. Planned comparisons were designed to detect the source of effects.

Tables 3 and 4. Distribution Tables for Metcalf's Model (U and G) and Kjellmer's Model (S1, M1, and O1) (Numbers in the cells: the number of words in each category; numbers in parentheses: the scores of the conditions according to the models.)

	U(0)	U(1)	U(2)
G(0)	17	8	11
G(1)	13	4	4
G(2)	11	2	8

	S1(-1)		S1(1)	
	M1(-1)	M1(3)	M1(-1)	M1(3)
O1(0)	13	14	6	9
O1(2)	23	6	4	4

4.3. ANOVA Results and Discussion

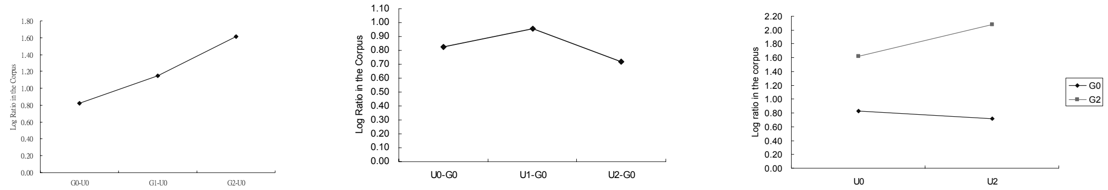
(1) One-way ANOVA showed a significant difference between the 2006 log normalized ratios of G0, G1, and G2 [$F(2, 38)=8.00$, $p<.01$, $\alpha=0.05$] when these words all belong to unobtrusiveness level zero. Planned comparisons suggest significant differences in G0-G2 and G1-G2 [$t(38)=-4.00$, $p<0.001$; $t(38)=-2.24$, $p<0.05$] but not in G0-G1 [$t(38)=-1.708$, $p=0.10$]. The results showed that words with different degrees of productivity, namely the ability of generating other forms and meaning, are different in their log normalized ratios in 2006 when all of these words are on unobtrusiveness level zero; generation, thus, would be a good predictor in Metcalf's model.

(2) One-way ANOVA showed no significant effect in the log normalized ratios of U0, U1, and U2 [$F(2, 32)=0.43$, $p=0.66$] when the words all belong to generation level zero, suggesting that putting prestigious (exotic, borrowed) words at a lower grade and non-conspicuous words at a higher grade may not lead to a good predictor of survival; as can be seen in the following Figures 1, 2, and 3, non-conspicuous words actually obtain lower log normalized ratios in the end, and U0 (prestigious) words got a little higher log normalized ratios than U2 words. This does not support Metcalf's hypothesis and implicates prestige may not take a disadvantage.

(3) Two-way ANOVA displayed a main effect in the log normalized ratios between G0 and G2 [$F(1, 43)=47.47$, $p<0.001$] but not between U0 and U2 [$F(1, 43)=1.29$, $p=0.26$]; no significant interaction was found [$F(1, 43)=3.32$, $p=0.08$]. This does not support Metcalf's idea concerning unobtrusiveness and reinforces our previous statement that generation (productivity) would serve as a good predictor, whereas putting prestigious words at a lower level may not be

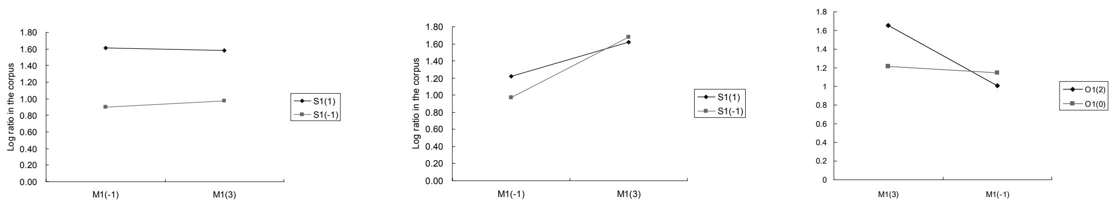
[10] The reason for not having other selections is that the scores in frequency were themselves obtained by categorizing frequency levels, and the items chosen here mostly denote long-enduring concepts (in terms of endurance of concepts), and diversity of users is currently not of interest to us. For Kjellmer's model, we select gap/prestige-related factors as shown here; phonological and graphematic gaps are not selected, both due to limits of tokens and to the lack of discernment in their scores.

appropriate for Chinese data. Figure 3 shows a main effect in G0/G2 but not in U0/U2; the interaction was not significant.



Figures 1, 2, and 3. 2006 Log Normalized Ratios: (3) G0, G1, and G2 (Left); (4) U0, U1, and U2 (middle); and (5) U0/U2 and G0/G2 (right).

(4) Three-way ANOVA showed a significant main effect in S1 [F(1, 69)=5.29, p<0.05], indicating that words filling up semantic gaps are different in their log normalized ratios from those not filling up a gap. However, no significant main effect was found either in M1, morphological gaps [F(1, 69)=3.05, p=0.09] or O1, prestige [F (1, 69)=0.41, p=0.53]. We also found no significant interaction in S1×M1, S1×O1, M1×O1, or S1×M1×O1 [S1×M1: F(1, 69)=0.39, p=0.53; S1×O1: F(1, 69)=2.95, p=0.09; M1×O1: F(1, 69)=2.62, p=0.11; S1×M1×O1: F(1, 69)=0.10, p=0.75]. Nevertheless, we suspect that the lack of significance would be due to our insufficient tokens, especially for words with score three in M1 and score two in O1. For one thing, the following figures suggest that semantic gap-filling words have higher log normalized ratios than those not filling up semantic gaps in non-borrowed words, but this is not obvious in borrowed words. Moreover, we further examined near-significant interaction effects and found a near-significant difference between O1, score two and O1, score zero, when the words are all scored as three for the M1 condition (i.e., when we do not take S1 into consideration) [F (1, 30)=2.33, p=0.14]. This would hint at the fact that words filling up morphological gaps (M1(3)) have higher log normalized ratios for borrowed words than for non-borrowed words:



Figures 4a/4b and 5. 2006 Log Normalized Ratios: (4a) S1×M1 at O1, score zero (left, non-borrowed words); (4b) S1×M1 at O1, score two (middle, borrowed words); (5) M1×O1 (right).

In all, we have shown significant effects in both productivity and semantic gaps; there is no support for Metcalf's statement that prestigious words are at a disadvantage. We cannot be sure that morphological gaps and prestige have an effect on log normalized ratios; however, the effect for morphological gaps is marginal (p<0.1), and from the graphs and further examination of the data, we can see prestige and morphological gaps may still have some effects under certain conditions, though this is not significant.

5. Towards a Model for the Prediction of Chinese Novel Verbs

So far in our discussion, we find at least three ways to improve the results: (1) select important factors from the two models (or, to be more conservative, to delete inappropriate factors); (2) combine certain factors in order to avoid overlappings; and (3) adjust the direction and the magnitude of the weightings. Here we adopt Kjellmer's model as our basic design, since it distinguishes different levels of factors clearly, except that the conditions might be wrongly weighted. We believe that after deleting inappropriate factors, solving overlappings, and adjusting the weightings, we should be able to make some improvement.

Firstly, we want to select conditions and solve the overlappings. In Metcalf's model, this would mean deleting unobtrusiveness (as previously discussed) and diversity (D2 overlaps with F2, and D0, specialized terms, overlaps with our operational definition for Kjellmer's semantic transparency). Frequency and endurance of concepts are kept because the two are complementary for Kjellmer's model. In Kjellmer's model, this would mean deleting Ph1, Ph2, and G1 (these factors generally have the same scores across items, as discussed in 4.1), and M4 and O3 (currently not of interest).

Secondly, we combine M1 and M2 to solve its overlapping. As noted before, filling up a morphological gap would imply following the morphological rule, and not following the rule would mean not filling up a gap. We, thus, adopt a dichotomy: the word either does not follow the rule or it follows the rule (including filling up a gap). Now the following ten factors remain: frequency, productivity (generation), semantic gaps, transparency, endurance of concept, morphological rules and gaps, productive affixes, agreement between spelling and pronunciation, prestige, and conciseness.

Thirdly, while we adopt identical operational definitions with those used previously for Metcalf's and Kjellmer's models, we merely want to adjust the weightings. Throughout the factors, we observe the principle, similar to Kjellmer's idea, that zero is the neutral score for a word to survive, while positive scores mean a beneficial effect and negative scores mean a disadvantage. The weightings in the factors are discussed as following (except for the first two factors, the others are all yes/no questions): (1) Frequency: We adopt the three frequency levels as those in Metcalf's model, except that we change the weightings to -1, 0, and 1, the reason being that low frequency may be a disadvantage; (2) Productivity: We adopt the three levels as those in Metcalf's generation factor, except that we change the weightings to -1, 0, and 2. The weighting for high productivity is 2 because both Metcalf and Kjellmer take this concept into consideration and even assign high scores to related concepts (e.g., productive affixes); (3) Semantic Gaps: Instead of the original 1/-1 pair of scores, we change the weightings to 1/0, since we deem the filling up of a gap as beneficial but not being gap-filling should not be negative; (4) Transparency: We adopt the original 0/-1 pair of scores as in Kjellmer's model; (5) Endurance of Concept: Different from Metcalf's model, here we only adopt a dichotomy, i.e., a neutral zero for long-enduring concepts, and -1 for nonce forms or words pertaining to historical events; (6) Morphological Rule Observation/Gap Filling: The two factors are combined and assigned for a 1/0 pair of scores, the idea being consistent with that in semantic gaps; (7) Productive Affixes: The weightings are changed from 3/-1 to 2/-1, in consistency with the previous productivity condition's weightings and to maintain Kjellmer's original heavy weight for words with highly productive morphemes; (8) Agreement between Spelling and Pronunciation: The pair of scores is changed from Kjellmer's 0/-1 to 1/0, since in our observation, words pronounced differently from the norm (e.g., in Taiwanese pronunciation) would not take a disadvantage; (9) Prestige: Since, as noted before, prestige appears to have some effects on the words, we would currently adopt it but reduce the weightings from 2/0 to 1/0; and (10) Conciseness: Since we consider conciseness a normal situation, and not being concise a disadvantage, instead of the original 1/0, we adjust the weightings to 0/-1.[11]

The results of the modified model are shown following in Table 5. As can be seen in the table, we improved both the overall and survival prediction correctness to around 70 percent, although the correctness ratio of failure prediction is lower:

[11] For a summary of the ten factors, please refer to Appendix V.

Table 5. Correct Prediction Ratios in the Three Models

	Real Case	Metcalf		Kjellmer		Current Model	
Survival	53	29	54.72%	29	54.72%	39	73.58%
Failure	24	20	83.33%	17	70.83%	16	66.67%
Overall	77	49	63.64%	46	59.74%	55	71.43%

6. Concluding Remarks

By focusing on Chinese novel verbs that appeared around ten years ago, we evaluate two word adoption models in this study. In addition to the finding that the two perform similarly in predicting survival words and in overall accuracy, we suggest that borrowings may not take a disadvantage in Chinese, and that words that fill up semantic gaps would be potentially accepted into the vocabulary. We also propose a model that is specifically designed for the group of words, and by deleting inappropriate factors and adjusting the weighting in the factors, we attempt to elevate the ratio of accurate predictions. As the improved results would imply a need for re-evaluated factors and weightings for new words in different languages, we would suggest, for future study, a test for other combinations of factors and weightings as well as an investigation on novel nouns and sense-neologisms, which, as we hope, would facilitate our understanding of the underlying mechanism for sense-creating and extension as well the fascinating process of creation.

References

Ahrens, K. 1999. The Mutability of Noun and Verb Meaning. *Chinese Languages and Linguistics* V. Interactions in Language, Y. Yin, I. Yang, and H. Chan (eds.), pp. 335–548. Taipei. Academia Sinica.

Algeo, J., and Algeo, A. 1991. *Fifty Years among the New Words: A Dictionary of Neologisms, 1941–1991*. Cambridge: Cambridge University Press.

Chang, Y. J. 2008. On the Rise and Fall of Chinese Neologisms in the Last Decade. Paper presented at *The Second Conference on Language, Discourse, and Cognition* (CLDC-2). National Taiwan University, Taiwan. May 17–18.

Fischer, R. 1998. *Lexical Change in Present-Day English: A corpus-based study of the motivation, institutionalization, and productivity of creative neologisms*. Tübingen: Gunter Narr Verlag.

Hsu, F. h. 1999. *台灣當代國語新詞探微 'Exploring the Contemporary Mandarin New Words in Taiwan.'* MA Thesis, National Taiwan Normal University.

Kjellmer, G. 2000. Potential Words. *Word* 51:205–28.

Klein, D. E., and Murphy, G. L. 2001. The representation of polysemous words. *Journal of Memory and Language* 45:259–82.

Metcalf, A. 2002. *Predicting New Words: the Mystery of Their Success.* Boston New York: Houghton Mifflin Company.

National Languages Committee, (國語會). 1998. *新詞語料彙編 1 'Collection of Neologisms I.'* Taipei: Ministry of Education.

Rey, A. 1995. The Concept of neologism and the evolution of terminologies in individual languages. In *Essays on Terminology*. Amsterdam: John Benjamins.

Sabino, R. 2005. Survey Says . . . Gameday. *American Speech* 80:61–77.