

# USAGE EFFECTS ON THE COGNITIVE ROUTINIZATION OF CHINESE RESULTATIVE VERBS

*BEN PIN-YUN WANG*

The Pennsylvania State University  
benpywang@gmail.com

## **Abstract**

The present study adopts a corpus-oriented usage-based approach to the grammar of Chinese resultative verbs. Zooming in on a specific class of *V-kai* constructions, this paper aims to elucidate the effect of frequency in actual usage events on shaping the linguistic representations of resultative verbs. Specifically, it will be argued that while high token frequency results in more lexicalized *V-kai* complex verbs, high type frequency gives rise to more schematized *V-kai* constructions. The routinized patterns pertinent to *V-kai* resultative verbs varying in their extent of specificity and generality accordingly serve as a representative illustration of the continuum between lexicon and grammar that characterizes a usage-based conception of language.

## **1. Introduction**

The specific linguistic structure in Chinese under the investigation of the present study has been most commonly referred to in the literature as “resultative verb compounds”, which can be defined as a two-element verb compound in which “the second element signals some *result* of the action or process conveyed by the first element” (Li & Thompson 1981:54, italics original). In his book, Shi (2002) also adopts this definition and further points out that it features the “action-result” relationship between the two constituents of resultative verb compounds, a distinctive feature that distinguishes this type of V-V complex verbs from other serial verb constructions in Chinese.

Still, some researchers cast doubts on the widely-accepted assumption that such V-V structures, which are words whose part of speech is verb, with an “action-result” relationship held between  $V_1$  and  $V_2$  are the output of a compounding process. While adopting the term of “resultative verb compound” in his study, Shi (2002) still deems the naming of “compound” misleading in that the term suggests that this subtype of V-V constructions involve non-productive lexical process. Instead, he argues in favor of a syntactic treatment of resultative verbs since it allows many types of collocation. Also, Starosta et al. (1997) treat  $V_2$ , or the postverb, as the suffix of  $V_1$ , an affixation rather than a compounding approach, which they believe can better capture the productivity as well as the regularity of the sense of the postverb. It can thus be concluded that the issue of whether resultative verbs involve lexical or syntactic operations remains unsettled.

To avoid the foregoing controversy over the morpho-syntactic status of the postverb and the gestalt complex verb as a whole, the cover label “resultative verb” (Packard 2000) is adopted throughout this paper. The term “resultative” is employed in this study in its broadest sense, which can refer to “the state, degree, accomplishment, achievement, or effect of the action” (Shi 2002:29). Such a general term is intended to characterize the wide range of meanings encoded in resultative verb constructions in Chinese.

Via a bottom-up approach, the present study empirically delves into the usage patterns of the specific class of *V-kai* resultative verbs from a usage-based theoretical point of view. Specifically, this paper aims to elucidate the roles of frequency in structuring the linguistic representations of *V-kai* resultative verbs. Besides, to methodologically enhance the usage-basedness of this study, collocational and frequency analyses have been applied to large corpus data. It is hoped that, via an exhaustive scrutinization of the empirical data of *V-kai* complex verbs, this study can draw an accurate picture of how resultative verbs are conventionally and creatively used in naturally-occurring language.

The corpus from which the data of this study are drawn is Academic Sinica Balanced Corpus of Mandarin Chinese (Sinica Corpus) (Version 4.0), which contains nearly five million Chinese words.<sup>1</sup> A total of 2747 tokens of *V-kai* resultative verbs were retrieved from the corpus, which consist of combinations of 138 distinct verbs with the postverb *-kai*.

## **2. A usage-based conception of language**

This section first reviews two characteristic theoretical assumptions shared by varied usage-based theories, mainly based on Croft and Cruse (2004), Diessel (2004), Kemmer and Barlow (2000), Mukherjee (2005) and Tummers et al. (2005). Moreover, the methodological applications of a usage-based model, specifically focused on the corpus-oriented approach, are also addressed.

### **2.1 Theoretical assumptions of a usage-based approach**

This section expounds the essential principles underlying the theoretical aspects of a usage-based view of language. In particular, two properties germane to the establishment of linguistic representations in the grammar of a language user will be highlighted: the meanings of linguistic expressions in actual language use as well as the frequency of occurrences of specific linguistic expressions (c.f. Croft & Cruse 2004).

---

<sup>1</sup> The corpus can be accessed online at: <http://dbo.sinica.edu.tw/SinicaCorpus/>.

### 2.1.1 An utterance-grounded model

The foremost characteristic of a usage-based approach to language consists in the precedence of language use. Basically, usage-based theories posit that the language system, or grammar, of a speaker is fundamentally grounded in utterances or usage events (Kemmer & Barlow 2000). In general, an utterance or a usage event refers to “a situated instance of language use which is culturally and contextually embedded and represents an instance of linguistic behavior on the part of a language user” (Evans & Green 2006:110). A speaker’s internal linguistic system is thus directly and intimately connected to the actual use of language in that “the grammar does not only constitute the knowledge repository to be employed in language use, but it is also itself the product of language use” (Tummers et al. 2005:228). In other words, grammar has an experiential basis: it is profoundly tied to one’s experience with language (Bybee 2006).

In view of their grounding relation with respect to grammar, usage events assume dual roles: they both stem from and (re)define the more abstract system of linguistic knowledge (Kemmer & Barlow 2000; Tummers et al. 2005). Firstly, usage events are actual instantiation or elaboration of the more general and less detailed symbolic representations they pertain to. In that case, usage events are necessarily more specific and complex in nature (Langacker 2008). Hence, analyzing usage data can illuminate the mental structure of a speaker’s grammar.

Secondly, more importantly, usage events can also be the empirical foundation of the language system. The usage-based theories of language share the assumption that “language structure emerges from language use” (Tomasello 2003:5), presuming that usage data can serve as input that constantly structures linguistic representations. In other words, “usage events drive the formation and operation of the internal linguistic system” (Kemmer & Barlow 2000:xi). Therefore, the grammar of a language user is far from fixed but dynamic in that linguistic units are “subject to creative extension and reshaping with use” (*ibid.*, p.ix).

Such an emergent view of linguistic structures common to usage-based theories of language acknowledges the interplay between language use and other cognitive processes such as schematization and entrenchment (Langacker 1987, 2000, 2008). The process of schematization, which involves generalization of patterns and abstraction of commonalities across all sorts of usage events, gives rise to schemas. A schema is therefore bound to be less detailed and precise than the specific structure that elaborates or instantiates it (Evans & Green 2006; Langacker 2000).

In this sense, in stark contrast to the generative models (e.g. Pustejovsky 1998), the usage-based model does not intend to draw a clear-cut boundary between lexicon and grammar (Diessel 2004; Langacker 2008; Mukherjee 2005). Instead, grammar is viewed as a continuum ranging from wholly idiosyncratic, specific lexical entries to maximally general grammatical assemblies. In other words, the usage-based approach treats both lexicon and grammar as symbolic units that vary largely in their extent of specificity or level of abstraction (Langacker 1987, 2000, 2008). All these conventionalized schemas or patterns with various degrees of symbolic complexity are conceptually organized into activation networks in the speaker’s mental grammar (Diessel 2004; Kemmer & Barlow

2000; Langacker 2008). In this ongoing process of establishing the network-like generalizations, the crucial role that usage events play is to link between the general schema and the specific instances.

However, for a schema to achieve a status of conventional unit easily accessed as an integrated whole, it is required to undergo sufficient entrenchment, a psychological phenomenon defined as “the cognitive routinization of linguistic units and structures on grounds of repetitive events in language use” (Mukherjee 2005:222). That is, linguistic units “emerge via the progressive entrenchment of configurations that recur in a sufficiently number of events to be established as cognitive routines” (Langacker 2008:220). It is therefore reasonable to hypothesize that the degree of entrenchment of a linguistic unit is positively correlated with the frequency of its actual occurrence in usage events.

### **2.1.2 A frequency-focused model**

Another principal feature of a usage-based conception of language is the crucial role of frequency in shaping the language system of a speaker. Frequency in actual language use is deemed as the quantitative assessment of the degree of entrenchment (Mukherjee 2005; Tummars et al. 2005). Usage-based theories of language generally distinguish two types of frequencies: token frequency and type frequency (Bybee 1985, 2007; Croft & Cruse 2004; Diessel 2004; Tummars et al. 2005). These two types of frequency have been identified to exert distinct effects on the storage, activation, and processing of linguistic units (Diessel 2004). It has been postulated that, while token frequency leads to the routinization of specific instances, type frequency gives rise to the entrenchment of general schemas (Evans & Green 2006; Rostila 2006).

Token frequency refers to the number of occurrences of a specific linguistic unit (such as a particular sound or word) in usage data. Linguistic expressions with higher token frequencies are discovered to be more likely to undergo sound change (i.e. The Reduction Effect) but to resist analogical change (i.e. The Conserving Effect) (Bybee 2006, 2007, 2010; Bybee & Hopper 2001). The two frequency effects can be attributed to the distinct psychological status of these linguistic units on account of repetition. Due to repetition, their memory representations are strengthened, enabling independent storage of the units in mental grammar and facilitating direct access of these units without activating a higher-level schema (Bybee 2007; Diessel 2004; Tummars et al 2005). In other words, such units enjoy a higher degree of autonomy in the language system of a language user.

In contrast, type frequency, peculiar to abstract patterns or schemas, is concerned with “the number of distinct items that can occur in the open slot of a construction or the number of items that exemplify a pattern” (Bybee 2007:14). Type frequency is generally envisaged as the primary indicator of the productivity of a grammatical pattern (Croft & Cruse 2004; Diessel 2004; Tummars et al. 2005). In the usage-based approach to language, the notion of *productivity* refers to “the likelihood of being selected as the active structure used to categorize a novel expression (Langacker 2000:26). The higher

the number of possible linguistic expressions that instantiate a schema, the more entrenched the abstract pattern is in mental grammar, which in turn enhances its productivity as a template for novel uses. The schema would then have a stronger representation in the language system, rendering it more available or accessible for constructing and interpreting new expressions (Bybee 2007).

Moreover, the interaction between the effects of token and type frequency on the cognitive routinization of linguistic units also demands our attention. Technically, the entrenchment of individual types is anchored in their token numbers; therefore, token frequency should be indirectly related to the entrenchment of a schema. Nevertheless, due to the high autonomous status of frequently occurring linguistic units in language representations, such units may not contribute to the productivity of general patterns or schemas associated with them (Bybee 2007; Diessel 2004). In summary, the cognitive routinization of an abstract pattern or schema is contingent on both the number of types instantiating a schema and the number of tokens filling the pattern.

## 2.2 Towards a corpus-oriented usage-based approach

In terms of the methodological aspects of usage-based theories of language, the priority given to usage data and frequency count calls for the accessibility of a large collection of samples of language use for comprehensive quantitative and qualitative analyses. The availability of large corpora and concordance software is thus regarded as a methodological breakthrough for linguistic theories that focus on language use (Bybee 2007). In general, corpus-based studies empirically analyze the actual patterns of use by utilizing balanced or representative natural texts and employing automatic techniques that enable a systematic and exhaustive analysis of usage patterns across varied frequencies (Biber 2000; Gries 2006).

Integrating corpus data into the usage-based model has two apparent advantages (Mukherjee 2005). For one thing, the authenticity of the usage data recorded in corpora makes the usage-based model truly empirical. For another, corpus data provide natural, diverse contexts in which a given linguistic unit is used, which is favorable to the depiction of usage patterns related to the unit and the identification of relevant contextual factors that motivate its occurrences.

Concerned mainly with the technical state of the arts of usage-based linguistics, Tummers et al. (2005) indicate that the attribute of corpus data as off-line products of non-elicited language use accounts for the tendency of usage-based studies to employ corpora rather than other methods of gathering usage data such as surveys and experiments. They further distinguish between *corpus-illustrated* and *corpus-based* linguistics and argue that the latter would be the more suitable methodology for a conception of language that is genuinely usage-based. To be exact, corpus data should not simply serve as a data set for the selection of instances and as evidence for the mere existence of a given linguistic expression, as in corpus-illustrated studies. Rather, quantification and statistical data such as frequencies and percentages, as adopted in

corpus-based research, are required for a more systematic and detailed analysis that identifies relevant features or patterns of language use.

In search for a quantitative usage-based method, Tummers et al. (ibid., p.240) further single out two predominant clusters of quantification techniques employed in usage-based studies: frequency and collocational analyses. Information drawn from frequency counts is empirically valuable for pinpointing such usage tendency as how a given linguistic expression is typically used. In addition, collocational analyses are extensively utilized to identify the lexical/grammatical and non-linguistic association patterns with specific linguistic expressions as well as to examine the distribution of near-synonymous linguistic units.

In the present study, both of the two quantitative methods based on corpus data are used to derive a usage-based account of *V-kai* resultative verbs. Frequency counts of each individual *V-kai* combination (i.e. token frequency) was conducted first. In addition, the frequency of different categories of verbal collocates with respect to the postverb *-kai* (i.e. type frequency) was also calculated. However, before a systematic collocational analysis involved in *V-kai* resultative verbs can be provided, an apposite categorization of the verbs that can appear in the slot preceding *-kai* was in order, a methodological issue to be dealt with in the next section.

### 3. A frame-semantic categorization of verbal collocates of *-kai*

Previous attempts to systematically categorize verbs rely either on their grammatical behaviors or on their meaning relations with one another. For instance, Levin (1993) classifies English verbs according primarily to their syntactic alternation patterns, based on the assumption that “verb behavior can be used effectively to probe for linguistically relevant pertinent aspects of verb meaning” (ibid., p.1). To illustrate, transitive verbs that involve putting and covering, including *spray*, *brush*, *load*, *pump*, *smear*, and *stuff*, can occur in the locative alternation (e.g. *Sophie sprayed water on the bushes* vs. *Sophie sprayed the bushes with water*), and thus belong to the same verb class (ibid., p.51).

While the approach of verb categorization deriving exclusively from alternations adopted in Levin (1993) is liable to generate semantically coherent verb classes, verbs with related meanings may also be split in different classes, or verbs with disparate meanings can also be lumped in the same class (Baker & Ruppenhofer 2002). Levin’s verb classes are only partially semantically motivated in that not all argument syntax reflects the inherent lexical semantics of verbs. Furthermore, the approach of verb classification on the sole basis of grammatical behaviors would be expected to have limited applicability for languages with more impoverished syntactic alternations like Chinese.

In contrast to the syntactic approach employed in Levin’s (1993) study of verb classes, the frame-based approach groups words (not merely verbs) according to the underlying conceptual structures that support and motivate them, namely, *cognitive frames* or *knowledge schemata* (Fillmore & Atkins 1992). The concept of *frame*, as a fundamental construct of Frame Semantics (Fillmore 1976, 1977, 1982, 1985), refers to

“schematic representations of the conceptual structures and patterns of beliefs, practices, institutions, images, etc. that provide a foundation for meaning interaction in a given speech community” (Fillmore, Ruppenhofer & Baker 2004:26).

Additionally, Frame Semantics holds an encyclopedic view of meaning (cf. Evans & Green 2006; Fillmore 1985), contending that each linguistic expression (a lexical, phrasal, or constructional unit) evokes a schematized semantic frame or instantiates particular elements of it; thus, the frame functions as a knowledge structure against which the meaning of the expression is understood (Croft & Cruse 2004). That is, linguistic units are relativized to a structured inventory of encyclopedic knowledge that provides the cognitive underpinnings to license their grammatical behaviors such as valence or argument structure. It follows that the frame semantic approach of verb categorization resorts mainly to lexical semantic information and exempts itself from the necessity to count on syntactic criteria for their grouping; for that reason, a frame-based classification would have the advantage of allowing semantically related verbs to be grouped together despite their discrepant syntactic behaviors (Boas 2006).

For the purpose of this study, the verbs found to collocate with the postverb *-kai* are preferably classified on a semantic basis, for the two verbs in Chinese resultative verbs are generally recognized to have a semantic rather than a syntactic collocational relationship. It has been suggested that any two verbs capable of prompting for an “action-result” interpretation can be combined together to form a resultative verb (Shi 2002). In addition, the interpretation of the gestalt resultative verb derives from the meanings of the two constituents that are integrated in the semantically most natural way (Thompson 1973). It can therefore be inferred that the semantic properties of the verbal collocates are the principal factor that determines whether they can co-occur with a given postverb to form a resultative verb.

In the actual grouping procedure executed in the present study, an intuition-based categorization of the verbal collocates according to the meanings these verbs denote as they co-occur with *-kai* in resultative verbs was performed first. Such preliminary collocational categorization was then confirmed or modified based on the frames that the English equivalents of these verbal collocates are associated with. As expected, verbal collocates in the same categories usually pertain to frames that are conceptually related to each other.

#### 4. Frequency and productivity of *V-kai* resultative verbs

This section presents the results of data analysis concerning the usage patterns of *V-kai* resultative verbs in the corpus. The distribution of individual *V-kai* complex verbs will be presented first. Then the major categories of verbal collocates of *-kai* and its different meanings involved when co-occurring with distinct verb groups in resultative verbs will also be discussed.

#### 4.1 Distribution of V-*kai* resultative verbs in the corpus

As mentioned, the corpus on which this study is based yields a total of 2747 tokens of V-*kai* resultative verbs and 138 distinct verbs were found to collocate with the postverb *-kai*. Table 1 presents the most frequent ten V-*kai* resultative verbs, all of which occur at least 45 times in the corpus data. As can be seen, the distribution of V-*kai* constructions in the corpus is fairly skewed. To illustrate, the four most frequent V-*kai* combinations alone, namely, *lí-kāi*, *zhǎn-kāi*, *dǎ-kāi*, and *fēn-kāi*, with each of their token frequencies exceeding one hundred, account for more than three-fifths (60.4%) of the data.

Rank	V- <i>kai</i> Exemplar	Tokens	Percentage
1	<i>lí-kāi</i> 'to leave'	628	22.9%
2	<i>zhǎn-kāi</i> 'to unfold; to launch'	601	21.9%
3	<i>dǎ-kāi</i> 'to open; to switch on'	309	11.2%
4	<i>fēn-kāi</i> 'to separate'	121	4.4%
5	<i>bì-kāi</i> 'to avoid'	88	3.2%
6	<i>jiē-kāi</i> 'to uncover'	69	2.5%
7	<i>zhāng-kāi</i> 'to open'	63	2.3%
8	<i>fān-kāi</i> 'to open; to turn over'	57	2.1%
9	<i>lā-kāi</i> 'to pull apart'	48	1.7%
10	<i>jiě-kāi</i> 'to untie'	45	1.6%

**Table 1: Top ten most frequent V-*kai* resultative verbs in the corpus**

The finding that V-*kai* complex verbs are not equally frequent in the corpus data conforms to the probabilistic nature of language use. The different token frequencies of co-occurrences between V<sub>1</sub> and *-kai* are not idiosyncratic, but motivated, mainly by semantic rather than syntactic compatibility. Shi (2002) points out that a verb tends to co-occur with its most natural result in Chinese resultative verbs. The above four extremely common cases of V-*kai* complex verbs indeed have an internal semantic structure of natural action-result relationship. The significantly high frequencies of occurrences of particular V-*kai* complex verbs in actual language use would also have linguistic and cognitive consequences, namely, lexicalization and entrenchment, which will be discussed in detail later in Section 5.1.

#### 4.2 Categories of verbal collocates within V-*kai* resultative verbs

Based on a frame-semantic approach to verb classification, some verb categories can be observed to collocate more productively and frequently with the postverb *-kai* than others. Table 2 summarizes the number and percentage of both verb types and tokens of V-*kai* resultative verbs accounted for by each of these productive categories of verbal collocates with *-kai*. Overall, the six major categories represent almost 90 percent of the



verb types found to co-occur with *-kai* and nearly 95 percent of the tokens of *V-kai* resultative verbs in the corpus data. The remaining types of verbal collocates of *-kai* and tokens of *V-kai* resultative verbs are represented by other minor, far less productive verb categories.

Categories of Verbal Collocates	Verb Types		<i>V-kai</i> Tokens	
	Number	Percentage	Number	Percentage
Verbs of Motion	41	29.7 %	898	32.7 %
Verbs of Manipulation	29	21.0 %	589	21.4 %
Verbs of Separating	27	19.6 %	284	10.3 %
Verbs of Dispersal	13	9.4 %	51	1.9 %
Verbs of Expansion	8	5.8 %	666	24.2 %
Verbs of Avoiding	5	3.6 %	108	3.9 %
Total	119	89.1 %	2596	94.4 %

**Table 2: Productive categories of verbal collocates with *-kai***

In fact, these major categories of verbal collocates involve interrelated verb frames to which many verb types belong (i.e. high type frequency). The verbs in these categories also tend to co-occur with *-kai* more frequently than those in other categories (i.e. high token frequency). It is worth noting that the percentage of verb types accounted for by a given category of verbal collocates does not necessarily correspond to its proportion of tokens of *V-kai* resultative verbs. For example, though the category of verbs of expansion represents only six percent of the verb types found to co-occur with *-kai*, the *V-kai* complex verbs in which  $V_1$  belongs to this category stand for about one-fourth of the data. The mismatch between type and token frequency in the distribution results from the markedly high token frequency of the *V-kai* exemplar *zhān-kāi* ‘to unfold; to launch’.

It is also observed that there exist regular patterns of meaning integration between  $V_1$  and the postverb in *V-kai* resultative verbs. Take the verb category with the highest type frequency, namely, verbs of motion, as an illustration. Verbs classified in this category are all associated with semantic frames involving translational motion and share core frame elements such as Theme, Path, Source, and Goal. When co-occurring with verbs of motion, *-kai* appears to convey the meaning of (an entity’s) moving away from a reference point, as in *yí-kāi* ‘move-away’ and *pǎo-kāi* ‘run-away’.

For another example, the second most productive category of verbal collocates is verbs of manipulation. Since most manipulation verbs pertain to frame elements that can be construed as containing objects, in *V-kai* complex verbs where  $V_1$  belongs to this category, the postverb generally designates the resultant open state of an entity, as in *zhuǎn-kāi* ‘turn-open’ and *qiāo-kāi* ‘knock-open’. In short, when *-kai* collocates with a specific class of verbs, its meanings are accordingly modulated in a systematic fashion. The emergence of foregoing patterning points to the abstraction of schemas of *V-kai* resultative verb construction with different levels of generality, a cognitive process to be addressed in the following section.

## 5. Usage effects on the entrenchment of *V-kai* resultative verbs

This section discusses the effects of usage frequency on the linguistic representation of resultative verbs in Chinese grammar. It will be argued that token and type frequencies exert distinct effects on the formation and entrenchment of Chinese resultative verbs. Specifically, while high token frequency gives rise to more lexicalized *V-kai* complex verbs, high type frequency results in more schematized *V-kai* constructions. Finally, I will also touch on the contribution of the corpus-oriented usage-based approach adopted in this study to the understanding of establishing *V-kai* resultative verbs as cognitive routines.

### 5.1 The effect of token frequency on the lexicalization of *V-kai* resultative verbs

The frequency of co-occurrence of two linguistic units is strongly related to their constituency: “the more often two elements occur in sequence, the tighter will be their constituent structure” (Bybee & Hopper 2001:14). Moreover, the co-occurring units with high frequency can further undergo linguistic change such as phonetic weakening, semantic alteration, functional shift, and the reanalysis of their internal grammatical structure (Biq 2007). Similarly in the cases of Chinese resultative verbs, the more frequently a verb co-occurs with a particular postverb, the more likely the verb and postverb are subject to fusion into a single lexical item (Shi 2002). In that case, the recurrent usage of a specific *V-kai* complex verb would serve as the impetus of its entrenchment as a lexicalized unit.

The corpus data concerning the usage of *V-kai* resultative verbs with relatively high token frequencies (e.g. *zhǎn-kāi*, *dǎ-kāi*, and *fēn-kāi*) indicate that these complex verbs display symptoms of lexicalization. It is generally agreed upon in the literature on lexicalization that semantic and grammatical reductions are often the concomitant features of lexicalized units (cf. Liberman & Sproat 1992). Moreover, as lexicalization occurs, the internal information in connection with individual word components, be it phonological, syntactic, or semantic, is likely to be less available to the operation of grammar as a whole (Packard 2000).

As for resultative verbs, Packard (2000:250) specifically points out that “the more lexicalized the gestalt verb is, the less the argument structure of the  $V_2$  non-head becomes part of the argument structure of the gestalt verb.” In other words, while the argument structure of resultative verbs is mostly a composite of the valence of  $V_1$  and  $V_2$ , the process of lexicalization would render the syntactic information of argument structure less accessible. This study of *V-kai* constructions will further show that not only  $V_2$ 's but  $V_1$ 's argument structure information can also be lost when the gestalt verbs undergo the process of lexicalization.

The analysis of the most frequently used *V-kai* complex verbs illustrates that the argument structure is indeed less compositional in terms of the respective valence of individual verbs. Take both arguments of the *zhǎn-kāi* complex verb in example (1) for

example, *xiǎo-yīng* ‘little-eagle’ and *chibāng* ‘wing’ pertain to  $V_1$  only, i.e. *zhǎn* and thus neither of the arguments is predictable from *-kai*. Likewise, in example (2), the two arguments of the *V-kai* complex verb, *dǎ-kāi*, i.e. *fùqīn* ‘father’ and *táng-hé* ‘candy box’, could be analyzed to be selected merely by the postverb *-kai*. The observation that  $V_1$ , *dǎ*, does not contribute to the argument realization of the gestalt verb could be ascribed to the semantic bleaching of its original meaning of *hitting* so that the verb denotes only a general cause of the opening of entities.

- (1) *xiǎo-yīng*            *bùyóudé*            ***zhǎnkāi***    *le*            *chibāng*<sup>2</sup>  
 little-eagle            cannot.but            spreadASP            wing  
 ‘The eaglet had no choice but to spread its wings.’
- (2) *fùqīn*    ***dǎkāi***    *táng-hé*,    *qǔ-chū*    *shí-kē*    *táng*            *lái*  
 father    open    candy-box    take-out    ten-CL    candy come  
 ‘The father opened the candy box and took out ten pieces of candy.’

Another source of evidence supporting the lexicalized inclination of considerably frequent *V-kai* complex verbs is the meaning extensions these gestalt verbs undertake. For instance, the original meaning of the complex verb *zhǎn-kāi* is considered ‘spread out’, as in (1). Nonetheless, the corpus data show that *zhǎn-kāi* is now predominantly used in the contexts where the gestalt verb denotes the meaning of ‘launch’, as in (3). Take *fēn-kāi* as another example; in addition to its grammatical role of a complex verb meaning *separate*, the complex word en bloc can also be used as an adverb, with its meaning extended to *separately* or *independently*, as in (4). The usage of extremely frequent *V-kai* resultative verbs with semantic extensions and functional shift is an indicator of these lexical patterns to be highly entrenched as autonomous gestalt verbs.

- (3) *gè*            *hòuxuǎnrén*    *yǐ*            *bàn-tuǒ*            *jìngxuǎn*    *dēngjì*  
 every candidate            already            do-proper            run.for            registration  
*bìng*    *fēnfēn*            ***zhǎnkāi***            *jìngxuǎnhuódòng*  
 and    one.by.one    launch            campaign  
 ‘Every candidate has already finished their registrations for the campaign and launched their campaigns one by one.’
- (4) *duì*    *yuánzhùmín*    *xuéshēng*    *yīnggāi*            *jízhōng*            *zài*    *yìqǐ*  
 to    aborigine            student            should            gather.up    in    together  
*huò*    ***fēnkāi***            *shàng-kè*    *de*            *wèntí*  
 or    separately attend-class    NOM question  
 ‘As for the question of whether aboriginal students should be put together or separately for the classes’

Delving into the usage data of noticeably frequent *V-kai* complex verbs confirms the previous assumption that token frequency is a reliable quantitative signpost of the degree of entrenchment. The repetitive co-occurrence between a specific verb with the

<sup>2</sup> Abbreviations and transcription conventions adopted in this study: ASP = aspect marker; CL = classifier; NOM = nominalizer.

postverb *-kai* establishes a gestalt status for lexicalized units with their own argument structure and semantic extension. It follows that the cases of *likāi*, *zhǎnkāi*, *dǎkāi*, and *fēnkāi* are more likely to be reanalyzed as independent gestalt words and represented in the language user's mental lexicon.

## 5.2 The effect of type frequency on the schematization of *V-kai* resultative verbs

In addition to token frequency, the factor of type frequency also pertains to the usage of *V-kai* resultative verbs, for this type of complex verbs are characterized by the semantically-constrained open slot that precedes the postverb. Further, the verbal collocates of the postverb *-kai* can be semantically classified into several groups and the verb instances in each group share meaning similarities, pointing towards the establishment of subcategories of *V-kai* constructions with their own structures.

The semantic coherence within the meaning integration between distinct verb categories and the postverb in *V-kai* complex verbs, along with their formal resemblance, suggests that a network of *V-kai* constructional schemas that are organized hierarchically, from the most schematic to the most substantive, should be represented non-reductively in the linguistic system, as displayed in Figure 1. As can be observed in the diagram, a more general morpho-syntactic *V-kai* schema as one specific case of resultative verb constructions in Chinese can be abstracted. Moreover, the more abstract *V-kai* schema could be further elaborated by some more specific schemas, all of which are represented by square boxes in the diagram. For instance, the sub-schema [ $V_{\text{MOTION-kai}}$  'away'] can be instantiated by the exemplar linguistic forms such as *zǒu-kāi* 'walk-away' and *rào-kāi* 'detour-away', instances of *V-kai* complex verbs represented in oval-shaped boxes in the diagram. In short, all of the *V-kai* schemas differing in their schematicity or specificity are redundantly represented in the language user's mental grammar.

In addition, various degrees of productivity regarding distinct categories of verbal collocates pinpointed in this study also reveal how well different *V-kai* sub-schemas are entrenched in grammar. In Figure 1, the degree of entrenchment of *V-kai* schemas and their instances is indicated by the extent to which the square or oval-shaped box is emboldened. In this case, the *V-kai* complex verbs in which the  $V_1$  belongs to the major categories of verbal collocates are deemed as more conventional routines, which are more likely to be stored as productive templates for generating and interpreting novel uses, and thus constitute the core usage of *V-kai* constructions. For instance, more instances of manipulation verbs are found to collocate with *-kai* than, say, expansion verbs, in real language use, so it can be predicted that the schema [ $V_{\text{MANIPULATION-kai}}$  'open'] is more routinized and enjoys a stronger representation in grammar.

Still, the combinations between less productive verbal collocates and the postverb showcase the more creative, peripheral usage of *V-kai* complex verbs in that they illustrate how speakers can draw on the more conventionalized routines and further deviate from them for innovative usages. For example, since fewer usage events that involve the distinct types of chatting verbs are found in the corpus data, it could be inferred that the [ $V_{\text{CHATTING}}-kai$  ‘inceptive’] schema would be far less entrenched than, for instance, the [ $V_{\text{MANIPULATION}}-kai$  ‘open’] schema in the linguistic system. The patterns of collocation in the usage of *V-kai* resultative verbs substantiate the fact that “while language users stick to various kinds of routines in the vast majority of communicative events they are able to deviate from these routines and produce...novel or unusual linguistic forms and structures” (Mukherjee 2005:215).

Moreover, the interface between token and type frequency can further shed light on the category formation of *V-kai* resultative verbs. The semantic similarities shared by all the instances of *V-kai* complex verbs that elaborate a particular *V-kai* pattern point to the development of different categories of *V-kai* schemas. Specifically, the properties of the verb types associated with a specific category of verbal collocates of *-kai* delimit the boundary of the ensuing *V-kai* schema. To illustrate, what the  $V_1$  in the exemplars that instantiate the constructional schema of [ $V_{\text{MOTION}}-kai$  ‘away’] have in common is the meaning component involving translational motion. It follows that these semantically-related verbs constitute an exemplar cluster or category that display prototype effect.

Precisely, this more productive category of *V-kai* schema [ $V_{\text{MOTION}}-kai$  ‘away’] is further organized with respect to members that may be central or peripheral to the category. In this example, the instance of *lì-kāi*, which entertains extremely high token frequency to be well-routinized, would be perceived as more prototypical, whereas the instance of *rào-kāi*, which occurs with very low token frequency and undergoes less entrenchment, would be recognized as a more peripheral member of this category. The present study hence subscribes to the position that “[g]rammar is built up from specific instances of use that marry lexical items with constructions; it is routinized and entrenched by repetition and schematized by the categorization of exemplars” (Bybee 2006:730).

### 5.3 From corpus to cognition

The quantitative information on both the token and type frequency of verbal collocation in *V-kai* complex verbs extracted from the corpus data suggest distinct patterns of cognitive routinization of *V-kai* resultative verbs in the speaker’s linguistic system. High token frequency of individual *V-kai* combinations is indicative of the entrenchment of particular *V-kai* resultative verbs as lexical items. In contrast, high type frequency of co-occurrence of verb instances classified in different categories with the postverb *-kai* illuminates the routinized patterns of *V-kai* resultative verbs as constructional schemas. This study thus corroborates the “From-Corpus-to-Cognition Principle” (Schmid

2000:39), which postulates that “frequency in text instantiates entrenchment in the cognitive system.” In other words, the investigation of real usage data in a corpus sheds much light on how *V-kai* resultative verbs are represented and organized in the speaker’s mental grammar.

With the premise that corpus data comprise large samples of actual use of the linguistic system, it follows that the “from-corpus-to-cognition” methodology adopted in this study also bridges the two extreme senses of “usage-based” mentioned in Dickinson and Givón (2000:151), namely, “user-based models” and “usage-produced data.” The former more theoretical sense involves providing linguistic descriptions to represent actual mental operation of speakers. In contrast, the latter methodological sense is concerned with descriptions strictly deriving from the empirical data of actual language use. This study therefore follows the spirit of a usage-based conception of language by acknowledging the significant role of usage data in describing language structures and constructing linguistic theories (c.f. Kemmer & Barlow 2000).

## 6. Concluding remarks

The present study adopts a corpus-oriented usage-based approach to a subclass of Chinese resultative verbs, *V-kai* constructions, employing the quantitative techniques of collocational and frequency analysis for their distribution and productivity. A skewed distribution of various *V-kai* combinations has been sketched in the corpus data; some *V-kai* complex verbs (e.g. *lí-kāi* and *zhǎn-kāi*) pertain to notably high token frequency. The frequency occurrence of these resultative verbs can be explained by the observation that  $V_1$  and *-kai* have a natural “action-result” semantic relationship.

In terms of the productivity of *V-kai* patterns, a frame-semantic categorization of  $V_1$  indicates that different verb categories co-occur with the postverb in varying degrees of type frequency. The major categories of verbal collocates of *-kai* include motion verbs, manipulation verbs, separating verbs, dispersal verbs, expansion verbs, and avoiding verbs. The high productivity of these verb groups signals their higher semantic compatibility with the postverb *-kai*. Additionally, when collocating with distinct verb categories, *-kai* yields different interpretations in the ensuing resultative verbs. In other words, when *-kai* collocates with a specific class of verbs, its meaning would be modulated to cater to the most plausible result reading in the given class of resultative verbs.

Moreover, this paper has also explicated the essential roles of frequency in usage events in shaping the linguistic representations of *V-kai* resultative verbs. The *V-kai* resultative verbs with comparatively high token frequencies in the corpus have been found to exhibit signs of lexicalization. The finding that their argument structure is less compositional and that they can undertake meaning extensions shows that these lexical units are highly entrenched as autonomous gestalt verbs. The corpus-based usage data in this study therefore reveal the close relationship between the frequency of two linguistic units occurring in sequence and their ensuing constituent structure.

In addition, the observation that verbal collocation within *V-kai* resultative verbs comes in different strengths accords with the experience-driven, probabilistic view of grammatical generalizations. Besides, the finding that various types of verbs co-occur with *-kai* with different frequency of co-occurrence attracts our attention to both conventional and creative usage of *V-kai* resultative verbs. Based on large samples of real usage data of *V-kai* constructions, this study demonstrates that cognitive routines and creative uses should be in equilibrium in actual language use.

Finally, the frequency analysis undertaken in this study indicates that the effects of token and type frequency effects are both at work in structuring the usage of *V-kai* constructions, which suggests that the formation of Chinese resultative verbs should be conceived as a lexical as well as a syntactic matter. Accordingly, the routinized patterns of *V-kai* constructions that vary in their extent of specificity and generality serve as a representative illustration of the continuum between lexicon and grammar that characterizes a usage-based conception of language.

## References

- Baker, Collin F. & Josef Ruppenhofer 2002. "FrameNet's frames vs. Levin's verb classes". *Berkeley Linguistics Society*, 28: 27-38.
- Biber, Douglas 2000. "Investigating language use through corpus-based analyses association patterns". In Michael Barlow & Suzanne Kemmer (Eds.), *Usage-based models of language* (287-313). Stanford, CA: CSLI Publications.
- Boas, Hans C 2006. "A frame-semantic approach to identifying syntactically relevant elements of meaning". In Petra Steiner, Hans C. Boas & Stefan Schierholz (Eds.), *Contrastive studies and valency: Studies in honor of Hans Ulrich Boas* (119-49). Frankfurt: Peter Lang.
- Bybee, Joan L. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.
- Bybee, Joan L. 2006. "From usage to grammar: The mind's response to repetition". *Language*, 82: 711-33.
- Bybee, Joan L. 2007. *Frequency of use and the organization of language*. Oxford: Oxford University Press.
- Bybee, Joan L. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Bybee, Joan & Paul Hopper (Eds.) 2001. *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.
- Biq, Yung-O 2007. "Lexicalization and phrasalization of *na* collocates in spoken Taiwan Mandarin". *Contemporary Linguistics*, 9: 128-36.
- Croft, William & D. Alan Cruse 2004. *Cognitive linguistics*. Cambridge: Cambridge University Press.
- Dickinson, Connie & Talmy Givón 2000. "The effect of the interlocutor on episodic recall: An experimental study". In Michael Barlow & Suzanne Kemmer (Eds.), *Usage-based models of language* (151-96). Stanford, CA: CSLI Publications.

- Diessel, Holger 2004. *The acquisition of complex sentences*. Cambridge: Cambridge University Press.
- Evans, Vyvyan & Melanie Green 2006. *Cognitive linguistics: An introduction*. Edinburgh: Edinburgh University Press.
- Fillmore, Charles J. 1976. "Frame semantics and the nature of language". In Stevan R. Harnad, Horst D. Steklis & Jane Lancaster (Eds.), *Origins and evolution of language and speech* (20-32). New York: Annals of the New York Academy of Sciences.
- Fillmore, Charles J. 1977. "The need for a frame semantics in linguistics". In Hans Karlgren (Ed.), *Statistical methods in linguistics* (5-29). Stockholm: Skriptor.
- Fillmore, Charles J. 1982. "Frame semantics". In *Linguistics in the morning calm* (111-37). Seoul: Hanshin Publishing.
- Fillmore, Charles J. 1985. "Frames and the semantics of understanding". *Quaderni di Semantica*, 6: 222-54.
- Fillmore, Charles J. & Beryl T. Atkins 1992. "Towards a frame-based organization of the lexicon: The semantics of RISK and its neighbors". In Adrienne Lehrer & Eva Feder Kittay (Eds.), *Frames, fields, and contrasts: New essays in semantics and lexical organization* (75-102). Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Fillmore, Charles J., Josef Ruppenhofer & Collin F. Baker 2004. FrameNet and representing the link between semantic and syntactic relations. In Chu-Ren Huang & Winfried Lenders (eds.), *Computational linguistics and beyond*: 19-64. Taipei: Institute of Linguistics, Academic Sinica.
- Gries, Stefan Th. 2006. Introduction. *Corpora in cognitive linguistics: Corpus-based approaches to syntax and lexis*, ed. by Stefan Th. Gries and Anatol Stefanowitsch: 1-17. Berlin: Mouton de Gruyter.
- Kemmer, Suzanne & Michael Barlow 2000. Introduction: A usage-based conception of language. *Usage-based models of language*, ed. by Michael Barlow and Suzanne Kemmer: vii-xxviii. Stanford, CA: CSLI Publications.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar, vol. 1*. Stanford, CA: Stanford University Press.
- Langacker, Ronald W. 2000. "A dynamic usage-based model". In Michael Barlow & Suzanne Kemmer (Eds.), *Usage-based models of language* (1-64). Stanford, CA: CSLI Publications.
- Langacker, Ronald W. 2008. *Cognitive Grammar: A basic introduction*. Oxford: Oxford University Press.
- Levin, Beth 1993. *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago Press.
- Li, Charles N. & Sandra A. Thompson 1981. *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.
- Liberman, Mark & Richard Sproat 1992. "The stress and structure of modified noun phrases in English". In Ivan A. Sag & Anna Szabolcsi (Eds.), *Lexical matters* (131-81). Stanford, CA: CSLI Publications.
- Mukherjee, Joybrato 2005. *English ditransitive verbs: Aspects of theory, description and a usage-based model*. Amsterdam: Rodopi.



- Packard, Jerome L. 2000. *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge: Cambridge University Press.
- Pustejovsky, James (1998). *The generative lexicon*. Cambridge, MA: MIT Press.
- Rostila, Jouni 2006. "Storage as a way to grammaticalization". *Constructions 1/2006* (www.constructions-online.de).
- Schmid, Hans-Jorg 2000. *English abstract nouns as conceptual shells*. Berlin: Mouton de Gruyter.
- Shi, Yuzhi 2002. *The establishment of modern Chinese grammar: The formation of the resultative construction and its effects*. Amsterdam: John Benjamins.
- Starosta, Stanley, Koenraad Kuiper, Siew-ai Ng & Zhi-qian Wu 1997. "On defining the Chinese compound word: Headedness in Chinese compounding and Chinese VR compounds". In Jerome L. Packard (Ed.), *New approaches to Chinese word formation: Morphology, phonology and the lexicon in modern and ancient Chinese* (347-70). Berlin: Mouton de Gruyter.
- Thompson, Sandra A. 1973. "Resultative verb compounds in Mandarin Chinese: A case for lexical rules". *Language*, 49: 361-379.
- Tomasello, Michael 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge, Mass.: Harvard University Press.
- Tummers Jose, Kris Heylen & Dirk Geeraerts 2005. "Usage-based approaches in Cognitive Linguistics: A technical state of the art". *Corpus Linguistics and Linguistic Theory*, 1: 225-261.

#### About the author

**Ben Pin-Yun Wang** is currently a Ph.D. candidate in the dual-title program of Applied Linguistics and Asian Studies. In addition, he was a visiting student researcher at the University of California, Berkeley (2008-2009) and taught as a lecturer in the Academic Writing Education Center at National Taiwan University in 2010-2011. His research interests include cognitive-functional linguistics and text/discourse analysis as well as their applications to language pedagogy.