

18 *Phonetic fidelity vs. suggestive semantics: variations in Chinese character choice in the writing of loanwords*

MARK HANSELL

1 Introduction

The Chinese writing system has long been thought to be qualitatively different from other writing systems. In Western scholarship, early theories that described hanzi (Chinese characters) as ideographs have given way to theories that acknowledge the role of sound in the structure and use of hanzi.¹ Yet even these more informed and sophisticated approaches typically draw a sharp distinction between ‘logographic’ (or pleremic) Chinese and other ‘phonographic’ (or cenemic) writing systems. The former refers to a type of writing system in which individual graphs represent meaningful elements, and represent sound only secondarily if at all; the latter refers to a system in which individual graphs represent only sound, and represent meaningful elements such as morphemes or words only secondarily, as surrogates for spoken forms.

The one exceptional situation, in which almost everyone will agree that hanzi are used phonographically, is in the writing of foreign loanwords. Loanwords are typically described as using the same set of hanzi as native vocabulary, but with the hanzi ‘emptied’² of their meaning (in the sense that the reader is intended to ignore a hanzi’s usual meaning, and read it only for sound.) Some claim that there is a particular subset of hanzi used in the transcription of loanwords, constituting a sort of syllabary within the total set of hanzi; others maintain that the hanzi used in the writing of loanwords are selected arbitrarily from the set of all hanzi at the whim of whomever is the first to fix the loanword

¹ For in-depth discussion of the concept of ideography as applied to Chinese and Japanese see the various articles in Erbaugh 2001.

² See Haas (1976, 1983) for the origin of this term, which is also used in Hansell (1989b).

David Bradley, Randy LaPolla, Boyd Michailovsky and Graham Thurgood eds. *Language variation: papers on variation and change in the Sinosphere and in the Indosphere in honour of James A. Matisoff*, 277-290

Canberra: Pacific Linguistics, 2003.

Copyright in this edition is vested with Pacific Linguistics.

in a written form.³ In either case, the writing of loanwords is taken to be a totally distinct, cenic process, as opposed to the normally pleremic writing of Chinese native vocabulary.

Of the three possible types of lexical loans (semantic, graphic, and phonetic), the loanwords referred to above belong to only the phonetic. In semantic loans (loans in which the Chinese morphemes are chosen on the basis of their semantic similarity to the morphemes in the original), the connections between the individual hanzi and their usual Chinese meanings are unbroken. For example, in 熱狗 *règǒu* 'hot dog' (<English), the usual connections between the hanzi 熱 and the meaning 'hot' and between the hanzi 狗 and the meaning 'dog' respectively remain unchanged, the only innovation introduced by the semantic loan is the combining of the two morphemes into a new compound with a new meaning. In graphic loans, the choice of the hanzi is controlled by the source language original, and the resulting morphemes are those that are associated with those graphs in the recipient language (e.g. 社會 *shèhuì* 'society' <Japanese 社會 *shakai*.) Although the Chinese loan is created on the basis of graphic similarity to the Japanese, the connections between the particular morphemes and the hanzi are not any different from their usual connections in native Chinese lexical items. In a phonetic loan, however, no connection between meaning and graph is necessary. In 麥克風 *màikèfēng* 'microphone' (<English), any reference to the individual meanings of the constituent hanzi —'wheat', 'conquer', and 'wind' respectively—is at best irrelevant and at worst misleading in interpreting the word. Choice of hanzi seems to be strictly on the basis of phonetic similarity to the English original.

A phonetic loan ideally represents the closest approximation of source-language pronunciation that recipient-language phonology can construct. Since there is no possible motivation for recipient-language speakers to deviate from source-language pronunciation, it is normally assumed that differences between model and replica are the result of discrepancies between the phonological structures of the two languages. How such differences come into play in the borrowing process can be formalised, for instance in Silverman's (1992) and Yip's (1993) treatments of English loans into Cantonese. Yip proposes a set of ordered constraints on the phonological form of loanwords, with 'FAITHFULNESS' (fidelity to the phonetic shape of the source-language model) ranking second, behind only the syllable structure constraints of the recipient language. Once the most faithful rendition of the model that also obeys the recipient-language syllable structure constraints is found, a new loanword is created, consisting of a string of Chinese syllables. Then, as stated or implied in Lou (1992), Hansell (1989b), French (1976), and Godwin (1979), it is given written form through the choice, for each syllable, of one Chinese character from the homophonous sets of characters that could possibly represent that syllable.

Perhaps the most familiar example (used in French 1976, Haas 1976 and 1983, Sampson 1985, among others) is the proper name *Marx*. Assuming a British r-less pronunciation, it can be adapted to Mandarin by first mapping English phones onto Chinese phonemes, then resyllabifying with epenthesis. The result is three Mandarin

³ Kratochvil (1968), French (1976) and to a lesser degree Sampson (1985) support the idea of a small subset of characters being used for transliteration. Novotna (1968) finds much more variability, and Lou (1992) points out that Mainland renditions of phonetic loans use a much smaller and more regular set of hanzi than loans in Taiwan and Hong Kong.

syllables unspecified for tone: *ma*, *ke*, and *si*. For the first syllable, one hanzi (馬) is chosen from the set of homophones:

摩媽嗎麼麻痲馬嗎碼碼罵 etc.,

all pronounced *ma* (18 in all, according to Liang 1992). For the second syllable, 克 is chosen from among the 34 hanzi pronounced *ke*, and for the third syllable, 思 is chosen from among 37 syllables pronounced *si*. The result is the written form 馬克思 *mǎkèsī*, with the tones of the spoken form being dictated by the normal pronunciation of the hanzi used to represent each syllable.

Despite the clarity of the model presented above, and the clear distinctions drawn between the three types of loan, the distinction between semantic and phonetic loan does not hold up in the face of actual data. In the preceding example, the choice of 馬 is undoubtedly due to the fact that it is a common Chinese surname, and therefore contributes to the 'meaning' of the word by marking it as a proper name. Various scholars (Gelb 1963, Novotna 1968, Hansell 1989b, Tang 1989, Yao 1992, among others) have commented on a significant class of loans into Chinese which resemble the source language model in both sound and (though often to a lesser extent) meaning. These loans, called 'semanticised transcriptions' by Hansell (1989) and 音中有義 by Yao (1992), resemble folk etymologies. For example:

- 1) 維他命 *wéitāmìng* 'vitamin' ('support' + 'him/other' + 'life')
- 2) 聲納 *shēngnà* 'sonar' ('sound' + 'receive')
- 3) 香吉士 *xiāngjǐshì* 'Sunkist' ('fragrant' + 'lucky' + 'scholar')

In all three examples, the Chinese pronunciation is quite close to the English original, but there is also some relevant semantic content attached to the particular hanzi that have been chosen to represent those sounds. The meaning may be quite explicit (1), somewhat vague (2), or nothing more than positive connotations associated with the hanzi chosen (3), nevertheless it is difficult to attribute it to pure chance.⁴

The large number of homophones expressed by hanzi gives a possible explanation for this phenomenon. For example, the second syllable of (1) is *ta*. Since tone is not a feature of English, it can be safely ignored in the process of adapting English words into Chinese. There are about 30 different hanzi that are pronounced *ta* in Mandarin (as well as about 70 *wei* and 17 *ming*.) The choice of those particular hanzi can be attributed to the post-phonological-adaptation writing strategy described above: first adapt the source language model through a phonological sinicisation process; then take the output of that process (three toneless syllables), and choose, from among the homophonous hanzi with those pronunciations, the three whose morphemes can singly or in combination express the meaning of the original.

Attractive though it may be, this explanation only works if the details of phonological adaptation are not considered. None of these three examples is the most faithful adaptation of the English model into Mandarin syllables. In (1), given that Mandarin [w] is the usual

⁴ Chinese is by no means unique in having different connotations attached to different graphic means of representing the same sound. Bolinger (1946) gives examples from English in which different connotations are attached to one spelling variant or another, from <grey> vs. <gray> to the agentive suffixes <-er> vs. <-or>, to <old shop> vs. <Olde Shoppe>. Such examples are relatively rare exceptions, however.

substitute for English [v] in loanwords, we would expect *wai* rather than *wei* in the first syllable⁵ and *min* instead of *ming* in the third. In (2) the first syllable should be *song* rather than *sheng*, and in (3) the first syllable should be *seng* or *sun* rather than *xiang*. The phonetic distortions in (1)–(3) make it clear that the hanzi were not simply chosen from lists of homophonous characters after phonological adaptation was complete, but that the quest for meaning has influenced the ultimate phonological form.

Deviation from the ideal phonological adaptation is not based simply on a vague impulse towards meaning in general, but often is directed towards a very particular meaning. For example, the familiar (4) could just as easily have been rendered into Chinese as (5), which would be phonologically more faithful to the English, and equally meaningful:

4) 可口可樂 *kěkǒu kělè* ‘Coca-Cola’ (‘tasty’+ ‘cola’)

5) *口渴可樂 *kǒukě kělè* (‘thirsty’+ ‘cola’)

Naturally, the marketers of the product who were responsible for coining the Chinese name preferred (4) over (5), despite its lower degree of phonological faithfulness. The choice of hanzi in semanticised loans can be manipulated at will to express the writer’s attitude toward the referent, even when sales of a commercial product are not at stake. For example, (6) and (7) are two common written variants of the same English word.

6) 雅輩 *yǎbèi* ‘yuppie’ (‘elegant’+ ‘generation’)

7) 雅痞 *yǎpǐ* ‘yuppie’ (‘elegant’+ ‘scoundrel’)

Examples (1)–(7) demonstrate that there is a process by which phonetic loans into Chinese are assigned hanzi in a nonrandom, meaningful way. They also show that phonological faithfulness can be sacrificed in some way to enhance expression of meaning. The remainder of this paper will be devoted to exploring how, why, and by whom this alignment of sound and meaning is arrived at, based on the analysis of three different types of loanwords: commercial brand names, geographic place names, and general common vocabulary. Conclusions drawn from that analysis will be applied to larger questions about hanzi and loanwords in general.

2 Characteristics of the corpora

Semanticised phonetic loans are particularly numerous in the brand names of foreign products. In order to better understand the phenomenon by analyzing this particularly rich concentration, I collected a corpus of 537 foreign brand names used in Taiwan from advertising signs, print and electronic ads, and product labels. Brand names are particularly interesting as linguistic data because of the transparency of their origins and motivations. They are carefully designed by specialists with the goal of maximising the memorability of the name and the desirability of the product. Their etymologies are unlikely to get lost in the mists of the receding past, as happens to so many common lexical items, and their forms are standardised (by law!)

Of course, many of these advantages turn into decided disadvantages when one wishes to generalise from brand name data to other types of borrowing. Such clarity of

⁵ Though if this word were originally borrowed into Cantonese, and later relayed by graphic loan into Mandarin, the first syllable would indeed be a regular adaptation.

motivation, careful calculation of creation, and standardisation of finished product are not usually found in other types of loanword. Brand names therefore cannot be used as a simple model for other loanwords, and phenomena observed in brand names cannot be directly generalised to all loanwords; they need to be seen more as artificially bred laboratory animals, whose behavior under carefully controlled conditions reveal the basic mechanics of the model, which once established can form the basis for observation of animals in the wild.

To complement these laboratory rats, I have assembled two other corpora which reflect different sets of motivations and constraints. One is a set of phonetically borrowed place names (toponyms) from a recent map of the world published in Taiwan. They are mostly the names of cities and physical features, chosen from all areas of the world at random. The toponyms were almost exclusively pure phonetic loans, and are very useful for comparison with the highly semanticised brand names.

The other corpus is a set of 163 phonetic loans culled from a much larger corpus of lexical borrowings of common nouns. All are either fully phonetic or semanticised phonetic loans, and all are clearly identified with a foreign source (usually English).

Of 537 brand names, 137 or 25.5 per cent were not phonetic loans (they were either semantic loans, graphic loans, or not loans at all.) Though no such thorough count was done on the other two corpora, it is clear that nearly all toponyms are phonetic loans, while less than 50 per cent of lexical loans in common vocabulary are phonetic loans.

3 Types of phonetic loans

The phonetically borrowed brand names can be divided into three categories based on the degree and type of semanticisation.

3.1 *Coherently semanticised loanwords (CSL)*

These are loans in which the hanzi chosen to approximate the sound of the source language model represent morphemes that can be construed in a semantically compositional way to form a coherent phrase of some sort. For example:

8) 倍耐力 *bèinàilì* ('multiple durability') 'Pirelli' (tires)

9) 雅仕 *yǎshì* ('elegant official') 'Astor' (cigarettes)

10) 力多精 *lìduōjīng* ('essence of great strength') 'Lactogen' (baby formula)

CSLs most resemble folk etymologies, in that the source language sound sequence ends up encoded in a way that resembles as closely as possible a native word or phrase. Definitions of folk etymology differ, with some emphasising the unconscious and erroneous nature of the assumed derivation (Bolinger 1975:406–407, Crystal 1993), and others emphasising the reshaping of the word or phrase (Trask 1993:105). Whatever definition is used, there is always an element of seeking for meaning, and the idea that folk etymology transforms a lexical item into something that better satisfies a craving for meaningfulness. CSLs are consciously designed and therefore lack the unconscious, accidental origin that characterises folk etymology, but they pander to the same desire for transparent meaningfulness in the lexicon.

3.2 Randomly semanticised loanwords (RSL)

Here the hanzi chosen may be semantically related to the product referred to, or may be chosen simply on the basis of positive connotations or auspiciousness. The morphemes normally represented by the hanzi do not form a coherent or well-formed phrase in Chinese:

- 11) 沛綠雅 pèilǜyǎ ('copious' + 'green' + 'elegant') 'Perrier'
- 12) 豐力康 fēnglìkāng ('abundant' + 'strong' + 'healthy') 'Femcare' (formula)
- 13) 喜美 xǐměi ('happy' + 'beauty') (Honda) 'Civic'

In both (11) and (12), most of the hanzi refer directly or obliquely to the product. In (11), 沛 'copious' carries connotations of water, both because of the water radical in the hanzi itself, and because of the compound 沛沛 'copiously flowing, a great flow of water'. 綠 'green', besides its connotations of coolness, is the color of the Perrier bottle. In (12), abundance is associated with having plenty to eat (good nutrition), and strength and health are the qualities that parents want to nurture in their babies. Example (13), on the other hand, illustrates that hanzi may be chosen that have no direct relationship to the product, but that only bring pleasant associations.

While the examples in (11)–(13) are all fully semanticised, in that every hanzi is associated with the meaning or connotation that is to be conveyed, there are also RSLs that are only partially semanticised. Some of their hanzi are clearly semantically motivated, while others seem to be only phonetically motivated:

- 14) 雲絲頓 yúnsīdùn ('cloud' + 'silk' + 'pause') 'Winston' (cigarettes)
- 15) 蜜絲扛 mīsīkǎng ('honey' + 'silk' + 'Buddha') 'Max Factor' (cosmetics)
- 16) 吉比 jībǐ ('auspicious' + 'compare') 'Skippy' (peanut butter)

In (14) both 'cloud' and 'silk' are associated with the smooth flavor that cigarettes like to boast of, while 'pause' is hardly relevant. 'Honey' and 'silk' in (15) resonate with desirable characteristics of cosmetics, but 'Buddha' is from a completely different arena. 吉 'auspicious' has no particular association with peanut butter, but provides overall positive connotations for (16) in a way that 比 'compare' does not.

3.3 Purely phonetic loanwords (PPL)

Loanwords that are rendered using hanzi that carry no special connection to the referent, and no special positive connotations, can be presumed to be chosen strictly on the basis of phonological fidelity. This is not to imply that RSLs must necessarily have been consciously designed to contain meaningful hanzi; it is certainly conceivable that some of the RSLs mentioned above might have acquired their meaningful hanzi purely by chance, or unconsciously. Since it is impractical to track down and interrogate the inventor of each brand name, and since it is always safer to err on the side of caution, it makes sense to simply restrict the category PPL to brand names with no semantically relevant hanzi at all, and avoid the sticky issue of chance vs. intent.

- 17) 米其林 mǐqílin ('rice' + 3pr.pron + 'forest') 'Michelin' (tires)
- 18) 奧斯摩比 àosīmóbǐ ('mysterious' + 'this' + 'rub' + 'compare') 'Oldsmobile'

Out of a total of 400, the proportions of the different types of phonetic loan in the brand names corpus is shown in Table 1.

Table 1: Loan types (brand names)

Coherently Semanticised Loans	141		35.25%
Randomly Semanticised Loans	209		52.25%
Fully Semanticised		117	
Partially Semanticised		92	
Purely Phonetic Loans	50		12.5%

Compare the distribution in the 163 phonetic loans from common vocabulary shown in Table 2:

Table 2: Loan types (common loans)

Coherently Semanticised Loans	58		35.6%
Randomly Semanticised Loans	28		17.2%
Fully Semanticised		5	
Partially Semanticised		23	
Purely Phonetic Loans	77		47.2%

It is striking how similar the percentages of CSLs are in the two corpora. Equally striking is that the preponderance of RSLs over PPLs in the brand names is reversed in the common loans. Combining these two facts with the observation that in the third corpus, the toponyms, CSLs are virtually nonexistent, it becomes clear that different strategies of phonetic borrowing are used in these three areas of the lexicon.

The semanticisation of phonetic loans can be expected to have two different effects on the selection of hanzi, that of increasing phonetic distance from the original, and of increasing the frequency of certain hanzi. The former is the phenomenon seen in examples (1)–(5) above: that the written form of a phonetic loan may have a pronunciation that deviates from the ideal phonological adaptation of the source language model. If none of the homophonous hanzi corresponding to a given syllable have a desirable meaning or connotation, a non-homophone may be chosen on the basis of its more desirable semantics. It makes sense to hypothesise that the most semanticised words (CSLs and fully semanticised RSLs) have undergone this process to the greatest extent, and that they therefore would experience the most phonetic deviation from the source language originals.

As for frequency of hanzi, it also makes sense to assume that positive connotations make certain hanzi more likely to be chosen under a semanticising strategy. Since certain areas of the lexicon use the semanticising strategy more than others, the frequency distribution of hanzi in the different areas should reflect the difference: in brand names semantically potent hanzi should have the highest frequency; in toponyms semantically neutral ones should rank highest; and common loans should be in between. In the following section, both of these hypotheses will be tested.

4 Quantitative analysis: phonetic fidelity

The phonetic fidelity of a given word can be estimated using a scoring system that measures the distance between an actual phonetic loan and the ideal phonological adaptation of the source language model. The basis for comparison is the ideal adaptation arrived at by replacing each of the source-language segments with the Chinese segment most closely matching it in place and manner of articulation, breaking up consonant clusters with an epenthetic vowel ([ə] or [ɔ], depending on the initial, except for sibilant initials, which use the apical vowel), and syllabifying non-nasal final consonants.⁶ Tone is ignored.⁷ The ideal adaptation is then compared to the existing Chinese loan, and discrepancies are rated according to the following scoring scheme:

- 3 pts. missing or extra CVC syllable: 歐舒康 *ōu* __ *shūkāng* 'Orthoxicol'
 2 pts. missing or extra smaller syllable: 愛馬 *àimǎ* __ 'Amana'
 wrong front/backness of vowel: 固齡玉 *gùlǐngyù* 'Kolyinos'
 1 pt. missing consonant: 蘭寇 — *lánkòu* 'Lancome'
 wrong place or manner of articulation of C: 禮蘭 *lǐlán* 'Dearland'
 1 pt. for each deviation in degree of vowel height, or in lip rounding:
 賓士 *bīnshì* 'Benz' (1 pt.)

Allowable exceptions:

- | | |
|---|--------------------------|
| missing post-vocalic /r/ or /l/ | /l/ for /r/ |
| retroflex-dental confusion in sibilants | confusion of /n/ and /ŋ/ |

Allowable conventional Chinese adaptations:

- | | | |
|------------------|---------------------------|--|
| wrong aspiration | <i>jia</i> for /ga/, /ka/ | <i>chu-</i> , <i>tsu-</i> for /dr/, /tr/ |
|------------------|---------------------------|--|

The allowable exceptions and allowable conventional adaptations refer to frequently observed deviations from ideal adaptations. Allowable exceptions are those that are potentially explainable in terms of dialectal phonology (either Taiwan Mandarin or Cantonese).⁸ The conventional adaptations are correspondences that have become so common that they are used regardless of phonetic dissimilarity. For instance, despite the fact that aspiration is an extremely important part of the phonological systems of both Chinese and English, aspiration in Chinese loanwords from English is 'consistently inconsistent'—it seems to be randomly assigned, with no attention to whether or not the English model is aspirated (as noted in Novotna 1968). Likewise, Mandarin *jia* is often used for English /ga/ or /ka/, because a large number of Cantonese-based loanwords from English use hanzi such as 加 (Mandarin *jia*, Cantonese *ka*) for English /ga/ or /ka/, and the

⁶ Novotna (1968) gives a thorough account of the phonological adaptations that phonetic loans into Mandarin undergo.

⁷ Lou (1992) points out that tone is assigned more or less randomly in Mandarin loanwords, due to the influence of transliteration practices. Because in Hong Kong borrowing takes place in a much more oral, face-to-face borrowing situation, Cantonese is the opposite: in loanwords from English, there are very regular correspondences between tone patterns and the stress patterns of the English models (Kiu 1977, Silverman 1992).

⁸ For a systematic analysis of the differences between Taiwan and Mainland Mandarin, see Cheng (1985).

correspondence has become conventional despite the lack of strict phonetic similarity in Mandarin. The *chu-*, *tsu-* for /tr/ conventional adaptation may seem strange in terms of strict phoneme mapping, but is acoustically quite good.

One more adjustment to the coding scheme was the consideration of Cantonese pronunciation. A great many foreign consumer goods available in Taiwan first entered the Greater Chinese market through Hong Kong. Their brand names would therefore be sinicised according to Cantonese pronunciation, with the Mandarin version used in Taiwan created simply by graphic loan from Cantonese. For example:

19) 高露潔 'Colgate' Mand. [kaw lu tɕjɛ] Cant. [kow low kit]

20) 立頓 'Lipton' Mand. [li tuən] Cant. [lip tøn]

In both (19) and (20), missing consonants in the Mandarin pronunciation (the /t/ in *Colgate* and the /p/ in *Lipton*) are present in the Cantonese pronunciation. The Cantonese is therefore much more phonetically faithful than the Mandarin, and has a lower phonetic distance score. In such cases, the lower of the two scores is used in this analysis.⁹

The full operation of phonetic distance scoring is shown in two examples below:

21) English original: 'Gatorade' Ideal adaptation: *gei te lei d(e)*

Actual Chinese: 開特力 *kāitèlì*

Scoring by syllable:

gei - kai 1 point (vowel height). (Aspiration irrelevant)

te - te 0 points (identical)

lei d(e) - li 2 points (vowel height, missing consonant)

Total score: 3

22) English original: 'Quaker' Ideal adaptation: *kui ke*

Actual Chinese: 桂格 *guìgé*

Scoring by syllable:

kui - gui 0 points (Aspiration irrelevant)

ge - ge 0 points (identical)

Total score: 0

Table 3 gives the average phonetic distance scores for the different types of loans (score per hanzi is given to eliminate the possibility of word-length effects):

⁹ Robert Sanders (p.c.) suggests that Shanghai dialect was also a major source of loans, and should be included in the scoring. While I agree in principle, practical matters prevent it at this time.

Table 3: Phonetic distortion scores

		# of words	# of hanzi	word avg.	hanzi avg.
Brand names					
	CSLs	138	354	2.17	0.876
	RSLs	196	508	1.74	0.671
	PPLs	50	126	0.64	0.254
	Total	384	988	1.78	0.691
Toponyms					
		100	316	0.48	0.152
Common loans					
	CSLs	58	127	1.07	0.488
	RSLs	28	69	1.11	0.449
	PPLs	77	175	0.56	0.246
	Total	163	371	0.83	0.367

A brief look at the subcategories within brand names shows that there is certainly a trade-off between phonetic fidelity and semantics. Semanticised loans (CSLs and RSLs) show significantly greater phonetic distance from the foreign model than unsemanticised loans (PPLs). Major sacrifices of phonetic fidelity are being made to accommodate semanticisation. There is also a major discrepancy between fully semanticised and partially semanticised RSLs, as shown below:

Brands RSL Breakdown

Fully semanticised	113	279	2.23	0.903
Partially semanticised	83	229	1.07	0.389

Fully semanticised RSLs show the same high degree of phonetic distance as CSLs, while the partially semanticised are closer to PPLs.

The trade-off is also present in common loanwords, though not nearly to the same extent. PPL phonetic distance for brands and for common loans are virtually identical, while CSLs and RSLs are much lower for common loans than for brands. Toponyms have by far the lowest phonetic distance score, lower than any of the other PPLs.

This numerical analysis of phonetic fidelity in loans, though based on a rough and ready coding system, yields clear-cut results. The three corpora show sharp differences, differences that are easily correlated to their sociolinguistic status under the trade-off hypothesis. Brand names show the greatest phonetic distance, because they are created in a situation that puts a premium on meaningfulness, with phonetic accuracy being a secondary consideration. Common loans are much less market-driven, there is no such imperative to put the item being named in a positive light, though if a memorable and appropriate name should present itself, it will be used. Speakers will accept phonetic distortions compensated for by semantic suggestiveness, but they won't stretch too far for them. Toponyms represent the opposite end of the scale. The highest premium is placed on accuracy, since there is no accessible semantic content to most place names, and there is no motivation for injecting semantic content into them. The prime motivation is to render them in a way that will best approximate the sounds of the name of the place in a foreign language.

4 Hanzi frequency

A comparison of the highest frequency hanzi in the three corpora should also show differences. The existence of a large number of RSLs in the brand names should increase the frequency of hanzi with positive connotations. On the other hand, the phonetic fidelity of toponyms should produce no such effect, and lack of attention to semantics would preclude use of many different homophones for the same syllable.

The high frequency hanzi for each corpus are listed below:

Brands: (out of 325 types, 988 tokens)

Rank	Number	Hanzi	Pronunciation	Meaning
1	23	達	dá	'arrive, attain'
2	22	樂	lè	'joy', 'happy'
	22	美	měi	'beautiful'
4	17	愛	ài	'love'
5	16	麗	lì	'beautiful'
	16	利	lì	'advantage'
	16	士	shì	'scholar, worthy'
8	15	力	lì	'strong'
9	13	雅	yǎ	'elegant'
10	12	百	bǎi	'hundred'
	12	保	bǎo	'treasure'
	12	斯	sī	'this'

The highest frequency hanzi in the brand name corpus almost all refer to the attributes of the good life that consumers (indeed, everyone) wish to attain—love, beauty, strength, plenty, etc. It is not until 10th place that a hanzi with no positive connotations (indeed, no connotative or denotative content at all), 斯 *sī* 'this', is seen.

Toponyms: (240 types, 851 tokens)

Rank	Number	Hanzi	Pronunciation	Meaning
1	40	斯	sī	'this'
	40	拉	lā	'pull'
3	39	亞	yǎ	'inferior, second'
4	27	克	kè	'to overcome'
5	26	加	jiā	'to add'
6	24	巴	bā	'to hope anxiously'
	24	利	lì	'advantage'
8	20	特	tè	'special'
9	19	馬	mǎ	'horse'
	19	爾	ěr	3prs. pron.

The toponym list is strikingly different from the brands list. Only two hanzi appear on both: 利 *lì* 'advantage, benefit' appears in roughly the same position in both; and 斯 *sī* 'this' is first in toponyms (tied for 10th in brand names). Otherwise, the toponym list is almost entirely devoid of positive connotations, and in some cases devoid of any

connotative or denotative meaning whatsoever (爾 ěr 3prs. pron., 斯 sī 'this'). The high frequency hanzi also make up a much higher percentage of the total tokens in toponyms as opposed to brands. The 10 highest frequency hanzi account for 32.67 per cent of all tokens in toponyms, but only 17.4 per cent in brands.

Common loans: (211 types, 371 tokens)

Rank	Number	Hanzi	Pronunciation	Meaning
1	18	克	kè	'to overcome'
2	10	卡	kǎ	'stuck'
3	8	沙	shā	'sand'
4	6	斯	sī	'this'
5	5	特	tè	'special'
	5	司	sī	'preside over'
	5	力	lì	'strong'
	5	達	dá	'arrive, attain'

(8 tied for 9th place)

Once again, the common loans are intermediate between brands and toponyms. Though semantically neutral hanzi predominate, there are a few with positive connotations lower down in the frequency list ('special', 'strong', 'arrive, attain'). The 10 most frequent hanzi account for 18.87 per cent of all tokens, a figure much more similar to brands than to toponyms.

6 Discussion

The above data can be brought to bear directly on questions about the process of hanzi choice in phonetic loans. They indicate clearly that hanzi are not chosen randomly from a set of homophones, since random choice would result in fairly equal distribution of hanzi in loanwords, and is inconsistent with the high frequencies that some hanzi achieve. However, any attempt to find a regular, conventional set of hanzi reserved syllabary-style for transliteration is equally doomed to failure. Such a model would predict that no two homophonous hanzi would appear at high frequencies in loanwords, a claim easily disproven by 利, 麗, and 力 (all pronounced *lì*) which are 5th, 5th (tied), and 8th respectively in frequency for the brand name corpus. This phenomenon is by no means uniform over the three different corpora, however: the fact that the ten highest frequency hanzi in the toponym corpus make up 32.67 per cent of all tokens, compared to just over half that amount for the other two corpora, show that the syllabary model works better for that particular type of loan than for others.

Just as clearly, the semanticisation of phonetic loans cannot be purely a product of sporadic folk etymology. In brand names, this is true because of the circumstances of their creation and dissemination: they are carefully crafted by specialists, and protected from any kind of change by trademark laws. It is true for the other categories of loanword for two reasons: first of all, the different distribution in the different corpora indicates that semanticisation cannot be the result of a random process, but is rather related to the sociolinguistic functions served by loanwords. Secondly, the degree of phonetic distortion

in semanticised loanwords, and especially in RSLs, indicates that in the trade-off between phonetic faithfulness and meaningfulness, speakers are willing to stretch quite a ways to accomplish some kind of meaning, even if it is not a meaning which would conventionally be associated with the referent.

The trade-off involved indicates that semanticisation is not a post-adaptation process, as suggested in Hansell (1989b) and Lou (1992), but is present throughout the borrowing process. Only such a model can explain the large degree of phonetic distortion in so many loans.

The implications of this finding go beyond the study of loanwords, to the conception of what hanzi are and how writing systems in general work. The common tendency to divide writing systems into abstract categories of 'pleremic' vs. 'cenemic' (Haas 1976, 1983) or 'logographic' vs. 'phonographic' (Sampson 1985) ignores the flexibility that they can display in actual use. Even appealing to the notion of cenemic use of an otherwise pleremic writing system (Coulmas 1989, Hansell 1989b) fails to do justice to the facts, since we have seen so many examples above of simultaneous consideration of sound and meaning in hanzi choice. These facts point up the need for a type of analysis of writing that does not depend on purely structural considerations, but on a functional consideration of the communicative goals of the writer and reader as well.

The data presented here also serve as a warning that the role of writing in lexical borrowing in general, and in Chinese loanword phonology in particular, cannot be ignored.¹⁰ The study of writing, and of borrowing phenomena, need to be situated in a framework of broader understanding of the social meaning of linguistic forms, the communicative goals of language users, and the strategies that language users employ to achieve those goals.

References

- Bolinger, Dwight, 1946, Visual morphemes. *Language* 22:333–340.
- 1975, *Aspects of language*. New York: Harcourt, Brace Jovanovich.
- Cheng, Robert, 1985, A comparison of Taiwanese, Taiwan Mandarin, and Peking Mandarin. *Language* 61:352–377.
- Coulmas, Florian, 1989, *The writing systems of the world*. Cambridge, MA: Basil Blackwell.
- Crystal, David, 1993, *An encyclopedic dictionary of language and languages*. London: Blackwell.
- French, M.A., 1976, Observations on the Chinese script and the classification of writing systems. In William Haas ed. 1976a:101–129.
- Erbaugh, Mary S., ed., 2002, *Difficult characters: interdisciplinary studies of Chinese and Japanese*. Columbus, OH: Ohio State University/National East Asian Languages Resource Center.
- Gelb, I.J., 1963, *A study of writing*. Chicago: University of Chicago Press.
- Godwin, Christopher D., 1979, Writing foreign terms in Chinese. *Journal of Chinese Linguistics* 7:246–267.

¹⁰ As acknowledged in Yip (1993: 265).

- Haas, William, ed., 1976a, *Writing without letters*. Manchester, Eng.: Manchester University Press.
- 1976b, Writing: the basic options. In William Haas ed. 1976a:131-208.
- 1983, Determining the level of a script. In Florian Coulmas and Konrad Ehlich eds *Writing in focus*, 15-29. New York: Mouton.
- Hansell, Mark, 1989a, Lexical borrowing in Taiwan. Unpublished UC Berkeley PhD dissertation.
- 1989b. Non-logographic Chinese and the non-alphabetic alphabet. *Berkeley Linguistics Society* 15:102-113.
- 2002, Functional answers to structural problems in thinking about writing. In Mary Erbaugh ed., 124-176.
- Kiu, K.L., 1977, Tonal rules for English loan words in Cantonese. *Journal of the International Phonetics Association* 7:17-22.
- Kratochvil, Paul, 1968, *The Chinese language today*. London: Hutchinson & Co.
- Liang, Shiqiu, 1992, *Far East Chinese-English dictionary* (concise edition). Taipei: Far East Book Company.
- Lou Chengzhao, 1992, Transliterating non-Chinese proper nouns into Chinese: a comparative study of usage in mainland China, Taiwan, and Hong Kong. *International Journal of the Sociology of Language* 97:121-133.
- Novotna, Zdenka, 1968, Contributions to the study of loan-words and hybrid words in modern Chinese, part II. *Archiv Orientalni* 36:295-325.
- Sampson, Geoffrey, 1985, *Writing systems*. Stanford, CA: Stanford University Press.
- Silverman, Daniel, 1992, Multiple scansion in loanword phonology: evidence from Cantonese. *Phonology* 9:289-328.
- Tang Ting-Chi, 1989, Xincichuangzao yu Hanyucifa [The creation of new words and Chinese morphology]. In Tang Ting-Chi ed. *Hanyucifajufa xuji [Further writings on Chinese morphology and syntax]*, 93-146. Taipei: Student Book Company.
- Trask, Robert L., 1993, *A dictionary of grammatical terms in linguistics*. London: Routledge.
- Yao Rongsong, 1992, Taiwan xianxing wailaiyude wenti [Problems of current loan vocabulary in Taiwan]. *Bulletin of National Taiwan Normal University* 37:329-362.
- Yip, Moira, 1993, Cantonese loanword phonology and optimality theory. *Journal of East Asian Linguistics* 2:261-291.