Phonology and the Chinese Lexicon

San Duanmu, University of Michigan
Huili Zhang, Peking U & U of Michigan
CUNY Word Conference
January 14, 2010

1

Overview

- We argue that words are not well defined in Chinese.
- Instead, phonology determines the length of an expression in a given position
- In a 'short position', a disyllable (compound) can be truncated to a monosyllable (word)
- In a 'long position', a monosyllable (word) can be stretched to a disyllable (pseudo-compound).
- The phonological effect has created a dual-vocabulary in Chinese, where 40% of the disyllabic words have a short (monosyllabic) form

2

Outline

1. Traditional views of the word in Chinese
2. Problem: many morphemes are not free
3. Expedient solutions in dictionaries
4. The dual vocabulary
5. Phonology-driven word length choices
6. Some consequences
7. Conclusions

3

1. Traditional views of the word in Chinese

- Chinese words are monosyllabic (Jespersen 1922)
- Modern Chinese has some disyllabic words, mostly compounds, owing to loss of syllable inventory
- Examples:
  - tou-fa '(head)-hair'
  - ma-lu '(horse)-road'
  - qi-che '(gas)-car'
  - yu-yan 'language-(speech)'

4

Chinese morphemes are monosyllabic

- Some disyllabic morphemes, e.g.
  - pu-tao 'grapes'
  - ma-nao 'amber'
  - bo-li 'glass'
- Some polysyllabic foreign names, e.g.
  - jia-li-fu-ni-ya 'California'
- But such polysyllabic morphemes are rare. Most Chinese morphemes are monosyllabic

5

Chinese words are monosyllabic

Zhong-guo you hen duo qi-che
china-country have very many gas-car
'China has many cars.'

6

Chinese words are monosyllabic?

Zhong-guo you hen-duo qi-che
china-country have very-many gas-car
'China has many cars.'

Some apparent compounds are not really compounds, because half of it is semantically empty.

7

2. Problem: many morphemes are not 'free'

- Traditional definition: *a word is a free morpheme*
- 'Free': *a word is free if it can be used alone*.
- 'Used alone':
  - In isolation (as in answering a question)
  - As a major syntactic unit, e.g. subject or object
- Problem: many Chinese morphemes cannot be used alone.

8

Non-free morphemes

- In isolation:
  - What did you see?
  - *Hu 'tiger'
  - Lao-hu '(old)-tiger'
- As major syntactic unit, e.g. subject
  - *Hu lai-le 'The tiger is coming'
  - Lao-hu lai-le 'The tiger is coming'
- What does a non-free morpheme need?

9

The 'root' analysis

- Sproat & Shih (1996): morphemes like *hu* 'tiger' are roots.
- Problem: a root usually needs an affix in a given direction (left or right)
- A Chinese 'root' can go with any 'root', on either side
  - meng hu 'fierce tiger'
  - hu shan 'tiger mountain'
  - che meng 'car door'
  - xiao che 'small car'

10

Still monosyllabic 'words' in nature?

- When there is another syllable, a disyllabic word would often prefer to shed the redundant syllable.
- Examples
  - *meng hu* 'fierce tiger'
  - *??meng lao-hu* 'fierce (old)-tiger'
  - *hu shan* 'tiger mountain'
  - *??lao-hu shan* '(old)-tiger mountain'

11

So what is a word in Chinese?

- hu 'tiger'
  - is not free in isolation
  - is free in combination
- lao-hu '(old)-tiger'
  - is free in isolation
  - is not free in combination

12

What is a word?

- If a bound form and its host make a word, then a word can be very long
- Examples (bound form underlined)
  - *[song hua]* 'pine flower'
  - *[song hua] jiang* '[pine flower] river'
  - *[song hua jiang] hu* '[pine flower river] tiger'
  - …
- What should a dictionary collect?

13

3. Expedient solutions in dictionaries

- Include disyllabic free units, e.g.
  - *lao-hu* '(old)-tiger'
  - qi-che '(gas)-car'
- Avoid trisyllabic or longer units
- Include monosyllabic roots, e.g.
  - *hu* 'tiger'
- Result: in several modern lexicons, 60%-70% Chinese words are disyllabic.

14

Example

- XDHYCYCB (2008): newest and largest corpus-based modern Chinese lexicon
- Corpus: 270 million characters
- Segmentation: favor 2 (favor disyllabic units)
  - Example (+ indicates segmentation boundary): qi-che + chang '(gas)-car + factory'
- Monosyllabic words result from cutting [2+1] and [1+2]

15

Result

Length	Count	%
S	3,181	5.7 %
SS	40,351	72.0 %
SSS	6,459	11.5 %
SSSS	5,855	10.5 %
SSSSS+	162	0.3 %
All	56,008	100.0 %

16

4. The dual vocabulary

- Many pairs of equivalent words are included
- Example

From	We get (by 'favor 2')
qi-che + chang	*qi-che*
'(gas)-car + factory'	'(gas)-car'

From	We get (by 'favor 2')
ri-ben + che	*che*
'Japanese+ car'	'car'

- Dual-vocabulary: a long and a short form of the same word

17

Question

- How many words are repeated?
- How many words have both a long form and a short form?

18

Method

- For each disyllabic word, check whether it can be use without one of the two syllables in *some* environment
- Example (semantically empty syllable underlined):
  shu-mian yu-yan shu-mian yu
  'written language-(speech)' 'written language'
- The two expressions are equivalent, so *yu-yan* and *yu* are the same word
- One graduate student coding the data

19

Three types of dual vocabulary

- Right-redundant
  – *yu-(yan)* 'language-(speech)'
- Left-redundant
  – *(lao)-hu* '(old) tiger'
- Either useable (repetitive)
  – *xu-yao* 'need-want'

20

Result

- Over 80% of text coverage

Type	Count	%	Example
right-redun.	671	18%	中国, 倒闭
left-redun.	429	12%	学生 , 北京
either	355	10%	穿越 , 需要
neither	2,183	60%	所有 , 问题
All	3,638	100%	

21

Left-redundant examples

- 没有 not (have)
- 已经 already (experience)
- 中国 China (country)
- 可以 can
- 这个 this (one)
- 现在 now (present)
- 开始 start (origin)
- 世界 world (boundary)
- 因为 because (for)
- 时间 time (interval)

22

Right-redundant examples

- 学生 (study) student
- 北京 (north) capital
- 情况 (situation) situation
- 精神 (spirit) spirit
- 重要 (heavy) important
- 出现 (out) appear
- 历史 (experience) history
- 之后 (this) after
- 原因 (origin) because
- 回答 (return) answer

23

Either-reducible examples

- 领导 lead-guide
- 希望 hope-expect
- 发生 occur-grow
- 方面 direction-surface
- 需要 need-want
- 表示 express-show
- 决定 cut-decide
- 选择 choose-select
- 朋友 companion-friend
- 改变 alter-change

24

Neither-reducible examples

- 公司 company public-company
- 这样 this way this-way
- 社会 society association-community
- 同时 at the same time same-time
- 发现 discover occur-appear
- 过去 previously pass-go
- 进行 carry out enter-move
- 非常 very not-common
- 因此 therefore because-this
- 最后 final most-back

25

5. Phonology-driven word length choices

- Duanmu (2008, and references therein)
- Given the dual vocabulary or elastic word length
  - (shou-)biao 'watch'
  - (gong-)chang 'factory'
  - zhong(-zhi) 'to plant'
  - (da-)suan 'garlic'
  - (shu-)cai 'vegetable'
- A two-word expression ought to have four length patterns: 2+2, 2+1, 1+2, 1+1
- However, not all of them are always good

26

Word Length Problem

- [NOUN noun]
  - 2+2 shou-biao gong-chang 'watch factory'
  - 2+1 shou-biao chang
  - *1+2 biao gong-chang
  - 1+1 biao chang
- [verb OBJECT]
  - 2+2 zhong-zhi da-suan 'to plant garlic'
  - *2+1 zhong-zhi suan
  - 1+2 zhong da-suan
  - 1+1 zhong suan

27

Generalization (for elastic words)

- [NOUN noun]
  - NOUN cannot be shorter than noun
- [verb OBJECT]
  - OBJECT cannot be shorter than verb

28

Preliminary generalization

- Stressed word cannot be shorter than an unstressed word
- Both English and Chinese (uppercase indicates prominence)
  - [NOUN noun]
  - [verb OBJECT]

29

Foot Binarity

- Binarity: A foot should have two syllables
- Foot: Every stress implies a (trochaic) foot
- Cyclicity: compound and phrasal stress is assigned cyclically (Chomsky, Halle, & Lukoff 1956)
- In a compound [A B], A has compound stress
  - 2+2 (AA)(BB)
  - 2+1 (AA) B
  - *1+2 (A)(BB)
  - 1+1 (AB)

30

VO phrases

- In a VO phrase, O has compound stress

2+2	(VV)(OO)	
*2+1	(VV)(O)	
1+2	V (OO)	
1+1	(VO)	treated as a compound

31

6. Some consequences

- Some 'words' have little 'meaning'
- Some 'compounds' may not have sophisticated syntax

32

'words' with little 'meaning'

Zhong-(guo) you (hen)-duo (qi)-che
china-(country) have (very)-many (gas)-car
'China has many cars.'

33

'words' with little 'meaning' (underlined)

Ta hen gao
He very tall
'He is tall'

lao hu
old tiger
'tiger'

bei zi
cup son
'cup' (not 'small cup')

xie-xie
thank-thank
'Thank you'

qi-che
gas-car
'car'

da shu
big uncle
'uncle'

34

'Sophisticated' syntax

- Li (1990): Theta-role merger in serial verbs
  X kick-run Y 'X kicked Y and Y ran (away)'
  [X kick Y] & [Y run] → [X kick-run Y]

  X ride-tire Y 'X rode Y and X is tired'
  [X ride Y] & [X tired] → [X ride-tire Y]
- Hale & Keyser (1993): N-to-V raising
  to shelf the books
  to [V books [P shelf]] → to [shelf books [P t]]

35

Sophisticated syntax?

- Some may be due to theta-role merger
  ti-pao 'kick-run'
  [X kick B] & [B run] → [X kick-run B]
  jie-dai 'receive-treat'
  [X receive B] & [X treat B] → [X receive-treat B]
- Some may not
  xue-xi 'study-practice'
  shou-gou 'collect-buy'
  mei-tan 'coal-charcoal'
  jie-gei 'lend-give'
  song-zou 'send-walk'
- Possible solution: use non-dual vocabulary

36

7. Conclusions

- Words are not well defined in Chinese.
- Phonology, in particular Foot Binarity, determines the length of an expression in a given position
- A foot needs two syllables, either two monosyllabic words or one disyllabic pseudo-compound.
- Truncation and stretching are common methods to adjust word length to satisfy metrical requirements
- Truncation and stretching have created a dual-vocabulary in Chinese, where 40% of the words have both a long (disyllabic) and a short (monosyllabic) form.

37

References

Chomsky, N., M. Halle, and F. Lukoff. 1956. On accent and juncture in English. In *For Roman Jakobson*, ed. M. Halle et al., 65-80. The Hague: Mouton.

Duanmu, S. 2008. The phonology of Standard Chinese. 2nd Edition. Oxford.

Hale, Kenneth, and Samuel Jay Keyser. (ed.) 1993 The view from Building 20: Essays in honor of Sylvain Bromberger. Cambridge, MA: MIT Press.

Jespersen, Otto. 1922. Language: its nature, development and origin. New York: Macmillan.

Li, Yafei. 1990. On V-V compounds in Chinese. *NLLT* 8.2: 177-207.

Sproat, R., and C. Shih. 1996. A corpus-based analysis of Mandarin nominal root compound. *Journal of East Asian Linguistics* 5: 49-71.

XDHYCYCB. 2008. 《现代汉语常用词表》 [Lexicon of Common Words in Contemporary Chinese]. Beijing: Shangwu.

38