This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use (https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms), but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: http://dx.doi.org/10.1007/978-3-319-73573-3_51

Transitivity Variations in Mandarin VO Compounds

--- A Comparable Corpus-based Approach

Menghan Jiang¹ and Chu-ren Huang²

Abstract. This paper adopts a comparable corpus-based statistical approach to VO compound Variations in two varieties of Mandarin Chinese and examines the variations from a transitivity perspective. In recent years, more and more VO compounds are observed to have transitive usages. Previous studies categorize the transitivity of VO compound in a dichotomy way, while we argue that each VO actually differs in their degree of transitivity, especially when the variations between different variants of Mandarin are taken into consideration. The degree of transitivity can be measured by both transitivity frequency and its semantic/syntactic properties (follow the theory of Hopper and Thompson [1]). In our study, we compare the transitivity difference between Mainland and Taiwan Mandarin by adopting a corpus-based statistical approach. For both transitivity frequency and semantic/syntactic properties study, the results clearly show that Taiwan VO compounds have a higher degree of transitivity than the Mainland counterparts. We further argue that the higher transitivity degree in Taiwan also illustrates the conservatism of Taiwan Mandarin. This observation is consistent with the earlier study of transitivity variations of light verbs (Jiang et al. [2]) and follows the established null hypothesis in language changes that peripheral varieties tend to be more conservative.

Keywords: VO Compound, transitivity, Language Variation.

1 Introduction

In Mandarin Chinese, some VO compound verbs are observed to have transitive usages, i.e. these VO compounds can take external objects and yield expressions in the configuration of $[VO_1+O_2]$. The interesting thing for this phenomenon is that the verbal morpheme V in VO_1 has already taken an object. Syntactically it is not allowed to take another object. However, in fact, this kind of configuration has become more and more popular in the actual usage, and numbers of VO compounds began to be used transitively.

The transitivity of VO compound has attracted the interests of many linguists. Most of the studies categorize VO in a dichotomy way: VO compound is either transitive or intransitive (e.g., Liu and Li [3]). Her [4] has made a tripartite division in terms of the transitivity: VO can be intransitive (e.g., 得意 'be proud'); semi-

transitive (e.g., 在行 'be good at') and transitive (e.g., 留意 'pay attention to'). However, in our study, we have observed that the VOs which are considered to be transitive actually are very different in their degree of transitivity. Hence, it is more appropriate to view the transitivity of VO in a continuous way. For example, for the word 关心 'be concerned with' and 留心 'be careful', it is very natural for them to take another object, as in the construction 关心别人/留心他的表现 'be concerned with someone/be careful about his performance'. For the word 观光 'tour', it is not very natural if we say 观光意大利 'tour Italy' but it is still comprehensible without ambiguity. Nevertheless, it is not acceptable when the word like 出丑 'make a fool of oneself' is used transitively.

The degree of transitivity of VO compound becomes more complicated if the variations between different variants of Mandarin Chinese are considered. It has been observed that the transitivity of VO compound is likely to differ in different varieties of Chinese. Wang [5] lists about 50 VO compounds (e.g., 登陆 'landing', 涉嫌 'be suspected of', 加盟 'join in', 入围 'finalist') which are commonly used as transitive verbs in Singapore and climes that the transitivity degree of VO compounds in Singapore is much higher than that of in Mainland Mandarin. Diao [6] also points out that the reason behind the increasing popularity of VO₁+O₂ is the influence of Taiwan Mandarin. Although some previous studies have already observed the variation difference, they investigate the issue by only listing some examples extracted from daily newspapers or novels. In addition, their studies are still at the descriptive level. Hence, there lacks of systematic and comprehensive study for the VO compounds variations between different variants of Mandarin, especially based on large corpora. A similar problem is, in previous studies, the variation differences are also examined in a dichotomy way (i.e. Taiwan or Singapore VO can be used transitively while Mainland VO cannot). As the communication between different variants become more and more frequent, the variation differences tend to be a tendency/preference difference instead of the dichotomy difference. For example, although 中意 'like' has also become transitive in Mainland usage, the frequency of being used transitively is much lower than that of in Taiwan or Hong Kong Mandarin. Moreover, Taiwan and Mainland VO compounds can also differ in the type of objects they are taking (e.g., 帮忙他 'do him a favor' can only be observed in Taiwan Mandarin) or the context they can occur in (e.g., 曝光 'exposure' can only be used in negative context in Mainland but not necessary in Taiwan), this kind of variations are not easy to be observed by a traditional methodology.

2 Methodology and research questions

In our current study, in order to investigate the syntactic variations between different variants of Mandarin Chinese in actual usages, a comparable corpus-based statistical approach is applied. The variants we focused on are Mainland and Taiwan Mandarin.

171 VO compounds collected from previous studies (e.g., Wang, 1997; Diao, 1998; Liu and Li, 1998) were examined. The corpus used in this paper is the annotated Chinese Gigaword corpus, which contains over 1.1 billion characters, including 700 million characters from Taiwan Central News Agency and 400 million characters from Mainland Xinhua News Agency (Huang [7]).

In our study, two research questions are addressed:

- 1). For each VO compound, does its Mainland and Taiwan usages differ in their transitivity frequencies (i.e. the frequency of being used transitively)? What is the pattern of the tendency?
- 2). What kind of distributional differences in terms of semantic/syntactic features do they have between Mainland and Taiwan Mandarin? Does this refer to their differences in transitivity? (According to Hopper and Thompson [1])

3 Variations between Mainland and Taiwan Mandarin

3.1 Variation in transitivity frequency

To investigate the transitivity frequency differences of VO compounds between Mainland and Taiwan Mandarin, we calculated the relative frequency for each VO in each variant (relative frequency=transitive tokens/all the tokens). The 171 VO compounds were first intersected with the wordlist of Gigaword corpus and Sinica Corpus, in order to exclude the words that are only used in Mainland or Taiwan Mandarin. 16 words such as Mainland unique words 叫板儿 'challenge', 试水 'test the water' were excluded after the intersection. A list of 155 VO compounds was used for further analyses. To test whether the transitivity frequencies of Mainland and Taiwan have significant difference or not, Z-test is utilized. When P-value < 0.05, the difference between two varieties is considered to be significant.

The result of Z-test shows that among all the 155 VO compounds, 33 words show non-significant transitivity differences between Mainland and Taiwan (e.g., 驰 誉 'be famous'/调任 'be transferred' /放眼 'expand one's horizon'). 46 words are actually used as intransitive verbs in both varieties (the transitivity frequencies in both variants are smaller than 1‰, e.g., 参展 'join an exhibition'/出丑 'make a fool of oneself'/对话 'have a dialogue with'/联网 'networking'). Therefore, about half (76/155) VO compounds show statistically significant difference in transitivity frequency between these two variants.

Among these 76 words, 53 Taiwan VO compounds show significant higher transitivity frequency than their Mainland counterparts, while 23 VO compounds have significantly higher transitivity frequency in Mainland. Obviously, Taiwan VO compounds tend to be more likely to have higher transitivity usages.

It should also be noted that, although all these 76 words show statistically significant frequency differences, their degrees of differences vary. The degrees of differences can be measured by the likelihood ratio (=higher frequency/lower frequency), meaning that the higher the ratio is, the more obvious variation differences they have. For example, the transitivity frequency of 过境 'transit' in Taiwan is 0.341 (=341/1000) while in Mainland is 0.033 (=33/1000). The likelihood ratio of Taiwan to Mainland is 10.33, meaning that Taiwan 过境 'transit' is about 10 times more likely to be used as a transitive verb than the Mainland counterpart. If the 76 words are further filtered by the likelihood ratio, the transitivity tendency between Mainland and Taiwan becomes more obvious. When the ratio>=10, it is considered to have prominent significant difference in transitivity frequency between Mainland and Taiwan. Table 1 shows that for all the 8 words whose ratio is larger than 10, their frequencies in Taiwan are all prominently higher than that of in Mainland. Combined with the result of Z-test, it can be summarized that in terms of their transitivity frequency, Taiwan VO compounds are more likely to be used in a transitive way, especially when the variation difference is prominent significant.

Table 1. Words have prominent difference (ratio>=10).

Words	Taiwan	Mainland	Ratio
媲美	727/1021	28/1030	26.19
中意	192/540	8/1337	59.42
把关	182/743	11/1547	34.45
过境	341/1000	33/1000	10.33
献计	6/84	2/1000	35.71
移民	455/2000	1/1000	227.5
接壤	34/922	1/2269	83.67
撤军	23/1000	1/1000	23

More importantly, 7 words have been observed to have contrast transitivity difference (i.e. transitivity usages only detected in one variety) between Mainland and Taiwan. For these 7 words (as shown in Table 2), all of them are used as transitive words frequently and naturally in Taiwan while no transitive usages can be detected in Mainland corpus.

Table 2. VOs have contrast frequency difference

Words	Taiwan	Mainland	Examples in TW
撤兵	1/197	0/46	撤兵 <u>西岸地区</u>
垂爱	5/37	0/2	老天特別垂爱 钟岳岱
领航	76/810	0/169	有能力领航 国家发展
观光	4/1000	0/5224	观光 意大利
转行	18/392	0/167	转行 <u>影视界</u>
失望	3/1000	0/1000	我很失望 他未全力处理问题
过目	22/317	0/65	过目 所有的展品幻灯片

Therefore, referring to the transitivity frequency, Taiwan VO compounds are more likely to be used in a transitive way, and the differences between Mainland and Taiwan are obvious.

3.2 Variation in semantic/syntactic properties

To address the 2nd question, the distributional differences in terms of syntactic and semantic properties should be examined. As mentioned above, we argue that the variation differences usually lie in the presence/absence of a kind of trend; thus, a statistical method based on annotated data is carried on to investigate their preference differences in collocation. 9 features which help to characterize Mainland and Taiwan variants are selected. These features cover semantic (e.g., the semantic role of the taken object), syntactic (e.g., POS of the object; take aspectual marker or not), as well as discourse levels (e.g., the polarity of the context it occurs in; structural parallelism). The detailed annotated schema is shown in Appendix.

The 109 VO compounds used for Z-test are also included in this step. About 200 transitive tokens are randomly collected for each VO compound, half from the Mainland Gigaword Xinhua Agency sub-corpus and the other from the Taiwan Gigaword Central News Agency sub-corpus. It should be noted that some of the VO compounds do not have enough 100 transitive tokens. In that sense, we collect all their transitive usages for annotation. The selection principle is to check whether it can cover all different uses of each VO compound. All examples collected for analysis are manually annotated based on these 9 features. The annotator is a trained expert on Chinese linguistics. Any ambiguous cases were discussed with another two experts in order to reach an agreement.

To examine the variation differences between Mainland and Taiwan, Chi-square test is used to determine the significance of the co-occurrence of each feature with each variety for each individual VO compound. The tool we used is SPSS v. 22. The two variables are considered to have significant difference at P<0.05 level. 24 words which violate the assumption of Chi-square test are first excluded for further analyses (the cells do not have expected counts greater than or equal to five).

Among all the 85 (=109-24) VO compounds, 14 words show no significant distributional difference in any of these 9 features: 入境 'enter a country', 驰誉 'be famous', 留学 study abroad', 入籍 'naturalize', 定都 'establish a capital', 登场 'come on stage', 迁居 'move', 中意 'like', 操心 'worry about', 满意 'be satisfied with', 致函 'write to', 出土 'be unearthed', 变身 'shape shifting', 更名 'rename'. For other 71 words which show significant variation differences in at least one feature, we will discuss their tendency differences in details by each feature.

1) Syntactic type of object

First, about the syntactic type of the object taken by VO compound: among 85 words, 30 words demonstrate significant differences in this feature. Among these 33 words, 15 words present the tendency that the Taiwan VO tends to take event-denoting objects (deverbal noun, VP and clause) while the Mainland counterpart prefer to take common NP as the object (投身 'throw oneself into', 插手 'have a hand in', 动员 'mobilize', 投诉 'complain', 控股 'hold controlling interest', 寄语 'send word', 寄望 'depend on', 签约 'sign a contrast', 参选 'stand for election', 接手 'take over', 聚焦 'focus', 插足 'participate' and 涉嫌 'be suspected'). Only 4 words (立足 'base oneself upon', 起诉 'sue' and 帮忙 'do a favor') display the opposite tendencies (Mainland is more likely to take the event-denoting objects). Several examples are shown in Table 3:

Table 3. Example of syntactic difference

Words	Taiwan	Taiwan Examples	Mainland	Mainland examples
插手	VP	插手经营家族企业	NP	插手地方事务
投诉	Clause	投诉本地导游所提供的 服务欠佳	NP	投诉有关部门
接手	VP	日本接手举办这项比赛	NP	他很愿意接手这 个职务

Therefore, in terms of the syntactic type of the taken object, Taiwan VO compounds present a very strong tendency of taking event-denoting objects while Mainland have more preference of taking common NP as the object.

Another tendency which is not very strong but still can be seen from the result is that, when VO compounds can take both common NP and quantity NP or time NP as the object (撤军黎巴嫩/三十万 'troop withdrawal from Lebanon/withdrawal 30000 troops'; 执教中国男篮/五十年 'coach Chinese National Basketball team/coach for fifty years'), Taiwan VO compounds more prefer the common NP (e.g., 撤军黎巴嫩 'troop withdrawal from Lebanon') while the Mainland may have more preference for the quantity and time object (e.g., 投资三千万 'invest 30 million').

2) Pronominality

For the pronominality property, only 4 words (接手 'take over', 起诉 'sue' and 获赠 'be awarded') display significant differences in this feature. But variation differences still exist, although sometimes are not statistically significance. In general, Taiwan VO are more likely to take person pronoun as the object, the constructions as 帮忙他 'do him a favor', 满意他 'be satisfied with him' can only be detected in Taiwan corpus.

3) Structural parallelism

Referring to the feature of structural parallelism, the tendency between Mainland and Taiwan is quite clear. Among all the 10 words, which show significant differences, 7 of them (把关 'guarantee, 取信 'win trust', 称霸 'seek hegemony', 聚焦 'focus' and 挑战 'challenge') demonstrate the tendency of using structural parallelism strategy in Mainland while only 3words (立足 'base oneself upon' and 放眼 'expand one's horizon') display the opposite trend. In other words, Mainland VO+O construction is more likely to be appeared in the parallelism structures (e.g., 他们坚持把关审核、强化责任、规范操作、配套协调 'They insist on guarantee the checking, strengthen the responsibility, standardize the operation and supporting the coordination).

4) Headline

For this feature, the tendency is very obvious that Mainland VO+O construction is more likely to appear in the headline while the Taiwan ones prefer the normal texts (15 words). For example, 16 out of 104 VO+O constructions of 登陆 'land' (e.g., 登陆上海 'landing Shanghai') are shown in the headline in Taiwan while in Mainland the frequency is 36/104.

5) Complexity of object

Here, the complexity of the object is measured by "whether the object has modifier or not". The variation difference between Mainland and Taiwan is quite clear. Among 27 words which show significant difference in this feature, 18 words show preference over this feature in Taiwan (e.g., 媲美 'rival with', 放眼 'expand one's horizon', 致力 'devote oneself to'). Complicated examples as 任职国界委员会与大陆礁层委员会所辖海洋事务部 'hold a post in Boarders Committee and Ministry of Ocean Affairs under the jurisdiction of the Continental Reef Commission' are not very commonly used in Mainland Mandarin. In contrast, Mainland VO prefers the bare object.

6) Polarity of context

8 words show variation difference in this feature. The difference in preference between Taiwan and Mainland is not very obvious. For Taiwan variety, 2 words prefer neutral context, 3 words have preference for negative context and 4 words favor the positive context (words can show significance in more than one polarity). While in Mainland, only 3 words prefer the negative context, and the other 5 all prefer the neutral context. For instance, 插手 'have a hand in' in Mainland are very likely to be shown in negative context (e.g., 各方势力经常插手流通环节,造成市场秩序十分混乱 'the parities often intervene in the flow of links, resulting in the disruption of market order'), while Taiwan 插手 'have a hand in' are frequently used in both neutral (e.g., 她很少插手姐姐的事情,鼓励她自己做决定 'she rarely intervene the business of her sister, and encourage her to make decision by herself') and positive context (e.g., 各个部门热切希望他可以插手运营,稳定市场秩序 'Every departments are intent on his intervenes on the business operation and help to stabilize the market order').

7) Proper noun

Although more than half (37) words show significant variation difference in this feature, there is no obvious preference difference observed. Among these 37 words, 18 VO compounds in Taiwan show the preference of choosing a proper noun as the object, while in Mainland the other 19 words present the same tendency.

8) Aspectual marker

Aspectual marker taken by VO compounds is also considered. Only 6 words display significance variation difference in this feature, but the difference of tendency is absolute. All of the 6 words (出席 'attend', 出版 'publish', 得罪 'offend', 命名 'name' and 执教 'coach') show significant favor of aspectual marker (either "了" LE or "过" GUO) in Mainland while their Taiwan counterparts show dis-favor over aspectual marker. This tendency is consistent with the findings showed in light verb variation study (Huang et al. [8]).

4 Transitivity in VO Variations

Based on their distributional differences between Taiwan and Mainland Mandarin, we argue that the VO compound itself may differ in its degree of transitivity. Taiwan VO compound have a higher degree of transitivity compared to its Mainland counterpart. The most obvious evidence is that Taiwan VO compounds tend to have higher transitivity frequency (as the result of Z-test presented in Section 3.1), and also Taiwan VO compounds have the tendency of taking more types of NPs and VPs as the complements with less collocation constraints (as has shown in section 3.2). It is more transitive in the sense that it is more likely to be used transitively in different contexts.

Besides this, some other properties as we discussed in section 3.2 also demonstrate the higher transitivity of Taiwan VO compounds. According to Hopper and Thompson [1], 10 parameters can be used to identify the degree of transitivity. VO compounds in two varieties do show variations in some of the parameters (mainly in individuality and affectedness of object). There are mainly three evidences supports this argument.

As mentioned in section 3.2, Taiwan VO compound has the preference of taking VP or clause as the complement (e.g., 投诉澳娱分红不公平 'complain that the dividends of STDM are not unfair'; 插手经营家族企业 'intervene the management of family enterprise') while the Mainland counterparts may prefer deverbal noun (e.g., 插手电商的经营 'intervene the management of electricity business') or common NP (e.g., 插手中亚问题 'intervene the issue of Central Asia'). VP, compared to deverbal noun, is obviously higher in individuality as the patient is overt. Also, as the patient is already in the construction, the affectedness of the objects in Taiwan is also higher than that of in Mainland. The higher individuality and affectedness of object both indicate the higher transitivity of Taiwan VO compounds.

We have shown that Taiwan VO compounds more prefer common NP (e.g., 撤军黎巴嫩 'troop withdrawal from Lebanon') while Mainland may have more preference for quantity and time object (e.g., 投资三千万 'invest 30 million'; 执教五十年 'coach for fifty years'). This should be considered as another evidence to show the higher transitivity of Taiwan VO compounds. Common NP, compared to the object denoting time and quantity, is obviously higher in individuality, and of course, higher in transitivity.

The third evidence is, the Chi-square test shows that Taiwan usages tend to prefer complicated object (object has modifier with DE) while Mainland may prefer bare object. This contrast actually corresponds to the distinction of endurant/perdurant variations (proposed by Huang [9]). According to Huang [9], DE-insertion is allowed only when the object has perdurant properties (a perdurant is an entity which has a time element crucially associated with its meaning). On the contrary, bare object without DE, which is preferred by Mainland Mandarin, may refer to a generic

concept, which is endurant (an endurant, is an entity that has spatial component but does not depend on a specific time of occurrence). Therefore the referentiality of bare object is less, which demonstrates a lower individuality, and also indicates a lower transitivity of VO compound in Mainland.

By showing that Taiwan VO has higher degree of transitivity, we further argue that the higher transitivity of Taiwan VO may indicate the conservatism of Mandarin in Taiwan, since Taiwan VO compounds remain more verbal properties, especially the ability of taking the object. Remaining the transitivity is an evidence of being conservative. In contrast, Mainland VO is introducing a new verb class that can take the object, which is transferring from the intransitive to transitive verbs. The evidence of Mainland usages being innovative is that, we have found for the $VO_1 + O_2$ construction, the frequency of being appeared in headline is much higher in Mainland than in Taiwan Mandarin. The headline, as the summaries of a normal text, has length limit and also has more flexibility. It is usually forced to be innovative and new constructions often emerge from the headline.

The observation that Taiwan VO compounds are being more conservative is consistent with Jiang et al.'s [2] finding in Light Verb Variations.

5 Conclusion

In our current study, a corpus-based statistical approach is adopted to compare the transitivity difference between Mainland and Taiwan Mandarin. Different from previous researches, our study clearly shows that in both frequency and distributional constrains, Taiwan VO compounds have a higher degree of transitivity. We further argue that the higher degree of transitivity is the consequence of attempting to preserve the transitive nature of the verbal construction during incorporation. In this interpretation, Taiwan Mandarin is more conservative, hence consistent with the null hypothesis in historical linguistics that peripheral varieties away from the main speaking community are more conservative.

Appendix

Appendix. Annotation schema of $VO_1 + O_2$ construction

Feature	Values (example)	Feature	Values (example)
1. syntactic type of the object taken	Common NP: 获赠 一部手机	6. The semantic polarity of context	Polarity: 各方都热 切期盼他能插手
	Event noun: 投身大游行		Neutral: 她一般不插 手妹妹的事

	Deverbal noun: 插手		Negative: 多方插手
	电商的运营		农资流通,市场秩 序混乱
	Clause: 投诉澳娱分 紅不合理		777形山
	VP:插手经营企业		
	Quantity: 驻军十万		
	Time: 执教五十年		
2. proper noun or not:	Yes: 移民美国	7.Whether the object of VO has modifier	No: 媲美中亞
noun of not.	No: 移民发达国家		Yes: 媲美这个世界 一流名声大噪的科 技新产品
3. Whether the VO compounds are followed by pronouns	Yes: 帮忙他 No: 帮忙家人	8.Whether the VO compound is affixed with zhe.le.guo	ASP.le: 出版了一套 系列从书
	Tion III IE 307		ASP.zhe: 媲美着世 界闻名的技术
			ASP.guo: 执教过中 国队
4. Whether the VO+O construction	Yes: 经委协调控制,厅局把关审批,铁路监督保证	9. Semantic role of the object	Patient; theme; result; target; beneficient; goal; location
has parallel structure or not	No: 把关服务质量		location
5. Whether the constructions are appeared in headline	Yes: 海地警察头目 要求出国避难太子 港		
	No: 当年避难上海的3万多犹太人		

References

- 1. Hopper, P. J., and Thompson, S. A. 1980. Transitivity in grammar and discourse. *Language*, 251-299.
- 2. Jiang, M. H., Shi, D. X., and Huang, C.R. 2016 Transitivity in light verb variations in Mandarin Chinese- A comparable corpus-based statistical approach. *Proceedings of the 30th Pacific Asia Conference on language, information and computing*. Seoul, Korea.
- 3. Liu, Y., and Li, J. X. 1998. The transformation of VO₁+O₂ construction and the semantic type of O₂ ("动宾式动词+ 宾语" 的变换形式及宾语的语义类型). *Journal of Jinghan University* (江汉大学学报), 5, 011.
- 4. Her, O. S. 1997. Interaction and variation in the Chinese VO construction. Crane Publishing Company.
- 5. Wang, H. D. 1997. The rule of VO₁+O₂ construction ("动宾式动词+ 宾语" 规律 何在?). *Language Planning* (*语文建设*), (8), 30-31
- 6. Diao, Y. B. 1998. Discussion on VO₁+O₂ construction (也谈 "动宾式动词+ 宾语"形式). *Language Planning* (*语文建设*), (6), 39-41.
- 7. Huang, C. R. 2009. *Tagged Chinese Gigaword Version 2.0*. Philadelphia: Lexical Data Consortium, University of Pennsylvania. ISBN 1-58563-516-2
- 8. Huang, C. R., Lin, J. X., Jiang, M. H., and Xu, H. Z. 2014. Corpus-based Study and Identification of Mandarin Chinese Light Verb Variations. *COLING Workshop on Applying NLP Tools to Similar Languages, Varieties and Dialects*. Dublin, August 23.
- 9. Huang, C. R. 2015. Notes on Chinese grammar and ontology: the endurant/perdurant dichotomy and Mandarin DM compounds. *Lingua Sinica*, *1*(1), 1.