Corpus-based Comparison of Verbs of Separation “Qie” and “Ge”

Nga-In Wu¹, Chu-Ren Huang², and Lap-Kei Lee³

¹College of Professional and Continuing Education, Hong Kong Polytechnic University
ngain.wu@cpce-polyu.edu.hk
²Department of Chinese and Bilingual Studies, Hong Kong Polytechnic University
churen.huang@polyu.edu.hk
³School of Science and Technology, The Open University of Hong Kong
lklee@ouhk.edu.hk

Abstract

This paper predicts that Chinese Synonyms qie and ge are verbs of separation and uses a variety of Chinese Word Sketch (CWS) functions to distinguish them. Several subtle differences are demonstrated in modifying relation and noun-verb relation, showing that the use of the two target words differ mainly in terms of the purpose of separation. Developing history is one of the factors why the use of qie and ge differ across the straits. These findings are more detailed when comparing with the work focusing on dictionary study. Obviously, traditional dictionary is no longer enough to Chinese language learners. This study is expected to provide some insights for Chinese dictionary editors and hence Chinese teachers.

1 Introduction

Many studies were done on near synonyms in Mandarin Chinese, and verb has been particular interesting to scholars (Wang and Huang, 2018). Qie 切 and ge 割 is one of the interesting pair of near synonyms. As a native speaker, semantic difference between the two words is not clear at the first glance. It is interesting to note that 切 割 is acceptable, while 割 切 sounds strange. This implies that there should be a semantic difference between two words, because pure coordination usually allows reversed order.

Studies on qie 切 and ge 割 have been done by Lian (2005) based on dictionary before. Yet, she failed to identify unique features of the two words. It may be because polysemy of words is not supported in a traditional dictionary. If we simply look at the definitions provided, it is not feasible to distinguish the difference between their usages, especially in different part of speech (Fillmore and Atkins, 1992). To fill the gap, Chinese Word Sketch (CWS) will be used in this article; CWS is a combination of Word Sketch Engine (Kilgarriff et al., 2005) and Chinese GigaWord Corpus (Huang et al., 2005). With the help of “computer-aided armchair linguistics” (Fillmore, 1992), it is believed that some common and unique features of the two words will be found, as the observations are based on large amount of authentic data. This method should be more efficient than relying on researchers’ background knowledge merely to process the data (Li et al., 2018) and more reliable than studying the dictionary.

Our contribution. This paper tries to find out grammatical and collocational relations of qie 切 and ge 割, hoping to identify the differences and similarities between these two synonyms so as to figure out unique features and core meanings of the two words. Cross-strait comparison is also done, which aims to see how the use of two words differ in Mainland and Taiwan in view of different developing history and time. We expect that this study will provide insights to dictionary editing and writing.

Organization of paper. Section 2 states the research questions. Section 3 examines the meanings of the two words in dictionaries, the significant claim form Lain (2005), the classification suggested by Lian (2005) based on dictionary, and the frequency distribution in the Chinese GigaWord corpus. Section 4 and 5 are a cross-strait comparison and a summary.
2 Research Questions

This paper explores the research questions below:
(1) What are the grammatical and collocational relations of the target words found based on the Chinese Word Sketch results? (2) Are there any unique features and core meanings for the two words? If yes, what are they? (3) What are the differences on the usage of the two words in Mainland and Taiwan?

3 Dictionary-based and Corpus-based analysis

Dictionary-based analysis. As shown in Table 1, both qie 切 and ge 割 mean “to cut” in English. Ambiguity between two words is found when we refer to the definitions of the Contemporary Chinese Dictionary.

| 切 qie | 割 ge |
|---|---|
| Pronunciation 1: qie to cut / to slice | to cut / to cut apart |
| Pronunciation 2: qie definitely / absolutely (not) / (scoffing or dismissive interjection) |  |
| Pronunciation 2: qie 符合 / 貼近；親近 / 急切；殷切 / 切實：務必 | 用刀截斷：分割：捨棄 |

Table 1: Explanation of qie and ge in dictionary

Lian (2005) has tried to figure out the features of the two words (see Table 2) based on dictionary. However, the study failed to give real explanation to the two words. Lian (2005) used other near synonyms to explain and distinguish qie 切 and ge 割. Lian (2005) used fen 分 to paraphrase qie 切 and zhe 截 to paraphrase ge 割. Clearly, definitions in dictionary are not sufficient to tell the unique features of the two words.

Although Lian (2005) failed to give the real explanation of the two verbs, her claim gives a great implication to this paper (i.e. distinguishing meaning of words by using different paraphrases). When we paraphrase the verb duan 斷 in the Chinese classical poem choudao dianshui shui gen liu 抽刀斷水水更流, ge 割 is acceptable. It is found that water is not really cut by knife, but separated. Therefore, this paper predicts that qie 切 and ge 割 are verbs of separation instead of just verb of cutting.

To see how the corpus data is useful on capturing the features so as to modify the definitions, the classification proposed by Lian (2005) will be adopted and discussed in this paper.

| 切 qie | 割 ge |
|---|---|
| 1. Tool used for the action – Dao 刀 knife | 1. Tool used for the action – Dao 刀 knife |
| 2. Process involved in the action – Fen 分 separate | 2. Process involved in the action – Zhe 截 cut |
| 3. Final state of object being cut because of the action – Cheng ruogan bufen 成若干部分 become several pieces | 3. Final state of object being cut because of the action – Duan 斷 separate |
| 4. Object being cut in the action – Wupin 物品 product |

Table 2: Analysis of qie and ge in the work of Lian (2005)

Gigaword corpus via CWS is used in this paper. We will present details of Gigaword corpus and the frequency distribution of the two words in this section.

---

1 According to the Contemporary Chinese Dictionary, there are two pronunciations for the word qie 切. Only qie 切 in the first tone giving similar meaning as 割 ge will be discussed in this paper.
**Corpus-based analysis.** Chinese Gigaword corpus data consists of three sub-corpora which are corpora coming from Central News Agency in Taiwan (CNA, 501,456,000 words), Xinhua News Agency in Mainland (XIN, 311,660,000 words) and Lianhe Zaobao in Singapore (Gigaword2zbn, 18,632,000 words). Table 3 shows the overall frequency and frequency of the two words in Gigaword2cna and Gigaword2xin. It is found that the overall frequency of qie 切 per million words is almost four times higher than ge 割. Also, the frequency of qie 切 is four times higher than ge 割 in Mainland and China. Based on the results, it is found that the use of qie 切 is dominant across the straits. Mainland and Taiwan share the same preference on the usage of qie 切.

| Corpora      | freq. | freq./million | freq. | freq./million |
|--------------|-------|---------------|-------|---------------|
| Gigaword2all | 1352  | 1.63          | 831   | 6.09          |
| Gigaword2cna | 540   | 1.08          | 2019  | 4.03          |
| Gigaword2xin | 750   | 2.41          | 2821  | 9.07          |

Table 3: Frequency of ge 割 and qie 切 in corpora

The following sections find out the similarities and differences between qie 切 and ge 割 in terms of lexical grammatical relations, and the features are discussed and categorized according to the classification proposed by Lian (2005).

4 Grammatical Patterns Through Word Sketch

The Word Sketch function helps to illustrate the relations the target word has and the salient words within the relation. The *minimum frequency* is set at 5. Clicking *Show Word Sketch* and then inputting each word generate the result in Table 4.

|        | PP_到 | Subject | Object | SentObj | et_of | Modifier | Modifies |
|--------|-------|---------|--------|---------|--------|----------|----------|
| qie    | ✓     | ✓       |        | ✓       |        | ✓        | ✓        |
| ge     | ✓     | ✓       |        | ✓       | ✓      |          |          |

Table 4: Grammatical patterns of qie 切 and ge 割

As shown in Table 4, there are more grammatical patterns for qie 切 than ge 割. This may explain why the frequency of the use of qie 切 is higher than the use of ge 割 as mentioned in the corpus-based analysis in Section 3. It is found that PP_給 and modifiers are relations absent for ge 割. When we set *minimum frequency* to 2, only *youshouwan* 右手腕 right wrist appears in the modifies relation for ge 割. Yet, PP_給 still does not appear in the prepositional relation for ge 割.

It is interesting to note that PP_將 appears for ge 割 at the *minimum frequency* of 2. The instance given below suggests that ge 割 is done for a particular goal. Feeding cow is the purpose. This observation suggests that feature “purpose of the action” should be added to the classification proposed by Lian (2005).

(1) 乾脆 分批 分期 將 麥子 割 了 喂牛。

gān cuì fèn pī fèn qī jiāng mài zi gē le wèi niú

Simply in batches JIANG wheat cut ASP fed cow

‘Simply harvest wheat by stages to feed cow.’

PP_到 and PP_把 are the other two prepositional relations appearing for ge 割 at the *minimum frequency* of 2. However, it is found that two instances containing PP_把 generated by the SkE are mismatches to ge 割. Ba 把 set does not appear as a prepositional relation collocating with ge 割. Ba 把 set is a classifier modifying dao 刀 knife. The two instances generated are as follows.

(2) 如同 三把 刀 強行 割 佔 了 中國 1 0 0 0 多 平方

Rú tíng sānbǎ dāo qiáng xíng gē zhàn le zhōngguó 1 0 0 0 duō píng fāng

Like three CLASSIFIER knife forcibly cut ASP China 1000 square meters.

Take 1000 square meters from China forcibly like three knives.

(3) 他們 再 用 一把 刀 割 我 的 長褲。

Tāmen zài yòng yī bā dāo gē wǒ de cháng kù
They again use one CLASSIFIER knife cut I DE trouser
They cut my trousers with a knife again.

5 Common Patterns and Only Patterns via Sketch Diff

The Sketch Diff function can compare and contrast two words in one time. It can help to find the common patterns and exclusive patterns of the pair of words. Table 5 is the Word Sketch Differences Entry Form. The default setting is used: the minimum frequency is 5; the maximum number of items in a grammatical relation of the common block is 25; the maximum number of items in a grammatical relation of the exclusive block is 12. After generating the results, this section will further explore the unique and common features of target words. To ensure the accuracy, mistakes like instances of PP_ 把 in Section 4 are removed from the results.

Table 6: Common patterns of qie and ge

As shown in Table 6, qie 切 and ge 割 are similar in three aspects, i.e. they can have ren 人 person as a subject; they can have shangkou 傷口 wound as an object; and they can be modified by ye 也 also, yao 要 necessity, bu 不 negation, and zai 再 again. These results imply that the two words share a core meaning (i.e. a wound made by someone).

6 Common patterns of qie and ge

The colour chain generated can show the tendency (see Table 6). The words are highlighted in red and green. The greener the word means it has a higher tendency to collocate with qie 切. While the words are in red colour, it means that they have higher tendency to collocate with ge 割.

7 Only patterns of qie and ge

After focusing on the common pattern of the two words, this section focuses on exclusive patterns to figure out the unique features of the two words. As shown in Table 7, it is noticed that qie 切 and ge 割 differ in five grammatical relations including subject, object, modifier, sentObject_of and modifies. To facilitate the analysis, the five grammatical relations are categorized into two categories which are noun-verb relation and modifying relation. The following section focuses on noun-verb relation first.
imply that "ge" 割 is used especially with knives for specific purpose. In contrast, we only know that "qie" 切 collocates with "dao" 刀 knife as a subject according to Table 6, and it collocates with "yonglai" 用來 used for for the SentObject_of relation suggesting that a tool should be used. It implies that "qie" 切 can be used with any knives as a subject.

**Process involved in the action.** It is noticed that food such as "dangao" 蛋糕 cake, "cai" 菜 vegetable and "shougao" 壽糕 birthday cake are objects collocated with "qie" 切. These objects imply that "qie" 切 refers to a fixed cutting method. On the other hand, organs or tissues (e.g. "shuangyanpi" 双眼皮 double eyelid, "baopi" 包皮 foreskin and "hou" 喉 throat) and crops (e.g. "maizi" 麦子 wheat and "daozi" 稻子 paddy) are objects tend to collocate with "ge" 割. "Shuangyanpi" 双眼皮 double eyelid and "hou" 喉 throat suggests that the cutting method is flexible which can be horizontal cutting, ring cutting and diagonal cutting.

**Object being cut in the action.** Meanwhile, the collocates (e.g. "cai" 菜 vegetable and "shougao" 壽糕 birthday cake) with "qie" 切 which are served as objects imply that the target being cut should be placed horizontally on a surface. The target of cutting should not be too small, as they can be cut into several pieces. In contrast, the collocates such as "shuangyanpi" 双眼皮 double eyelid, "baopi" 包皮 foreskin for "ge" 割 acting as objects suggest that the cutting target can be small. "Maizi" 麦子 wheat and "daozi" 稻子 paddy suggests that "ge" 割 can be used when the target of cutting is standing upright.

**Final state of object being cut.** "Dangao" 蛋糕 cake, "cai" 菜 vegetable and "shougao" 壽糕 birthday cake act as object for "qie" 切 suggest that the actual amount of the cutting target should remain unchanged after the cutting process. The cutting target is cut into several pieces. In contrast, "shuangyanpi" 双眼皮 double eyelid, "baopi" 包皮 foreskin served as objects for "ge" 割 suggest that a part is be removed and taken away. The actual amount of the cutting target should be different after the process.

**Purpose of the action.** Based on findings above, obviously, "qie" 切 is especially used in cooking context aiming to cut the target into several pieces and "qie" 切 can be done with any

7.1 Noun-verb relation

This category of grammatical relation refers to the relation between collocated words and target words which are served as subjects or objects and verbs respectively. As mentioned earlier, the modified Lian (2005)'s classification (i.e. with the new feature) is adopted for further analysis in the following sections.

**Tool used for the Action.** As shown in Table 6, it is found that "ge" 割 tends to collocate with knives as a subject such as "ren" 刀 blade, "liandao" 镰刀 sickle and "meigongdao" 美工刀 utility knife. It
kinds of knives or tools with a fixed cutting method. *Qie 切* is a verb which is result-oriented. On the contrary, it is found that *ge 割* should be done with specific knives for a particular purpose. Also, the aim of *ge 割* can refer to the removal of a small part of the cutting target with a more flexible cutting method. The verb *ge 割* is purpose-oriented. We can clearly see that this is a unique feature to the two target words, and it is rather abstract which can be explained and supported by other four features.

| Subject | Object | SentObject |
|---------|--------|------------|
| 大家 We | 蛋糕 cake, 菜 vegetable, 壽糕 birthday cake | 用來 Used for |
| 費斯 Fisichella | 梅爾金 Chemerkin, 戈夫州 Chernihiv oblast, | |
| 顱刀 sickle, 美工刀 Utility knife, 刀子 knife | 雙眼皮 double eyelids, 麥子 wheat, 腕 wrist,喉 throat, 包皮 foreskin, 尾巴 tail, 頸腱 cecum, 稻子 paddy, 肉 meat, 麥 wheat | 助 help |

Table 8: Only patterns of noun-verb relation

7.2 Modifying relation

Different features can be found from their modifier and collocated modifies listed in Table 7. The features of the two words are as shown below. 

**Process involved in the action.** Two collocates appear in the modifiers relation for *qie 切*, which are *yiqi 一起* *together* and *yitong 一同* *together* implying that the process can be done by more than one person at the same time, while *ge 割* does not have such words in the modifiers relation. On the other hand, *qiangxing 強行* *forcibly* appear in the modifier relation whereas *qie 切* does not have such words in the modifiers relation. It implies that *qie 切* is a more well accepted action, while people are forced to do the action when *ge 割* is used. Moreover, *ge 割* has a collocated modifier *yizhi 一直* *continue* which implies that *ge 割* is a continuing or repeating action, while *qie 切* does not have such words. It implies that it is an action completed at once.

**Purpose of the action.** Similar to the previous section, it is found that *qie 切* should be a result-oriented verb while *ge 割* is a purpose-oriented verb. *Qu 去* *for* appears in the modifier relation whereas *qie 切* does not have such words in the modifiers relation. It infers that people do the action for a particular goal when *ge 割* is used.

For modifies, while *gesheng 歌聲* *song, xin 信 letter, banfa 辦法* *method* and *qiuyuan 球員* *player* are present in the modifies relation, *ge 割* does not have this relation. This result is consistent with the findings in Section 4.

| Modifier | Modifies |
|----------|----------|
| 一起 together, 一同 together, 愈 more, 更 more, 共 sum, 各 separate, 並 also, 難 no | 歌聲 song, 信 letter, 辦法 method, 球員 player |

Table 9: Only patterns of modifying relation

8 Cross-strait Comparison of *ge 割* and *qie 切* in CNA and XIN

Due to the frequent communication across straits, cross-strait comparison is worth discussing, and (Hong and Huang 2008; Hong and Huang, 2007) have already done some related studies. The use of vocabularies always depends on the context of texts, and the context may differ because of different culture, history, living habit and customs across the straits. To examine the actual use of
vocabularies, we can make use of Word Sketch function. As mentioned in Section 3, Chinese
Gigaword corpus data is composed of three sub-
corpora. Now, we would like to make use of two of
them. They are corpora coming from Central News
Agency in Taiwan and Xinhua News Agency in
Mainland. First, we input ge 割 and qie 切 by using
CNA and set the minimum frequency at 5. Then,
we click Show Word Sketch. After that, same
procedures are done again using XIN. The findings
are then generated as shown in Table 9 and 10.

It is noticed that ge 割 and qie 切 differ mainly
in collocations of the noun-verb relation. Therefore,
only noun-verb relation is discussed in this section.
This section focuses on qie 切 first. It is found that
many subjects and objects of qie 切 in Mainland
are transmigration; therefore, they will not be
discussed in this paper. On the contrary, there is an
intriguing finding which deserves more detailed
analysis and explanation. Shengyupian 生魚片
sashimi is a dish, hence ge 割 should be preferred
as this action is purpose-oriented. Yet, qie 切
includes shengyupian 生魚片 sashimi acted as
object. It is believed that it may be because part of
the noun (i.e. pian 片 slice) seems to require qie 切.
Qiepian 切片 is acceptable, while gepian 割片 is
unacceptable. There is a similar usage such as
qiezhangyupian 切章魚片 sashimi. Pian 片 slice
also requires qie 切 in this case

| Noun-verb relation | Mainland (XIN) | Taiwan (CNA) |
|---------------------|---------------|--------------|
| Subject: 費斯 | Object: 梅爾金 | Object: 蛋糕 cake, 菜 vegetable, 塊 piece, 禮礼 cake birthday cake, 麵店 noodle shop, 生魚片 sashimi, |
| Fischella |             | Subject: 她 she, 他 he, 人 person |

Table 10: Comparison of qie between Mainland (XIN) and Taiwan (CNA)

As for ge 割, objects of it in Mainland are more
diverse and much richer than Taiwan (see Table
10). The additional collocations in Mainland acted
as objects are crops and classifier for fields (i.e. mu 穬 classifier for fields, maizi 麥子 wheat, daozi 稻子 paddy and jiucui 韭菜 leek), place (i.e. Taiwan 台灣 ), ideology (i.e. weiba 尾巴 ideology) and love (i.e. qinqing 親情 family love and qing 情 love). However, ge 割 includes only tissues or organs acted as objects: hou 喉 throat, wan 腕 wrist, baopi 包皮 foreskin, mangchang 盲腸 cecum, shetou 舌頭 tongue in Taiwan. Shuangyanpi 雙眼皮 double eyelid is the only collocation shared by both of them.

For subjects, liandao 雙刀 sickle is the most frequent subject in Mainland, whereas it is meigongdao 美工刀 utility knife in Taiwan. Liandao 雙刀 sickle is the tool for harvesting corps, but meigongdao 美工刀 utility knife are tools mostly used for crafts. It reflects that love, ideology, One-China principle, harvest and farming are still very important to Mainland. On the other hand, Taiwan people focus more on arts development and they are more open-minded. They are more willing to accept doing surgery. The differences can be attributed to the different political history, developing time and political stands of Mainland and Taiwan. Mainland is still more traditional, while Taiwan is more commercial.

Also, it is interesting to note that a new feature of ge 割 is found when place (i.e. Taiwan 台灣 ), ideology (i.e. weiba 尾巴 ideology) and love (i.e. qinqing 親情 family love and qing 情 love) acted as subjects in Mainland. Clearly, ge 割 is used for a particular purpose. More important, it is observed that the separation is not done by knife when these collocations appear. It simply implies that a part is separated from one entity. This result echoes to the prediction we make in section 3.

| Noun-verb relation | Mainland (XIN) | Taiwan (CNA) |
|---------------------|---------------|--------------|
| Subject: 費斯 | Object: 麥子 wheat, 尾巴 tail, 雙眼皮 double eyelid, 肉 meat, 稻子 paddy, qie 切, 菜菜 leek, 穩 classifier for fields, 情 love, 台灣 Taiwan |
| Fischella |             | Subject: 美工刀 utility knife |

Table 21: Comparison of ge between Mainland (XIN) and Taiwan (CNA)
9 Conclusion

Our study shows that CWS is a powerful tool which can help discriminate the pair of synonyms ge 割 and qie 切. Huge amount of authentic linguistics data are generated by using various functions of SKE in an efficient way. Because of the rich data, detailed analysis and cross-strait comparison can be demonstrated in this paper. To conclude, it is predicted that ge 割 and qie 切 are verbs of separation implied by Lian (2005). Several differences are demonstrated in noun-verb relation and modifying relation. The subtle differences can show that the two target words differ mainly in terms of the purpose of separation. Ge 割 is a purpose-oriented verb, while qie 切 is a result-oriented verb. For similarities, both can be used when a wound is made by someone. These findings are far better than Lian (2005)’s work which focuses on dictionary study. Furthermore, the cross-strait comparison can help understand how Mainland and Taiwan differ from each other. It is believed the meaningful results obtained will facilitate cross-strait communication. Obviously, traditional dictionary is no longer enough to Chinese language learners. This study is expected to provide some insights for Chinese dictionary editors, and hence Chinese teachers. One possible future work direction is to figure out ways to improve the accuracy of results generated by CWS so as to provide more reliable sources for analysis and show how these findings are organized in a modern dictionary. The other possible way is to compare the results generated by CWS and SKE obtained in this study with the interesting results achieved with the semantic decomposition approach stated by Gao (2001).

References

Fillmore, C. J. (1992). “Corpus Linguistics” or “Computer-aided armchair linguistics”. In J. Svartvik (Ed), Directions in corpus linguistics. (pp. 35-60). Berlin, Germany: Walter de Gruyter.

Fillmore, C. J., & Atkins, B. T. (1992). Toward a frame-based lexicon: The semantics of RISK and its neighbors. In A. Lehrer & E. F. Kittay (Eds), Frames, fields, and contrasts: New essays in semantic and lexical organization, (pp. 75-102). New York, NY: Routledge.

Gao, H. (2001). The physical foundation of the patterning of physical action verbs. (Doctoral dissertation, Lund University, Sweden). Retrieved from https://lup.lub.lu.se/search/publication/693c9995-4857-4a77-b972-8852a3c93be3

Hong, J. F., & Huang, C. R. (2008). A corpus-based approach to the discovery of cross-strait lexical contrasts. Language and Linguistics, 9(2), 221-238.

Hong, J. F., Huang, C. R., & Xu, M. (2007). A study of lexical differences between China and Taiwan based on the Chinese gigaword corpus. Paper presented at 19th Conference on Computational Linguistics and Speech Processing, ROCLING 2007, Taipei, Taiwan.

Huang, C. R., & Hong, J. F. (2005). Deriving conceptual structures from sense: A study of near synonymous sensation verbs. Journal of Chinese Language and Computing, 15(3), 125-136.

Kilgarriff, A., Huang, C.-R., Rychly, P., Smith, S., & Tugwell, D. (2005). Chinese Word Sketches. ASIALEX 2005: Words in Asian Cultural Context. June 1-3. Singapore.

Li, L., Huang, C. R., & Gao, X. (2018, May). A SkE-assisted comparison of three “prestige” near synonyms in Chinese. In J. F. Hong, Q. Su & J.S. Wu (Eds.), Chinese Lexical Semantics. (pp. 256-266). Chiaiyi, Taiwan: Springer, Cham.

Lian, X. (2005). To understand the semantics fuzziness of verbs from “qie, ge, jie”. China Academic Journal, 19(3), 62-64.

Wang, X., & Huang, C. R. (2018, May). From Near Synonyms to Power Relation Variations in Communication: A Cross-Strait Comparison of “Guli” and “Mianli”. Chinese Lexical Semantics (pp. 155-166). Chiaiyi, Taiwan: Springer, Cham.