



Faculty Publications

2008-03-15

Analysis of Canonical Chinese Antonym Co-occurrence

Eric K. Ringger
ringger@cs.byu.edu

Guohui Liu

Shiping Liu

Xingfu Wang

Follow this and additional works at: <https://scholarsarchive.byu.edu/facpub>



Part of the [Computer Sciences Commons](#)

Original Publication Citation

Xingfu Wang, Eric Ringger, Guohui Liu, and Shiping Liu. "Analysis of Canonical Chinese Antonym Co-occurrence". Presentation given at the Conference of the American Association for Corpus Linguistics (AACL 28). Provo, Utah. March 15, 28.

BYU ScholarsArchive Citation

Ringger, Eric K.; Liu, Guohui; Liu, Shiping; and Wang, Xingfu, "Analysis of Canonical Chinese Antonym Co-occurrence" (2008). *Faculty Publications*. 1296.
<https://scholarsarchive.byu.edu/facpub/1296>

This Presentation is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Faculty Publications by an authorized administrator of BYU ScholarsArchive. For more information, please contact ellen_amatangelo@byu.edu.

Analysis of Canonical Chinese Antonym Co-occurrence

Xingfu Wang, Eric Ringger,
Guohui Liu, Shiping Liu

Chongqing University, China/BYU

Contents

1. Aims of this study
2. Comparison between English and Chinese
3. English constructions of antonym pairs
4. Chinese examples to be examined
5. Methods of investigation
6. Cases in CCL
7. Conclusion
8. Future work

1. Objectives of this study

- ① To set the canonical Chinese antonym pairs list
- ② To analyze whether there are constructions in the semi-idiomatic four-word Chinese chunks
- ③ To assess the strength and correlation of interpolated words in antonym pairs
- ④ To evaluate the probability of Chinese antonym pairs expanded by other words

2. Comparison between English and Chinese

老子《道德经》	Thomas Cleary's English Version of Daoism*
<p><u>有无相生</u>， <u>难易相成</u>， <u>长短相形</u>， <u>高下相倾</u>， <u>音声相和</u>， <u>前后相随</u>。</p>	<p><u>Being and nonbeing</u> produce each other, <u>difficulty and ease</u> complement each other, <u>long and short</u> shape each other, <u>high and low</u> contrast with each other, <u>voice and echoes</u> conform to each other, <u>before and after</u> go along with each other.</p>

Lyons (1977) claimed that there exists a 'general human tendency to categorize experience in terms of dichotomous contrast'.

*<http://www.davemckay.co.uk/philosophy/taoism/taoism.php?name=taoteching.cleary>

Google scholar search

antonymy, co-occurrence, corpora corpus	422
反义、同/共现、语料库	47

3. English constructions of antonym pairs

wildcard-first frame	wildcard-second frame
<i>* and Adj alike</i>	<i>Adj and * alike</i>
<i>between * and Adj</i>	<i>between Adj and *</i>
<i>both * and Adj</i>	<i>both Adj and *</i>
<i>either * or Adj</i>	<i>either Adj or *</i>
<i>from * to Adj</i>	<i>from Adj to *</i>
<i>* versus Adj</i>	<i>Adj versus *</i>
<i>whether * or Adj</i>	<i>whether Adj or *</i>

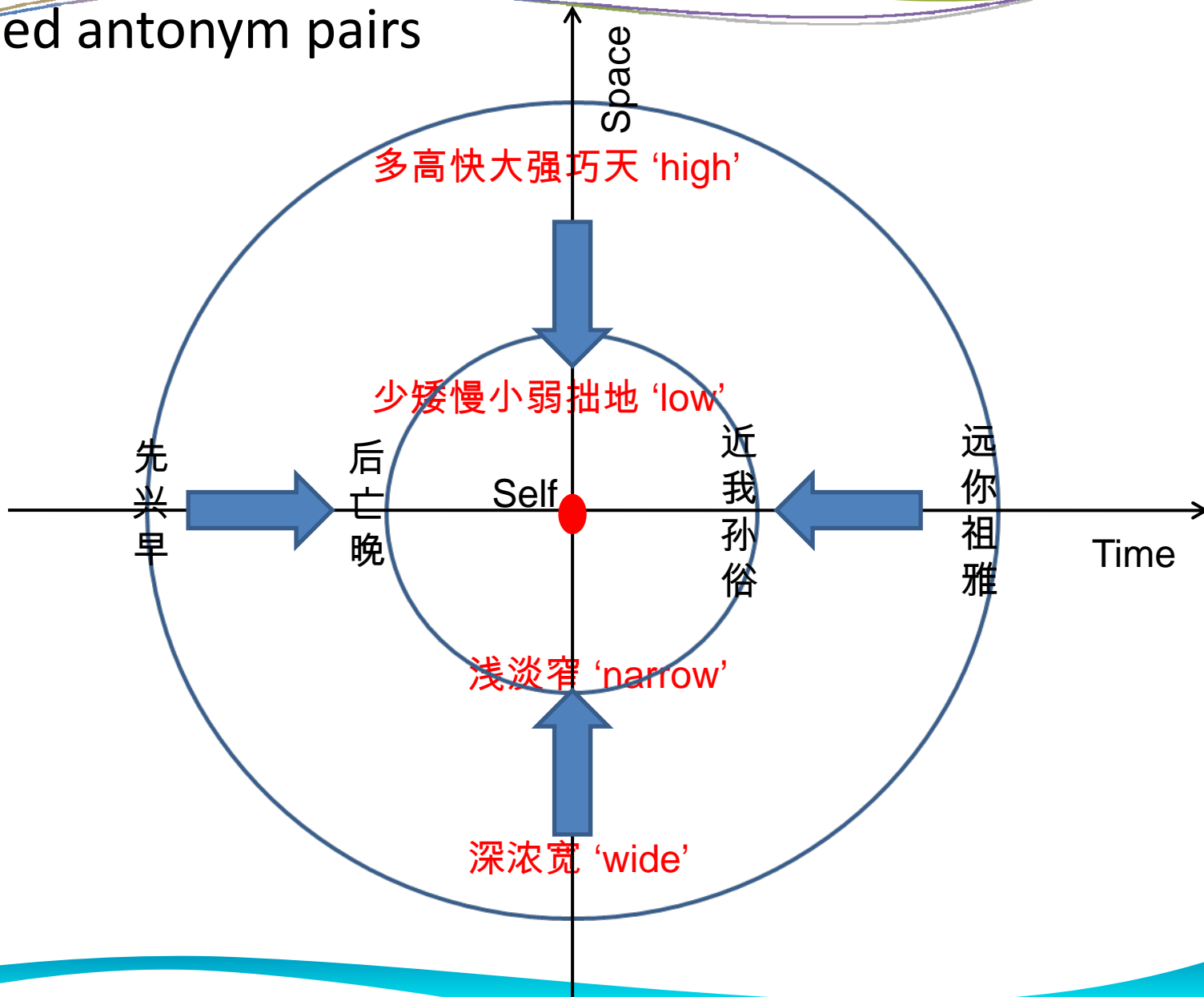
- (Steven Jones, M. Lynne Murphy, Carita Paradis and Caroline Willners, 2007)
- The term **'canonical'** was therefore reserved for word pairs that were found to retrieve one another on at least two occasions in at least ten of the fourteen frames.

Sequence rules of antonym pairs

- Normal sequence among antonymous pairs is not arbitrarily established. The vast majority of pairs observe a sequence which is determined by one or more relevant factors. (Jones, 2002:123)

Sequence rules	English	Chinese
Morphology	√	X
Positivity	√	√
Magnitude	√	√
Chronology	√	√
Gender	√	√
Phonology (word length)	√	?
Idiomacity	√	√
Frequency and markedness	√	√
Courtesy	?	√
Height	?	√

Graded antonym pairs



A special case

Gender \ Age	Male	Female
Older	兄(哥)	姐
Younger	弟	妹

Morphologically derived antonyms

active/passive
agree/disagree
attack/defend
badly/well
boom/recession
confirm/deny
difficult/easy
discourage/encourage
disprove/prove
dry/wet
fact/fiction
failure/success
fast/slow
feminine/masculine
guilt/innocence
hard/soft
heavy/light
illegal/legal
long/short
major/minor
new/old
old/young
optimistic/pessimistic
permanent/temporary
private/public
punishment/reward
right/wrong
rural/urban

advantage/disadvantage
alive/dead
bad/good
begin/end
cold/hot
correct/incorrect
directly/indirectly
dishonest/honest
drunk/sober
explicitly/implicitly
fail/succeed
false/true
female/male
gay/straight
happy/sad
hate/love
high/low
large/small
lose/win
married/unmarried
officially/unofficially
optimism/pessimism
peace/war
poor/rich
privately/publicly
quickly/slowly
rightly/wrongly
strength/weakness

- 9 out of 56 in Jones' research

- Rare in Chinese:

否, 歪, 甬 and 孬

They contain the prefix
不 'not'.

Courtesy

(BYU Corpus of American English and CCL)

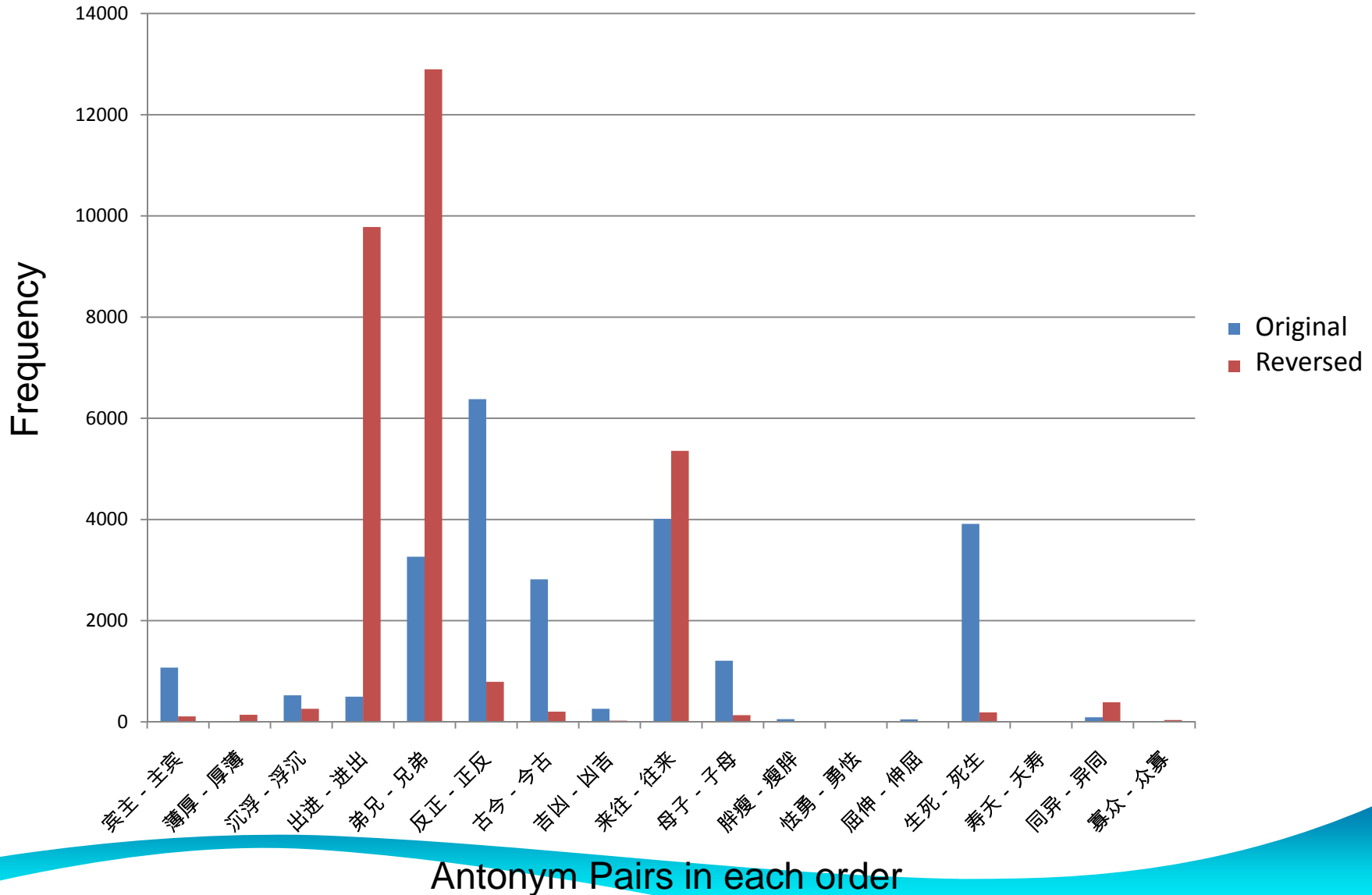
Co-occurrence of some antonym pairs	numbers
young and old	499
Old and young	118
老少，老小，老幼	2663
?少老，小老，幼老	?
Hosts and guests	17
Guests and hosts	6
主宾	109
宾主	1074

“Old is the first-mentioned antonym in only 29.0 per cent of old/young sentences.”
(Jones (2002:127) puts it into the rule of chronology.)

CCL Introduction

- CCL was set up by Peking University
 - raw corpus without segmentation and tagging
 - [http://ccl.pku.edu.cn:8080/ccl_corpus/]
- CCL contains more than 838 million bytes
 - contemporary part contains more than 264 million tokens [without punctuation marks]
 - 9552 Chinese types
- CCL (modern part) contains 7,818,863 sentences
 - counts acquired by searching 。 ? !
 - average sentence length: about 34 Chinese characters
- CCL contains 718,577 bare antonym pairs in the modern database
 - One antonym pair per 11 sentences, not counting the case when antonym pairs are separated by other characters.

The frequencies of ordered pairs



17 pairs in both orders

	gloss	original	reversed	difference
宾主 - 主宾		1074	109	965
薄厚 - 厚薄		15	140	-125
沉浮 - 浮沉		526	258	268
出进 - 进出	'come in', 'come out'	499	9780	-9281
弟兄 - 兄弟		3265	12897	-9632
反正 - 正反		6377	793	5584
古今 - 今古		2818	202	2616
吉凶 - 凶吉		258	24	234
来往 - 往来		4013	5357	-1344
母子 - 子母		1208	134	1074
胖瘦 - 瘦胖		55	0	55
怯勇 - 勇怯		0	3	-3
屈伸 - 伸屈		49	12	37
生死 - 死生		3914	188	3726
寿夭 - 夭寿		13	6	7
同异 - 异同		92	387	-295
寡众 - 众寡		0	36	-36

4. Chinese examples of antonym pairs

Alphabetic order	numbers	repeated	The first part of 371 Chinese antonym pairs	New pairs suggested
A	3		爱憎 安危 凹凸	
B	9	褒贬	爸妈 褒贬 悲喜 悲欢 本末 彼此 表里 宾主 薄厚	
C	20		裁缝 操纵 长短 唱和 畅滞 朝野 沉浮 成败 城乡 乘除 迟早 出入 出进 出没 出纳 春秋 雌雄 粗细 脆韧 存亡	
D	19	东西 多少	大小 大细 单双 旦夕 得失 敌友 嫡庶 余粟 弟妹 弟兄 点面 爹娘 东西 冬夏 动静 断续 对错 多少 多寡	堵疏 敌我
E	2		恩怨 儿女	
F	12		繁简 反正 方圆 肥瘦 分合 夫妻 夫妇 浮沉 父母 父子 俯仰 分总	废立

Types and Examples of Chinese expanded antonym pairs

1. Interleaved with 2nd antonym pair: x+y+!x+!y

- 厚此薄彼 喜新厌旧 天南地北
- 死去活来 古往今来 凶多吉少
- 今是昨非 深入浅出 外柔内刚
- 无独有偶 出生入死 弃恶扬善
- 破旧立新 此起彼伏 好逸恶劳

2. Two juxtaposed antonym pairs: x+!x+y+!y

- 古今中外 轻重缓急 盛衰荣辱
- 悲欢离合 纵横捭阖 喜怒哀乐
- 是非曲直 亲疏远近 春夏秋冬
- 生死存亡

Types and Examples of Chinese expanded antonym pairs

3. Interleaved with synonymous words:

x+y1+!x+y2

- 生离死别 冷嘲热讽 南腔北调
- 东张西望 天长地久 横冲直撞
- 长吁短叹 东藏西躲 天诛地灭
- (争长论短 瞻前顾后 同甘共苦
- 说长道短 欢天喜地)

4. Followed by fixed phrase: x+!x+m+n

- 新陈代谢 悲喜交集 赏罚分明
- 生死攸关 雅俗共赏 本末倒置
- 表里如一 劳逸结合 左右逢源
- (权衡利弊 明辨是非 颠倒黑白
- 无足轻重 人心向背 不知死活
- 自相矛盾 坐观成败 晓以利害)

5. Interleaved with grammatical particles:

半/忽/亦+x+半/忽/亦+!x

- 半明半暗 半新半旧 半死不活
- 半生不熟
- 忽冷忽热 忽左忽右
- 亦中亦西 亦喜亦忧 不破不立 若有若无

6. Repetition: x+x+!x+!x

- 日日夜夜 朝朝暮暮 浮浮沉沉
- 吞吞吐吐 进进出出 断断续续
- 老老少少 虚虚实实 真真假假

5. Method of investigation

- Perl scripts were written to download the data and aid the analysis.
- 374 antonym pairs $X+!X$ and their expanded forms $X+1+!X$ have been retrieved from CCL.
- **pre- X +mid+! X -post**
Pre, mid, post, pre&mid, mid&post are filtered for frequency counting and combining.

6. Cases in CCL

Chinese four-word constructions retrieved containing antonym pairs

我年轻的住的哦西单，那个时候儿讲呢东城西城的话也不一样。

后来我看到他们的旗帜东倒西歪，车辙也乱七八糟，才相信他们阵势全活
儒生经不起拷打，又东拉西扯地供出一大批人来。

后来，他打败了攻进兖州（今山东省西南部和河南省东部，兖音 yǎn）的黄巾军
曹操采纳了荀攸的意见，来个声东击西。

过了年，诸葛亮采用声东击西的办法，传出消息，要攻打彭城（今陕西
人马也到了彭城（今江苏徐州市），从东到西一万多里长的战线上，前秦水陆两路进军，
叛军又在城东城西堆起两座土山，想从土山上攻进城去，城里

李密没办法，只好改姓换名，东躲西藏，几次险点儿被官府抓去。

颉利东奔西逃，最后带着几个亲兵躲在荒山里，被他自

李训东奔西逃，走投无路，在路上被杀。

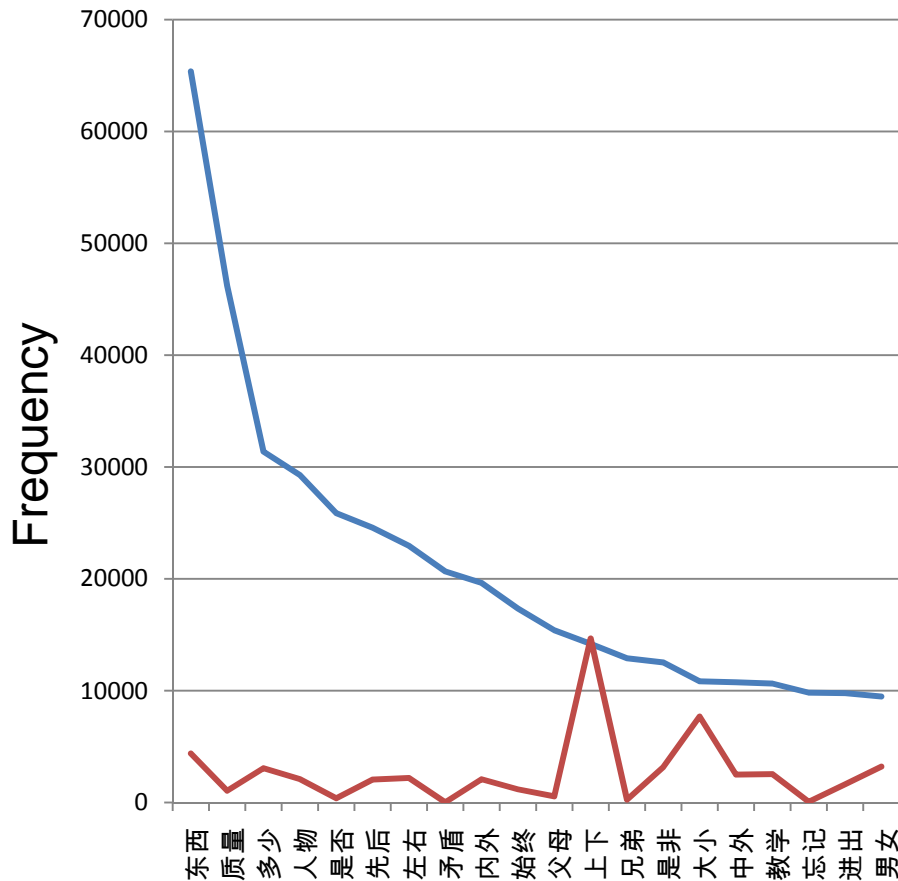
他手拿一支长枪，带头冲进敌阵，东挑西杀。

尼堪外兰东奔西窜，最后逃到了鄂勒琿（今齐齐哈尔附近）

这些平日仗势欺人的兵士吓得东奔西窜，想逃出群众的包围。

'east' 'west'

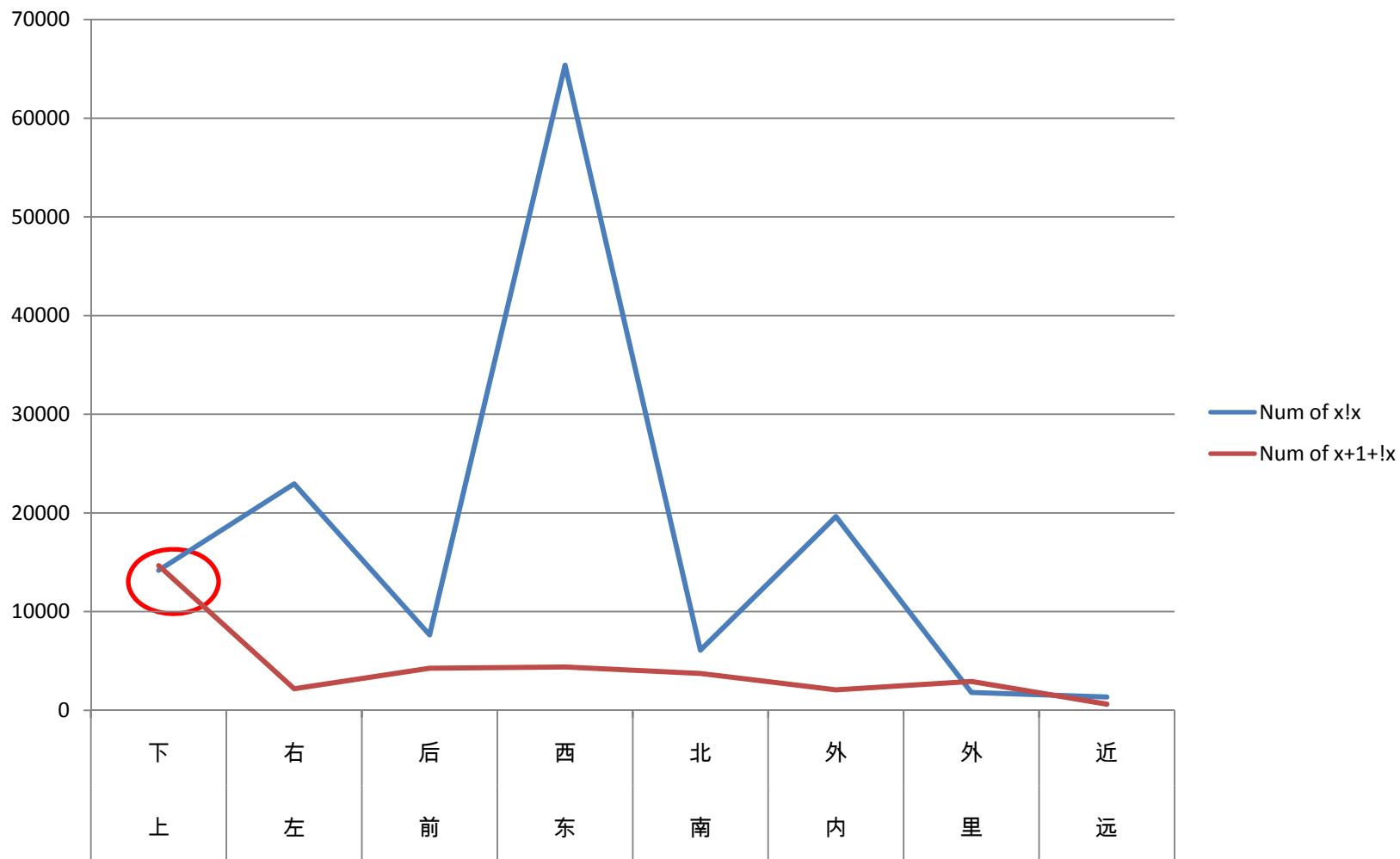
Top 20 Chinese Antonym Compounds



Order of X!X	Chinese pairs	X!X	X+1+!X	Order of X+1+!X
1	东西	65365	4391	上下
2	质量	46196	1045	大小
3	多少	31386	3069	东西
4	人物	29289	2098	男女
5	是否	25880	383	是非
6	先后	24565	2047	多少
7	左右	22956	2192	教学
8	矛盾	20668	22	中外
9	内外	19636	2085	左右
10	始终	17345	1183	人物
11	父母	15401	553	内外
12	上下	14193	14669	先后
13	兄弟	12897	245	进出
14	是非	12522	3173	始终
15	大小	10840	7689	质量
16	中外	10758	2496	父母
17	教学	10637	2539	是否
18	忘记	9828	62	兄弟
19	进出	9780	1627	忘记
20	男女	9481	3217	矛盾

Spatial antonym pairs in red

Spatial antonym pairs



Expected co-occurrence rate of 南北 'north-south'

x	Num	!x	Num	x-!x	Num of x!x	Num of x+1+!x
南	211804	北	279876	-68072	6101	3743

Total sentences in CCL: 7,818,863

X/Total=0.02708 (=37)

!X/Total=0.03579 (=28)

One co-occurrence of '南北' among $37 \times 28 = 1036$ sentences.

Expected co-occurrences= $7,818,863 / 1036 = 7547$

Observe co-occurrences= $6101 + 3743 = 9844$

Pre- and mid-words with

南北 'south-north'

走		闯	262
天		地	214
东		西	208
从		到	153
天		海	146
由		向	82
一		一	45
山	南	海	35
从		向	30
由		到	29
是		辕	28
自		向	21
河		河	21
的		水	21
自		而	20

Pre- and mid-words with

南北 'south-north'

Top10 most frequently used characters in CCL:

的,一,是,了,
不,在,人,有,
国,他

走		闯	262
天		地	214
东		西	208
从		到	153
天		海	146
由		向	82
一		一	45
山	南	海	35
从		向	30
由		到	29
是		辕	28
自		向	21
河		河	21
的		水	21
自		而	20

Tagging of the interpolated words

走/v

天/n

东/f

从/p

天/n

由/v

一/m

山/n

从/p

由/v

~~是/v~~

自/p

河/n

~~的/u~~

自/r

闯/v

地/n

西/f

到/v

海/n

向/p

一/m

海/n

向/p

到/v

~~辕/Ng~~

向/p

河/n

~~水/n~~

而/c

Conclusion

- The canonicity of Chinese antonym pairs should be set according to the numbers of their extended cases instead of original antonym compounds; and their expanded form are expected to be filled with symmetric information.
- Not many clear signs to represent the constructions of semi-idiomatic expressions in Chinese besides 半/忽/亦+x+半/忽/亦+!x, the best way to do this is to tag the words and categorize the collocations in order to get the possible patterns .

Future work

- To filter the noise by segmenting and tagging the sample Chinese sentences first;
- To categorize the canonical Chinese antonym pairs and extract the possible constructions for them;
- To make comparisons between English and Chinese to figure out the same cognitive mechanism in this respect.

The slide features decorative wavy lines at the top and bottom. The top section consists of a dark red-to-brown gradient wave, a white wave, and a thin multi-colored wave. The bottom section is a solid cyan wave.

Thank you!