Word Ordering in Chinese Opposite Compounds

Jing Ding[(⊠)] and Chu-Ren Huang

Department of Chinese and Bilingual Studies,
The Hong Kong Polytechnic University, Hung Hom, Hong Kong
amanda.ding@connect.polyu.hk, Churen.huang@polyu.edu.hk

Abstract. The smantic factor is considered to play an important role in deciding the word order of compounds. Previous studies on Chinese opposites generally agree with the Pollyanna Principle, but do not offer detailed statistical analysis, especially for opposite compounds. In this paper, we go through 315 opposite pairs in Sinica Corpus and the result shows that, for Chinese opposite compounds, prosodic factor and cultural hierarchy are also crucial for ordering.

Keywords: Chinese opposite compound · Pollyanna Principle · Corpus

1   Introduction

According to the **Pollyanna Principle** (Matlin & Stang, 1978), the positive or evaluatively (more) positive members are more favored than the negative ones and therefore preferred to be placed in the front of the negative ones.

Such a hypothesis is supported by some early dated observations. Tang (1979) mentioned that the word order of "antonymous coordinate construction" (:18) features having the positive meaning one in the front position and the negative, in the latter position. That was, he interpreted, because the positive one also carried the neutral, or unmarked, meaning.

For the exception of 轻重 [*qing1zhong4*] (*weight*), for example, 问题的轻重 [*wen4ti2 de0 qing1zhong4*], Tang thought it was equal to "the importance (of the issue)" and referred to a state, rather than a physical measurement. And therefore it did not follow the general order. Similar cases like *qing1zhong4* are （态度的）冷热 [*(tai4du4 de0) leng3re4*] (*attribute*), and （人间的）冷暖 [*(ren2jian1 de0) leng3 nuan3*] (*social snobbery*) (:18-20). However, such an interpretation is not convincing enough. As we have mentioned above, *qing1zhong4* is also commonly used as 'weight' in both ancient and modern Chinese. The meaning of 'importance' is actually derived from the initial meaning of 'weight'. So do the cases of *leng3re4* and *leng3nuan4*. But as to the other compounds, which are combined by opposite pairs and refer to the hypernym, it is more often to have the positive ones in prior position: 长短 [*chang2duan3*] (*length*), 高低/高矮 [*gao1di1*]/[*gao1ai3*] (*height*), 多少 [*duo1shao3*] (*quantity*). And it is also seen as to provide a move from the higher point of the scale to the lower point.

In his book, Tang listed out fifteen special word orders in Chinese (:20-26):

(1) '天'>'人' (the concept of *tian1* 'sky' prior to that of *ren2* 'human being')

(2) '人'>'兽'或'物' (the concept of *ren2* 'human being' prior to that of *shou4* 'animal' or *wu4* 'things')

(3) '公'>'私' (the concept of *gong1* 'public' prior to that of *si1* 'private')

(4) '家'>'人' (the concept of *jia1* 'family' prior to that of *ren2* 'human being')

(5) '长'>'幼' (the concept of *zhang3* 'senior' prior to that of *you4* 'junior')

(6) '尊'>'卑' (the concept of *zun1* 'respected' prior to that of *bei1* 'humble')

(7) '亲'>'疏' (the concept of *qin1* 'close' prior to that of *shu1* 'distant')

(8) '男'>'女' (the concept of *nan2* 'male' prior to that of *nv3* 'female')

(9) '优'>'劣' (the concept of *you1* 'good(_quantity)' prior to that of *lie4* 'bad(_quantity)')

(10) '盈'>'亏' (the concept of *ying2* 'surplus' prior to that of *kui1* 'deficit')

(11) '主'>'副' (the concept of *zhu3* 'host' prior to that of *ke4* 'guest')

(12) '鸟'>'（鱼）兽' (the concept of *niao3* 'avifauna' prior to that of *yu2* 'ichthyfauna' or *shou4* 'quadruped')

(13) '上'>'下' (the concept of *shang4* 'up' prior to that of *xia4* 'down')

(14) '软'>'硬'>（或者，'流体'>'固体'）(the concept of *ruan3* 'soft' prior to that of *ying4* 'hard', or, *liu2ti3* 'liquid' prior to that of *gu4ti3* 'solid')

(15) '里'（'进'）>'外'（'出'）(the concept of *li3* 'inside' prior to that of *wai4* 'outside', or, *jin4* 'come in' prior to that of *chu1* 'go out')

Tang's list was not for bi-syllabic compounds only, but served as general rules for the word ordering of related Chinese concepts, be they for compounds, phrases or others.

Besides, Xu also generally mentioned that, in the opposite compounds, it is usual to have the words which bear the positive meaning in front, and the negative one behind. So he demonstrated that the word order of opposite compounds does have relation with their meanings. (2000: 450-1) In that sense, his finding also agreed with that of the Pollyanna Principle.

2   Research Question

Tang's analysis offers a general rule for Chinese opposite compounds. However, his rules do not fully explain the existence of compounds such as 轻重 [*qing1zhong4*] (*weight*). Therefore, corpus-based statistical data is necessary for testing whether the fact agrees with Pollyanna Principle or Tang's list. For this paper, we want to answer following questions: is Pollyanna Principle sufficient to explain all the compound orderings; if not, then, what would be the other deciding factors?

3   Corpus and Method

There are two main steps in our experiment: first, the opposite pairs are selected from modern Chinese dictionaries; then, all the compounds that contain opposite pairs are extracted from the corpus for later analysis.

Notice, in the experiment, we settle a window size of extraction from 2 characters to 4 characters, to avoid the noises which are not considered as compounds. And, all the data are extracted automatically but then manually checked for higher accuracy.

《新华反义词词典》(*Dictionary of Opposites*) (2003) is selected as the base of our candidate list. It is the most authorized dictionary for Chinese opposites and its definition of what an opposite pair is, is based on the lexical meanings. That is to say, for the two members of an opposite pair: the assertion of one member implies the negation of the other,, and vice versa, like 生: 死, [*sheng1: si3*] (*alive: dead*); or, the meanings of two members are contrasts to each, and the negation of one does not necessarily imply the assertion of the other, like 大: 小, [*da3: xiao3*] (*big: small*); or, converse pairs like 买: 卖, [*mai3: mai4*] (*buy: sell*), 敌: 友, [*di2: you3*] (*enemy: friend*). This rule convers most of the Chinese opposite compounds. However, the book does not include the pairs like 夫: 妻, [*fu1: qi1*] (*husband: wife*), or 父: 子, [*fu4: zi3*] (*father: son*), neither the Chinese cultural contrast pairings (对举词 [*dui4ju3 ci2*]), like 水: 火, [*shui3: huo3*] (*water: fire*), which are also very frequently used as opposites in Chinese. In that case, we manually add extra 39 pairs to the initial list hence the final seed list contains 315 pairs, which are supposed to cover most of the opposite pairings in Chinese.

We chose Sinica Corpus because: it is one of the largest balanced corpora for Modern Chinese and it is fully segmented. By using it, we mean to reflect the nature of Modern Chinese, and, for practical aspect, to avoid the problems of word segmentation as well as the argument of what the word is, in Chinese.

4   Data Result

There are four basic kinds of patterns in our data. The first kind of pattern contains only one opposite pairing, which is generalized as [A][-A], such as 美丑[*mei3: chou3*] (*beauty: ugliness*), 胜负[*sheng4: fu4*] (*victory: failure*), 爱恨 [*ai4: hen4*] (*love: hate*) and 男女 [*nan2: nv3*] (*man: woman*). This basic pattern [A][-A] sometimes derives into patterns of [A][-A][A][-A] (彼此彼此 [*bi3ci3bi3ci3*] (*there and here*)), [A] [A][-A][-A](来来往往 [*lai2lai2wang3wang3*] (*come and go*)) and [A][-A][A](里外里 [*li3wai4li3*] (*in all*)).

The second pattern is that of one opposite pair explicitly representing a relation between the opposite pairing members within the phrase. The relation may appear in front/behind the pairing, like [A][-A][X] (松紧度 [*song1jin3du4*] (*degree of tightness*), 生死与共 [*sheng1si3yu3gong4*] (*share the same destiny*)) or [X][A][-A](见高低 [*jian4gao1di1*] (*show the result*), 不相上下 [*bu4xiang1shang4xia4*] (*be equally good*)), and also in the middle position of the pattern like [A][X][-A](老来少

[*lao3lai2shao4*] (*have a young heart at an old age*)), or, in some of the cases, be separated into two parts interjecting the two opposite members like [X-][A][+X][-A] (反客为主 [*fan3k4wei2zhu3*] (*guest acts like host*)) and [A][X-][-A][+X](今非昔比 [*jin1fei1xi1bi3*]   (*things change with time*)).

The third is to have two pairs of opposites combined together. For example, [A1][A2][-A1][-A2], such as古往今来 [*gu3wang3jin1lai2*] (*of all ages*), and, 优胜劣汰 [*you1sheng4lie4tai4*] (*select the fittest*).

The forth one is combined by one opposite pair and one synonymous pair, like 东躲西藏 [*dong1duo3xi1can2*] (*hide hard*) and 阴错阳差 [*yin1cuo4yang2cha1*] (*a mistake caused by fate*) are both marked as [A][S][-A][S']. In some of the examples, the added word repeats in the same patterns, such as 时冷时热 [*shi2leng3shi2re4*] (*sometimes cold sometimes hot*) ([S][A][S][-A]) or 明里暗里 [*ming2li3an4li3*]   (*in all places*)

([A][S][-A][S]). Such examples of patterns [A][S][-A][S] and [S][A][S][-A] are also considered as one of the patterns in the second category.

Theoretically peaking, there are more patterns which may, for example, bear three of the above pairs at the same time, but the number of them is not very large in daily language. So they are not considered as the most **basic** patterns, although we will still keep them for discussion in this study.

Table 1. Hits of patterns for four basic categories

category	pattern	hits	total
one opposite pairing	[A][-A]	218	
	[A] [A][-A][-A]	38	
	[A][-A][A][-A]	3	
	[A][-A][A]	1	
			260
One opposite pairing with one relationship	[A][-A][X]	459	
	[X][A][-A]	120	
	[X-][A][+X][-A]	63	
	[A][X][-A]	62	
	[A][X-][-A][+X]	6	
			710
Two opposite pairings	[A1][A2][-A1][-A2]	136	
			136
One opposite pairing with one synonymous pairing	[S][A][S][-A]	216	
	[A][S][-A][S']	129	
	[S][A][S'][-A]	118	
	[A][S][-A][S]	17	
			480
			1586

5   Discussion

As we have mentioned above, previous studies agree with the Pollyanna Principle in general. Our research however, shows that the Pollyanna Principle has its limitation when applied to Chinese opposites as a whole, and the real situation is more complicated than Pollyanna Principle or Tang's list can cover.

Adopting the general rule and list, we went through the 218 bi-syllabic opposite compounds, which are restricted to the [A][-A] pattern. Only 113 out of them agree with the rules. The rest of the examples, including 往来 [*wang3lai2*] (*come and go*), 问答 [*wen4 da2*] (*questions and answers*), 经纬 [*jing1wei3*] (*horizontal and latitudinal*), are not applicable, since it is hard to define the positive or negative in the components.

In the 113 valid instances, 87 (77.0%) agreed with the above rules and only 26 were against it. Both lists contained some highly frequent words, such as 亲疏 [*qin1shu1*] (*close or distant (of social connection)*), 利害 [*li4hai4*] (*benefit and harm*), 美丑 [*mei3chou3*] (*beautiful and ugly*), as positive-in-front order, and, 死活 [*si3huo2*] (*death and living*), 悲喜 [*bei1xi3*] (*sad and happy*), 贫富 [*pin2fu4*] (*poor and rich*), as negative-in-front order.

These exceptions are also found in the experiment. The saying of 雌雄 [*ci2 xiong2*] (*male and female*) appears 158 times, is much better accepted than *xiong2ci2*, which does not appear once in the whole corpus. But as we will see in later this section, for the same concept of gender, *nan2: nv3* and *gong1:mu3* strongly follows the rules.

Moreover, when the reversed order can also be used, and the two word orders are both acceptable for speakers, it is usually the positive-in-front order more preferred in speaking.

We use the hits of each word in Chinese GigaWord (Simplified) to see which one is better accepted.

The word 真假 [*zhen1jia3*] (*real and fake*), it has 331 hits, while the reversed word 假真 [*jia3zhen1*] does not have any hit. 真真假假[*zhen1zhen1jia3jia3*] has 16 hits, which is three times higher than that of 假假真真[*jia3jia3zhen1zhen1*] (4 hits). 生死 [*sheng1si3*] (*living and death*) has 1916 hits, but 死生 [*si3sheng1*] has 0 hit and even for 死活 [*si3huo2*] (*death and living*) the hit number reaches is only 156. The gap between 男女 [*nan2 nv3*] (*male and female*) and 女男 [*nv3nan2*] is more obvious. 12576 instances contain the saying *nan2nv3* but none for *nv3nan2*. Similarly, 公母 [*gong1: mu3*] (*male: female*) is accepted but *mu3gong1* is not.

The variation of the [A][-A] pattern, that is, that the rest patterns is category One, and, the second category, as well as the forth category, follow the order of their according [A][-A] compounds.

The results show that the Pollyanna Principle does have an overwhelming effect on the applicable words but does not cover all the opposite compounds of the [A][-A] pattern. In fact, the word order related to their semantic meanings in different ways, within different kinds of opposites.

For the third category, that is, the one of the compounds having two pairs of opposites within one compound or idiom, the ordering rules are generalized as following:
1. The positive one(s) is/are preferred to take the prior place(s).

There are only three possible orderings in our corpus, and within them: [P(ostive)1][N(egative)1][P(ostive)2][N(egative)2] > [P1][N1][N2][P2](安危祸福 [*an1wei1huo4fu2*] (*fate*), 是非曲直 [*shi4fei1qu3zhi2*] (*right and wrong*), 是非黑白 [*shi4fei1hei1bai1*] (*right and wrong*))>[N1][P1][N2][P2](悲 欢 离 合 [*bei1huan1li2he2*] (*happiness and sadness*), 轻重缓急 [*qin1zhong4huan3ji2*] (*various situations*).

That is to say, [P1][N1][P2][N2] is the most common pattern.

2. For the ones which share the same natural domain, they are preferred to appear in close positions. Hence, in 男女老少 [*nan2nv3lao3shao4*] (*all the population*), *nan2* (*male*) *and nv3* (*female*) are both for gender, while *lao3* (*old*) and *shao4* (*young*) *are* both for age, is preferred to 男 老 少 女 [*nan2lao3shao4nv3*] or 男 老 女 少 [*nan2lao3nv3shao4*] (this compound may be possible when it means 'the male one is elder and the female one is younger', which differs from the meaning of *nan2nv3lao3shao4*). And, in 亲疏远近 [qin1shu1yuan3jin4], *qin1* (*close*) and *shu1* (*remote*) are both for relationship, while *yuan3* (*distant*) and *jin4* (*close*) are both for distance, 'close and distant', rather than 亲远近疏 [*qin1yuan3jin4shu1*]; in 利弊得失 [li4bi4de2shi1], *li4* (*advantage*) and *bi4* (*disadvantage*) mean 'cons and pros', and, *de2* (*win*) and *shi1*(*lose*) mean two possible results), 'all the aspects', rather than 利失得弊 [*li4shi2de2bi4*].

3. If the same domain has more than one way to be divided, then the ones which share the same manner of division should be placed together. For example, 前后左右 [*qian2hou3zuo3you4*] (*in four directions*), *qian2* (*front*) with *hou3*(*behind*), and *zuo3* (*left*) with *you4* (*right*). And, 加减乘除 [*jia1jian3cheng2chu2*] (four arithmetic operations), *jia1* (*addition*) with *jian3* (*subtraction*), and *cheng2* (*multiplication*) with *chu2*(*division*), according to the rules of arithmetic operation.

From the aspect of opposite sub-categories, different kinds of opposite perform differently with the rules. For antonym and complementary compounds, it is the truth that most of the instances prefer to have the positive ones in front: 亲疏 [*qin1shu1*] (*close and distant*), 优劣 [*you1lie4*] (*good(_quantity)* and *bad(_quantity)*)), 利害 [*li4hai4*] (*benefit and harm*), 善 恶 [*shan4e4*] (*kindhearted and evil*), 多寡 [*duo1gua3*] (*many and few*), 尊卑 [*zun1bei1*] (*respected and humble*).

For taxonomy words, there are so many compounds like文武 [*wen2wu3*] (*civil and military*), 经纬 [*jing1wei3*] (*horizontal and latitudinal*), 水火 [*shui3huo3*] (*water and fire*), 南北 [*nan2bei3*] (*south and north*) that it is hard to evaluate the polarity.

In other words, the hypothesis is not applicable to these examples. Some taxonomy compounds, such as 甘苦 [*gan1ku3*] (*sweet and bitter*), 爱恨 [*ai4hen4*] (*love and hate*), have the ones which are considered as culturally positive words in front. But, it also has compounds like 悲喜 [*bei1xi3*] (*sad and happy*) or 阴阳 [*yin1yang2*] (*Yin and Yang*), which have the culturally negative ones in front.

For the converse, the order of the two members is exactly the same as they are in the real event. That is to say, the event order decides the word order of the according converse compound. Let's take 买卖 [*mai3mai4*] (*buy and sell*) for example. In the trade activity, to sell something is always after the action of buying or possessing something. Hence the word order keeps the same as their temporal order.

When there is no temporal difference between the two components, the compounds prefer to agree on the positive-in-front order. 婚 [*hun1*] (*man marries woman*) and 嫁 [*jia4*] (*woman marries man*) describe a marriage from the groom's and bride's aspects, respectively. In the compound of this pairing, *hun1* is put in front of *jia4*, that is, *hun1jia4*, as the same order of rule (8) "male prior to that one of female". In our data, we did not find any instance which follows the temporal order but are against the Pollyanna hypothesis. From the examples of taxonomy group, we added some rules to Tang's list:

(16) '古'>'今' (the concept of *gu3* 'ancient' prior to that of *jin1* 'nowadays')

(17) '东'>'西' (the concept of *dong1* 'east' prior to that of *xi1* 'west')

(18) '南'>'北' (the concept of *nan2* 'south' prior to that of *bei3* 'north')

(19) '经'>'纬' (the concept of *jing1* 'horizontal' prior to that of *wei3* 'latitudinal')

(20) '昼'>'夜' (the concept of *zhou4* 'day' prior to that of *ye4* 'night').

(21) '头'>'尾' (the concept of *tou2* 'head' prior to that of *wei3* 'end')

(22) '左'>'右' (the concept of *zuo3* 'left' prior to that of *you4* 'right')

(23) '前'>'后' (the concept of *qian2* 'front' prior to that of *hou4* 'behind')

(24) '天'>'地' (the concept of *tian1* 'sky' prior to that of *di4* 'earth').

5.1   A Prosodic Explanation[1]

Apart from the above, the prosodic factor[2] is also possible for the explanation of word ordering in compounds. We compute the tone order of the 218 [A][-A] compounds. 113 examples have first or second tones in the front position and the third or fourth tone in the back position; 62 examples have the reverse situation. That is, with the third or fourth tones in the front position, and the first or second tone in back position; the rest of the 43 compounds have the same tone in both positions. That is to say,

---

[1] Thanks to Yu Shiwen, Chen Keh-Jiann, Su Xinchun and Yuan Yulin, for pointing out this factor and offering a discussion for us, in CLSW 2014, at University of Macau.

[2] Because of the limitations of time and knowledge, this paper only considers the initial tone of each word. The tonal modification, which means the influence and change in actual speaking, is not considered. From the aspect of historical linguistics, it is also possible for the word tone to change since some compounds were created. That kind of change, however, is not discussed in this paper either.

first or second tone (113>62) has an obvious leading position in compounds, than third or fourth tone.

Do the Pollyanna Principle and prosodic factors agree with each other? In our data, there are 84 compounds which are both available to the Pollyanna Principle and show a difference of tone in two positions:

Table 2. The distribution of 84 examples

Pollyanna Principle	Prosodic factor	Example number
+	+	40
+	-	18
-	+	19
-	-	7
		Total: 84

As we can see here, the examples in which the Pollyanna Principle agrees with the prosodic factor overwhelms the other three situations. When the two factors clash with each, the chances for any one of them to win out are almost equal. That is to say, from this table, it is hard to tell which factor is stronger than the other in deciding the order within opposite compounds.

However, there are 7 examples violating both of the two rules. They are: 死生 [*si3sheng1*] (*being dead or alive*), 死活 [*si3huo2*] (*being dead or alive*), 短长 [*duan3chang2*] (*shortness*), 祸福 [*huo4fu2*] (*unfortunate and fortunate*), 冷热 [*leng3re4*] (*cold and hot*), 苦乐 [*ku3le4*] (*unhappy and happy*), 浊清 [*zhuo2qing1*] (*muddy and clear*).

6   Conclusion

In this paper, we use 315 Chinese opposite pairs as the seed list to go through Sinica Corpus, in order to see their ordering rules in compounds from bi-syllabic to quad-syllabic. Based on the data result, we modified the initial observation by previous studies as follows: The Pollyanna Principle is available to around half of the Chinese opposite compounds. In these examples, most agree with the principle. Additionally, when two-word orders are both available, the positive-in-front ones are preferred. Temporal order is an influencial factor for the ones which are not available to the Pollyanna Principle. It is mostly proved in the opposite category of converse.

In general, the result shows that, in Chinese, cultural hierarchy and the Pollyanna Principle together determine the order of opposites in compounds. Moreover, detailed conceptual orders of such cultural hierarchy are also added to previous studies.

We must admit, due to the limit of time and space, this paper only focuses on the semantic factor for Chinese opposite compounds. Other possible factors, such as prosodic, would be considered in our future work.

References

Cruse, D.A.: Lexical Semantics. Cambridge University Press, Cambridge (1986)

Murphy, M.L.: Semantic Relations and the Lexicon: Antonyms, Synonyms and other Semantic Paradigms. Cambridge University Press, Cambridge (2003)

Paradis, C.: Ontologies and construals in lexical semantics. Axiomathes **15**, 541–573 (2005)

Matlin, M.W., Stang, D.J.: The Pollyanna Principle: Selectivity in Language, Memory, and Thought. Schenkman Publishing, Cambridge (1978)

Tang, T.-C.: Studies on Chinese Morphology and Syntax. Taiwan Student Book Press, Taipei (1979). (in Chinese)

Xu, W.: Chinese lexical. Chinese Lexicology in 20th Century. Book Sea Publisher, Taiyuan (2000). (in Chinese)

Chinese Knowledge Information Processing Group (CKIP): CKIP Technical Report No. 95-02/98-04. Academia Sinica, Taipei (1993)

Springer