Article

Andreas Liesenfeld*, Meichun Liu and Chu-Ren Huang

Profiling the Chinese causative construction with rang (讓), shi (使) and ling (令) using frame semantic features

https://doi.org/10.1515/cllt-2020-0027

Abstract: This behavioural profiling (BP) study examines the use of the near-synonyms rang (讓), shi (使) and ling (令), three ways to express cause-effect relationships in Chinese. Instead of using an out-of-the-box BP design, we present a modified approach to profiling that includes a range of frame semantic features that aim to capture variation of slot fillers of this construction. The study investigates the intricate semantic variation of rang, shi and ling through a comprehensive analysis of 38 contextual features (ID tags) that characterize the collocational, lexical semantic and frame semantic environment of the near-synonyms. Our dataset consists of around 100,000 data points based on the annotation of 1002 sentences of Mandarin Chinese of three varieties. The BPs of each near-synonym are compared using multidimensional scaling and hierarchical cluster analysis. The results show that rang, shi and ling are each characterized by a combination of distinctive features and how different feature types contribute to setting the near-synonyms apart based on their usage patterns. Methodologically, this study illustrates how behavioural profiling can be modified to include frame semantic features in accordance with the method’s emphasis on producing empirically verifiable results and how these features can aid a comparative analysis of near-synonyms.

Keywords: behavioural profiling, construction grammar, frame semantics, near-synonymy, causative constructions, Mandarin Chinese

*Corresponding author: Andreas Liesenfeld, The Hong Kong Polytechnic University, Hong Hum, Hong Kong, China, E-mail: amliese@polyu.edu.hk. https://orcid.org/0000-0001-6076-4406
Meichun Liu, The City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong, China, E-mail: meichliu@cityu.edu.hk. https://orcid.org/0000-0001-9471-7181
Chu-Ren Huang, The Hong Kong Polytechnic University, Hong Hum, Hong Kong, China, E-mail: churen.huang@polyu.edu.hk. https://orcid.org/0000-0002-8526-5520
1 Introduction

This article presents a behavioural profiling study of a Mandarin Chinese construction that is used to express a range of functions broadly located on the cause-effect continuum, such as resultatives, causation, permission, yielding. Specifically, we focus on *rang* (讓), *shi* (使) and *ling* (令), three ways to express To let/make. Each of these near-synonyms comes with a range of different uses that sets them apart from each other. This near-sameness of meaning that is characterized by both overlapping uses and subtle variation is the subject of this study. The topic that this paper addresses is thus, for one, how to account for similarity and difference in the use of *rang*, *shi* and *ling*, and how these subtle differences can be described in a corpus-based way?

Our method is behavioural profiling, a quantitative approach that is known for producing data-driven, verifiable results in the area of near-synonymy research (e.g. Divjak 2003, 2006; Gries 2006, 2010; Gries and Divjak 2009). The method allows a detailed comparison of the internal structure of *rang*, *shi* and *ling* through the identification of co-occurrence patterns, and enables a data-driven evaluation of how these patterns characterize the (dis)similarity of the near-synonyms.

However, applying an out-of-the-box BP approach to our case has proven problematic due to the method’s reliance on lexical semantic classification of slot fillers. We show that, in contrast to previously examined causative constructions in Russian and Dutch (and arguably most Indo-European languages), Chinese *rang*, *shi* and *ling* are hard to profile by focussing on analysing lexical semantic properties of noun phrases that appear in their causer and causee slots (cf. Divjak 2006; Divjak and Gries 2006; Levshina et al. 2014). Specifically, we show that this established approach to behavioural profiling is problematic in the case of Chinese causative constructions due to practical difficulties in systematically identifying causer slot fillers in Chinese. Instead, we propose a modified approach to profiling that draws on frame semantics (Fillmore 1985). This is reflected in our feature design that includes a range of frame semantic contextual features (ID tags) that aim to describe frame elements. Our approach aims to explore whether frame semantic features can aid the analysis of the Mandarin Chinese causative constructions in a corpus-driven, systematic way. We investigate this question by incorporating a wide range of collocational, lexical semantic and frame semantic features in our profile that describes the slot fillers of the causative construction in a manually annotated dataset. The following questions are posed:

1. Does the inclusion of frame semantic elements as features in the profile of *rang*, *shi* and *ling* aid the systematic description of usage differences of the near-synonyms?
(2) Do the three near-synonyms differ in their preference to select certain frame elements?

The paper is organized as follows: Section 2 briefly outlines major trends in synonymy studies in cognitive linguistics and reviews previous behavioural profiling studies on synonymy. The section also reviews previous cognitive-linguistic work on the Chinese near-synonyms rang, shi and ling. Section 3 discusses the dataset and method, focusing on how frame semantic features are incorporated in our profile and how they were systematically annotated. Section 4 presents the results of our behavioural profiling and discusses the role that different types of features play in describing the behaviour of the causative construction. We conclude with showing that some frame semantic features turned out to be amongst the top distinguishing features of the three near-synonyms and discuss some accidental findings with regards to verbs of emotion.

2 The study of synonymy in cognitive linguistics

Theoretically this study is grounded in the usage-based constructionist view and explores how knowledge of rang, shi and ling may be organized in taxonomic networks, and how the subtle variation that is characteristic of near-synonyms becomes inductively learnable based on a given exemplar space (Croft 2001; Goldberg 1995, 2006; Langacker 1987).

A well-established method in cognitive linguistics, the behavioural profiling approach is based on the view that the study of semantics involves lexical items as categories. Historically, this perspective can be traced back to work in philosophy of language on family resemblance (Wittgenstein 1953) and research on categorization (Rosch 1978). With regards to synonymy, the categorization approach raises questions related to the structure and organization of categories that are somewhat similar or “overlapping” as well as how categories and their differences can be identified and described (cf. Divjak and Gries 2008). The BP approach engages with these issues by adopting some form of representational formalism based on semantic network models (e.g. Brugman and Lakoff 1988; Lakoff 1982, 1987). Generally, these models consist of a range of encyclopedic semantic features (listemes), based on which senses are distinguished and linked to lexical items (Glynn 2014).

Behavioural profiling stands out amongst some other semantic network model-based approaches for its strong commitment to empiricism and inductive reasoning (Sandra and Rice 1995). BPs aim to enable insights about something intrinsically non-observable by collecting and relating aspects of it that are more
directly observable. In the case of synonymy, where the subject of study is to elucidate nuances and subtle variations in meaning, BP enable a degree of empirical verification of results and hypotheses (Berez and Gries 2008; Geeraerts 2010; Glynn 2010b, 2010c). In short, behavioural profiling is an approach to synonymy that measures meaning quantitatively, doing so as empirically as possible while producing operationalizable results.

2.1 Studying synonymy using behavioural profiles

This study employs the method of behavioural profiling to examine the (dis)similarity of rang (讓), shi (使) and ling (令). Following Cruse (2010), we view the three items as near-synonyms that exhibit overlapping meaning and senses, but that do not overlap completely. behavioural profiling offers a way to tackle long-standing problems in the field: reliance on subjective interpretation and difficult validation of results. The method enables the researcher to describe and delineate the subtle differences in meaning that characterize near-synonyms in a way that relies less on subjective interpretation and more on data-driven observation (see amongst others, Geeraerts 2006; Gibbs 2007; Glynn 2010a; Glynn and Fischer 2010; González-Marquez et al. 2007).

Corpus-driven, quantitative and distributional in nature, behavioural profiling is especially suitable to analyse the interface of different levels of linguistic analysis, based on the distributional hypothesis that holds that distributional similarity relates to semantic and functional similarity (Firth 1957; Harris 1954). Alongside polysemy and antonymy, synonymy and near-synonymy studies that use behavioural profiles have gained popularity following a series of seminal papers that laid out the method by Divjak (2003, 2006) and Gries (2006, 2010, 2017). Since its inception, behavioural profiling has been applied to a range of topics in this field. The method has been used to conduct exploratory studies of two or more lexical items with the aim to analyse similarities and dissimilarities based on their behaviour in a dataset (e.g. Desagulier 2014; Dosedlová and Lu 2019; Krawczak 2014; Levshina 2011; Levshina and Heylen 2014; Levshina et al. 2014; Liu 2010; Liu and Espino 2012). Other applications are comparative studies across near-synonyms in different languages or varieties (Čarapić 2015; Levshina 2012; Xu et al. 2020), studies that test specific hypotheses (Janda and Solovyev 2009) and studies on diachronic change (Jansegers and Gries 2020; Pettersson-Traba 2016, 2018).
The present study is conceived as an exploratory data analysis and applies the BP approach to the analysis of the Chinese causative construction with rang, shi and ling. To the authors’ knowledge, this is the first study on this topic that uses the BP approach. This raises two issues. Generally, existing BP studies have largely focussed on Indo-European languages and the design of these studies may not be readily transferable to the present case. As to causative constructions in particular, a series of BP studies has been conducted on Dutch periphrastic causatives (Levshina 2011, 2012; Levshina and Heylen 2014; Levshina et al. 2014). While these studies provide valuable insights on BP approaches to causative constructions, several issues arise when working with Chinese instead of Dutch.

2.2 The Mandarin causative construction with rang (讓), shi (使) and ling (令)

The Chinese near-synonyms rang (讓), shi (使) and ling (令) can, sometimes interchangeably, be used as constructions that express causative relationships including coercion, ordering, causation and permission (see amongst others Chao 1965; Huang and Shi 2016; Li and Thompson 1989; Lu 1982). Rang can also be used as passive and to express yielding (Weng 2007).

As a causative construction, the three near-synonyms have a three-slot structure in common. They are followed by a noun phrase, the causee (sometimes also called effect) and an affectee (also affected predicate). And they are preceded by a causer (also cause) that can take many forms, ranging from single noun phrases and zero-subjects to complex events. Following Stefanowitsch (2001), the analytic causative construction and its slots can be depicted as follows (see Figure 1):

![Figure 1: Structure of the Mandarin analytic causative construction with rang.](image-url)
Example (1) shows an instance of *rang* featuring two noun phrases (personal names) in the causer and causee slots, followed by a verb construction as the affectee.

(1) 月梅讓阿奇送她回家。
    *Yue-mei make A-ji bring she return home*
    “Yue-mei makes/lets A-ji bring her home.”

Previous lexicography studies on *rang, shi* and *ling* (and on Mandarin causatives in general) have described different senses that each item can have and how these senses differ (Hu and Yang 2015; Li 2003; Zhao and Shao 2009). Various different senses were described that populate a cause-effect continuum, such as resultatives, causation, permission and yielding. We adopt the generic term ‘causative construction’ to refer to the whole spectrum of these uses. Other studies on the construction explored grammaticalization (Niu 2007), typological aspects (Niu 2008) and lectal variation (Tian and Zhang 2020).

Using behavioural profiling, we aim to further explore characteristics of the construction by analysing the environment they occur in, made measurable through designing a range of contextual features (ID tags) that subsequently enable the analysis of co-occurrence patterns. We begin by examining collocational patterns, also known as Wordsketches (Kilgarriff and Tugwell 2001). Given a segmented and tagged corpus, Wordsketches can be extracted automatically and provide important preliminary insights of how the construction behaves across a dataset. Then we further enrich our dataset by manually annotating a range of contextual features that are designed to capture aspects of the main slot fillers (causer, causee and affected). In summary, our features roughly fall into three categories: collocational, lexical semantic and discourse, and frame semantic. A series of studies focussing on Dutch “periphrastic causative constructions” is similar in scope (Levshina 2011, 2012; Levshina and Heylen 2014; Levshina et al. 2014). Example (2) from Levshina and Heylen (2014) shows the Dutch causative construction *deed* and *liet* (“to let/to cause”) in action.

(2) De politie deed/liet de auto stoppen.
    *the police did/let.PAST the car stop*
    Causer – Auxiliary Predicate – Causee – Effected predicate
    “The police stopped the car (let the car stop).”

Levshina and Heylen identified noun phrases in the causer and causee slots (“the police” and “the car”) and then, inter alia, classified them by semantic type, e.g. the ‘Causer’ as an animate NP and the ‘Causee’ as an inanimate NP. This strategy yields good results as long as the causer and causee slots are consistently filled by a noun phrase. However, in Chinese, this may not be the case. Here, the
identification of what fills up the causer slot can be complex. In contrast to the previously discussed Example (1) that shows the use of rang with two single noun phrases in causer and causee slot, consider Example (3):

(3) 1993 end of the year, Ethiopian government announced PAR develop science and technology PAR national policy document, formulate need to cultivate, select, introduce, develop, popularize and apply suitable PAR Ethiopia PAR science and technology, allow/cause science and technology promote productivity PAR development.

“At the end of 1993, the Ethiopian government announced a national policy document for the development of science and technology, stipulating that it wants to cultivate, select, introduce, develop, popularize and apply science and technology suitable for Ethiopia, allowing/causing science and technology to promote the development of productivity.”

Here no single noun phrase (“Ethiopian government” or “national policy document”) fills the causer slot of the causative construction with rang. Rather, the event of the announcement of the policy document by the government (as well as its content) fills the causer slot for what happens in the causee slot, i.e. that science and technology promotes development of productivity. The occurrence of causer slot occupants other than single noun phrases is frequent in our dataset which poses a challenge to approaches that annotate semantic properties of the causer based on noun phrases only. A different approach to describing slot occupants is needed. This study explores the use of frame semantic features as a possible way to describe slot fillers. Our behavioural profile includes a range of frame semantic features that are designed to systematically describe characteristics of the slot fillers. We test whether the inclusion of this type of feature enables a richer analysis of near-synonyms through examining the characteristic of their slot fillers, including co-occurring frame semantic elements. Do the near-synonyms attract or repel specific frame semantic features? How does this type of feature interact with other features? Following Fillmore (1985), we adopt a theory of semantics based on world knowledge that aims to describe the use of causative constructions as related to causation scenarios (as common in frame semantic resources such as framenets (Baker et al. 1998; Fillmore and Baker 2010; Fillmore et al. 2004). Our behavioural profile reflects this through the inclusion of nine frame semantic features that are designed to describe the occurrence of various frame elements that occur in relation to the main slot fillers of the causative construction.
In total, our behavioural profile (BP) draws on 38 features that can roughly be divided into three types: 14 collocational features, 15 lexical semantic and discourse level features and nine frame semantic features. The profile design aims to provide corpus-based evidence of the (dis)similar behaviour of rang, shi and ling. Through the inclusion of frame semantic features we capture additional characteristics of their slot fillers and aim to describe characteristics of slot occupants other than single noun phrases. The following section describes the dataset we use and discusses how the hallmark emphasis of the profiling approach on producing verifiable results can be retained when including features based on the sometimes vague notion of frames.

3 Data and method: Profiling (讓), shi (使) and ling (令) using frame semantic features

The BP approach was chosen for two reasons. As a multi-factorial method, it is particularly suitable to investigate fine-grained differences in meaning such as those that characterize near-synonyms. Meaning is here treated as a graded phenomenon and investigated from a probabilistic point of view in terms of attraction and repulsion tendencies. Another strong suit of the BP approach is its high verifiability and potential to minimize subjective judgement. This is achieved in several steps typical of the BP approach (Divjak and Gries 2006; Gries 2010, 2017; Gries and Divjak 2009; Gries and Otani 2010).

First, a dataset is created by compiling a corpus of sampled target sentences that feature the lexical items under study (rang, shi or ling). This dataset is then annotated for a range of characteristics that aim to characterize the collocational environment, morphosyntax and semantics of each instance. This coding process can include a wide range of specific variables on different scales, but that all have in common that they can be annotated with a high degree of verifiability. The sum of all these annotated variables (also known as contextual features or ID tags) then yields the behavioural profile of an instance of the use of rang, shi or ling.

Based on the premise that distributional similarity relates to semantic and functional similarity, these behavioural profiles can then be compared using multivariate statistical methods. The result is a verifiable set of observations that describe the differences and similarities within the dataset based on the manually annotated contextual features, and that allows for probabilistic insights into how certain aspects of these features relate to each other. In addition to morphosyntactic and lexical semantic features, the dataset of this study also contains a range of frame semantic features. These were included to capture additional
characteristics of the behaviour of rang, shi and ling, such as the occurrence of various frame semantic elements related to the near-synonyms and their slots. Crucially, these features were designed and annotated with similar emphasis on verifiability, including only aspects of frames and frame elements that are observable with a high degree of objectivity.

This research design aims to offer corpus-based evidence for further theorizing of the relationship of rang, shi and ling as three ways to express causative relationships, and topics such as their synonymy-continuum, network representations of their meaning and the interface between morphosyntactic, lexical semantic and frame semantic aspects of these constructions.

3.1 Data preparation

The study analyses the behaviour of the three most frequent lexical items that instantiate the causative construction in the Chinese Gigaword corpus (Graff and Chen 2005; Huang 2009): rang, shi and ling. We extract a subcorpus of Gigaword in the form of 1002 sentences that each contain the use of rang, shi or ling as causative constructions, 334 instances each. This subcorpus was cleaned to only contain this use of the three words and then manually annotated for 103 contextual variables, producing a data frame of 103,206 data points (1002 × 103).

Chinese Gigaword is a corpus of news-wire Chinese that includes three varieties of Mandarin Chinese in the form of data from Taiwan Central News Agency, Xinhua News Agency and Singapore Zaobao Newspaper. Hence, this study is based on a cross-variety dataset that explores the usage of the near-synonyms in formal, journalistic writing in mainland China Putonghua, Taiwan Mandarin and Singapore Mandarin. The decision to base our investigation on a cross-variety dataset instead of focusing on a single variety was made after examining the Wordsketches (collocational patterns) of the near-synonyms across the three varieties using the tagged version of Chinese Gigaword 2.0 (see Appendix I). The top three most frequent causer slot fillers, causee slot fillers and collocating modifiers of the three near-synonyms overlap significantly across the three varieties. We take this as an indicator that their behaviour across the varieties is similar enough to warrant a cross-variety examination. For a study that explores lectal variation of this construction in more detail, see Tian and Zhang (2020).

We extracted a simple random sample of 1002 sentences from the tagged and segmented version of Chinese Gigaword Corpus 2.0 (Huang 2009). This extracted subcorpus reflects the composition of Chinese Gigaword 2.0 as a whole, around 55% of the data is Taiwan Mandarin, 35% is mainland China Putonghua and 10% is Singapore Mandarin. The dataset was then manually checked and uses of rang, shi
and *ling* other than as causative constructions were excluded. For instance, uses of *rang*, *shi* or *ling* as part of verb phrases (e.g. *shiyong* “to use”, *xingshi* “to exercise”, *rangbu* “give way” or noun phrases (e.g. *dashi* “ambassador”, *panling* “order”) and idiomatic expressions (e.g. *shi ge zhaoushu* “play a trick”) were excluded.

### 3.2 Feature selection and annotation

The annotation of the 1002 instances of *rang*, *shi* or *ling* was done by three expert annotators, double-blind and with disagreement resolved via majority vote. Inspired by Divjak and Gries (2006), we annotate for a large range of heterogeneous features, starting from formal collocational patterns, then gradually extending to lexical semantic, discourse level and frame semantic features. Table 1 provides an overview of the 38 features that each instance was annotated for. In total, the 38 features contain 103 variables as some features are binary and others have as many as nine categories. The design of the features and categories was driven by two principles. On one hand, we based our feature design on existing profiling studies on causative constructions, their modifiers, light verbs and other related verb types (Huang et al. 2014; Huang et al. 2017; Levshina et al. 2014; Xu et al. 2020). On the other hand, we had to take hapax legomena and the overall frequency distribution of our dataset into account and adjust the scope of some features to come up with categories that divide our dataset into chunks of appropriate size for further analysis. The Sections 3.2.1–3.2.4 provide an overview of the annotation procedure of different types of features. We briefly discuss collocational and lexical semantic features as they are already an established part of behavioural profiling design, but provide a more detailed account of how we include frame semantic features as well as some of the issues that arise from doing so. A description of all features, their categories and definitions, alongside example sentences, can be found in Appendix II.

#### 3.2.1 Annotation of collocating modifiers and light verbs of *rang*, *shi* and *ling*

Before examining the main slot fillers of the causative constructions, we looked at all the elements that co-occur with *rang*, *shi* or *ling*. The Wordsketches showed that the three near-synonyms collocate with a total of 31 different modifiers and light verbs in our dataset. Based on previous studies, we group these elements into a range of categories based on their function of modifying polarity, degree, status or time, as well as the occurrence of light verbs (Hong and Huang 2006; Huang and Lin 2012; Huang et al. 2005, Zhou 1987; Zhu 1985). Frequently occurring modifiers such as *ye rang* “also make/let” are included as a category of their own, while less
frequent modifiers and hapax legomena are grouped together under umbrella categories to avoid an excessive number of categories for further analysis. “Future
tense modifiers”, for instance, is an umbrella category that includes hui rang, jiang hui rang, and jiang rang “will make/let”.

Feature: ye
- includes: 也+讓/使/令

Feature: future tense modifier (occurrence of hui, jiang hui or jiang)
- includes: 會+讓/使/令, 將會+讓/使/令 or 將+讓/使/令

3.2.2 Annotation of causer and causee slot fillers

A second round of annotation examines the main slot filler of the causative constructions. We focus on the three core elements that the construction comes with: a causer (i.e. an actor, cause or event), an causee (i.e. an undergoer of the causer) and an affected (i.e. a predicate). Our annotation efforts aim to dissect this information into adequate categories beginning with the noun and verb phrases that fill up the causer, the causee and the affected slots of the causative construction. This is achieved by bringing in a taxonomy of lexical semantic types and verb classes. But what scope and level of detail is appropriate? Existing taxonomies for Mandarin and behavioural profile studies of near-synonyms in other languages provide a starting point (Russian: Divjak and Gries 2006, Dutch: Levshina et al. 2014, studies on Mandarin: (Liu 1996, 2020; Liu and Chang 2016; Liu and Chiang 2008; Tian and Zhang 2020). Our custom taxonomy is largely based on existing literature but was adjusted to fit the scope of our dataset. As to semantic types, we went for a relatively broad categorization: (1) human/animate/speaker, (2) organization/institution, (3) event/phenomenon of nature/abstract/social, (4) inanimate object/physical/body part/geographic. Here is an example of a human causer in form of the personal name Yue-mei:

Feature: semantic type: human/animate/speaker causer: Yue-mei
- 月梅讓阿奇送她回家 Yue-mei makes/lets A-ji bring her home.

Verb classes are categorized as follows: (1) motion/movement; (2) cognition; (3) communication, (4) emotion; (5) perception/mental activity; (6) social-interaction; (7) verbs of change-of-state; (8) verbs of events and (9) verbs of being/states. Here an example of a verb of motion/movement:

Feature: verb class: verb of motion/movement: song huijia “bring home”
- 送她回家 bring her home
In addition to semantic type and verb class, we also annotate voice via the occurrence of the passive marker *bei*, transitivity structure and semantic type of the affected in case of ditransitive structures. At the discourse level we annotate clause structure, co-reference, volitionality, possession and mood.

3.2.3 Annotation of frame elements

After annotating the main slot fillers that make up the skeleton of the constructions, one can now start filling in other elements that occur in relation to the slot fillers, so-called “non-core elements” that further specify the causation scenario. This includes elements of this frame other than the three core elements which we focussed on in the Section 3.2.2. However, the frame elements that different framenet resources include under causation frames vary and, to our knowledge, no specification of a causation frame and its elements exists for Mandarin Chinese. After consulting some of the most developed framenet projects for English, Brazilian Portuguese, Swedish and Japanese (Borin et al. 2010; Fillmore et al. 2004; Laviola et al. 2017; Ohara et al. 2004), we decided to design our frame elements in accordance with “general causation scenario” frame in the Berkeley Framenet of English. We annotate elements that specify place, time, frequency, means and manner of the causation scenario. Taking the Berkeley framenet as a reference, we include the following elements that this framenet lists under the “causation” frame.¹

  Included frame elements are
  - Frequency [fre]: How often the action denoted by the target occurs.
  - Manner [man]: Any holistic description of the event, including overall depictions (the same way) and descriptions pertaining to the Actor’s influence on the character of the event (eagerly, quietly).
  - Means [mea]: An action through which the Actor or Causer accomplishes the action indicated by the target.
  - Place [place]: Where the event takes place.
  - Time [time]: This FE identifies the Time when the event occurs.

However, some other elements that this framenet also lists are not included in our study, because we could not find a way to annotate these elements consistently across a large dataset:

  - Explanation [exp]: The Explanation denotes a proposition from which the main clause (headed by the target) logically follows. This often means that the Explanation causes the target’s proposition, but not in all cases.

¹ https://framenet.icsi.berkeley.edu/fndrupal/frameIndex.
Concessive [con]: This FE signifies that the state of affairs expressed by the main clause (containing the target) occurs or holds, and something other than that state of affairs would be expected given the state of affairs in the concessive clause.

Circumstances [cir]: Circumstances describe the state of the world (at a particular time and place) which is specifically independent of the event itself and any of its participants.

Example (4) shows a causation scenario with shi from our dataset. Here we annotated two non-core frame elements.

(4) 昨晚 9 時 20 分, 這名 40 多歲的馬來男子, 疑在武吉班讓信佳路第 1 6 1 座組屋 殴打一名 20 多歲女子，使她鼻 子流血。

Yesterday evening 9 o’clock 20 minutes, this CL 40 more years-old PAR Malay male, is suspected at Bukit Panjang Senja Road No 161 public housing block, beat a CL twenty more years-old woman, cause her nose bleed

“Yesterday evening 9:20 o’clock, this forty years-old Malay male, is suspected to have beaten a twenty year-old woman at Bukit Panjang Senja Road No 161 public housing block, causing her nose to bleed.”

In this example, “Yesterday evening 9:20 o’clock” is annotated as a time frame element, a specification of the time of the causation scenario. “At Bukit Panjang Senja Road No 161 public housing block” is a place frame element, a specification of the location of the causation scenario. Both elements here precede shi. Including frame elements as additional features in the behavioural profile enables a richer description of the contextual similarities and differences of rang, shi and ling. But can the frame elements be defined clearly and annotated consistently? How can these elements be included in the behavioural profile without jeopardizing the method’s hallmark high degree of objectivity?

3.2.4 Quality control and annotation consistency

Our dataset of 1002 sentences, 38 features and 103 variables (in total 103.222 data points) was manually annotated and checked independently by three expert annotators. For behavioural profiling studies this is a high number as many studies seem to be annotated by one person only (e.g. Divjak and Gries 2006). This is usually not a problem because BP studies, as discussed earlier, generally rely on features and variables that are identifiable with a high degree of objectivity. When bringing in the sometimes vague notion of frames and their elements, however, this topic deserves extra attention. We opt for a rule-based definition of frame elements that results in clear annotation guidelines and high inter-annotator
agreement, but that leaves open questions of how to appropriately define frame elements in practise.

We define each frame element as a list of rules. The place frame element, for instance, is defined as the occurrence of a location marker zai “at”, shang/xia “on/under”, li/wai “in/out” in conjunction with a noun phrase that specifies the place (see Example 4). Likewise, the time frame element is defined as a list of time-related lexical items including shi, dian “o’clock”, ri, yue, nian “day, month, year” etc. Based on such definitions, our three annotators consistently annotated around 95% of the data. Majority vote was used to resolve the remaining cases. Strong formal guidelines lead to high inter-annotator agreement, but do we adequately capture frame elements this way? During the annotation process two issues have emerged that we consider important practical limitations of this approach:

**Problem of scope:** We only consider frame elements that appear within the same sentence as the causative construction. This leaves out frame elements that appear in the previous sentence or paragraph and that are arguably relevant to the causation scenario. Relying on sentence boundaries as the cut-off point can be problematic given that discourse units may not align well with this punctuation mark in the formal, written Chinese we work with (Li et al. 2015; Xue and Yang 2011). But extending the scope beyond the sentence would require another established guideline that we were not able to identify in the relevant literature, so we went for the rather restrictive option of considering only inner-sentence frame elements.

**Problem of idiomaticity:** Most cases of disagreement between our annotators were related to idiomaticity. When annotating place frame elements, for instance, idiomatic uses of location markers such as shijie shang “in the world” (compared to: diqiu shang “on Earth”) caused diverging opinions on where to draw the line between literal and idiomatic uses. This is a much discussed topic in frame semantics (Fillmore et al. 1988; Kovecses and Szabco 1996; Ruppenhofer et al. 2006), which continues to pose a challenge for the consistent annotation of frame elements in practise.

Given these practical and theoretical limitations, we established the following quality control measures: (1) We defined frame elements as a set of lexical patterns as this allows consistent annotation across a large dataset; (2) We defined a boundary to what extend frame elements are taken into account with regards to the causative construction; (3) We implemented a minimum level of inter-annotator quality checking by having three independent annotators resolve disagreement via majority vote. In line with the guiding principles of the BP method, these measures aim to enable an analysis that is largely based on the consistent annotation of empirically verifiable observations.
3.3 Data analysis

The annotation results in a table that features 1002 instances of sentences with *rang*, *shi* and *ling* (rows) and their 38 features (columns). We convert these 38 features into 103 binary features. The resulting $1002 \times 103$ dataframe is high-dimensional for its size, we count 937 unique data entries. Next, we converted the dataframe into a distance matrix (a co-occurrence table). This table features the relative co-occurrences of each instance of *rang*, *shi* and *ling* with each of the annotated features. The relative co-occurrences of all contextual features, in sum, constitute the behavioural profile (BP) vector of each instance of *rang*, *shi* and *ling*. The BPs of *rang*, *shi* and *ling* were then compared via two multivariate statistical techniques.

3.3.1 Comparative analysis of BP vectors

First, we applied a multidimensional scaling solution to the dataset that plots the BP vectors in two-dimensional space. This allows for a closer look at how *rang*, *shi* and *ling* vectors compare to each other in our dataset.

Second, we performed a hierarchical agglomerative cluster (HAC) analysis to further explore the (dis)similarity between *rang*, *shi* and *ling* clusters. Specifically, the dataset was manually split into separate clusters for *rang*, *shi* and *ling*, respectively, to perform a comparative cluster analysis. The results are presented in form of a rootless hierarchical dendrogram that visualizes (dis)similarity as Euclidean distance between clusters in two-dimensional space.

Lastly, we present the results of this HAC in form of two snake plots that visualizes the ranked total effect size of individual contextual features to the overall cluster solution. This allows for a more detailed discussion of individual contextual features and their contribution to cluster overall (dis)similarity.

The data processing was done using R and Python, drawing on existing code by (Levshina 2015; Mancell & Deutsch 2019). To verify our results, we validate our cluster solution by computing its multi-scale bootstrap p-values (Approximately Unbiased [AU] and Bootstrap Probability [BP]), and provide alternative effect size measures for each feature in form of Cramer’s V test.

4 Results: Profiling *rang* (讓), *shi* (使) and *ling* (令)

This section presents the results of a comparative analysis of the behavioural profiles of *rang* (讓), *shi* (使) and *ling* (令). First, we show the results of a
multidimensional scaling (MDS) solution that allows preliminary observations by visualizing the (dis)similarity of the contextual features in relation to each of the three near-synonyms based on their distance matrix. Then we present the results of a HAC analysis that further explores the roles that specific contextual features play in distinguishing rang, shi and ling. Specifically, we show which contextual features in our cluster solution have the greatest effect on distinguishing the rang, shi and ling clusters, respectively, based on their absolute effect size. We conclude with discussing the theoretical implications of our findings.

4.1 General observations

As a first look at the co-occurrence data that were computed from our dataset we present a multidimensional scaling solution plotted in two-dimensional space, colour-coded for rang, shi and ling (Figure 2). The plot enables a rough estimate of around 80% of the variation in the dataset (stress 0.21%), plotted on two dimensions (MDS1 and MDS2). The instances of ling (in blue) in our dataset appear to populate the upper area of the plot, while rang and shi (in red and green) are spread across larger areas. This indicates a relatively distinct behaviour of ling in contrast to rang and shi. The plot also shows that shi (green) clusters around two areas in the lower part of the plot while rang (red) appears more evenly distributed, which indicates that behavioural differences also exist between rang and shi. The remainder of this section further explores these differences.

Next, we applied HAC analysis to further explore (dis)similarities between rang, ling and shi. Interested in a comparative analysis of the three near-synonyms,
we manually divided the dataset according to which of the three constructions is used and in which sentence structure the construction occurs. First, we split the dataset into three parts, based on whether *rang*, *shi* or *ling* is used. Second, we further divided these three parts into two groups each, based on whether *rang*, *shi* and *ling* follow a single noun phrase as the causer or not (see Example [1] and [3] for examples of this distinction). A sentence structure featuring a single noun-phrase as the causer is called NP-*rang*-NP (and NP-*shi* NP or NP-*ling*-NP). Cases where the causer slot is occupied by larger structures are referred to as X-*rang*-NP (or X-*shi*-NP, X-*ling*-NP). This means that our dataset was split into a total number of six sections, based on the occurrence of *rang*, *shi* and *ling* in one of two sentence structures. This split enables us to investigate both the behavioural differences between the three items, as well as differences between their occurrences in two sentence structures. This way, we also investigate to what extent sentence structure impacts the use of the three items. The behavioural profile vectors of all six sections were then compared via HAC analysis using Euclidean distance and the Ward hierarchical clustering method.

Figure 3 shows the results of the HAC analysis in form of a rootless hierarchical cluster tree. The distance between each of the six clusters represents the degree of

![Rootless hierarchical dendrogram of *rang*, *shi* and *ling* (distance = euclidean, method = ward).](image)
distinction within the dataset. The HAC analysis confirms our previous observation that ling behaves relatively distinctive from both other near-synonyms, and that this holds for instances featuring NP-only causes and instances of other causes. Rang and shi in both sentence structure configurations also form distinctive and reasonably stable clusters (based on the results of a cluster stability analysis using multiscale bootstrap p-values). Based on these results, the cluster solution is further evaluated by first comparing the ling cluster with both the rang and shi cluster (Section 4.2), and then comparing the less distinct rang and shi clusters to each other (Section 4.3) (see Figure 4).

4.2 Behavioural profile of ling (令) versus rang (讓) and shi (使)

This section compares the ling cluster (blue) to both other clusters (pink) in our dataset (Figure 4). We are interested in which contextual features set these two clusters apart, i.e. the effect size of their contextual features. The analysis was done in two steps. First, we computed the absolute differences using the co-occurrence table and identified the 10 features that have the largest effect on distinguishing the two clusters. In order to verify our results, we also computed and compared the effect size of each of these contextual features using Cramer’s V.

4.2.1 Cluster comparison

Figure 5 shows a so-called snake plot that shows a list of contextual features ranked by their total effect size (Divjak and Gries 2009). The features at the very bottom and top of the plot are those with the largest impact on distinguishing both

Figure 4: Steps of cluster analysis: (1) ling (blue) versus rest (pink); (2) rang (red) versus shi (green).
clusters. We are interested in the distinctive features of the ling cluster (the pink end of the snake plot).

Table 2 shows an overview of the top 10 contextual features ranked by total effect size that characterize the ling cluster. Each feature in the list tends to co-occur with ling, but not with rang and shi. In other words, ling attracts these features, while rang and shi repel them.

The snake plot shows that a relatively low number of contextual features are doing most of the work of pulling the clusters apart. The most distinctive feature of...
**Table 2:** Overview of top 10 contextual features that distinguish the ling (令) cluster from rang and shi clusters; based on total effect size and Cramer’s V test.

<table>
<thead>
<tr>
<th>Contextual feature type</th>
<th>Contextual feature</th>
<th>Effect size rank (total)</th>
<th>Cramer’s V</th>
</tr>
</thead>
<tbody>
<tr>
<td>ling (令) cluster</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Causee: Verb class</td>
<td>verb of emotion</td>
<td>1 (0.5798)</td>
<td>0.644</td>
</tr>
<tr>
<td>Causee: Transitivity</td>
<td>intransitive</td>
<td>2 (0.4999)</td>
<td>0.467</td>
</tr>
<tr>
<td>Affectee: Semantic type</td>
<td>no noun phrase</td>
<td>3 (0.4999)</td>
<td>0.463</td>
</tr>
<tr>
<td>Causee: Pronoun</td>
<td>ling ren (令人)</td>
<td>4 (0.4054)</td>
<td>0.364</td>
</tr>
<tr>
<td>Causee: Part of speech</td>
<td>pronoun</td>
<td>5 (0.3517)</td>
<td>0.474</td>
</tr>
<tr>
<td>Causee: Semantic type</td>
<td>speaker/human/animate</td>
<td>6 (0.2471)</td>
<td>0.292</td>
</tr>
<tr>
<td>General: Volitionality</td>
<td>negative</td>
<td>7 (0.1626)</td>
<td>0.235</td>
</tr>
<tr>
<td>Causee: Semantic type</td>
<td>abstract/event</td>
<td>8 (0.1432)</td>
<td>0.126</td>
</tr>
<tr>
<td>Causee: Frame element</td>
<td>manner</td>
<td>9 (0.1073)</td>
<td>0.217</td>
</tr>
<tr>
<td>Causee: Frame element</td>
<td>place</td>
<td>10 (0.0765)</td>
<td>0.099</td>
</tr>
</tbody>
</table>

*ling* is the occurrence of verbs of emotion in the causee slot. Ranked second and third are two features related to the transitivity structure of verbs in the causee slot. *Ling* is characterized by intransitive verbs in this slot, from which follows that no noun phrase occurs in the affectee slot. Example (5) illustrates this pattern.

(5) 台灣人民參與政治熱忱令外國記者驚訝

_Taiwan people participate politics passion make foreign journalists astonished_

“The passion with which people in Taiwan participate in politics astonishes foreign journalists (makes foreign journalist astonished)”

Features number 4, 5 and 6 provide information regarding the noun phrase in the causee slot. *Ling* attracts pronouns in this slot, especially the quasi-pronoun *ren* (“people, someone, one”), and generally attracts noun phrases of the semantic type ‘speaker/human/animate’. See the previous Example (5) for a noun phrase of type ‘speaker/human/animate’ in the causee slot (foreign journalists), and Example (6) for the use of *ling ren* “make people/one”.

Contextual feature no 7 describes volition between causer and causee. The use of *ling* is characterized by non-volition between the event, topic or agent that occurs in the causer slot and that in the causee slot. Related to this, feature no 8 describes the semantic type of the causer slot, showing that *ling* attracts noun phrases of type ‘abstract event’ in the causer slot (in contrast to agents or human/animate noun phrases). In combination, these two features show that *ling* attracts causes that are events instead of agents (that may act volitional). Example (6)
illustrates this pattern, also featuring the previously mentioned *ling ren* “make people/one”.

(6) *如今 有了讀大學的機會又輕易放棄，難免令人覺得可惜。*

*Nowadays have PAR study university PAR chance and easily give up, inevitably make people feel regrettable*

“Nowadays having a chance to attend university and then easily giving up on it, inevitably makes one feel regrettable”

Features no 9 and 10 are non-core frame elements that characterize frame semantic aspects of the construction. *Ling* attracts manner frame elements in the causee slot, i.e. specifications of the manner in this slot, often in form of an adverbial modifier (Example 7).

(7) *特別是許多國家的救難團體，在最短的時間內趕到災區，參與第一線的搜救工作，令我們深為感激。

*Especially is many country PAR rescue team in shortest time hurry to catastrophe area participate frontline PAR search and rescue work, makes us deeply grateful*

“Especially that the rescue teams of many countries hurried to the catastrophe area in the shortest time and participated in frontline search and rescue operations, makes us deeply grateful.”

*Ling* also attracts place frame elements in the causer slot. These are specifications of a geographic or abstract location, often featuring the Mandarin location marker zai (在). See (8) for an example of this pattern.

(8) *王國昌在醫院內昏睡，令警方捏了一把冷汗。

*Wang Guo-chang LOC-MKR hospital inside lethargic sleep, makes police break out in cold sweat*

“((That)) Wang Guo-chang sleeps lethargically in the hospital, makes the police break out in cold sweat.”

In combination, the above described contextual features provide an overview of the 10 most distinctive characteristics that set *ling* apart from *rang* and *shi*. These features describe various aspects of the environment in which *ling* occurs. Notably, since the features each differ in scope and design, a certain degree of overlap and redundancy is expected. The contextual features are not strictly delineated variables that can be neatly summed up. Rather, in combination, they enable a multifactorial linguistic analysis of *ling* in contrast to *rang* and *shi* in terms of probabilistic trends, attraction and repulsion. And, crucially, an analysis that is grounded in observable contextual features that can be verified in actual corpus data.
4.2.2 Interim conclusion: Distinctive behaviour of ling (令) in comparison to shi (使) and rang (讓)

The behavioural profile shows that ling is characterized most distinctively by occupants of its causee slot, especially by a range of co-occurring noun phrases and a specific verb class: verbs of emotion. Six out of the top 10 of the most distinguishing contextual features describe this slot. In contrast, the causer slot of ling is only weakly characterized in terms of the semantic type of the causer. Frame elements play a role in characterizing both the causer and causee slot. Ling also attracts non-volitional relationships between causer and causee.

These results show that a contrastive profile of ling contains not only collocational but also lexical and frame semantic features as some of its most distinctive contextual features. Taking all features into account, causative constructions with ling appear to be rather constrained to attracting a range of noun and verb classes in the causee slot, but tend to feature a wider range of structures appearing in the causer slot, broadly characterized as events consisting of more elements than a single noun phrase. Notably, frame semantic features such as place and manner frame elements appear to play a role in characterizing the causer and causee slots, respectively. This shows that frame semantic features are a useful discriminatory factor with regards to the three near-synonyms. We conclude that the distinctive behaviour of ling (in contrast to rang and shi) is driven by a few contextual features of the causee slot with strong discriminatory effect (the ling cluster contains four features with effect size 0.4 and above, see Table 2 for details). The Section 4.3 compares rang and shi.

4.3 Behavioural profile of rang (讓) versus shi (使)

This section compares the rang cluster (red) and the shi clusters (green), examining which contextual features set rang and shi apart (see Figure 4). We examined the total effect size of each contextual feature and measured effect size using the Cramer’s V test. We also validated the cluster solution using multi-scale bootstrap resampling to ensure that our solution is stable enough to permit an effect size analysis.2

---

2 Cluster validation using multi-scale bootstrap resampling consistently yielded p-values (AU and BP) between 0.95 and 1 for both the rang and shi cluster (see Figure 2). This indicates that the cluster solution is well-supported in the dataset.
4.3.1 Cluster comparison

Figure 6 shows a snake plot of the contextual features ranked by their total effect size. The features that most strongly discriminate *rang* and *shi* are located at both ends of the plot, coloured in red and green, respectively. On both ends, the total effect size is lower than in the previous analysis focussing on the *ling* cluster, indicating that *rang* and *shi* generally distinguish less clearly in the dataset.

Table 3 shows an overview of the top 10 distinguishing contextual features for both *rang* and *shi*. This time we are interested in both clusters and discuss the top contextual features of *rang* first, followed by a similar analysis for *shi*. In the *rang* cluster, the feature with the largest effect size by some margin is the semantic type of the noun phrase in the causee slot. *Rang* attracts human/animate noun phrases as effects. This is especially interesting when taking into account that the second most distinct feature are human/animate noun phrases in the causer slot. Thus, *rang* is characterized by human/animate causer and causee slots (feature nos 1 and 2), and the causative relationship is characterized by non-volitionality (no 3). See Example (9) for an illustration of this pattern.

(9) 他 現 在 已 放 心 不 下 她 ， 不能 讓 她 獨 自 回 家 。
*He now already feel relieved-NEG her, NEG-can allow her alone go home*

“He now already can’t stop worrying about her, can’t allow her to go home alone.”

Six out of the remaining seven features further describe the causee slot. *Rang* attracts pronouns (no 4), especially *rang* wo (讓我) “to make/let me” and *rang* ren (讓人) “to make/let people/someone” (no 8 and 9). *Rang* also attracts intransitive verbs (no 5 and 6) as well as verbs of cognition (no 10). Example (10) illustrates some of these patterns, showing an instance of *rang* with a pronoun in the causee slot followed by an (intransitive) verb of cognition. As the only modifier of *rang* and *shi*, *rang* attracts polarity modifiers such as *bu* (不) or *wei* (未) (see Example 9).

(10) 雖 然 我 不 是 同 性 愛 者 ， 但 這 樣 想 ， 可 以 讓 我 舒 服 一 點 。
*Although I NEG-be homosexual, but like this think, can make me comfortable little bit*

“Although I am not gay, but to think like this, can make me a little more comfortable.”

Now examining the *shi* cluster, many contextual features of the same type also appear as top distinguishing features again. This is not surprising given that the two clusters may well fall on opposing ends of dichotomous features such as volitionality or modifier occurrence. Like in the *rang* cluster, *shi* also features the semantic type of noun phrases in both the causer and causee slots as the most
distinguishing features. However, in contrast to human/animate causes and effects, *shi* attracts noun phrases classified as ‘abstract events’ in both slots.

Feature nos 3, 4 and 8 further characterize the causee slot as attracting the verb classes ‘verbs of change/change-of-state’ and ‘verbs of being/states’ as well as repelling pronouns. Example (10) shows an instance that exemplifies this pattern, here with a verb of change-of-state.

**Figure 6:** Overview of contextual variables in *rang* and *shi* clusters, ranked by effect size.
Table 3: Overview of top 10 contextual variables that distinguish the *rang* cluster from the *shi* cluster; based on total effect size and Cramer’s V test.

<table>
<thead>
<tr>
<th>Contextual feature type</th>
<th>Contextual feature</th>
<th>Effect size rank (total)</th>
<th>Cramer’s V</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>rang</em> (讓) cluster</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Causee: Semantic type</td>
<td>speaker/human/animate</td>
<td>1 (0.3455)</td>
<td>0.378</td>
</tr>
<tr>
<td>Causeer: Semantic type</td>
<td>speaker/human/animate</td>
<td>2 (0.2008)</td>
<td>0.206</td>
</tr>
<tr>
<td>General: Volitionality</td>
<td>negative</td>
<td>3 (0.1900)</td>
<td>0.259</td>
</tr>
<tr>
<td>Causee: Noun phrase</td>
<td>pronoun</td>
<td>4 (0.1717)</td>
<td>0.207</td>
</tr>
<tr>
<td>Causee: Transitivity</td>
<td>intransitive</td>
<td>5 (0.1142)</td>
<td>0.112</td>
</tr>
<tr>
<td>Affectee: Semantic type</td>
<td>N/A (no noun phrase)</td>
<td>6 (0.1142)</td>
<td>0.144</td>
</tr>
<tr>
<td>Modifier of rang/shi/ling</td>
<td>polarity modifier</td>
<td>7 (0.0963)</td>
<td>0.126</td>
</tr>
<tr>
<td>Causee: Pronoun</td>
<td>rang ren (讓人)</td>
<td>8 (0.0757)</td>
<td>0.251</td>
</tr>
<tr>
<td>Causee: Pronoun</td>
<td>rang wo (讓我)</td>
<td>9 (0.0738)</td>
<td>0.251</td>
</tr>
<tr>
<td>Causee: Verb class</td>
<td>verb of cognition</td>
<td>10 (0.0691)</td>
<td>0.402</td>
</tr>
<tr>
<td><em>shi</em> (使) cluster</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Causee: Semantic type</td>
<td>abstract/event</td>
<td>1 (−0.2430)</td>
<td>0.206</td>
</tr>
<tr>
<td>Causeer: Semantic type</td>
<td>abstract/event</td>
<td>2 (−0.2178)</td>
<td>0.378</td>
</tr>
<tr>
<td>Causee: Verb class</td>
<td>verb of change-of-state</td>
<td>3 (−0.2062)</td>
<td>0.402</td>
</tr>
<tr>
<td>Causee: Noun phrase</td>
<td>no pronoun</td>
<td>4 (−0.2045)</td>
<td>0.251</td>
</tr>
<tr>
<td>General: Volitionality</td>
<td>positive</td>
<td>5 (−0.1900)</td>
<td>0.259</td>
</tr>
<tr>
<td>Affectee: Semantic type</td>
<td>abstract/event</td>
<td>6 (−0.1621)</td>
<td>0.144</td>
</tr>
<tr>
<td>Causee: Transitivity</td>
<td>monotransitive</td>
<td>7 (−0.1246)</td>
<td>0.112</td>
</tr>
<tr>
<td>Causee: Verb class</td>
<td>verb of being/state</td>
<td>8 (−0.0995)</td>
<td>0.402</td>
</tr>
<tr>
<td>Modifier of rang/shi/ling</td>
<td>no polarity modifier</td>
<td>9 (−0.0963)</td>
<td>0.126</td>
</tr>
<tr>
<td>General: Sentence mood</td>
<td>declarative</td>
<td>10 (−0.0821)</td>
<td>0.177</td>
</tr>
</tbody>
</table>

(11) 亞洲電影業和時裝業隨意宣傳吸煙，將使吸煙成為亞洲的第一號殺手。

*Asia cinema scene and fashion scene careless promote smoking,*

*TENSE-MKR make smoking become Asia PAR first number killer*

“Asian cinema and fashion scenes carelessly promote smoking, ((which)) will make smoking become Asia’s No. 1 Killer.”

The *shi* cluster is characterized by a volitional relationship between causer and causee (no 5) as well as monotransitive verbs in the causee slot (no 6), attracting noun phrases of ‘abstract events’ in the affectee slot (no 7). Example (11) illustrates these patterns.

(12) 新的撥款使香港抗沙斯的撥款總額自3月以來超過18億元。

*New PAR fund allocation, makes Hong Kong fight SARS PAR fund allocation total amount from 3 month PAR exceed 18 billion currency*
“New fund allocation makes Hong Kong’s fund allocation to fight SARS exceed 18 billion Yuan since March.”

Furthermore, in contrast to rang, the shi cluster repels polarity modifiers (no 9) and tends to feature a declarative mood (no 10), i.e. repel imperative and interrogative sentence moods.

4.3.2 Interim conclusion: Distinctive behaviour of rang (讓) and shi (使)

Contrasting the profiles of rang and shi reveals that both causative constructions mainly distinguish in relation to the semantic types of the noun phrases that occupy their causer and causee slots. Rang attracts human/animate noun phrases in both slots while shi attracts abstract events. Other features that set rang and shi apart are volitionality of the causative relationship, attraction/repulsion of pronouns and particular verb classes including their transitivity structure in the causee slot, polarity modifiers of rang and shi, as well as sentence mood.

5 Conclusion

Using a behavioural profiling (BP) approach, this study examined the Chinese causative construction with rang (讓), shi (使) and ling (令), producing a range of insights regarding the use of these three ways to express cause-effect relationships. The study is situated alongside other fruitful BP studies of synonymy and near-synonymy, providing additional evidence that the BP approach is suitable to study fine-grained variation and partially overlapping aspects of meaning in a systematic and data-driven way (e.g. Divjak 2010; Divjak and Gries 2006).

Methodologically, this study illustrates a modified approach to profiling that, unlike many existing BP studies, incorporates frame semantic features and discusses how these features can be included in a data-driven, systematic way. In comparison to out-of-the-box BP approaches that typically only include collocational, lexical semantic and discourse level features, this study also includes contextual features (ID tags) that aim to describe frame elements related to the phenomenon under study. The study provides usage-based, quantitative evidence based on co-occurrence patterns for further cognitive linguistic theorizing related to near-synonymy, network representations of meaning and the interface between morphosyntactic, lexical semantic and frame semantic aspects of meaning variation.

The comparative analysis of the intricate behavioural differences of rang, shi and ling showed that each of the three ways of expressing cause-effect
relationships has a different behavioural footprint consisting of a specific range of collocational, lexical semantic and frame semantic features. *Ling* exhibits the most distinctive behaviour, driven by a relatively strong attraction to a small number of contextual features. *Ling* attracts event and place elements in the causee slot and animate/human causee slot fillers (see Tian and Zhang 2020 for similar observations). In more detail, the causee slot is characterized by noun phrases of the type abstract event and place frame elements. The causee slot attracts pronouns, noun phrases of the type human/animate, manner frame elements and verbs of emotion. These findings indicate that emotion verbs play a role in the behaviour of the near-synonyms. However, we leave much of this area unexplored, such as investigating the role of sentiment polarity or exploring whether different classes of emotion are attracted (Chang et al. 2000; Lee et al. 2013).

*Rang* and *shi* exhibit less distinctive behaviour but still form distinctive and stable clusters when compared with each other. The most discriminatory features that set *rang* and *shi* apart are lexical semantic. *Rang* attracts pronouns in the causee slot and noun phrases of type human/animate in both the causer and causee slot. *Shi* attracts abstract events in both slots. *Rang* and *shi* also exhibit divergent behaviour with regards to verb classes in the causee slot, transitivity, polarity modifiers, volitionality and sentence mood. *Rang* attracts intransitive verbs in the causee slot, especially verbs of cognition. *Rang* also attracts non-volitional causative relationships and more frequently collocates with polarity modifiers. *Shi*, on the other hand, attracts mono-transitive verbs in the causee slot, especially verbs of change-of-state and verbs of being or states. *Shi* also attracts volitional causative relationships and repels polarity modifiers. *Shi* attracts declarative sentence mood while *rang* appears in a wider range of different sentence moods. These observations show that each *rang*, *shi* and *ling* are characterized by distinctive behaviour and that the dissimilarities span across different types of features.

That frame semantic features appear as some of the top distinguishing features of the causative construction shows the value of including this type of feature in profiling studies on near-synonymy. It shows that frame semantic contextual features, which are not typically included in profiling studies, can provide additional insights and can play a significant role in distinguishing near-synonyms. Methodologically, the study illustrates how frame semantic contextual features can be included in accordance with the hallmark emphasis of the BP approach on empirical verification, and how issues that arise from their manual annotation can be dealt with. We show how one can bring in the, sometimes elusive, notion of frames in form of the systematic annotation of lexicalized frame elements based on existing frame semantic resources (Baker et al. 1998; Liu and Chiang 2008).
Operationalizing frames this way can contribute to a more comprehensive behavioural profile.

To sum up, the present study presents a behavioural profile of the three Chinese near-synonyms rang, shi and ling that have not previously been examined using this approach. The present study also shows how the scope and design of behavioural profiling can be expanded to include and examine frame semantic phenomena. Thus, this study, on one hand, produced corpus-based evidence with theoretical and methodological implications for further research in areas of Chinese causative constructions and near-synonymy, while also hopefully breaking a little new ground methodologically by illustrating how frame semantics and behavioural profiling can be fruitfully combined. In addition to related studies that explored Mandarin construction-frame relations (Chien and Liu 2010) and purely collocational studies on Mandarin causative constructions (Hu and Yang 2015), this study provides corpus-based evidence in unprecedented detail and scope regarding the (dis)similarity of these three ways of expressing causative relationship in Mandarin Chinese, including the relation of this type of construction to causation frames.

We hope that the present corpus study serves as a useful starting point for further explorations of Chinese causative constructions, their intricate semantics, behavioural differences and relation to different modifiers and verb classes.

**Acknowledgment:** We thank Dr. Shiwei Chen for her help with conceptualizing and preparing the dataset. We gratefully acknowledge the detailed and valuable comments of the reviewers. All remaining errors are our own.

**Appendix I:** Top collocates of rang, shi and ling in Chinese Gigaword 2.0 across three varieties of Mandarin Chinese

<table>
<thead>
<tr>
<th>Mandarin variety</th>
<th>Modifier of cxn</th>
<th>Causee slot(cxn+X)</th>
<th>Causer slot (X+cxn)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taiwan Mandarin: (Central News Agency)</td>
<td>1. 能讓</td>
<td>1. 讓人</td>
<td>1. 他讓</td>
</tr>
<tr>
<td></td>
<td>2. 也讓</td>
<td>2. 讓民眾</td>
<td>2. 政府讓</td>
</tr>
<tr>
<td></td>
<td>3. 要讓</td>
<td>3. 讓他</td>
<td>3. 我們讓</td>
</tr>
<tr>
<td></td>
<td>1. 將使</td>
<td>1. 使他</td>
<td>1. 事件使</td>
</tr>
<tr>
<td></td>
<td>2. 會使</td>
<td>2. 使人</td>
<td>2. 舉使</td>
</tr>
<tr>
<td></td>
<td>3. 能使</td>
<td>3. 使他們</td>
<td>3. 措施使</td>
</tr>
</tbody>
</table>
Appendix II: Annotation procedure

This is a list of all 38 features and their 103 variables, each with example.

Collocations of the causative constructions

Modifier: Polarity

Whether the causative construction appears with a polarity modifier: bu (不), mei (没), wu (無), wu (勿)

- 不讓 bu rang POL_MOD_RANG “not cause”
Modifier: ye (也)

Whether the causative construction appears with: ye 也 “also; as well, too”
- 也讓 ye rang YE_MOD_RANG “also cause”.

Modifier: dou (都)

Whether the causative construction appears with: dou (都) “also, just”
- 都讓 dou rang DOU_MOD_RANG “also cause”

Modifier: jiu (就)

Whether the causative construction appears with: jiu (都) “just”
- 就讓 jiu rang JIU_MOD_RANG “just cause”

Modifier: cai (才)

Whether the causative construction appears with: cai (才) “really, only, just”
- 才讓 cai rang CAI_MOD_RANG “really cause”

Modifier: Temporal/frequency

Whether the causative construction appears with a temporal (including frequency) modifier: zaici (再次), zai (再), chongxin (重新), tongshi (同時), zhengzai (正在), zai (在), you (又), zhongy (終於) chang (常), jingchang (經常)
- 經常讓 jingchang rang TEMP_MOD_RANG “often causes”

Modifier: Temporal: Past tense

Whether the causative construction appears with a past tense modifier: ceng (曾), yiqian (以前), yi (已) yijing (已經)
- 曾讓 ceng rang PAST_MOD_RANG “previously caused”

Modifier: Temporal: Future tense

Whether the causative construction appears with a future tense modifier: jiang (將), jiang hui (將會), hui (會)
Modifier: Capability and intention

Whether the causative construction appears with a modifier of capability or intention: neng (能), guyi (故意), ke (可), keyi (可以)

Modifier: Degree

Whether the causative construction appears with a modifier of degree: hao (好), hen (很), feichang (非常), tebie (特别), zhen (真), geng (更), gengjia (更加), ruci (如此), que (确), queshi (确实), shizai (实在), zuyi (足以), shi (实), bian (便)

Modifier: Contrast

Whether the causative construction appears with a modifier of contrast: er (二), erqie (而且), qishi (其实), reng (仍), hai (还), jing (竟), jingran (竟然)

Modifier: Doubt

Whether the causative construction appears with a modifier of doubt: nanyi (难以), buyiding (不一定), jihu (几乎), ying (应), bufang (不妨), keneng (可能), shifou (是否)

Modifier: Only

Whether the causative construction appears with a modifier of type “only”: jin (仅), jinjin (仅仅), zhi (只), zhihao (只好), zhineng (只能), shenzhi (甚至)

Modifier: Light verbs: Verbs of want/need

Whether the causative construction appears with a light verb of want/need: yao (要), xiang (想), xiangyao (想要)
– 想要讓 xiangyao rang WANT_MOD_RANG “want to cause”

**Slot filler and discourse level features**

**Causer slot: Noun phrase: Semantic type**

Whether the noun phrase in the causer slot is of type:

1. human/animate/speaker
   - 月梅讓 Yue Mei rang ANIM_CAUSER “Yue Mei(personal name) causes/lets”
2. organization/institution/country
   - 中國讓 zhongguo rang ORGA_CAUSER “China causes”
3. abstract event/social/phenomenon of nature
   - 地震讓 dizhen rang ABST_CAUSER “The earthquake causes”
4. inanimate object/physical/body part/geographic
   - 泰山讓 Taishan rang INANIM_CAUSER “Tai mountain causes”

**Causer slot: Verb phrase: Verb class**

Whether the verb phrase in the causer slot is of class:

1. motion/movement
   - 將孩子丟到家中 jiang haizi duidao jia zhong MOVE_CAUSER “Put the kids home”
2. cognition
   - 老人家聽不懂英語 laorenjia ting bu dong yingyu COGN_CAUSER “Older people don’t understand English”
3. communication
   - 亞洲電影業和時裝業宣傳 yazhou Dianshiye he shizhuangye suiyi xuanquan COMM_CAUSER “The Asian Television industry and clothing industry publicizes”
4. emotion
   - 爷爺不放心 yeye bu fangxin EMOT_CAUSER “Grandfather is not at ease”
5. perception (incl mental activity)
   - 我們看到這些小朋友 women kandao zhexie xiaopenyou PERC_-CAUSER “We see those kids”
6. social-interaction
   - 陳先生回答 chen xiansheng huida SOCI_CAUSER “Mr Chen replies”
(7) verbs of change-of-state
   – 中國變成 zhongguo biancheng CHAN_CAUSER “China becomes”

(8) verbs of events
   – 在阿富汗發生 zai afuhan fasheng EVEN_CAUSER “In Afghanistan happens”

**Causer slot: Verb phrase: Voice**

Whether the passive marker bei (被) appears in the causer slot:

– 洗衣被裹在一個袋子裡 yongyi bei guo zai yi ge daizi li PASS_CAUSER “swimsuit got wrapped in a bag”

**Causee slot: Noun phrase: Semantic type**

Whether the noun phrase in the causee slot is of type. Categories similar to semantic type of causer slot.

**Causee slot: Noun phrase: Co-reference**

Whether the causee slot contains reference to noun phrase of the causer slot (reiteration or pronoun).

– 中國讓它人口 zhongguo rang ta renkou PASS_CAUSER “China makes its population”

**Causee slot: Noun phrase: Pronoun**

Whether the noun phrase in the cause slot contains a pronoun, as well as type of pronoun:

(1) 1st person singular or plural
   – 讓我 rang wo PRON_EFFECT “cause me”
   – 讓我們 rang women PRON_EFFECT “cause us”

(2) 2nd person singular or plural
   – 讓你 rang ni PRON_EFFECT “cause you”
   – 讓你們 rang nimen PRON_EFFECT “cause you all”

(3) 3rd person singular or plural
   – 讓他 rang ta PRON_EFFECT “cause him”
   – 讓她 rang ta PRON_EFFECT “cause her”
   – 讓它 rang ta PRON_EFFECT “cause it”
Profile the Chinese causative construction

- 讓 他們 rang tamen PRON_EFFECT “cause them”
- 讓 她們 rang tamen PRON_EFFECT “cause them”
- 讓 它們 rang tamen PRON_EFFECT “cause them”

(4)
- 讓 人 rang ren PRON_EFFECT “cause one; cause people”

(5) No pronoun occurs (residual category)

**Causee slot: Verb phrase: Verb class**

Whether the verb phrase in the causer slot is of class. Categories similar to verb class of causer slot.

**Causee slot: Verb phrase: Voice**

Whether the passive marker bei (被) appears in the causee slot. Similar to passive marker in causer slot.

**Causee slot: Verb phrase: Transitivity structure**

Whether the verb phrase in the causee slot is monotransitive, ditransitive or complex transitive:

(1) monotransitive verb phrase in causee slot
   - 讓 阿奇 心慌 rang Aqi xinhuang MOTR_EFFECT_VP “make A-qi panic”

(2) ditransitive verb phrase in causee slot
   - 讓 孩子 產生 確認 rang haizi chansheng rentong DITR_EFFECT_VP “cause kids to develop a sense of belonging”

(3) complex transitive in causee slot
   - 讓 阿奇 送 她 回家 rang Aji song ta huijia COTR_EFFECT_VP “make/let A-ji bring her home.”

**Causee slot: Verb phrase: Affected noun phrase: Semantic type (if di- or complex transitive)**

The semantic type of the affected noun phrase in case the causee slot verb phrase is ditransitive. Like noun phrases in causer and causee slot, categorized into the same four semantic types.
**Sentence: Clause structure**

Whether causer and causee slots are part of one clause or span over more than one clause:

(1) causer and causee slots appear in one clause (including use of *de* phrases)
   - 月梅 讓 阿奇 送 她 回家 *Yuemei rang Aji song ta huijia* INCL_CAUSE_EFFECT “Yue-mei makes/lets A-ji bring her home.”
   - 那 笑容 俊朗 得 讓 人 无 法 呼吸 *na xiaorong junlang de rang ren wufa huxi* INCL_CAUSE_EFFECT “This smile is handsome that it causes one to be unable to breathe”

(2) causer and causee slots span over more than one clause (including pre-positioned clauses)
   - 阳光 出 奇 的 亮，每 一 寸 土地 被 投射 得 亮晃晃 的，让人 睜 不 开 眼。 *yangguang chuqi de liang, mei yi cun tudi bei toushe de lianghuanghuang de, rang ren zheng bu kai yan* CRCL_CAUSE_EFFECT “The sunshine is surprisingly bright, every inch of the land is projected brightly, making people unable to open their eyes.”
   - 令 我 吃 惊 的 是， 他 们 工 作 起 来 简 直 不 要 命 *ling wo chijing de shi, tamen gongzuo qilai jianzhi bu yaoming* CRCL_CAUSE_EFFECT “What causes surprise to me is that they simply work like crazy”

**Sentence: Volitionality**

Whether the causative construction is used to express a volitional or non-volitional causation scenario:

(1) volitional
   - 美國 人 為 什 麼 不 讓 松 鼠 登 堂 入室 成 為 一 種 寵 物 *meiguoren weishenme bu rang songshu dengtangrushi chengwei yi zhong chongwu* VOLI_CAUSE_EFFECT “Why Americans don’t allow squirrels to come inside and become a kind of pet”

(2) non-volitional
   - 不 標 准 的 客 語， 就 讓 許 多 老 人 家 聽 了 很 難 過 *bu biaozhu de keyu, jiu rang xuduo laorenjia ting le hen nanguo* COTR_EFFECT_VP “inaccurate Hokkien just causes many older people to listen very unhappily”
**Sentence: Possession**

Whether change of possession occurs between causer and causee slot fillers, i.e. whether causer or causee fillers gain or lose something:

- 讓 學習 有 困難 的 學生 提早 獲得 幫助 *rang xuexi you kunnan de xuesheng tizao huode bangzhu* **POSS_CAUSE_EFFECT** “allow students with learning problems to get help earlier.”

**Sentence: Mood**

Whether the sentence is of declarative (indicative), imperative or interrogative mood:

1. declarative (indicative)
   - 月梅 讓 阿奇 送 她 回家 *Yuemei rang Aji song ta huijia* **DECL_MOOD** “Yue-mei makes/lets A-ji bring her home.”

2. imperative
   - 不能 讓 他 活著！ *buneng rang ta huozhe* **IMPE_MOOD** “Cannot let him live!”

3. interrogative
   - 納粹式 政黨 而 讓 國家 覆亡？ *nacuishi zhengdang er rang guojia fuwang* **INTE_MOOD** “Does a Nazi-style ruling party cause the country to fall?”

**Frame elements**

**Causer slot: Place; Causee slot: Place**

Whether the causer slot contains a frame element that specifies the location of the causation scenario. Defined as the use of a location marker: *zai (在) “at”, zhong (中) “in, at”, shang (上) “on”, xia (下) “in, under”, nei (内) “in”, li (里) “in”, wai (外) “out, outside of”:

- 昨晚 9 時 20 分，這 名 40 多 歲 的 馬來 男 子， 疑 在 武吉班讓 信佳 路第 161 座 組屋 毆 打 一 名 20 多 歲 女 子， 使 她 鼻 子 流 血。 *zuowan 9shi 20fen, zhe ming 40duo sui de malai nanzi, yi zai wujibanrang xinjialu di 161 zuo zuwu ouda yi ming 20duo sui nvzi, shi ta bizi liuxue* **PLAC_FE** “Yesterday evening 9:20 o’clock, at Bukit Panjang Senja Road No 161 public housing block, this more than 40 years-old Malay male is suspected to haven beaten a more than 20 years-old woman, causing her nose to bleed”
**Causer slot: Time; Causee slot: Time**

Whether the causer slot contains a frame element that specifies the time of the causation scenario. Defined as the use of a time marker: shi (時) and dianzhong (點鐘) “o’clock”, ri (日) “day”, yue (月) “month”, nia (年) “year”, zhengzai (正在) and zai (在) “now”, zuo (昨) “yesterday”, qiantian (前天) “two days ago”, zhou (周) and xingqi (星期) and libai (禮拜) “week”:

From the example sentence of “Frame element: place”:

- 昨晚 9 時 2 0 分 zuowan 9shi 20fen TIME_FE “Yesterday evening 9:20 o’clock”

**Causer slot: Frequency; Causee slot: Frequency**

Whether the causer slot contains a frame element that specifies the frequency of the causation scenario. NOTE: This does not contain frequency modifiers of the causative construction itself. These are examined as part of the feature: modifier: frequency. Here we examine the use of frequency modifiers in the slots. Our dataset contains the following: quantifiers: ci (次), lun (輪), 遍; modifiers: zai (再) “again” and chongxin (重新) and you (又) “again, anew”, mei (每) “every”, chongxin (重新), tongshi (同時) “simultaneously”, zhongyu (終於) “finally”, chang (常) and jing-chang (經常) “often”, ouer (偶爾) “seldom”

Example:

- 希望大家能夠好好利用公園裡的設施，常到公園來做運動，使身體健康。 xiwang dajia nengguo haohao liyong gongyuan li de sheshi, chang dao gongyuan zuo yundong, shi shenti jiankang FREQ_FE “hope that everybody can make good use of the equipment in the park, and come to the park often to do sport, making the body healthy”

**Causer slot: Manner; causee slot: Manner**

Whether the causer slot contains a frame element that specifies the manner of the causation scenario. Following the Berkeley framenet description of this frame element, this is “any holistic description of the event, including overall depictions and descriptions pertaining to the Actor’s influence on the character of the event (eagerly, quietly)”. We capture this frame element through the use of adverbs in the slots. NOTE: Our dataset contains no instance of this pattern in the causee slot, producing an empty category.
Example:

- 這件婚紗完全用手工製作，這令其價值更為高昂。zhe jian hunsha wanguan yong shougong zhiwo, zhe ling qi jiazhi gengjia gaoang MANN_FE “This wedding dress is entirely made by hand, which causes its prize to be even higher”

Causer slot: Means; Causese slot: Means

Whether the causer slot contains a frame element that specifies a means related to the causation scenario. We examine this through the use of verbs of using and utilisation: yong (用), yunyong (運用), liyong (利用), jiyou (藉由), jizhe (藉著); verbs of means and making: touguo (透過), jiezhu (借助), zuo (做), tongguo (通過), dailing (帶領) and verbs of grabbing, holding or taking: na (拿)

Example:

- 波士頓市用鞭炮嚇走鴿子，卻發現那徒然浪費錢，只能讓鴿子離開一會兒 boxidunshi yong bianpao xiazou gezi, que faxian na turan langwei qian, zhi neng rang gezi likai yihuier MEAN_FE “Boston uses firecracker to scare away doves, but discovered that this only wastes money, only making the doves leave for a while”

References


on Applying NLP Tools to Similar Languages, Varieties and Dialects, Dublin, Ireland, 23 August, 1–10.


Sandra, Dominiek & Sally Rice. 1995. Network analyses of prepositional meaning: Mirroring whose mind—the linguist’s or the language user’s?" *Cognitive Linguistics* 6(1). 89–130.


