
A Corpus-based Lexical Semantic Study of Mandarin Verbs of “Tui” and “La”

Han SUN

Faculties of Humanities / Room GH801, 8/F,
Wing GH, The Hong Kong Polytechnic
University, Hung Hom, Kowloon, Hong
Kong

20092425g@connect.polyu.hk

Chu-Ren HUANG

Faculties of Humanities / Room GH801, 8/F,
Wing GH, The Hong Kong Polytechnic
University, Hung Hom, Kowloon, Hong
Kong

churen.huang@polyu.edu.hk

Abstract

This paper investigates the differences between Tui and La under the framework of MARVS. Using Gigaword2 Corpus and Sinica Corpus as corpora and Sketch Engine as the query tool, the study is conducted following three procedures. First, a preliminary observation on morphosyntactic differences in these two verbs is obtained by searching the explanation of the two terms Tui and La in the online Chinese English Dictionary of Modern Usage. Then, two data corpora are used in searching for statistical evidence, after which the results will be represented in the framework of MARVS. In conclusion, Tui is inherently a single occurrence that cannot be measured by time duration and can be identified with a temporal point, whereas La is an activity that has a time course that can be measured based on duration and has a clear boundary as the starting point. Also, distinctions in role module lie in that La indicates abstract process without clear direction or an endpoint while Tui is more concrete with one direction and an endpoint.

1 Introduction

Verbs provides the key to study the nature of lexical knowledge and sentence process (Liu et al.,

2000). It is believed that the “syntactic behaviors of verbs are semantically determined”(Chief et al., 2000, p. 57), so it is an effective way to explore semantic differences between verbs by comparing the syntactic differences of synonym verbs (Liu et al., 2000). And since a more semantically constrained system is needed for natural language processing purpose, Module-Attribute Representation of Verbal Semantics (MARVS) theory has been shown as a new, powerful approach to represent semantic representations of near synonyms (Liu et al., 2000). In the previous literature, Liu et al. (2000) explored semantic differences of pivotal near-synonymous members of “tou”, “zhi”, “diu”, and “reng” by using two kinds of endpoint and the observed distinctions are then incorporated into a representational paradigm MARVS. Wang and Huang (2018) examined the stance differences of “guli” and “mianli” in terms of implication of interpersonal relation variations and accounted for the results the in MARVS. However, little studies have been conducted on the near synonym “tui” and “la”. Thus, the focus of this study is to establish semantic-syntactic interdependences by observing the morphosyntactic behaviors of the verbs “tui” and “la” displayed in a large corpus and representing the results in the framework of MARVS.

2 Research questions

This paper investigates the following two questions:
(1). What are the differences between Tui and La?

(2). How are these differences represented in the framework of MARVS?.

3 Theoretical Framework

The Module-Attribute Representation of Verbal Semantics (MARVS) theory (Huang et al., 2000; Huang & Hsieh, 2015) is taken as the theoretical framework. It is a practical approach to interpret the grammatical behaviors of verbs. Under this framework, the semantic representation of a verb has two modules, namely, event modules and role modules, both bearing its internal attributes. Event Module is the aspectual compositions of the event such as Boundary [·], Punctuality [/], Process [/////], State [___] and stage [^^^^]. Inherent (Event-internal) Attributes refer to the semantics of the event itself such as Control, Change-of-state. Role Module refers to focused roles of the event such as Agent, Theme, Instrument, Manner, Goal and typically include all required arguments but can also include optional arguments and adjuncts. Role-internal Attributes refer to the internal semantics of a particular focused role such as Factive and Generic. All conditions being equal, a higher-level module or attribute is preferred for the sake of generality and greater explanatory power. There are five “atomic event structures” (Huang et al., 2000, p. 26):

(1) . Boundary (including a Complete Event)

Boundary is an event module that can be identified with a temporal point and that must be regarded as a whole.

(2) / Punctuality

Punctuality is an event module that represents a single occurrence of an activity that cannot be measured by duration.

(3) ///// Process

Process is an event module that represents an activity that has a time course.

(4) _____ State

State is a homogeneous event module in which the concept of temporal duration is irrelevant.

(5) ^^^^^ Stage

Stage is an event module consisting of iterative sub events. (Huang et al., 2000, p. 26)

4 Research Methodology

For the Methodology, Gigaword2 (Huang, 2009) Corpus and Sinica Corpus are used as corpora for

the study, and Sketch Engine (Huang et al., 2005) is used as the query tool. And generally speaking, the study is conducted following three procedures. First, a preliminary observation on morphosyntactic differences in these two verbs is obtained by searching the explanation of the two terms *tui* and *la* in online Chinese English Dictionary of Modern Usage. Then, two data corpora are used in searching for statistical evidence., after which these found results will be represented the in the framework of MARVS.

5 Analysis and Findings

5.1 Brief overview of “Tui” and “La”

First, by using the online dictionary resources, a brief overview of “tui” and “la” was obtained concerning their definitions and typical usages. According to the online Contemporary Chinese Dictionary, “tui” means “to push, push open, push down push on”, as shown in the example “推門 tuimen” “push door open”; “to reject, make excuse” as shown in the example “推病不去 tuibingbuqu” (did not go on pretext of illness); “Push forward, promote, elect, recommend” as shown in the example “推為第一 tuiweidiyi” (regard as No. 1); “to shake off, shirk, refuse” as shown in the examples “推到別人身上 tuidaobeirensbenshang” (put the blame on others) and “推卸責任 tuixiezeren” (shake off all responsibility); and “to trace origins, investigate, deduce, calculate” as shown examples “推究 (原因) tuijiu(yuanyin)” (investigate, study causes). At the same time, “拉 la” has several meanings. The first meaning of “拉 la” is “to pull, drag, draw” as shown in the examples “拉拉扯扯 lalacheche” (pull and drag this way and that), and “拉長方臉兒 lachangfangxinglianer” (pull a long face). The second meaning is “to hold” as shown in the example “手拉手 shoulashou” (hold one another's hand). The third meaning is “to solicit” as shown in the examples “拉生意 lashenyi” (solicit business) and “拉票 lapiao” (solicit votes). The fourth meaning is “to recruit, impress, look for” as shown in the examples “拉角 lajue” (recruit actors for play). The fifth meaning is “to cut, amputate, eliminate, move bowels” as shown in the examples “摧枯拉朽 cuikulaxiu” (like tearing down dried-up

trees or buildings in decay) and “拉肚子 laduzi” (have loose bowels).

Then, we continued to find the overall distribution of “Tui” and “La” in Gigaword2 and Sinica as shown in the following table. It is shown that the word frequency of “tui” in gigaword2 is far more than that of “la”, whereas the word frequency gap is not obvious in sinica. It seems that “tui” is more commonly used than “la” according to the above overall result.

Word	Freq. in gigaword2	Freq. in sinica
推(tui)	13501	316
拉(la)	4215	538

Table 1. Overall distribution of the word “tui” and “la” in gigaword2 and sinica

5.2 Distinction in Event Module

The observation concerns that the event module of “tui” is more likely to be punctuality, an event module that represents a single occurrence of an activity that cannot be measured based on duration (Huang et al., 2000). To be more specific, it is a complete punctuality, one of the bounded event representations having one atomic event and must be bounded at least one end (Huang et al., 2000). As shown in the examples in (1), the action following “tui” such as ‘to push the door’ and ‘to postpone’ refers to a single occurrence in that we cannot ‘push the door’ or ‘postpone’ for many rounds in one time and these actions are difficult to be measured in terms of temporal duration. However, we can observe the boundary in these actions. For example, the door remains stable at a certain place after being pushed and we can postpone an activity to an exact time.

Meanwhile, the event module of “La” is more likely to be process, an event module that represents an activity that has a time course and can be measured in terms of its temporal duration. To be more specific, it is an inchoative process, bounded in the beginning side. As shown in the examples in (1), the action following “la” such as ‘to solicit business’ and ‘to ask voters for support’ usually lasts for a certain period of time and is bounded at the beginning side by using ‘start’ to do it.

(1) TUI as Punctuality

- a. tui-men 推门 ‘to push the door’
tui-chi 推迟 ‘to postpone’

LA as Process

- b. la-shengyi 拉生意 ‘to solicit business’
la-piao 拉票 ‘to ask voters for support’

Based on the preliminary observation, we moved on to find the statistical evidence in corpus to support the previous assumptions. And the statistical evidence can be classified into the following aspects including concordances differences in gigaword2, contrast in terms of PoS in sinica corpus, and differences in right context for the two words in sinica corpus. First, after running CQL query in Sketch Engine, we got the concordances for “tui” and “la” collocated with “buduan” (continuously) and/or “yizhi” (constantly) in gigaword2. We also find that the event structure of “la” belongs to process, as it refers to a kind of activity that can be modified by adverbs signaling temporal duration, such as “buduan” (0.23 per thousand in gigaword2), and/or “yizhi” (1.42 per thousand in gigaword2). However, for “tui”, it is similar to “la” in that it is inchoative. But it tends to be a punctuality event, as the punctuality actions “tuichi 推迟”, “tuiyan 推延”, “tuituo 推脱”, and “tuixie 推卸”, which represent only single occurrence take up 77.1% of the total amount of usage of “tui”. It worth noting that although “tuichi 推迟” and “tuiyan 推延” can be modified by the adverb “yizhi” and “buduan”, it does not mean the action of “postpone” last for a period of time. It only means that the action happens for several times, so these examples cannot be taken into the account of the process module.

	Tuic hi 推 迟	Tuiy an 推 延	Tuitu o 推 脱	Tuix ie 推 卸	Total amount/Perce ntage
Fre q.	8046	99	108	2165	10409 / 77.1%

Table 2. Frequency of “tui” as punctuality in gigaword2

Second, from the following results, one main difference can be found that only 拉 la can be used as VH(Stative intransitive verbs) as can be seen from the following example. The darkness is

conceptualized as a curtain which can be pulled down. Personification is employed in this sentence as 拉 la is used as stative intransitive verbs serving the function of main predicative verb. It gives people an imaginary scene where the sky, like a person, pulls himself down and leads the world into darkness. It also indicates that “la” is an inchoative process.

Fig. 1: 推 tui in Sinica Corpus: PoS

Nv	1
VC	315

Fig. 2: 拉 la in Sinica Corpus: PoS

Nv	2
VB	5
VC	530
VH	1

Fig. 3: The sentence example of 拉 la1 used as VH

因為(Cbb) 是(SHI) 冬天(Nd), 夜幕(Na) 早早(D) 就(D)拉(VH) vrv 了(Di) 下來(VH)vrr。

Third, from the following results, we can see that the most frequent common noun follows “tui” in the right context is door(men) and the most frequent common nouns follow “la” in the right context is base(pi), certain two-stringed bowed instrument(huqin), violin(xiaotiqin), alarm clock(jingbao) and stringed instrument(qin). In this respect, “tui” is used to refer a concrete actions which usually occur for only one time, whereas 拉 la1 is used to refer to a more abstract and general actions which usually occur for more than one time. As indicated in the collocation 拉小提琴(laxiaotiqin), people won’t pull the violin, the concrete wood instrument with string. Instead, we use the bow to push and pull on the string and this action occur many times and continues for a period of time.

In addition, the collocation 拉坏(lapi) is used as a general term referring to one step of the potting process. When we take a verb to turn it into

a noun, we usually want to conceptually refer to something that is constant and stable. Hence VH (stative intransitive verbs) are more easily conceptualized as something constant, coincided with the previous finding in PoS. Thus, “la” is a kind of inchoative process because the duration can be characterized by time duration, and the beginning point of the process is clear while the ending point is not. For the internal attributes of the event, both events can be controlled and are used with “bie” as shown in the example “bie tuimeng” and “bie laqin”.

Fig. 4: Immediate neighbor words of 推 tui1 in the left context

MI值	freq(y)	freq(x,y)	y:詞/詞類
6.610	1197	25	門(Na)
4.534	2290	6	往(P)
4.245	3561	7	至(P)
4.187	2698	5	前(Nes)
3.981	22770	35	著(Di)
3.783	10473	13	得(DE)
2.630	62757	26	了(Di)
1.641	34284	5	而(Cbb)
1.346	54155	6	他(Nh)
1.249	69848	7	不(D)
0.471	105379	5	—(Neu)
3.192	923326	29	, (COMMATEGORY)
-1.725	561110	5	的(DE)

Fig. 5: Immediate neighbor words of 拉 la1 in the left context

MI值	freq(y)	freq(x,y)	y:詞/詞類
9.031	15	6	坏(Na)
8.495	47	11	胡琴(Na)
7.868	128	16	小提琴(Na)
7.383	65	5	警報(Na)
7.320	83	6	撒(VC)
6.659	134	5	琴(Na)
4.829	1003	6	白(VH)
4.621	22770	113	著(Di)
4.430	1494	6	愈(D)
4.241	10473	35	得(DE)
3.376	3561	5	至(P)
2.593	7786	5	向(P)
2.337	62757	33	了(Di)

From the above discussion, we may conclude with a tentative hypothesis that “tui” and “la” differ in their lexical specification of an event module and its internal attributes; that is “tui” is inherently a single occurrence which cannot

measured by time duration and can be identified with a temporal point, but “la” is an activity that has time course which can be measured based on duration and has a clear boundary as the starting point. Both “tui” and “la” the controlled and collocated with “bie”. The results can be shown in the table below.

Module/Attribute	TUI	LA
Event Module	Completive Punctuality • /	Inchoative Process • / / / /
Inherent Attributes	+ Control	+ Control

Table 2. Major Distinction between “tui” and “la” in Event Module

5.3 Distinction in Role Module

According to the comparison of the “sketch differences” results of “tui” and “la” in gigaword2 in Sketch Engine, only “tui” can be collocated with “jiang 将” and according to WordSketch of “把...推 ba...tui” shown in figure 9, “一干二净 yiganerjin” have the highest MI(18.291). It means to “shake off all the responsibility”. “婚期 hunqi” also have a high MI(16.565). It means to postpone and to put off the data. And “把...推 ba...tui” can be collocated to “至 zhi” and “到 dao” which indicates an endpoint. These usages of “推 tui” are very special which only appear in the wordsketch of “推 tui”, not “拉 la”.

Moreover, comparing the differences in object collocation, the theme of “la” is more concrete as can be shown in the examples “拉布条 labutiao” (MI: 60.8) and “拉警报 lajingbao” (MI: 33.1) whereas, the theme of “tui” is more abstract as can be shown in the examples “推责任 tuizeren” (MI: 18.1). As for the subject collected with “tui” and “la”, the only pattern of “la” is “差距 chaju” (MI: 19.3) while the only pattern of “tui” is “浪 lang” (MI: 46.8). In these process, “chajulakai 差距拉开” refers to abstract process without clear direction or an endpoint while “houlangtuiqianlang 后浪推前浪” is more concrete with one direction and an endpoint.

Fig.6 A comparison of the “sketch differences” results of “tui” and “la” in gigaword2 in Sketch Engine

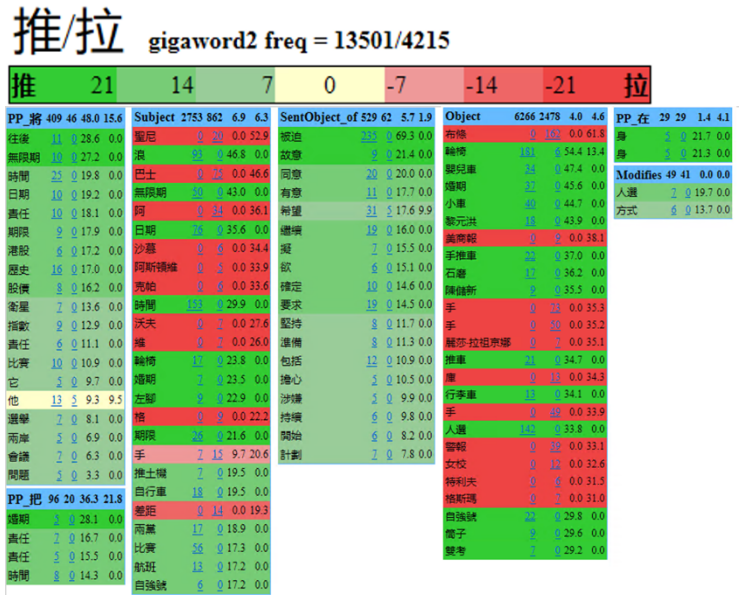


Fig.7 The “only patterns” of “tui” in gigaword2 in Sketch Engine



Fig.8 The “only patterns” of “la” in gigaword2 in Sketch Engine



Fig.9 Wordsketch of “把…推 ba…tui” in gigaword2 in Sketch Engine
Collocation candidates

PP 把	96	36.3		Freq	T-score	MI
婚期	5	28.09	p/n	一乾二淨	8	2.828 18.291
責任	7	16.71	p/n	婚期	5	2.236 16.565
責任	5	15.48	p/n	高到	3	1.732 15.518
時間	8	14.31	p/n	遲到	18	4.243 15.310
			p/n	下車	3	1.732 11.951
			p/n	遲	6	2.449 11.768

6 Discussion and conclusion

The distinctions between “tui” and “la” can be summarized and represented within the proposed MARVS framework as shown below.

Module/Attribute	TUI	LA
Event Module	Completive Punctuality • /	Inchoative Process • / / / /
Inherent Attributes	+ Control	+ Control
Role Module	+ endpoint	-endpoint
Role-Internal Attributes		

Table 3: MARVS Representation of the Semantic Differences among “tui” and “la”

In conclusion, the model of event-structure information help to identify and represent the crucial semantic factors that are syntactically relevant. In terms of theoretical and methodological aspect, it can be concluded that corpus data and computation may reveal some important generalization that might not be available from elicited data only. The clustering of morpho-syntactic patterns with lexical-semantic characteristics proves to be fruitful in differentiating near-synonyms and in systematically disentangling the complex interaction between syntax and semantics.

References

Chief, L.-C., Huang, C.-R., Chen, K.-J., Tsai, M.-C., & Chang, L.-L. (2000). What can near synonyms tell us. *International Journal of Computational Linguistics & Chinese Language Processing*, Volume 5, Number 1,

February 2000: Special Issue on Chinese Verbal Semantics.

Huang, C.-R. (2009). Tagged chinese gigaword version 2.0, ldc2009t14. Linguistic Data Consortium.

Chief, L.-C., Huang, C.-R., Chen, K.-J., Tsai, M.-C., & Chang, L.-L. (2000). What can near synonyms tell us. *International Journal of Computational Linguistics & Chinese Language Processing*, Volume 5, Number 1, February 2000: Special Issue on Chinese Verbal Semantics,

Huang, C.-R. (2009). Tagged chinese gigaword version 2.0, ldc2009t14. Linguistic Data Consortium.

Huang, C.-R., Ahrens, K., Chang, L.-L., Chen, K.-J., Liu, M.-C., & Tsai, M.-C. (2000). The module-attribute representation of verbal semantics: From semantic to argument structure. *International Journal of Computational Linguistics & Chinese Language Processing*, Volume 5, Number 1, February 2000: Special Issue on Chinese Verbal Semantics,

Huang, C.-R., & Hsieh, S.-K. (2015). Chinese lexical semantics. In *The Oxford handbook of Chinese linguistics* (pp. 290-305). Oxford Univ. Press.

Huang, C.-R., Kilgarriff, A., Wu, Y., Chiu, C.-M., Smith, S., Rychlý, P., Bai, M.-H., & Chen, K.-J. (2005). Chinese Sketch Engine and the extraction of grammatical collocations. *Proceedings of the fourth SIGHAN workshop on Chinese language processing*.

Liu, M.-c., Huang, C.-R., Lee, C. C., & Lee, C.-Y. (2000). When endpoint meets endpoint: A corpus-based lexical semantic study of Mandarin verbs of throwing. *International Journal of Computational Linguistics & Chinese Language Processing*, Volume 5, Number 1, February 2000: Special Issue on Chinese Verbal Semantics.

Wang, X., & Huang, C.-R. (2018). From near synonyms to power relation variations in communication: A cross-strait comparison of “guli” and “mianli”. *Workshop on Chinese Lexical Semantics*.