LANGUAGE AND LINGUISTICS 5.3:643-662, 2004
2004-0-005-003-000087-1

Basic Words and Language Evolution*

Feng Wang and William S.-Y. Wang
City University of Hong Kong

Swadesh (1952, 1955) proposed using basic-word lists for his studies in glottochronology. We have used these same lists to investigate language evolution. Chen (1996) distinguished two subgroups in Swadesh's 200-word list, placing half of the words in a high rank and half in a low rank; see §2 High-rank words are more stable through time and less likely to be loanwords; borrowed elements tend to occur more frequently among low-rank words. This characteristic of the Swadesh list can be usefully exploited to distinguish lexical retentions from borrowings; this kind of tool can be particularly useful when the evolution of a language has been affected by language contact, as was the case in the development of Middle Chinese entering tones in Pekinese. Finally we compare Dolgopolsky (1964)'s 15-word list, Yakhontov's 35-word list, and Swadesh, and conclude that Swadesh (1955)'s 100 basic words (high rank) are best for sub-grouping Chinese dialects.

Key words: Swadesh list, basic words, rank, Chinese dialect, genetic classification

1. Introduction

Ever since Swadesh (1952, 1955) came out with his 100- and 200-word lists of basic words[1] for use in glottochronology, various scholars have produced other basic word lists, such as Dolgopolsky (1964)'s with 15 items or Yakhontov's with 35.[2] Many arguments have focused on whether the rate of change in basic vocabulary is constant, or whether such lists are even suitable for historical-linguistic dating at all. Cavalli & Wang (1986) and Starostin (1991) have investigated variation in the rate of lexical replacement. In this paper we shall explore other uses for basic words. In Section 2, the

* This work was supported in part by grants from the City University of Hong Kong [#9010001], the Hong Kong Research Grants Council [#9040781], and Academia Sinica [Taiwan]. We thank Prof. Tongqiang Xu, who provided the data on some Chinese dialects. We would like to give many thanks to Prof. Baoya Chen and Dr. James W. Minett for kind comments and constructive suggestions. We also thank Prof. Dah-an Ho and Ms. Jinyun Ke for helpful discussion. Thanks are also due to two anonymous reviewers for helpful comments.

1 Some scholars call them kernel words or core words.

2 Yakhontov's list is cited from Starostin (1991:59-60).

ranks of basic words proposed in Chen (1996) will be taken as a baseline to distinguish borrowing from retention.[3] Chen (1996) splits Swadesh's 200-word list into two ranks: a **high rank** consisting of the 100 basic words proposed by Swadesh (1955), and a **low rank** based on the 200-word list in Swadesh (1952) with the high rank words removed.[4] Chen (1996) found that words in the high rank tend to be more stable and loan-resistant than those in the low rank; i.e., the high rank has more retentions, while the low rank tends to be influenced by more frequent borrowing. Based on this finding, we compare the regrouping patterns of Middle Chinese (henceforth MC) entering tones in Pekinese in order to distinguish retention and borrowing in Pekinese. In Section 3, we test the three important lists by Swadesh (1955), Yakhontov, and Dolgopolsky (1964), in order to find out which is best for sub-grouping Chinese dialects. In this test a morphological limitation is applied: If all corresponding words of a basic word are compounds in all the tested languages, that basic word will be discarded. Given this morphological limitation, Swadesh (1955)'s list generates the best genetic tree for Chinese dialects.

2. Ranks in basic words

As mentioned above, Chen (1996) has proposed splitting Swadesh's 200-word list into two subgroups: (1) a **high rank** group consisting of Swadesh (1955)'s 100-word list, in which diachronically items are relatively stable and loan-resistant; and (2) a **low rank** group consisting of the 200-word list (Swadesh 1952) minus the 100-word list (Swadesh 1955), in which vocabulary replacement occurs at a greater rate than for high-ranked words. Assuming different rates of change for the two ranks, Chen (1996) has devised a method for evaluating genetic relationships between languages. Words with sound correspondences between languages are called **related words**. Chen proposes that genetically related languages have a greater number of related words in the high rank than in the low rank. On the other hand, if the relationship were due to language contact, then we should expect that the number of related words in the high rank would be less than that in the low rank. This method was tested on data for languages from several well-established language families, including Indo-European and Chinese. The relationships established using this method concur with received opinion in these matters. Applying this method to other aspects of language evolution may shed further light on the origin of linguistic elements.

[3] Starostin (1991) proposed a similar treatment of dividing basic words into two sets according to their replacement rates. See also Comrie (1993) for further discussion.

[4] The remainder is a 107-word list, rounded off to 100. Cf. Chen 1996:297.

For a long time now, a very complex and intriguing problem in Pekinese has been how the MC entering tone (*ru sheng*) was regrouped into the three other tones,[5] a process succinctly described in Chinese as 入派三聲. Many a scholar has tried to determine whether the resultant regroupings have resulted from internal or contact-induced change; it is indeed difficult to distinguish what has been inherited from what has been borrowed in the process.

Bai (1931), Forrest (1950), Hirayama (1960, 1990), Stimson (1962), Hsieh (1971), Lin (1992), Ting (1998), and Chen (1999) have all attacked the problem in different ways. Hirayama (1960, 1990) and Stimson (1962) believe that the irregularities in Pekinese are due to a mixing of different dialects. Stimson (1962) devised four "strains" to explain the deposit in Pekinese. Arguing for lexical diffusion, Hsieh (1971) contends that the varieties are residues of uncompleted sound changes in different periods. However, Ting (1998) discounts such an interpretation as much too complicated to be taken seriously.

In discussing the evolution of MC tones, we refer to the traditional Song dynasty classification of initial consonants by which tonal changes are conditioned, using a romanized notation based on Wang (1996):

全清 uu = unvoiced, unaspirated
次清 ua = unvoiced, aspirated
全濁 vo = voiced, obstruent
次濁 vs = voiced, sonorant

As already mentioned, Stimson (1962) assigns Pekinese readings to four separate strains, labeled PA, PB, PC and PD, which are transcribed as below:

MC	PA	PB	PC	PD
vo	*yang ping*	*yin ping*	*qu sheng*	*yang ping*
vs	*qu sheng*		*yang ping*	*qu sheng*
uu	*shang sheng*			*yang ping*
ua			*qu sheng*	*qu sheng*

He then states (1962:383):

> A method for determining which of several strains is inherited in a language is suggested in an article by Isidore Dyen … Lg.32.83-7 (1956). This method is

[5] In MC, the four tones are 平 *ping sheng* 'level tone', 上 *shang sheng* 'rising tone', 去 *qu sheng* 'falling tone' and 入 *ru sheng* 'entering tone'.

quite straightforward: the strain most frequently represented in a short list of basic words is to be considered the inherited strain.

He found 20 readings for the strain PA among the 33 reflexes of MC[6] entering tone syllables in Swadesh's 200-word list. "Thus it is possible to identify the inherited strain as PA."

However, how frequently a strain is respresented among the basic words may not be the point, since it is well known that there is no limit to borrowing. It is possible that a borrowing strain is respresented more frequently in the basic vocabulary than the inherited one if contact had been heavy.

We now make use of Chen (1996)'s concept of high rank and low rank to distinguish retention and borrowing. According to this method, were we compare the different regrouping patterns of entering tones in the high and low ranks, a pattern representing an inherited strain should emerge. If a pattern had been borrowed, low rank words would have been affected first, and then high rank words. Thus a pattern occurring only among high rank words should have been inherited from the ancestral language. Conversely, if a corresponding pattern occurs only among low rank words, it must have been borrowed.

Below are two rank tables of correspondences between MC and modern Pekinese.

In high rank:

Initial in MC	Tone in Pekinese	Examples
vo	*yang ping*	舌白石
vs	*qu sheng*	月熱葉綠肉
uu & ua	*qu sheng*	血發不
	yin ping	殺吃說一虱膝黑
	shang sheng	骨腳給角

In low rank:

Initial in MC	Tone in Pekinese	Examples
vo	*yang ping*	活薄直
vs	*qu sheng*	獵[7]
uu & ua	*yin ping*	挖擦壓吸濕
	shang sheng	窄雪
	yang ping	結

[6] Stimson called MC Ancient Chinese (AC).

[7] "拉" is *yin ping* in Pekinese, not *qu sheng*. The reason for this irregularity is unknown.

In both ranks, vo initials cause the entering tones to become *yang ping* and vs initials to become *qu sheng*. The problem is that in Pekinese serveral tones correspond with the entering tone with unvoiced initials in MC without any sound condition. In both ranks, examples corresponding with *yin ping* and *shang sheng* can be found, but some syllables change into *qu sheng* in the high rank basic words, for example, 血發不, while there are no similar examples in the low rank. In low rank, an additional tone corresponding to the entering tone with unvoiced initials in MC is *yang ping*.

If we extend Chen (1996)'s method, a corresponding pattern appearing only in the high rank, not in the low rank, is to be considered as inherited. In Pekinese, this would be the case with *qu sheng* corresponding to MC entering tone with unvoiced initial. If it had been borrowed, there should be some examples to show the same pattern in the low rank, since borrowing will first influence low rank words. But there is no such trace in Pekinese. On the other hand, a corresponding pattern appearing only in the low rank, but not in the high rank, is very likely to have been borrowed. Such a case is *yang ping* corresponding to MC entering tone with unvoiced initial.

Therefore, if we acknowledge that the MC entering tone with unvoiced initial corresponding to *qu sheng* represents the inherited strain, then *yin ping* and *shang sheng* must have been borrowed from dialects in close contact with the Pekinese of that time. *Yang ping* variations, however, must have been borrowed from another dialect not having as strong an influence on Pekinese as the other two. The following table summarizes the evolution of the MC entering tone to the corresponding Pekinese categories:

Initial in MC	Inherited strain	Borrowed strain (D1)	Borrowed strain (D2)	Borrowed strain (D3)
uu & ua	*qu sheng*	*yin ping*	*shang sheng*	*yang ping*
vs	*qu sheng*	*qu sheng*	*qu sheng*	—
vo	*yang ping*	*yang ping*	*yang ping*	—

Ting (1998) has a different hypothesis, according to which four dialects of Mandarin overlap in Pekinese, resulting in complex correspondences between the MC entering tone and the tones of modern Pekinese. These four dialects are tabulated as follows:

Initial in MC	Jiaoliao Mandarin 膠遼官話	North Mandarin 北方官話 Shiji Pian 石濟片	Zhongyuan Mandarin 中原官話 Wuhe/Fengyang 五河/鳳陽	Southwest Mandarin 西南官話 Sichuan/Yunnan 四川/雲南
ua & uu	*shang sheng*	*yin ping*	*qu sheng*	*yang ping*
vs	*qu sheng*	*qu sheng*	*qu sheng*	*yang ping*
vo	*yang ping*	*yang ping*	*qu sheng*	*yang ping*

Ting (1998)'s proposal, however, has a problem. If his Pekinese pattern had been due to contact with the four dialects in his table, we should have expected different Pekinese correspondences to the MC entering tone with vo, vs, or unvoiced initial, because the dialectal patterns corresponding to the MC entering tone are different not only with MC unvoiced initials, but also with vo and vs initials. But, as we have previously explained, the only variation is in fact in the modern Pekinese pattern corresponding to the unvoiced initial.

Our analysis based on basic words suggests that in the inherited strain of Pekinese the entering tone with unvoiced or vs initials changed into *qu sheng*, while the entering tone with vo initials changed into *yang ping*. Pekinese then came into contact with two dialects (D1 and D2), which had developed different tones corresponding to the MC entering tone with unvoiced initials, while having identical correspondences with other initials. A third dialect (D3) may also have interfered with Pekinese at an early date, but its influence would not have weighed so heavily as was the case with D1 or D2, since it has left no trace among the high rank words of Pekinese.

Our interpretation is corroborated by other sources. First of all, Guo (1986, 1997) and Chen (1999) have shown that all entering tones with unvoiced initials in earlier Pekinese changed into *qu sheng* before the Ming dynasty. Chen (1999) points out that according to the statistics and analysis of Guo (1986, 1997), the entering tone with unvoiced initials in MC has variant reflexes during the Ming dynasty: literary pronunciations were *qu sheng*, while colloquial pronunciations were *yin ping*, *shang sheng*, or *yang ping*. The literary system of Pekinese was inherited since Pekinese had been the prestige dialect ever since the Yuan dynasty (1206-1367), while the colloquial systems were borrowed. All 2,738 characters with unvoiced entering tones have a literary reading with *qu sheng* in the Ming dynasty (1368-1644). This supports our hypothesis concerning the identification of the inherited strain in Pekinese.

Secondly, historical records of population movements into modern-day Pekinese-speaking areas lend support to our hyphothesis concerning the borrowed strains. According to Cao (1997:216-243), people had been migrating from Shandong Province into the area concerned from at least the beginning of the Ming dynasty. According to Lin (1987), around the time of the Qing dynasty (1616-1911), whole populations from

Shandong province were forced to migrate there. These mass migrations certainly had important effects on Pekinese. The following table gives some indication of the dialect situation in Shandong Province, after Qian et al. (2001:21):

Initial in MC	Eastern dialects	Most of Western dialects
uu & ua	*shang sheng*	*yin ping*
vs	*qu sheng*	*qu sheng*
vo	*yang ping*	*yang ping*

Obviously, patterns in the eastern and western dialects of Shandong coïncide with D2 and D1 respectively. D3 may be a Southwestern Mandarin variety, brought into the Pekinese-speaking region during the Ming dynasty. The following rough map sketches out this interpretation of population movements.

Figure 1: Migrations to the Pekinese-speaking area
The two thicker lines indicate strains from east and west Shandong; the thinner line indicates a population movement from a southwestern Mandarin region.

3. Basic words for genetic classification

The 200-basic-word list was originally proposed in Swadesh (1952) for use in glottochronology. Swadesh's initial proposal was that this vocabulary would have a stable enough rate of replacement to allow reliable dating, but he eventually boiled it

down to 100 words (Swadesh 1955), since half the original list proved insufficiently stable. Since then, many scholars have even questioned the reliability of this second, shorter list, casting into doubt the value of the 100-word list for historical linguistics. Consequently, linguists have constructed other basic-word lists, such as Dolgopolsky (1964)'s 15-word list or Yakhontov's 35-word list. The purpose of these more recent lists remains the same: genetic classification. Their appropriateness for genetic classification will be a key criterion in evaluating such lists.

To do just this—test the fitness of different basic-word lists—we have devised an algorithm. Chinese dialects from ten regions were selected for testing: Beijing (B), Yingshan (Y), Suzhou (S), Shanghai (H), Shuangfeng (F), Changsha (C), Nanchang (N), Guangzhou (G), Meixian (M), and Xiamen (X). Note that B and Y are both Mandarin dialects, that S and H are both Wu, and that C and F are both Xiang. The genetic relationships among these three pairs of dialects can be taken for granted. Any newly proposed classification must result in these three pairs being placed in the correct sub-group; and any classification that fails to do so must be rejected. These three pairs are therefore be taken as a basic index of the fitness of basic words in genetic classification. PHYLIP software is used to draw genetic trees based on different lists.[8] The first step is to determine which words in the various dialects are cognate, so that a similarity matrix can be constructed. This similarity matrix, however, must be transformed into a distance matrix, since the branches on a genetic tree must show distance, not similarity. This is done by deriving each distance d from the negative logarithm of each similarity s: $d = -\log s$. The input for the PHYLIP software is the distance matrix, and the output is a genetic tree.[9]

Before proceeding with discussion of the test, we shall restate our morphological criterion for applying the basic-word list to individual languages. It is commonly assumed that root-words are more basic than compound words. In a given language, for instance, the notion 'moon' might be rendered by a descriptive such as "eye of the night". Since it is composed of the more basic, irreducible forms 'eye' and 'night', this 'moon' compound must be discarded from our basic-word list. Cf. Sapir (1916:434):

> One of the most useful principles for the determination of the age of a word is a consideration of its form; that is, whether it can be analysed into simpler elements, its significance being made up of the sum of these, or is a simple irreducible term. In the former case we suspect, generally speaking, a secondary or relative late formation, in the latter considerable antiquity.

[8] A somewhat similar idea is discussed in Wang (1997).

[9] Details about the application are discussed in Saitou & Nei (1987) and Wang (1997).

Our strategy in constructing a basic word list is to exclude any word that fails to satisfy this morphological criterion according to evidence from the various languages under consideration. We then obtain a list of **relative basic words** for different languages. E.g., the 'bark (of a tree)' would have to be excluded from any Chinese list, since every dialect has 'tree-skin' (樹皮) for this notion.

The five basic-word lists enumerated below are the ones we shall be considering.

List 1: Dolgopolsky (1964) has investigated the stability of 15 meanings among the language families of Northern Eurasia. His list consists of: *first person marker*, *two*, *second person marker*, *who*/*what*, *tongue*, *name*, *eye*, *heart*, *tooth*, *verbal negation*, *finger-nail*/*toe-nail*, *louse*, *tear* (noun), *water*, and *dead*. Because *who/what* corresponds to two words in every Chinese dialect, we have split this item, actually resulting in a 16-word list.

List 2: The two words *tear* and *nail* in the preceding 16-word list are compounds in every Chinese dialect, and are thus discarded according to the morphological criterion. We therefore obtain a 14-word list.

List 3: Yakhontov has proposed these as the 35 most stable meanings: *blood*, *bone*, *die*, *dog*, *ear*, *egg*, *eye*, *fire*, *fish*, *full*, *give*, *hand*, *horn*, *I*, *know*, *louse*, *moon*, *name*, *new*, *nose*, *one*, *salt*, *stone*, *sun*, *tail*, *this*, *thou*, *tongue*, *tooth*, *two*, *water*, *what*, *who*, *wind*, *year*. None of these items needs discarding from a Chinese dialect list, since they would all satisfy the morphological criterion.

List 4: Swadesh's 100 basic-word list.

List 5: According to the morphological criterion, the five words *hair*, *swim*, *woman*, *man*, and *bark* in the fourth list will be excluded. This fifth list will then consist of the remaining 95 items.

And these then are the two tasks to be applied to the Dolgopolsky, Yakhontov, and Swadesh lists: a test of the fitness of the word lists for use in historical-comparative linguistics; and the application of the morphological criterion to restrict the lists to simple irreducible terms.

To represent genetic relationships among language varieties, we shall be using a special notation; e.g., given languages A, B, and C, ((A,B)C) means that A and B are closer to each other than either is to C. Applying the PHYLIP program to the lists by Dolgopolsky and Yakhontov, we obtain the following genetic relationships:

List	Result
14-word	((S,H)((M,X)((N(C,F,Y))(G,B))))
16-word	((((G(((C,Y)F)(B,N)))(S,H))(M,X))
35-word	(((B,G)(M,X))((S,H)((C,F)(N,Y))))

Note that for all three of these lists, B and Y are separated, an unacceptable outcome going against common opinion. The results show that the 14-, 16- and 35-word lists do not satisfy our expectation of a sub-grouping (B,Y). We must conclude that neither Yakhontov's word list nor Dolgopolsky's is suitable for the sub-grouping of Chinese dialects.

On the other hand, our three pairs of dialects do form subgroups as expected when we input either list 4 (Swadesh 100) or list 5 (Swadesh 95). In order to compare them, the additional parameter of **stability** is introduced to measure results. In this test, the standard representatives (B, N, C, S, G, M, and X) of the seven major Chinese dialects (Mandarin, Gan, Xiang, Wu, Yue, Hakka, and Min) are used as fixed items; and the three remaining dialects (F in Xiang, H in Wu and Y in Mandarin), whose genetic positions are well known, are taken as optional items.[10] We assume the adding of optional items does not affect the topology of the fixed items very much, as long as the right tree is generated based on a certain list. Inputting each list, we obtain a group of topologies for the fixed items by adding optional items to fixed items one by one. The distances between the topologies in each group are calculated as the index of stability of topology, with the smallest indicating the best fitness of a list. The results are as follows:

(1) 100-word list:

Optional item	Result
+F	(((B(S(N(C,F))))G)(M,X))
+H	(((B(S(C,N)))G) (M,X))
+Y	(((((B,Y)C)(S,N))G) (M,X))
+F,+H	((((B(S,H))(N(C,F)))G) (M,X))
+F,+Y	(((((B,Y)(C,F))(S,N))G) (M,X))
+Y,+H	((((B,Y) C) ((S,H)N) G) (M,X))
+F,+H,+Y	(((((B,Y)(C,F))(N(S,H)))G) (M,X))

(2) 95-word list:

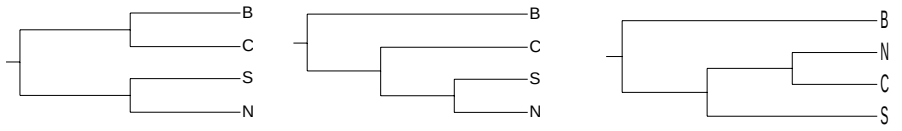
Optional item	Result
+F	(((B((S,N)(C,F)))G) (M,X))
+H	(((B((S,H)(N,C)))G) (M,X))
+Y	(((((B,Y)C)(S,N))G) (M,X))
+F,+H	((((B(S,H))(N(C,F)))G) (M,X))
+F,+Y	(((((B,Y)(C,F))(S,N))G) (M,X))
+Y,+H	(((((B,Y)C)((S,H)N))G) (M,X))
+F,+Y,+H	(((((B,Y)(C,F))((S,H)N))G) (M,X))

[10] We wish to thank the following who provided us with data on the Chinese dialects: Mei Fang, Xiaofan Li (and his student Yan Xiong), Eric Zee, Yun Mai, Xiuhong Yan, Baokui Ye (and his student Ruiyuan Xu).

Comparing the topologies of the seven fixed dialects, the positions of G, M, and X are always unchanged. This means that they do not provide any diagnostic information about topologies, and they are be ignored when differences between topologies are calculated. For the 100-basic-word list, we obtain three different types: 1. (((N,C)S)B); 2. ((B,C)(S,N))); 3. ((B,S)(N,C)). For the 95-word list, we obtained four types: 1. (((S,N)C)B); 2. (B(S(N,C))); 3. ((B,S)(N,C)); 4. ((B,C)(S,N)). The minimum movements from one topology to another will measure the topological distance. For example:

Dist { ((B,C)(S,N)) , (((N,C)S)B) }:

((B,C)(S,N)) ---> (B(C (S,N))) ---> (B (S (C,N)))=(((N,C) S) B)



2 movements => distance = 2

According to this algorithm, the sum of the distances between all topologies based on the 100 words is 24, while the sum for the 95 words is 22. That is to say, the 95 words result in a more stable topology for the seven major Chinese dialects. This suggests that the 95-word list is more suitable for Chinese dialects.

In order to test the effectiveness of deleting items from a list in accord with the morphological criterion, we randomly removed five words and repeated the procedure to classify Chinese dialects described above. Three experiments were conducted. The three groups of five words are: (1) *skin*, *knee*, *ash*, *stone*, *I*; (2) *nose*, *smoke*, *walk*, *seed*, *dog*; (3) *bird*, *grease*, *star*, *all*, *cloud*. For the three cases of random deletion, the sums of the distances between the topologies are 24, 24, and 34, respectively. All these figures are larger than the sum 22, which is the result when the deletions were based on the morphological criterion. The tests prove that Swadesh's 100 basic words adjusted by the morphological criterion result in the best fitness in the genetic classification of Chinese dialects.

4. Discussion

Basic words are an important window on language evolution. In this particular case, we explore the distribution of basic words in a group of Chinese dialects. The distinction between high vs. low rank (Chen 1996) in basic words is used to distinguish retention from borrowing, with high-rank words showing a fitness in application to the genetic classification of Chinese dialects. However, each basic word may have a

particular replacement rate, maintained on average in different language groups; cf. Cavalli & Wang 1986. There is still no convincing answer as to why the boundary between high vs. low words should be set as in Chen (1996). The sorting of basic words according to their rates of change is a problem needing much more substantive research. It may be necessary to look for **factors causing variation** in the replacement of words. Under "well-controlled" conditions, we might find out if there is a universal boundary between high and low rank basic words in the world's languages, and how to determine it. Whatever the answers to these questions may be, it is exploration and discovery that will shed more light on language evolution.

Appendix A: High rank words **(Words with entering tone are underscored.)**

1 I 我	2 you 你	3 we 我們	4 this 這	5 that 那
6 who 誰	7 what 什麼	8 not 不	9 all 全部	10 many 多
11 one 一	12 two 二	13 big 大	14 long 長	15 small 小
16 women 女人	17 man 男人	18 person 人	19 fish 魚	20 bird 鳥
21 dog 狗	22 louse 蝨子	23 tree 樹	24 seed 種子	25 leaf 葉子
26 root 根	27 bark 樹皮	28 skin 皮膚	29 flesh 肉	30 blood 血
31 bone 骨頭	32 grease 脂肪	33 egg 雞蛋	34 horn 角	35 tail 尾巴
36 feather 羽毛	37 hair 頭髮	38 head 頭	39 ear 耳朵	40 eye 眼睛
41 nose 鼻子	42 mouth 嘴	43 tooth 牙齒	44 tongue 舌頭	45 claw 爪子
46 foot 腳	47 knee 膝蓋	48 hand 手	49 belly 肚子	50 neck 脖子
51 breasts 乳房	52 heart 心臟	53 liver 肝	54 drink 喝	55 eat 吃
56 bite 咬	57 see 看見	58 hear 聽到	59 know 知道	60 sleep 睡
61 die 死	62 kill 殺	63 swim 游水	64 fly 飛	65 walk 走
66 come 來	67 lie 躺	68 sit 坐	69 stand 站	70 give 給
71 say 說	72 sun 太陽	73 moon 月亮	74 star 星星	75 water 水
76 rain 雨	77 stone 石頭	78 sand 沙子	79 earth 土地	80 cloud 雲
81 smoke 煙	82 fire 火	83 ash 灰	84 burn 燒	85 path 路
86 mountain 山	87 red 紅	88 green 綠	89 yellow 黃	90 white 白
91 black 黑	92 night 晚上	93 hot 熱	94 cold 冷	95 full 滿
96 new 新	97 good 好	98 round 圓	99 dry 幹	100 name 名字

Appendix B: Low rank words

1 and 和	2 animal 動物	3 back 背	4 bad 壞	5 because 因爲
6 blow 吹	7 breathe 呼吸	8 child 孩子	9 count 數	10 cut 砍
11 day 天	12 dig 挖	13 dirty 髒	14 dull 呆、笨	15 dust 塵土
16 fall 掉	17 far 遠	18 father 父親	19 fear 怕	20 few 少
21 fight 打架	22 five 五	23 float 漂浮	24 flow 流	25 flower 花
26 fog 霧	27 four 四	28 freeze 結冰	29 fruit 水果	30 grass 草
31 guts 腸子	32 he 他	33 here 這裏	34 hit 打	35 hold/take 拿
36 how 怎麼	37 hunt 打獵	38 husband 丈夫	39 ice 冰	40 if 如果
41 in 在	42 lake 湖	43 laugh 笑	44 leftside 左邊	45 leg 腿
46 live(alive) 活的	47 mother 母親	48 narrow 窄	49 near 近	50 old 老
51 play 玩	52 pull 拉	53 push 推	54 rightside 右邊	55 correct 對
56 river 江	57 rope 繩子	58 rotten 腐爛	59 rub 擦	60 salt 鹽
61 scratch 抓	62 sea 海	63 sew 縫	64 sharp 尖	65 short 短
66 sing 唱	67 sky 天空	68 smell 聞	69 smooth 平	70 snake 蛇
71 snow 雪	72 spit 吐	73 split 撕裂	74 squeeze 壓	75 stab 刺
76 stick 棍子	77 straight 直	78 suck 吮	79 swell 腫	80 there 那兒
81 they 他們	82 thick 厚	83 thin 薄	84 think 想	85 three 三
86 throw 扔	87 tie 捆	88 turn 轉	89 vomit 嘔吐	90 wash 洗
91 wet 濕	92 where 哪里	93 wide 寬	94 wife 妻子	95 wind 風
96 wing 翅膀	97 heavy 重	98 woods 森林	99 worm 蟲	100 year 年

Appendix C: 100 Basic Words in the Chinese dialects

Notes: (1) Words requiring rare Chinese characters or having no associated character at all are represented by upper-case roman letters. (2) For two dialects to be considered as having cognate forms, the latter must correspond exactly; e.g., s.v. *black*, F with 黑/青 will be counted as different from H with 黑.

	G	S	M	N	Y	B	C	F	H	X
all	鹹	通	完	都	下	整	下	下	全	全
ash	灰	灰	灰	灰	灰	灰	灰	灰	灰	灰
bark	樹皮	樹皮	樹皮	樹皮	樹皮	樹皮	樹皮	樹皮	樹皮	樹皮
belly	肚	肚	肚	肚	肚	肚	肚	肚	肚	腹肚
big	大	大	大	大	大	大	大	大	A	大
bird	雀/鳥	鳥	鳥	鳥	雀	鳥	鳥	鳥	鳥	A
bite	咬	咬	咬/齧	咬	咬	咬	咬	咬	咬	咬
black	黑	黑	烏	烏/青/黑	黑	黑	黑/青	黑/青	黑	烏
blood	血	血	血	血	血	血	血	血	血	血
bone	骨	骨	骨	骨	骨	骨	骨	骨	骨	骨
breasts	A	奶	奶	奶	媽	咂/媽	奶	奶	奶	奶
burn	燒	燒	燒	燒	燒	燒/著	燒	燒	燒	燒/熱
claw	爪	腳爪	腳爪	爪/腳爪	爪	爪	爪	爪	腳爪	爪
cloud	雲	雲	雲	雲	雲	雲	雲	雲	雲	雲
cold	凍/冷	冷	冷	冷	冷	冷	冷/清	冷/清	冷	冷/寒
come	來/嚟	來	來	來	來	來	來	來	來	來
die	死	死	死	死	死	死	死	死	死	死
dog	狗	狗	狗	狗	狗	狗	狗	狗	狗	狗
drink	飲	吃	食	吃	喝	喝	吃	吃/呵	吃	啉/啜
dry	乾	乾/燥	A	乾	乾	乾	乾	乾/A	乾	焦
ear	耳	耳	耳	耳	耳	耳	耳	耳	耳	耳
earth	地	地	地	地	地	地	地	地	地	地
eat	吃	吃	食	吃	吃	吃	吃	吃	吃	食
egg	春	蛋	卵	蛋	蛋	雞子/蛋	蛋	蛋	蛋	卵
eye	眼	眼	目	眼	眼	眼	眼	眼	眼	目睞
feather	毛	羽毛	毛	毛	毛	毛	毛	毛	毛	毛
fire	火	火	火	火	火	火	火	火	火	火
fish	魚	魚	魚	魚	魚	魚	魚	魚	魚	魚
flesh	肉	肉	肉	肉	肉	肉	肉	肉	肉	A
fly	飛	飛	飛	飛	飛	飛	飛	飛	飛	飛

	G	S	M	N	Y	B	C	F	H	X
foot	腳	腳	腳	腳	腳	腳	腳	腳	腳	骹
full	滿	滿	滿	滿	滿	滿	滿	滿	滿	A
give	畀	撥	分	把/給	把	給	把	B	撥	A
good	好/A	好/美/贊	好/A	好	好	好/強	好	好	好/靈光	好
grease	油/膏	油	油	油	油/膘	大油/葷油	油	油	油	油
green	綠	綠	青	綠	綠	綠	綠	綠	綠	青
hair	頭髮	頭髮	頭顱毛	頭髮	頭髮	頭髮	頭髮	頭髮	頭髮	頭毛
hand	手	手	手	手	手	手	手	手	手	手
head	頭	頭	頭顱	頭	腦	腦	腦	腦	頭	頭
hear	聽	聽	聽	聽	聽	聽	聽	聽	聽	聽
heart	心	心	心	心	心	心	心	心	心	心
horn	角	角	角	角	角	角	角	角	角	角
hot	熱/慶	熱	燒	熱	熱	熱	熱	熱	熱	熱/燒
I	我	我	我	我	我	我	我	我/卬	我	我
kill	殺	殺	殺	殺	殺	殺	殺	殺	殺	殺
knee	膝	膝饅頭/腳饅頭	膝	舌頭蓋	膝	膊楞蓋兒	膝	膝	腳饅頭	骹頭 A
know	知	曉	知	曉	曉	知	曉	曉	曉	知
leaf	葉	葉	葉	葉	葉	葉	葉	葉	葉	箬
lie	睏	睏/A	眠	睏	睡	躺	困	睏	睏	倒
liver	肝	肝	肝	肝	肝	肝	肝	肝	肝	肝
long	長	長	長	長	長	長	長	長	長	長
louse	虱	虱	虱	虱	虱	虱	虱	虱	虱	虱
man	男人/佬	男/男子客	男子人	男個	男的/男將	男的/爺們	男的/男人家	男人家	男人	A 夫/A 夫儂
many	多	多/交關	多	多	多	多	多	多	多	A
moon	月	月	月	月	月	月	月	月	月	月
mountain	山	山	山	山	山	山	山	山	山	山
mouth	嘴/口	嘴	啜	嘴	嘰	嘴	嘴	嘴	嘴	喙
name	名	名	名	名	名	名	名	名	名	名
neck	頸	頸/頭根	頸	頸	頸	脖	頸	頸	頸	頷管
new	新	新	新	新	新	新	新	新	新	新
night	夜晚/晚黑	夜	夜	夜	黑	夜/黑	夜	夜	夜	下昏時/冥時/暗時
nose	鼻	鼻	鼻	鼻	鼻	鼻	鼻	鼻	鼻	鼻
not	唔/冇	勿	唔	不	不	不	不	不	勿	呣/無
one	一	一	一	一	一	一	一	一	一	一
path	路	路	路	路	路	道/路	路	路	路	路
person	人	人	人	人	人	人	人	人	人	人
rain	雨	雨	雨	雨	雨	雨	雨	雨	雨	雨

	G	S	M	N	Y	B	C	F	H	X
red	紅	紅	紅/赤	紅	紅	紅	紅	紅	紅	紅
root	根/蓪	根	根	根	根	根	根/筧	根/莧	根	根
round	圓	圓	圓	圓	團	圓	圓	欒	圓	圓
sand	沙	沙	沙	沙	沙	沙	沙	沙	沙	沙
say	講/話	說/講	講/話	話	說	說	講	話/講	講	講
see	睇	看/望	看	看/望/妻*	看	看/瞧/瞅/A	看	看/相	看	看
seed	種	種	種	籽	種	種	種	種	種	種/籽
sit	坐	坐	坐	坐	坐	坐	坐	坐	坐	坐
skin	皮	皮膚	皮	皮	皮	皮	皮	皮	皮	皮
sleep	睏	睏	睡	睏	睡	睡	困	睏	睏	睏
small	細	小	細	小/細	小	小	細/小	細	小	細/小
smoke	煙	煙	煙	煙	煙	煙	煙	煙	煙	薰
stand	徛	立	徛	徛/站	站	站/戳	站/企	徛	立	徛
star	星	星	星	星	星	星	星	星	星	星
stone	石	石	石	石	石	石	石	石	石	石
sun	頭	太陽	日	日	日	太陽	太陽/日	日/太陽	太陽	日
swim	游水	游水	泅水/洗身	玩水/洗澡	抹汗	凳水/游泳	洗冷水澡/游泳	洗冷水澡	游泳	泅水
tail	尾	尾	尾	尾	尾	尾	尾	尾	尾	尾
that	個個	歸個/A 個	B 個	C	那	那個	那個/那只	啱只	伊	許
this	爾個	該個/哀個/A 個	B 個	C	D	這個	F 個	咯只	G	即
tongue	脷	舌	舌/利	舌	舌	舌	舌	舌	舌頭	舌
tooth	牙齒	牙齒	牙齒	牙齒	牙齒	牙	牙齒	牙齒	牙	喙齒
tree	樹	樹	樹	樹	樹	樹	樹	樹	樹	樹
two	二	二	二	二	二	二	二	二	兩/二	二
walk	行	走/跑	行	走	走	走	走	行	走	行
water	水	水	水	水	水	水	水	水	水	水
we	我哋	佢	我兜人	我裏/我們	我們	我們	我們	我哩/叩哩	我 A/阿拉	阮
what	乜	啥	A	什	麼	什麼	麼	麼	啥 B 事體	什物
white	白	白	白	白	白	白	白	白	白	白
who	邊個	啥人/A 格	瞞人	哪個	哪個/啥個	誰	哪個	哪個	啥人	啥人/AB
woman	女人/婆嫲	女 A	婦人家	女個	女的/女將	女的/娘們	女的/女人家/堂客	女人家/堂客們	女人	查某/女的
yellow	黃	黃	黃	黃	黃	黃	黃	黃	黃	黃
you	爾	汝	爾	爾	爾	爾	爾	爾	儂	爾

References

Bai, Dizhou (白滌洲). 1931. Beiyin rusheng yinbian kao [北音入聲音變考]. *Xueshu Yuekan* [學術月刊] 2.2. Beijing: Beijing Nüzi Shifan Daxue.

Cao, Shuji (曹樹基). 1997. *Zhongguo Yiminshi* [中國移民史]. Fuzhou: Fujian Renmin Chubanshe.

Cavalli-Sforza, L. L., and William S.-Y. Wang. 1986. Spatial distance and lexical replacement. *Language* 62.1:38-55.

Chen, Baoya (陳保亞). 1996. *Lun Yuyan Jiechu Yu Yuyan Lianmeng* [論語言接觸與語言聯盟]. Beijing: Yuwen Chubanshe.

Chen, Baoya (陳保亞). 1999. *Ershi Shiji Zhongguo Yuyanxue Fangfalun* [20 世紀中國語言學方法論]. Ji'nan: Shandong Jiaoyu Chubanshe.

Comrie, B. 1993. Review of Starostin (1991). *Language* 69.3:828-832.

Dolgopolsky, A. B. 1964. A probabilistic hypothesis concerning the oldest relationships among the language families in Northern Eurasia. Original Russian translated into English and published in *Typology, Relationship and Time* (1986), ed. by V. V. Shevoroshkin and T. L. Markey. Ann Arbor: Karoma Publishers.

Dyen, Isidore. 1956. The Ngaju-Dayak 'Old Speech Stratum'. *Language* 32.1:83-7.

Forrest, R. A. D. 1950. The ju-sheng tone in Pekingese. *Bulletin of the School of Oriental and African Studies* 13.2:443-447.

Guo, Li (郭力). 1986. *"Chongding Shimawen'gong Dengyuntujing" Yanjiu* [《重訂司馬溫公等韻圖經》研究]. Beijing: Peking University MA thesis.

Guo, Li (郭力). 1997. Gu qingruzi zai *Hebingzixuejiyun* zhong de guidiao [古清入字在《合併字學集韻》中的歸調]. *Yuyanxue Luncong* [語言學論叢] 19:72-73.

Hirayama, Hisao (平山久雄). 1960. Zhonggu rusheng zai Beijinghua li de duiying tongze [中古入聲在北京話裡的對應通則]. *Riben Zhongguoxuehuibao* [日本中國學會報] 12:139-156.

Hirayama, Hisao (平山久雄). 1990. Zhonggu Hanyu de qingruzi zai Beijinghua li de duiying guilü [中古漢語的清入字在北京話裡的對應規律]. *Beijing Daxue Xuebao* [北京大學學報] 1990.5:72-79.

Hsieh, Hsin-I. 1971. *The Development of Middle Chinese Entering Tone in Pekinese*. Berkeley: University of California dissertation.

Lin, Tao (林燾). 1987. Beijing guanhua suyuan [北京官話溯源]. *Zhongguo Yuwen* [中國語文] 1987.3:161-169.

Lin, Tao (林燾). 1992. "Rupaisansheng" bushi ["入派三聲" 補釋]. *Yuyanxue Luncong* [語言學論叢] 17:3-18.

Qian, Zengyi (錢曾怡) et al. 2001. *Shandong Fangyan Yanjiu* [山東方言研究]. Ji'nan: Qilu Shushe.

Saitou, N., and M. Nei. 1987. The neighbor-joining method: A new method of reconstruciting phylogenetic trees. *Molecular Biology and Evolution* 4:406-425.

Starostin, S. 1991. *Altajskaja Problema i Proisxoždenie Japonskogo Jazyka* [*The Altaic Problem and the Origin of the Japanese Language*]. Moscow: Nauka.

Sapir, E. 1916. Time perspective in aboriginal American culture: A study in method. *Canada, Department of Mines, Geological Survey, Memoir 90, Anthropological Series* No.13. Ottawa: Government Printing Bureau. Also in *Selected Writtings of Edward Sapir*, Univeristy of California Press, 1949.

Stimson, H. M. 1962. Ancient Chinese -p, -t, -k endings in the Peking dialect. *Language* 39.4:376-384.

Swadesh, M. 1952. Lexico-statistic dating of prehistoric ethnic contacts. *Proceedings of the American Philosophical Society* 96:452-463.

Swadesh, M. 1955. Time depths of American linguistic groupings. *American Anthropologist* 56.3:361-364.

Ting, Pang-hsin (丁邦新). 1998. Hanyu fangyan jiechu de jige leixing [漢語方言接觸的幾個類型]. *Yuyanxue Luncong* [語言學論叢] 20:149-165.

Wang, William S.-Y. 1996. Linguistic diversity and language relationships. *New Horizons in Chinese Linguistics*, ed. by C.-T. James Huang and Y.-H. Audrey Li. Dordrecht, Boston: Kluwer Academic Publishers.

Wang, William S.-Y. 1997. A quantitative study of Zhuang-Dong languages. *In Memory of Mantaro J. Hashimoto*, ed. by Anne O. Yue and Mitsuaki Endo, 81-96. Tokyo: Uchiyama.

[Received 30 January 2003; revised 25 September 2003; accepted 13 October 2003]

Feng Wang
Language Engineering Laboratory
City University of Hong Kong
Tat Chee Road, Kowloon
Hong Kong
50003680@student.cityu.edu.hk

William S.-Y. Wang
Language Engineering Laboratory
City University of Hong Kong
Tat Chee Road, Kowloon
Hong Kong
eewsyw@cityu.edu.hk

基本詞彙與語言演變

汪　鋒　　王士元
香港城市大學

本文利用 Swadesh (1952, 1955) 提出的基本詞彙表來探討語言演變中的一些問題。根據陳保亞 (1996) 劃分詞階的方法，Swadesh 的 200 詞可一分爲二： 第 100 詞稱爲高階詞；餘下的 100 詞稱爲低階詞。高階詞比低階詞更穩定，更難以借用。因此，高階詞的同源保留率比低階詞高，而低階詞的借用率比高階詞高。本文第 2 節根據這一規律來釐清北京話入聲字中的早期遺存和晚近的借用成分，並藉以說明這一規律在語言演變研究中的重要作用。本文第 3 節比較了幾種影響較大的基本詞彙表——Dolgopolsky (1964) 15 詞，Yakhontov 35 詞和 Swadesh (1955) 100 詞，觀察它們在漢語方言分區中的功效，發現根據 Swadesh (1955) 100 詞得出的結果更可信。同時，複合詞應該從基本詞中剔除的看法得到了驗證。

關鍵詞：基本詞彙，階，漢語方言，系屬分類