

A Comparable Corpus-based Study of three DO verbs in Varieties of Mandarin: 從事 *congshi*, 做 *zuo* and 搞 *gao*

Menghan Jiang and Chu-Ren Huang

Department of CBS, The Hong Kong Polytechnic University, Hong Kong, China

menghan.jiang@connect.polyu.hk;
churen.huang@polyu.edu.hk

Abstract. In this study, we adopt a comparable corpus-based approach to investigate variations of three DO verbs in Mandarin Chinese: 做 *zuo* ‘do’, 搞 *gao* ‘do’ and 從事 *congshi* ‘be engaged in’. Mandarin Chinese is unique in having three light verbs with bare meaning. The interesting and challenging facts about these three DO verbs are that: first, their usages can be differentiated even though they share the bare minimal meaning of ‘to do’; and second, their ranges of usages vary in different varieties of Chinese. How can the complex differentiations of these three verbs within one variety and across different varieties be accounted for with the minimal shared meaning? We tackle this challenge applying functions from Chinese Word Sketch to effectively identify the subtle differences among near-synonyms and their usage variations among different varieties with explicit semantic cues. This study thus underlines the contribution of empirical approaches when there is very little intuition to rely on.

Keywords: Do Verbs; Corpus-based Approach; Language Variation

1 Introduction

In Modern Chinese, the verbs 做 *zuo* 搞 *gao* and 從事 *congshi* share the function of English verb DO and are typically used to define one another in dictionary glosses [1, 2]. This is quite unique since a DO verb is supposed to be the lighted verb without any substantive meaning and the eventive meaning comes from its complements without (much) contribution from the light verb. For instance, 做研究 is considered to be (almost) totally synonymous to the verb 研究. Given the lack of semantic content, the uses of these three verbs are often interchangeable, as predicted. E.g. 從事 / 做 / 搞 研究 *congshi/zuo/gao yanjiu*, ‘do research’. However, they are not interchangeable all the time. E.g. 從事 / 搞 / *做 養殖業, *congshi/gao/*zuo yangzhiye* ‘engaged in breeding’, *從事 / ? 搞 / 做 買賣 *congshi/*gao/zuo maimai* ‘engaged in

business’ [3, 4]. Given the ‘bare’ meaning of these three verbs, then, what is the lexical semantic motivation that allows speakers to use them differently?

Furthermore, the usages of these three verbs are not identical in different Mandarin. Previous scholars [1, 5-6] observed many differences in usages of these DO verbs in different Mandarin varieties. For instance, 從事 *congshi* in Taiwan Mandarin can take complements denoting negative events such as 勾當 *goudang* ‘shady business’ and 性交易 *xingjiaoyi* ‘sex trade’, while such complements are seldom found in Mainland Mandarin. Assuming that such distributional and collocational differences arose from the different meanings and semantic constraints, how can we account for the full range of complex variations when the lexical meaning(s) of these verbs are already bare?

2 Previous Studies on 做 *zuo*, 搞 *gao* and 從事 *congshi*

Although some studies can be found on these three words respectively [1, 2], very few has focused on the differentiation and identification of them, let alone the verb variations between different variants of Mandarin. Wang [4] describes the (dis-)similarity of 做 *zuo* and 搞 *gao*. He summarizes that both 做 *zuo* and 搞 *gao* can take a very wide range of complements, including verbal and nominal. The syntactic collocations of these two verbs are very similar, except for the position of the theme of taken complement. The difference between 做 *zuo* and 搞 *gao* mainly lies in that 做 *zuo* can use preposition to introduce the theme, as in 對數據做分析 *dui shuju zuofenxi* for_data_do_analyze ‘analyze the data’, while 搞 *gao* cannot (e.g., *對數據搞分析 **duishuju gaofenxi* for_data_do_analyze ‘analyze the data’).

In this study, we are taking a comparable corpus-based approach, aiming at (a) differentiating these three verbs in Mainland Mandarin; (b) identifying the similarities and differences of each verb in two varieties, i.e., Taiwan Mandarin and Mainland Mandarin.

In terms of the semantic property of these three verbs, since these three verbs themselves are semantically impoverished, they themselves usually do not contain eventive or contentive information. The most important predicative information mainly comes from their taken complements. With respect to syntactic collocation, we find these verbs are very similar in syntactic behavior (i.e. their syntactic function, whether it can take aspectual marker, the sentential type they can occur in), while vary a lot in their taken complements (i.e. what complements can they take). Therefore, the focus of our study is to look at the taken complements of 從事 *congshi*, 做 *zuo* and 搞 *gao* in both Mainland Chinese and Taiwan Mandarin Chinese.

3. Data Analysis

The data for this study was extracted from Annotated Chinese Gigaword corpus [7], which was collected and available from LDC and contains over 1.1 billion Chinese words, consisting of 700 million characters from Taiwan Central News Agency and

400 million characters from Mainland Xinhua News Agency. Variety internal comparison (i.e. the differentiation among these three verbs in one variety) as well as variety external comparison (i.e. differentiation between different varieties) will be conducted.

3.1 Differentiation of 做 *zuo* 搞 *gao* and 從事 *congshi* in Mainland Mandarin

The function of word sketch in the sketch engine is used to examine the syntactic collocation for each word. The word sketch is a one-page, automatic, corpus-derived summary of a word's grammatical and collocational behavior. Its collocates are grouped according to grammatical relations in which they occur. In our study, we selected the 50 most frequently occurred complements for each verb. Appendix 1 shows the 50 most frequently occurred complements/objects taken by Mainland 從事 *congshi*.

Through an overview of the data, it is obvious to see that the complements of 從事 *congshi* are very limited in number and type. Most of them are nouns referring to profession and career, therefore the complement usually has the suffix 業 *ye* 'trade'.

For these collocations, if we replace 從事 *congshi* by 做 *zuo*, more than 50% (29/50) of the replaced constructions are not natural. Different from 從事 *congshi*, the complements taken by 做 *zuo* are usually verbs/deverbal nouns which refer to specific event or action like 研究 *yanjiu* 'research'/設計 *sheji* 'design'/交易 *jiaoyi* 'trade', as shown in Appendix 2.

Different from 做 *zuo*, all the 50 most frequently complements taken by 從事 *congshi* can also be taken by 搞 *gao*, without changing the meaning of the whole construction. (e.g., 搞電子商務/從事電子商務 *gao/congshi dianzi shangwu do_electronic_commerce* 'engage in e-commerce'; 從事水產養殖業/搞水產養殖業 *congshi/gao yanzhiye do_aquaculture* 'engage in aquaculture').

Based on the syntactic collocation shown by the large-scale corpus, we can summarize as 從事 *congshi* and 搞 *gao* are syntactically more similar, while 做 *zuo* tends to co-occur with different types of complements.

As can be seen from Appendix 3, the complements can only be taken by 做 *zuo* usually refer to specific telic events which have an endpoint and always requires certain outcome. For example, some of the constructions are accomplishment events like 做記錄 *zuo jilu* 'make record' and 做對比 *zuo duibi* 'make comparison', while most of the taken complements are activity like 報告 *baogao* 'report', 計畫 *jihua* 'plan', 論述 *lunshu* 'dissertate'. In other words, the complements taken by 做 *zuo* usually refers to dynamic process instead of state.

It would be clearer if we compare the usages of 做 *zuo* and 搞 *gao*, even when taking the same complements, their event structures are different. As the example given by Diao [1] in (1):

- (1) 做錄音---他在現場做錄音---? 他在現場搞錄音
zuo luyin --- ta zai xianchang zuo luyin --- ?ta zai xianchang gao luyin

‘make sound recording ---he is recording sound on site’
搞錄音---他是搞錄音的---? 他是做錄音的
gao luyin – ta shi gao luyin de – ?tashi zuo luyin de
‘make sound recording – he is a sound recordist’

The construction 做錄音 *zuo luyin* ‘make sound recording’ refers to a specific event that the subject is making record and this event requires certain outcome (like a CD or tape). In contrast, 搞錄音 *gao luyin* just implicates that the profession of the subject is a sound recordist. A locative modification can be inserted (e.g., 在現場 *zai xianchang* ‘on site’) before 做錄音 to specify the event (e.g., 他在活動現場做錄音 *ta zai huodong xianchang zuo luyin* ‘He is making sound recording on site’).

Moreover, when the three verbs all take the same complement 研究 *yanjiu* ‘research’, the meaning of the whole construction are also different. 做研究 *zuo yanjiu* ‘conduct research’ refers to a specific event while 搞研究 *gaoyanjiu* ‘conduct research’ and 從事研究 *congshi yanjiu* ‘conduct research’ usually refers to long-term career or profession (i.e. He is a researcher). Therefore, the progressive marker 正在 *zhengzai* and location marker 辦公室 ‘office’ can be used to test their event structure, since 正在 *zhengzai* is usually compatible with dynamic process (shown in (2)).

(2) 他正在辦公室做研究
ta zhengzai bangongshi zuo yanjiu
‘He is doing research in the office now’
? 他正在辦公室搞研究
ta zhengzai bangongshi gao yanjiu
‘He is doing research in the office now’
*他正在辦公室從事研究
ta zhengzai bangongshi congshi yanjiu
‘He is doing research in the office now’

As examples shown above, 做研究 *zuo yanjiu* as an activity event can co-occur with 正在 *zhengzai* and location, while 從事研究 *congshi yanjiu* as a state cannot compatible with it.

With respect to 搞 *gao*, the corpus data shows that it can take a much wider range of complements, the most frequently occurred ones are shown below.

Among all the 50 most frequently occurred complements of 搞 *gao*, 23 examples cannot be taken by 從事 *congshi* while 13 tokens cannot be taken by 做 *zuo*. This may help to explain why In Chinese traditional grammar, 搞 *gao* is called 萬能動詞 *wanneng dongci* ‘universal verb’. It is ‘universal’ in the sense that it can be used to replace a variety of verbs under different contexts. It has a wider range than 從事 *congshi* and 做 *zuo* in taking complements.

Summary:

Based on the analysis above, it is clear to see that 從事 *congshi* and 搞 *gao* are often interchangeable. For complements taken by 從事 *congshi*, almost all of them can also be taken by 搞 *gao*. 搞 *gao* has a much wider range in usage in the sense that it can take a much wider range of complements.

做 *zuo* behaves the most differently from 搞 *gao* and 從事 *congshi* in the sense that it seldom take nouns which refer to profession or career (a state). Instead, it usually takes verb or deverbal noun which refers to specific event or action, with a dynamic process. This event usually requires telic point or an outcome.

3.2 Comparison between Taiwan and Mainland Mandarin

To identify the variation differences between different variants of Chinese, a random sample of 200 sentences for each of the three verbs was extracted from both Mainland Mandarin sub-corpus and TW Mandarin sub-corpus. Specific usages (i.e. usages can only be detected in one variety) in each variety have been annotated

3.2.1 從事 *congshi* in Mainland and Taiwan Mandarin

TW 從事 *congshi* and Mainland 從事 *congshi* are very similar in usage. Specific usages can only be found in Taiwan corpus. For example, 從事 *congshi* in TW can take informal event complements such as 休閒娛樂 *xiuxian yule* “entertainment”, 散步等運動 *sanbu deng yundong* “do exercises like take a walk”, as well as complements denoting negative events such as 性交易 *xing jiaoyi*, 色情行業 *seqing hangye*, “sex trade”, etc. In contrast, 從事 *congshi* in Mainland can only take formal or positive event as its complement (e.g., 從事國事訪問 *congshi guoshi fangwe* ‘have a state visit’), and this is consistent with the result shown in Huang et al. [8].

3.2.2 做 *zuo* in ML and TW Chinese

Our data shows that the event complements of Mainland 做 *zuo* are limited. Among 200 sentences we have annotated, 72 are 做工作 *zuo...gongzuo* ‘do...work’. The only specific usage is 做...努力 *zuo...nuli* ‘take efforts’.

Different from Mainland 做 *zuo*, TW 做 *zuo* can take a much wider range of complements. Among the 200 sentences, only 7 are 做工作 *zuo...gongzuo*. Plenty of specific usage can be found, 做答覆 *zuodafu* ‘to reply’, 做以上的表示 *zuo yishang biaoshi* ‘make statement’, 做人身攻擊 *zuo renshengongji* ‘to make a personal attack’, 做澄清 *zuo chengqing* ‘to clarify’, 做接觸 *zuojietchu* ‘to touch’, 做衝刺 *zuo chongci* ‘to sprint’, 做競賽 *zuo jingsai* ‘to participate in the competition’.

The comparison of TW and ML 做 *zuo* shows that TW 做 *zuo* is semantically more bleached (light in meaning) in the sense that it can take a much wider range of complements. Therefore, for most of the specific complements taken by TW 做 *zuo*,

they are often taken by other light verbs in ML. For example, 做答覆 *zuo dafu* is seen as 進行答覆 *jinxing dafu* in ML Chinese, 做人身攻擊 *zuo renshengongji* is 進行人身攻擊 *jinxing renshen gongji*, 做澄清 *zuo chengqing* is 進行澄清 *jinxing chengqing*, 做考慮 *zuo kaolv* is always used as 加以考慮 *jiayi kaolv*, 做接觸 *zuo jiechu* is 進行接觸 *jinxing jiechu*, 做競賽 *zuo jingsai* is 進行競賽 *jinxing jingsai* in ML Mandarin. This may show that 做 *zuo* in TW tend to be used as the default light verb.

3.2.3 搞 *gao* in ML and TW Chinese

When it comes to 搞 *gao*, our annotated results show that Mainland 搞 *gao* can take a variety of complements, and plenty of specific usages can be found. The semantic domains of these complements are also very diverse, which is consistent with our findings in the above section. According to SUMO (Suggested Upper Merged Ontology) [9], 傾銷 *qingxiao* ‘dump’, 直銷 *zhixiao* ‘direct sales’, 批發 *pifa* ‘wholesale’, 銷售 *xiaoshou* ‘sales’ and 貿易 *maoyi* ‘trade’ belong to the domain of selling; while 運輸 *yunshu* ‘transportation’, 販運 *fanyun* ‘transport goods for sale’, 航運 *hangyun* ‘shipping’ belong to the domain of transportation, and also 製作 *zhizuo* ‘manufacture’, 加工 *jiagong* ‘process’, 設計 *sheji* ‘design’ belong to the concept of making.

In contrast, the complements of TW 搞 *gao* are very limited in types. More than 90% complements are political event (and also related to ML context). The specific complements mostly refer to specific political event of Taiwan and can only be appeared in the negative context. For example, 搞形式主义/和平演变/分裂/抹黑 *gao xingshizhuyi/heping yanbian/fenlie/mohei* do_formalism/peaceful evolution/dispute/throw-mud ‘take formalism approach/make peaceful evolution/cause state disruption/throw mud’.

In general, for 做 *zuo* and 搞 *gao* in both Mainland and Taiwan, we can summarize as the usage of Taiwan 做 *zuo* is semantically much lighter than its Mainland counterparts, in the sense that it can take a much wider range of complements under different context. While in Mainland, 搞 *gao* may be semantically more bleached than its Taiwan counterparts, because Taiwan 搞 *gao* can barely be used in other context except for political context with the negative meaning.

4 Conclusion

In this study, we differentiate three DO verbs in Mandarin Chinese 從事 *congshi*, 做 *zuo* and 搞 *gao*, both in terms of their collocational constraints and their usage variations in different varieties of Chinese. By examining their differences in the taken complements as well as the event structure, we also identify the variation differences for each corresponding verb between Mainland and Taiwan. We show that, with the bare meaning of these light verbs, they differ from either other in terms of the

event types they select. Similarly, by alternations of the groupings, different event types, varieties of Chinese also use these DO verbs differently. We show that, with the assistance of large-scale comparable corpus, a study of near synonyms share bare meaning is possible even though there is very little ‘meaning’ or intuition to rely on.

References

1. Diao, Yanbin. *Research on Delexical Verb in Modern Chinese* (現代漢語虛義動詞研究). Dalian: Liaoning Normal University Press (2004)
2. Du quner. Form Verbs in Modern Chinese (現代漢語形式動詞研究), Master’s dissertation, *Shanghai Normal University*. (2010)
3. Huang, C. R., Lin, J. X., Jiang, M. H., and Xu, H. Z. Corpus-based Study and Identification of Mandarin Chinese Light Verb Variations. *COLING Workshop on Applying NLP Tools to Similar Languages, Varieties and Dialects*. Dublin. (2014).
4. Wang Jian. Comparison and Collocation of “Gan, Gao, Zuo”, (“干, 搞, 做” 的詞語搭配範圍及其詞義對比), *Chinese Teaching Overseas* (海外華文教學), 2, 48-53. (2002).
5. Huang, Chu-Ren and Jingxia Lin. The ordering of Mandarin Chinese light verbs. *Proceedings of the 13th Chinese Lexical Semantics Workshop. D. Ji and G. Xiao (Eds.): CLSW 2012, LNAI 7717*, pp. 728-735. Heidelberg: Springer. (2013).
6. Jiang, M. H., Shi, D. X., and Huang, C.R. 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. (2016).
7. Huang, C. R.. Tagged Chinese Gigaword Version 2.0. Philadelphia: Lexical Data Consortium, University of Pennsylvania. ISBN 1-58563-516-2. (2009)
8. Huang Chu-Ren, Jingxia Lin, and Huarui Zhang. World Chinesees based on comparable corpus: The case of grammatical variations of jinxing. 《澳門語言文化研究》, 397-414. (2013).
9. Niles, I., & Pease, A.. Towards a standard upper ontology. In *Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001* (pp. 2-9). ACM. (2001)

Appendix:

研究 <i>Yanjiu</i>	第三產業 <i>disanchanye</i>	活動 <i>huodong</i>	業務 <i>yewu</i>	工作 <i>gongzuo</i>	第二職業 <i>di'er zhiye</i>
運動 <i>yundong</i>	種植業 <i>zhongzhiye</i>	農副產品 <i>nongfuchanpin</i>	服務業 <i>fuwuye</i>	房地產 <i>fangdichan</i>	創作 <i>chuangzuo</i>
交易 <i>Jiaoyi</i>	職業 <i>zhiye</i>	農業 <i>nongye</i>	建築業 <i>Jianzhu ye</i>	科研 <i>keyan</i>	公務 <i>gongwu</i>
漁業 <i>Yuye</i>	作業 <i>zuoye</i>	商業 <i>shangye</i>	貿易 <i>maoyi</i>	農牧業 <i>nongmuye</i>	建築 <i>Jianzhu ye</i>
旅遊業 <i>lyyouye</i>	設計 <i>sheji</i>	事業 <i>shiyeye</i>	產業 <i>chanye</i>	文學 <i>wenxue</i>	養殖業 <i>Yangzhiye</i>
社區服務 <i>shequfuwu</i>	訓練 <i>xunlian</i>	教育 <i>jiaoyi</i>	副業 <i>fuye</i>	種養業 <i>zhongyangye</i>	生意 <i>shengyi</i>
加工業 <i>Jiagongye</i>	客運 <i>keyun</i>	工商業 <i>gongshangye</i>	家政服務 <i>jiazheng fuwu</i>	行業 <i>hangye</i>	運輸業 <i>Yunshuye</i>
製造業 <i>zhizaoye</i>	電子商務 <i>dianzi shangwu</i>	工種 <i>gongzhong</i>	畜牧業 <i>xumuye</i>	庭院經濟 <i>tingyuan jingji</i>	商貿 <i>Shangmao</i>
體育活動 <i>tiyu huodong</i>	環保產業 <i>huanbaochanye</i>				

Appendix 1. The 50 most frequently occurred complements taken by 從事 *congshi*

手術 <i>Shoushu</i>	文章 <i>wenzhang</i>	生意 <i>shengyi</i>	貢獻 <i>gongxian</i>	工作 <i>gongzuo</i>	報告 <i>baogao</i>	實驗 <i>shiyan</i>
試驗 <i>Shiyan</i>	嘗試 <i>changshi</i>	規定 <i>guiding</i>	小買賣 <i>xiaomaimai</i>	宣傳 <i>xuanchuan</i>	記錄 <i>jilu</i>	買賣 <i>Maimai</i>
交易 <i>Jiaoyi</i>	學問 <i>xuewen</i>	演講 <i>yanjiang</i>	保證 <i>baozheng</i>	解釋 <i>jieshi</i>	研究 <i>yanjiu</i>	人工呼吸 <i>rengonghuxi</i>
調查 <i>Diaocha</i>	翻譯 <i>fanyi</i>	統計 <i>tongji</i>	評價 <i>pingjia</i>	評論 <i>pinglun</i>	小本生意 <i>xiaoben shengyi</i>	總結 <i>Zongjie</i>
訓練 <i>xunlian</i>	限制 <i>xianzhi</i>	報導 <i>baodao</i>	決策 <i>juece</i>	規劃 <i>guihua</i>	對比 <i>duibi</i>	保障 <i>baozhang</i>
計畫 <i>Jihua</i>	決定 <i>jueding</i>	尿檢 <i>niaojian</i>	後勤 <i>houqin</i>	論述 <i>lunshu</i>	抵押 <i>diya</i>	設計 <i>Sheji</i>
分析 <i>Fenxi</i>	測試 <i>ceshi</i>	建設 <i>jianshe</i>	科研 <i>keyan</i>			

Appendix 2. The 50 most frequently occurred complements taken by 做 *zuo*

手術 <i>shoushu</i>	文章 <i>wenzhang</i>	貢獻 <i>gongxian</i>	報告 <i>baogao</i>	嘗試 <i>changshi</i>	規定 <i>Guiding</i>
記錄 <i>Jilu</i>	演講 <i>yanjiang</i>	保證 <i>baozheng</i>	解釋 <i>jieshi</i>	評價 <i>pingjia</i>	評論 <i>Pinglun</i>
總結 <i>zongjie</i>	限制 <i>xianzhi</i>	決策 <i>juece</i>	對比 <i>duibi</i>	保障 <i>baozhang</i>	計畫 <i>Jihua</i>
決定 <i>jueding</i>	論述 <i>lunshu</i>	抵押 <i>diya</i>			

Appendix 3. Complements can only be taken by 做 *zuo*