UC Santa Barbara

Title

Transfer five ways: applications of multiple distinctive collexeme analysis to the dative alternation in Mandarin Chinese

Permalink

https://escholarship.org/uc/item/2zn4075v

Journal

Corpus Linguistics and Linguistic Theory, 0(0)

ISSN

1613-7027

Authors

Liao, Shengyu

Gries, Stefan Th

Wulff, Stefanie

Publication Date

2024

DOI

10.1515/cllt-2024-0033

Peer reviewed

Article

Shengyu Liao*, Stefan Th. Gries and Stefanie Wulff

Transfer five ways: applications of multiple distinctive collexeme analysis to the dative alternation in Mandarin Chinese

Abstract: The dative alternation has been extensively studied in the world's languages, and the meanings of the verbs participating in the alternation have been shown to play a key role in determining its argument realization options. The present paper presents a multiple distinctive collexeme analysis approach to the dative alternation in Mandarin Chinese, which involves a choice of one of five functionally similar alternants, and it does so by also discussing several ways to improve how this has been done statistically in most previous analyses. Linguistically, we identify the core semantic differences of the five constructions based on which verbs statistically prefer to occur in which pattern, focusing on semantic potential and direction of transfer. Methodologically, this study contributes to the slowly growing body of studies that use collexeme strengths that are not only less related to frequency than the traditional methods (i.e., association is measured in a less diluted way) and that are directional (i.e., we can focus on one direction of association from the verb to the construction).

Keywords: multiple distinctive collexeme analysis; association; frequency; dative alternation; Mandarin Chinese

1 Introduction

When we survey existing research using collostructional methods, two trends seem quite obvious: First, while the family of collostructional analyses includes simple collexeme analysis (CA), distinctive collexeme analysis (DCA), multiple distinctive

*Corresponding author: Shengyu Liao, University of Florida, Gainesville, USA,
E-mail: shengyuliao@foxmail.com
Stefan Th. Gries, University of California, Santa Barbara, Santa Barbara, USA; and Justus-Liebig-
Universität Giessen, Giessen, Germany, E-mail: stgries@linguistics.ucsb.edu
Stefanie Wulff, University of Florida, Gainesville, USA, E-mail: swulff@ufl.edu

collexeme analysis (MDCA), and co-varying collexeme analysis, the vast majority of studies have employed the first two of the above rather than the last two. There are likely several reasons for this: First, there might have been more interest in the kinds of questions CA and DCA are tailored to address. A second reason could be that they are easier to compute than MDCAs and co-varying collexeme analysis; and relatedly, their output is more straightforward to interpret. Taking these observations as a starting point, the present paper seeks to make both a linguistic/conceptual and a methodological contribution. As for the linguistic contribution, we offer an analysis of the dative alternation in Mandarin Chinese, a language that, at least compared to English, has been studied much less especially from a quantitative corpus-based perspective. Interestingly and unlike in English, where we usually either study double-object constructions (the ditransitive) on their own or contrast them with the prepositional dative construction (usually with *to*, rarely also with the benefactive *for*-construction), Mandarin Chinese has five constructional variants to choose from. That means, we need to use the rather rarely employed method of MDCA, which targets alternations of more than two variants. The methodological contribution of this paper, then, is to use the opportunity to compare three recently proposed alternatives to calculating MDCA with the original method. The three newer alternatives are much less cumbersome computationally and the results that they produce are much less difficult to understand. We walk the reader through a systematic comparison that highlights overlapping and differing results between all four ways of computation with the aim that readers may choose the computational method that best fits their data and research goals.

This paper is structured as follows. After a brief summary of how MDCA has been applied thus far, we turn to the dative alternation in Chinese specifically. We then explain how we retrieved and processed the Chinese data, and give brief explanations of how a MDCA can be computed in one of four ways, contrasting the traditional way and the three newer alternatives. In Section 3, we walk the reader through the results we obtained using all four methods, before we discuss implications, both methodological and linguistic, in Section 4.

1.1 Previous applications of MDCA

As mentioned above, MDCA is one member in the family of collostructional analyses first developed by Stefanowitsch and Gries (Gries and Stefanowitsch 2004a, 2004b; Stefanowitsch and Gries 2003, 2005). All types of collostructional analyses determine the strength of association between two kinds of constructions, most typically a lexical and a syntactic construction (although the methods can, and have been, applied also to determine lexeme-lexeme, morpheme-lexeme, lexeme-variety, and

other association strengths). (M)DCA, in particular, "looks at how much constructions (usually words/lemmas) (dis)like to occur in a slot of a usually more schematic construction as opposed to alternative, functionally similar constructions" (Gries 2023: 352). In other words, (M)DCA is used to zoom in on a variable slot in a limited number of functionally very similar constructions and typically asks the questions whether the word types that occur in a variable slot shared by the alternatives show any bias towards one constructional variant and what that might mean for how the constructions differ in terms of their functions (usually, their semantics).

DCA contrasts two constructional variants (or alternations) with each other, and it has been applied extensively. MDCA is a conceptual extension of DCA, allowing the comparison of three or more functionally similar constructions (e.g., Gilquin 2006, 2010, 2013). Compared to the number of studies that have applied DCA, the number of studies using MDCA is much smaller, but MDCA has arguably been applied to a wider range of phenomena, which we take to be indicative of its usefulness. Most studies using MDCA focused on differences between a set of near-synonymous syntactic or argument structure constructions (Béchet 2015; Gilquin 2006, 2010, 2013; Schönefeld 2013a, 2015; Su and Chen 2019). For example, Gilquin (2006) uses MDCA to uncover the differences in meaning between 10 English periphrastic causative constructions with regard to the verbs that are distinctively associated with the non-finite verb slot of each construction. Less typically, MDCA has also been applied to the comparison of near-synonymous words (Desaguliler 2014; Rajeg 2014, 2019, 2020) and morphemes (Van Goethem and Norde 2020).

As the above examples illustrate, the most typical application of MDCA is to compare three or more constructions and identify word types in a variable slot in these constructions that aid in identifying systematic semantic and/or functional differences between these multiple constructions. However, the 'multiple' part of MDCA has also been reconceptualized such that rather than looking at word types in multiple constructions, researchers have examined word types or constructions in multiple registers or in multiple time periods. For instance, Hilpert (2006) applies MDCA to the diachronic study of the infinitive complements of the English auxiliary *shall*. This method, which Hilpert (2008) refers to as 'diachronic distinctive collexeme analysis', has been widely used to track changes in the preferred collocates of individual constructions across multiple time periods (Fonteyn and Hartmann 2016; Hilpert 2006, 2008, 2012a, 2012b; Rens 2017; Zhou 2023). These shifting collocational preferences can be interpreted as changes in the meaning/function of the construction (Hilpert 2006, 2008). Yet another application of MDCA is to identify the register-specific behavior of a given construction (Schönefeld 2013b; Zhou 2023). For example, Schönefeld (2013b) uses MDCA to identify what types of co-occurring elements are preferred by the English *go un*-V-*en* construction in the four registers (academic prose, newspaper texts, fiction and conversation) in the British National

Corpus. Less typically, MDCA has also been applied to determine the lexeme-morpheme (Wulff et al. 2009) and lexeme-variety (Mukherjee and Gries 2009) association strengths.

As mentioned above, existing applications of (M)DCA reveals that most studies targeted Indo-European languages (one of those most targeted being English) – only few studies exist that examine non-Indo-European languages such as Chinese (Su and Chen 2019) and Indonesian (Rajeg 2014, 2019, 2020). The present paper, however, is concerned with an alternation phenomenon in a language that has been studied collostructionally much less: dative alternation in Chinese, to which we turn now.

1.2 Dative alternation in Mandarin Chinese

Argument structure alternations refer to the phenomenon in which pairs/sets of formally distinct argument structure constructions (called the *variants* of the alternation) share substantial aspects of their semantics such that they can accommodate a common set of lexical items to fulfill similar functions (Perek 2015: 145–146). The dative alternation is a classic example of argument structure alternation in the world's languages. In English, specifically, the double-object construction (the ditransitive) (*Stefan gave Shengyu the cookie*) alternates with the prepositional *to*-dative construction (*Stefan gave the cookie to Shengyu*).

While extensively studied in English and other Germanic/Western languages (e.g., Ambridge et al. 2012; Bresnan and Ford 2010; Bresnan et al. 2007; Colleman 2009; Colleman and Bernolet 2012; de Marneffe et al. 2012; De Vaere 2020; De Vaere et al. 2021; Gries 2003; Gries and Stefanowitsch 2004a; Gropen et al. 1989; Reali 2017; Theijssen 2012; Valdeson 2021), comparatively little is known about the status of dative alternation in Mandarin Chinese. According to previous studies, there are at least five variants of the dative alternation in Mandarin Chinese that share substantial aspects of their semantics (Chang 2014; Lu 2007; Ma 2013; Peyraube and Lü 2019):

a.   post-verbal prepositional dative construction: SBJ[1]$_{agent}$ V OBJ$_{theme}$ *gei* OBJ$_{recipient}$

我	送-了	一	本	书	给	他。
Wǒ	*sòng-le*	*yī*	*běn*	*shū*	*gěi*	*tā.*
1SG	give.as.a.present-ASP	one	CLF	book	to	3SG

'I gave a book to him as a present.'

---

1 The following abbreviations are used in the gloss: 1 = first person, 2 = second person, 3 = third person, ASP = aspect marker, BA = the *ba*-construction with a preposed object, CLF = classifier, GEN = genitive, NEG = negation, NMLZ = nominalizer, OBJ = object, PL = plural, PN = proper name, PRT = particle, SBJ = subject, SG = singular.

b.   V-*gei* double-object dative construction: SBJ$_{agent}$ V-*gei* OBJ$_{recipient}$ OBJ$_{theme}$

我	送-给	他	一	本	书。
Wǒ	*sòng-gěi*	*tā*	*yī*	*běn*	*shū.*
1SG	give.as.a.present-to	3SG	one	CLF	book

     'I gave him a book as a present.'

c.   double-object construction (the ditransitive): SBJ$_{agent}$ V OBJ$_{recipient}$ OBJ$_{theme}$

我	送-了	他	一	本	书。
Wǒ	*sòng-le*	*tā*	*yī*	*běn*	*shū.*
1SG	give.as.a.present-ASP	3SG	one	CLF	book

     'I gave him a book as a present.'

d.   *ba/jiang* dative construction: SBJ$_{agent}$ *ba/jiang* OBJ$_{theme}$ V(-*gei*) OBJ$_{recipient}$

我	把	这	本	书	送-给-了	他。
Wǒ	*bǎ*	*zhè*	*běn*	*shū*	*sòng-gěi-le*	*tā.*
1SG	BA	this	CLF	book	give.as.a.present-to-ASP	3SG

     'I gave the book to him as a present.'

e.   pre-verbal prepositional dative construction: SBJ$_{agent}$ preposition OBJ$_{recipient}$ V OBJ$_{theme}$

我	给	他	送-了	一	本	书。
Wǒ	*gěi*	*tā*	*sòng-le*	*yī*	*běn*	*shū.*
1SG	to	3SG	give.as.a.present-ASP	one	CLF	book

     'I gave him a book as a present.'

However, previous usage-based constructionist studies on dative alternation in Mandarin Chinese either included only a subset of the five constructions or did not distinguish among them at all. For example, Hsiao and Mahastuti (2020) carried out a distinctive collexeme analysis of constructions a, b, and c (they grouped constructions b and c as expressing the same construction, in contrast to construction a) and found that they differ in the number and completion of transfer events. However, according to the Principle of No Synonymy (Goldberg 1995), formally distinct constructions should also differ functionally. In other words, variants of an alternation are seen as related yet separate constructions. By contrast, Peyraube and Lü (2019) proposed that construction b and construction c differ in the types of verbs that can occur in them and the semantic roles of the recipient-like argument (recipient, source, or beneficiary) involved in the construction. Thus, it might be problematic for Hsiao and Mahastuti (2020) to group variants b and c as expressing the same construction. Zhang and Xu (2019) conducted a multifactorial corpus analysis of constructions a, b, c, and e – grouping them into patterns a and e versus b and c – and found with a binary logistic regression that the definiteness of theme, the

concreteness of theme, and the animacy of recipient are strong predictors of speakers' choice of construction. Recently, Zhang and Xu (2023) examined the factors determining speakers' choice between constructions a and c with a mixed-effects logistic regression analysis and reported that it was significantly affected by the animacy, pronominality, and definiteness of the recipient, the accessibility and concreteness of the theme, and the length difference between the theme and the recipient. Only Ma (2013) included instances of all five constructions, yet performed only a collostructional analysis (CA) on the combined data rather than running a MDCA.

In summary, it is yet unclear what distinguishes the five variants. The studies referenced above have shown that intra-lingual factors (the length, animacy, pronominality, and definiteness of the theme and the recipient, or the sense of the verb) impact speakers' choices. However, there are also several relevant extra-lingual factors that have been shown to matter in the dative alternation, such as the language variety (e.g., Bernaisch et al. 2014; Bresnan and Ford 2010; Röthlisberger 2018; Röthlisberger et al. 2017), mode/register (e.g., Buysse 2012; Theijssen 2008; Wang 2013), as well as gender and age of the speaker (e.g., Jenset et al. 2018; Lorente Sánchez 2018; Röthlisberger 2021; Theijssen 2012). Syntactic priming has also been proven to play a role (e.g., Bresnan and Ford 2010; de Marneffe et al. 2012; Gries 2005). While a comprehensive study of all these factors in the dative alternation in Mandarin Chinese is out of scope for the present study, this is the first study to identify verb-specific preferences of all five constructions.

2 Methods

In the following, we describe how we extracted a sample of the five constructional variants from corpus data before we turn to a brief description of (M)DCA and the four different ways to calculate distinctive collexemes values.

2.1 The corpus data

Our data come from the ToRCH2014 (Texts Of Recent CHinese) corpus (Xu 2017), a balanced corpus of written Mandarin Chinese, which consists of 983,651 tokenized words and 68,453 word types in 657 files. For the most part, the corpus was built following the sampling scheme of the Brown corpus (Kučera and Francis 1967), covering texts (mostly published in 2014) from 15 text categories, but there are also two differences: First, Brown's category N (western and adventure fiction) was replaced with a category of martial arts fiction, following the Lancaster Corpus of

Mandarin Chinese (LCMC, McEnery et al. 2003). Second, the sampling frame of 500 text segments of approximately 2,000 words was not strictly replicated for text files in categories A–E. There are more text files with varying length belonging to those categories in ToRCH2014 than in the Brown corpus, leading to a total of 657 files. The genres covered in ToRCH2014 are shown in Table 1.

For spoken Chinese corpus data, we decided to use the CallFriend-Mainland Mandarin corpus (Canavan and Zipperlen 1996), which consists of 43 files transcribed from casual Mandarin telephone conversations between friends, totaling 273,327 word tokens and 12,705 word types. Since only seven of the transcripts were segmented, we removed all the whitespaces between all words and then tokenized all files using the Stanford Word Segmenter[2] (Tseng et al. 2005) with the Chinese Penn Treebank word segmentation standard. Since the texts in ToRCH2014 was originally segmented using a different segmenter, the ToRCH2014 corpus was resegmented after all white spaces were removed, following the same procedure to maintain

Table 1: Genres covered in the ToRCH2014 corpus.

Category	Genre	No. of samples	No. of samples in Brown
A	News reportage	103	44
B	News editorials	79	27
C	News reviews	30	17
D	Religious writing	24	17
E	Skills/trades/hobbies	61	36
F	Popular lore	48	48
G	Biographies and essays	75	75
H	Reports/official documents	30	30
J	Academic prose	80	80
K	General fiction	29	29
L	Mystery/detective fiction	24	24
M	Science fiction	6	6
N	Martial arts fiction	29	29
P	Romantic fiction	29	29
R	Humour	10	10
	Total	657	500

consistency. After word segmentation was done, all the spoken and written texts were tagged using the Stanford Log-linear Part-Of-Speech Tagger[3] (Toutanova et al. 2003).

We adopted a verb-centered approach in the retrieval of all five constructions. A total of 354 verb candidates were identified by combining the verb lists of Mandarin ditransitive verbs studied in previous research (Hsiao and Mahastuti 2020; Liu 2006; Ma 2013; Yao and Liu 2010). The searches and retrieval are based on word form/PoS-tag combinations (c.f., Stefanowitsch 2019). Given the wide-spread cross-PoS polysemy in Mandarin Chinese, making the retrieval PoS sensitive seemed more sensible than basing the retrieval purely on word forms. For example, the word 还 can be a verb (还 *huán* 'return') or an adverb (还 *hái* 'still'), and is tagged as VV or AD accordingly (Xia 2000). There are altogether 1939 instances of 还_AD and 5 instances of 还_VV, so searching by word form/PoS-tag combinations can help us filter out those irrelevant uses.[4] Since 给 *gěi* 'to' can be attached to some ditransitive verbs in variant b and variant d, the query term is formed as follows: Verb_VA|Verb_VC|Verb_VE|Verb_VV|Verb给_VA|Verb给_VC|Verb给_VE|Verb给_VV. Specifically, the search term for 还 *huán* 'return' is 还_VA|还_VC|还_VE|还_VV|还给_VA|还给_VC|还给_VE|还给_VV. All the sentences in which the candidate verb/PoS-tag combinations occur were extracted using AntConc (Anthony 2020). Retrieving all instances of these verbs yielded 26,523 attestations in the ToRCH2014 corpus and 9,825 attestations in the CallFriend corpus for a total of 36,348 attestations.

Since (i) this initial sample is too large to comb through manually, (ii) the written and the spoken corpus are of dramatically different sizes, and (iii) a substantive number of verbs occur in either corpus only rarely (the frequencies of the verbs display the expected Zipfian distribution), we developed a sampling procedure that would ensure reasonable coverage but also minimum frequencies of occurrence that allow for the MDCAs to be applied meaningfully. We decided to set a frequency threshold such that only verb types occurring at least 5 times in each of the written and the spoken corpus would be included in the final sample, which resulted in 76 verb types (with a combined token frequency of 28,434, consisting of 18,743 in ToRCH2014 and 9,691 in the CallFriend corpus). That in turn means that these 76 verb types account for over three quarters of all the tokens of the entire data sample initially retrieved.

In order to minimize possible effects of speaker (approximated by the corpus file) and sample maximally diversely across the two corpora, the following sampling

---

3 Version 4.2.0, downloaded November 17, 2020, from https://nlp.stanford.edu/software/tagger.shtml.
4 For the Stanford Log-linear Part-Of-Speech Tagger (Version 4.2.0) used in the present study, its overall accuracy was reported to be 93.65 % and the unknown word accuracy 84.84 %. Accessed November 17, 2020, from https://nlp.stanford.edu/projects/chinese-nlp.shtml.

procedure was implemented: the complete data set was sorted by verb, then by file (with the files in random order), then by occurrence in the file (with the occurrences in random order). The first author then checked each attestation in a given file to identify a true hit (that is, an instantiation of either of the five constructions) until a true hit was found. Once a true hit was found in a file, Author 1 moved on to the next file. This process was continued until the end of the concordance was reached. A total of 1,015 instances from 50 different verb types were sampled this way. The frequency distributions of each constructional variant are as follows, ordered from the most frequent variant e to the least frequent variant a with their frequency in parentheses: construction e (584), construction c (232), construction d (123), construction b (44), and construction a (32).

2.2 Multiple distinctive collexeme analysis

Distinctive collexeme analysis (DCA, Gries and Stefanowitsch 2004a) identifies lexical items that show a strong preference for one member of a group of two or more seemingly synonymous constructions as opposed to the other(s). This method has been frequently applied in the context of syntactic alternations, such as the dative alternation. To calculate the association strength between a particular lexical item l and a member of two semantically/functionally similar constructions cx_1 and cx_2, the following frequencies are needed: the frequency of l in cx_1, the frequency of l in cx_2, the frequency of lexical items that are not l in cx_1, and the frequency of lexical items that are not l in cx_2, as shown in Table 2 (adapted from Stefanowitsch 2014).

   This contingency table can then be submitted to a contingency test that determines how much the observed distribution differs from the expected one. Typical association measures (AMs) that are used to express this difference are the Fisher-Yates exact test (p_{FYE}), G^2, or a Mutual Information (MI) score. Once this procedure has been repeated for all members of L ($l_{1...n}$) occurring in cx_1 or cx_2 in a given corpus, lexical items are ranked by their association strength. Lexical items that occur (significantly) more frequently than expected in one of the constructions are called (significantly) distinctive/differential collexemes of that construction (Stefanowitsch 2013, 2014, 2020).

Table 2: Distinctive collexeme analysis.

	Construction cx_1	Construction cx_2	Row totals
Word l	Cell a: Frequency of l in cx_1	Cell b: Frequency of l in cx_2	$a + b$: Frequency of l
Other words	Cell c: Frequency of $\neg l$ in cx_1	Cell d: Frequency of $\neg l$ in cx_2	$c + d$: Frequency of $\neg l$
Column totals	$a + c$: Frequency of cx_1	$b + d$: Frequency of cx_2	Sum total

As we discussed earlier, conceptually speaking, multiple distinctive collexeme analysis (MDCA) is a rather straightforward extension of a DCA to cases where more than two constructions are compared to one another, but mathematically speaking, the departure from the simple binary contrast requires a different statistical approach. In the following, we briefly outline how MDCAs were originally computed (exact binomial tests), and how they can alternatively be computed using multiple distinctive log odds ratios, Pearson residuals, or Kullback-Leibler divergence scores.

2.2.1 The traditional way: Exact binomial tests

As in a DCA, the first step in a MDCA is to calculate the observed and expected frequencies of each lexical item in each construction. However, rather than comparing these with Fisher–Yates exact tests, MDCAs use separate binomial tests to compute the probability of a particular observed frequency or anything more extreme given the expected frequency for each verb in each construction. As with FYE in most collostructional studies, this probability is then logged to the base of 10 and signed such that positive values indicate that the verb occurs more often than expected in the construction (that is, it is attracted to the construction), negative values indicate that the verb occurs less often than expected (that is, it is repelled by the construction). Gries (2004) wrote an R-script called Coll.analysis 3, which outputted a distinctiveness value for each verb in each construction. The higher the absolute value, the stronger the association strength (attraction or repulsion) between the verb and the construction, with a threshold level of statistical significance of 1.30103 (which corresponds to $-\log_{10}$ of $p = 0.05$). The script also provided a measure called 'SumAbsDev', the sum of all absolute distinctiveness values for a particular verb, which expresses the extent to which the verb deviates from its expected distribution (a high value indicates a large deviation). The construction with the largest deviation was indicated in the output file by 'LargDev', which can correspond to the highest degree of attraction (for positive distinctiveness values) or repulsion (for negative distinctiveness values).

2.2.2 Alternative 1: Pearson residuals

The first alternative we will discuss here is one that has been proposed in Gries (2023) and amounts to a huge simplification in terms of computational costs. This simplification is based on using Pearson residuals as an association measure. This approach just generates one complete co-occurrence table of all constructions c_{1-5} and all words l_{1-n} and then computes the Pearson residual for each cell. The

Pearson residuals are what would theoretically go into a stepwise computation of an overall chi-squared test and are computed as follows:

$$\frac{\text{obs} - \text{exp}}{\sqrt{\text{exp}}} \tag{1}$$

Because of the numerator of Equation (1), positive and negative values of the Pearson residuals indicate whether a word 'likes' or 'dislikes' to occur in a specific construction. Given that this approach avoids having to generate potentially hundreds or thousands of 2×2 tables like Table 2 above but computes just a single table on all co-occurrences, it is extremely simple to do (once the data are imported, it requires literally a single line of code, and it could theoretically even be done in just a spreadsheet software), it is extremely fast, it avoids the frequent problem of Fisher-Yates or exact binomial tests of probabilities becoming too small to compute with regular floating point operations, yet it is extremely highly correlated with the traditional but much more cumbersome methods.

2.2.3 Alternative 2: multiple distinctive log odds ratios

The second alternative to be discussed here is the log odds ratio, but in a way that generalizes the traditional method of (logged) odds ratios. Traditionally, an odds ratio is computed from a 2×2 table like Table 2 above with a focus on the a cell, which represents the relevant co-occurrence count of l and cx_1 and gets contrasted with the one other row (representing other words) and the one other column (representing cx_2). However, for our data, one would usually have to compute a 2×5 table: a certain word versus not the word against five constructions. In our generalization of the odds ratio, we essentially again compute the complete 50×5 table of all verb types against all constructions (like in the previous section), but then treat each cell of that frequency table as the a-cell of a traditional 2×2 table, determine the remaining frequencies from the rest of the big table, and collect the logged odds ratio for the combination of each construction $c_{1\text{-}5}$ and each word $l_{1\text{-}n}$ from that.

    Such log odds ratios fall into the interval $[-\infty, +\infty]$ with, as always, positive and negative values indicating the word 'likes' or 'dislikes' to occur in a specific construction. One theoretical advantage of (logged) odds ratios compared to measures such as exact binomial tests (or Fisher–Yates exact tests, log-likelihood scores, or the Pearson residuals) is that the (logged) odds ratio is an effect size measure (rather than a p-value expressing statistical significance), which means they are less 'tainted' by an extremely high correlation with co-occurrence frequency and, therefore, much better at expressing actual association (see Evert 2009; Gries 2022b).

2.2.4 Alternative 3: Kullback-Leibler divergence (*KLD*)

All of the above AMs share the characteristic that they are bidirectional: They output one single value that expresses the mutual association between a word and a construction and are, thus, unable to distinguish between the association of a word to a construction and the association of a construction to a word. Thus, bidirectional measures come with a loss of information because these two directions of associations can be differently strong. To use examples from collocational research (from Gries 2013), in collocations such as *according to* or *upside down*, the first word attracts the second one strongly, but not vice versa, whereas in collocations such as *of course* or *for instance* it is the second word that attracts the first rather than vice versa – bidirectional measures cannot see such differences.

One directional measure that has been proposed for such applications is Delta *P* (Ellis 2006; Gries 2013), a difference of conditional probabilities, but we will here follow Gries (2024) and use the Kullback-Leibler divergence (*KLD*), a measure from information theory which can return directional associations but which is also much less correlated with the observed co-occurrence frequencies. The general formula to compute *KLD* is represented here in, where *P* represents a posterior/observed distribution (e.g., the frequencies of each of 50 words in each of the five construction or the frequencies of the five constructions with each of the 50 words, depending on the direction one is interested in) and *Q* observes a prior/expected distribution (e.g., the overall frequencies of the five constructions or the 50 words, depending on the direction one is interested in).

$$\Sigma\left(P \times \log\frac{P}{Q} \right) \qquad (2)$$

Since our data involve what would be a multiple distinctive collexeme analysis – we have 50 verb lemmas and five constructional slots – our interpretation will be based on what one might call the contributions to *KLD* (by analogy to the Pearson residuals); the contributions to *KLD* are simply the individual summands in the above equation, with the posterior being the proportions of the 50 verbs over the five constructions and the prior being the proportions of the five constructions in our data. The multiple distinctive collexeme analyses were performed by means of Coll.analysis 4.0 (Gries 2022a).

3 Results

All four analyses provide quite similar results but they do differ in terms of the ranked order of these collexemes according to their association measure values. The
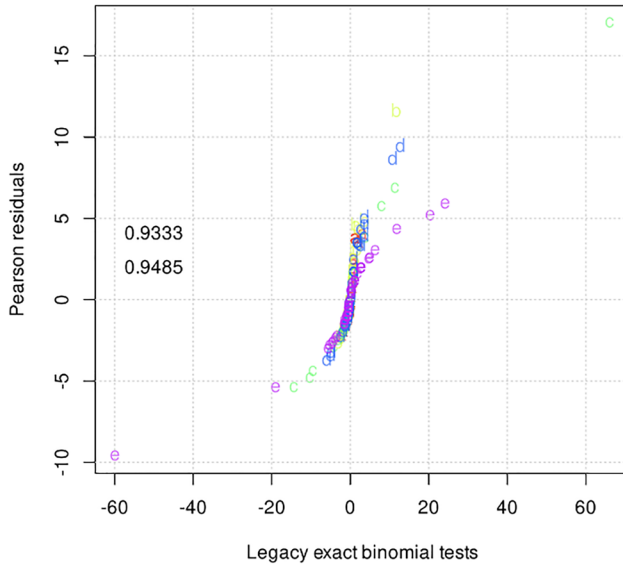
Figure 1: Comparison of the values obtained through exact binomial tests with the values obtained through Pearson residuals. The colors indicate constructional variants; the two numbers plotted into the text on the left represent the R^2s of generalized additive models predicting Pearson residuals from the binomial results (top) and the reverse (bottom).

legacy exact binomial tests and Pearson residuals provide almost identical results, which can be seen from the pairwise comparison displayed in Figure 1.

As one might expect from their formulae, both methods tend to attribute high scores to verbs that frequently occur with the respective constructional variant; in that regard, they are similar to the Fisher-Yates exact test that has been so frequently used in many collostructional studies and this conflation of frequency and association makes for a good exploratory tool. In contrast, multiple distinctive log odds and contributions to *KLD* keep frequency more out of the equation and, therefore, also permit words with low co-occurrence frequency to score fairly highly. That also means that all four methods seem to produce highly similar results when a minimum co-occurrence frequency 2 is required. Since the contributions to *KLD* are much less correlated with the observed co-occurrence frequencies – in other words, this is a cleaner/purer measure of association only – and since *KLD*-based measure is the only directional one, our interpretation will be based on the co-occurrence frequency, association as measured by the contributions to *KLD*, and the dispersion of the distinctive collexemes over the five constructions (see Olguín Martínez and Gries to appear, for a similar application), ordered from the most frequent variant e

to the least frequent variant a. The distinctive collexemes of each constructional variant, as identified by each of the four methods, can be accessed at https://osf.io/ds7zk/.

Variant e attracts the widest range and highest number of collexemes, mostly to do with short-distance physical transfer (e.g., 带来 *dàilái* 'bring.come', 留下 *liúxià* 'leave', 提供 *tígōng* 'provide', 买 *mǎi* 'buy', 要 *yào* 'ask.for', and 发 *fā* 'send.out') and short-distance verbal transfer/communication (e.g., 讲 *jiǎng* 'tell', 说 *shuō* 'say', and 介绍 *jièshào* 'introduce'). Besides, variant e also attracts 打 *dǎ* 'make.a.phone.call',[5] 寄 *jì* 'send', and 发 *fā* 'send.out', which imply some distance between the agent and recipient. See Figure 2. These verbs, which Peyraube and Lü (2019: 222) refer to as 'lexical dative' verbs, inherently carry a sense of 'transfer'.
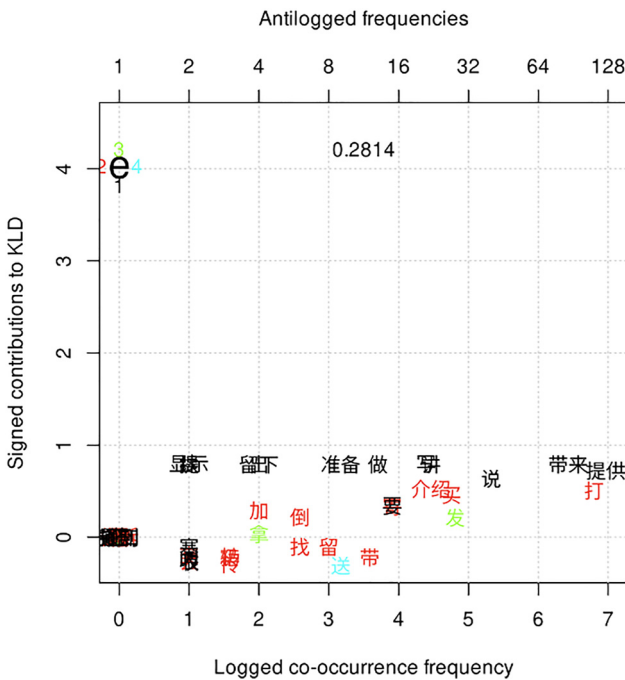

Figure 2: Distinctive collexemes of variant e by frequency and association (The numbers around the big bold letter on the left margin indicate what the color means, in this case the number of constructions a verb is attracted to; the 3- or 4-decimals number plotted into the graph is the R^2 from predicting the association measure from frequency with a generalized additive model (Gries 2022b)).

---

5 Though 打 *dǎ* 'beat/make.a.phone.call' has more than one sense, 92.13 % (117/127) of its dative use goes with the sense 'make.a.phone.call' and occurs with 电话 *diànhuà* 'phone.call'.

The two most frequent co-occurring short-distance physical transfer verbs, 带来 *dàilái* 'bring.come' and 提供 *tígōng* 'provide', however, tend to occur with an abstract object, as shown in (1) and (2).

(1) 奥林匹克   梦想   给   人们   带来   希望   和
     Àolínpǐkè *mèngxiǎng* *gěi* *rénmen* *dàilái*   *xīwàng* *hé*
     Olympic   dream   to   people   bring.come   hope   and
     力量。   (ToRCH2014: C07)[6]
     lìliàng.
     strength
     'The Olympic dream brings hope and strength to people.'

(2) 中方   愿   随时   向   韩方   提供
     Zhōngfāng *yuàn*   *suíshí*   *xiàng*   *hánfāng*   *tígōng*
     China   be.willing.to at.any.time   towards   South.Korea   provide
     必要
     bìyào
     necessary
     的   支持   和   帮助。   (ToRCH2014: A30C)
     de   *zhīchí*   *hé*   *bāngzhù.*
     GEN   support   and   assistance
     'China is willing to provide necessary support and assistance to South Korea at any time.'

Apart from these 'lexical dative' verbs, variant e also attracts a set of 'extended dative' verbs (Peyraube and Lü 2019: 222), i.e., verbs of which 'transfer' is not an intrinsic semantic component, such as 做 *zuò* 'make', 写 *xiě* 'write', 准备 *zhǔnbèi* 'prepare', 出 *chū* 'produce', 加 *jiā* 'add', 倒 *dào* 'pour', and 拿 *ná* 'take'. When 'extended dative' verbs like 写 *xiě* 'write' occur in variant e, 给 *gěi* 'to' arguably contributes more to the 'transfer' reading than the verbs do. In such case, the construction describes an event that is composed of two sub-events (Zhu 1979), as can be seen in (3): the first part in which the agent intends to complete the action of writing a letter, immediately followed by the second part in which the action of 'giving' is intended.

---

6 送 *sòng* 'give.as.a.gift' is the only verb in our study that is attested at least once in all five constructional variants and the only collexeme that is distinctive for four variants (variants a, b, c, and d). It is associated more strongly with variants b and d than with variants a and c (no matter which method we choose). Although it is not distinctively associated with variant e, it is still attested 8 times in variant e.

(3) A: 要不然   过   两   天,   我   再   给   她   写
       Yàobùrán *guò*   *liǎng*   *tiān,* *wǒ*   *zài*   *gěi*   *tā*   *xiě*
       otherwise spend   two   day   1SG   again to   3SG   write
       一   封   信.   (CALLFRIEND Mandarin-Mainland: 5975)
       yī   *fēng*   *xìn.*
       one   CLF   letter
       'Otherwise, I will write her another letter in two days.'
       Lit. 'Otherwise, I will write another letter and send it to her in two days.'

Variant e can not only describe the movement of the theme from the agent to recipient, but also express the same type of 'transfer' action with the opposite direction, i.e., from the source to an agent. (4) provides an example that describes the movement of the phone number from the source to the agent.

(4) B: 他们   那个   roomate@s   也   跟   我
     Tāmen nèigè   *roommate*   *yě*   *gēn*   *wǒ*
     3PL   that   roommate   also   towards   1SG
     要   电话.   (CALLFRIEND Mandarin-Mainland: 4447)
     yào   *diànhuà.*
     ask.for phone
     'Their roommate also asked me for my phone number.'

The second most frequent constructional variant, variant c, attracts the lowest number of collexemes, including verbs that denote verbal transfer/communication (e.g., 问 *wèn* 'ask', 教 *jiāo* 'teach', and 告诉 *gàosu* 'tell') and verbs that denote short-distance physical transfer (e.g., 给 *gěi* 'give', 喂 *wèi* 'feed', 收 *shōu* 'receive', 付 *fù* 'pay', and 送 *sòng* 'give.as.a.gift'[7]); see Figure 3. Examples (5) and (6) show the uses of variant c with 告诉 *gàosu* 'tell' and 给 *gěi* 'give', respectively.

(5) 那   就   好,   我   还   要   告诉   你   一
     Nà   *jiù*   *hǎo,*   *wǒ*   *hái*   *yào*   *gàosu*   *nǐ*   *yī*
     that   just   good   1SG   also   want   tell   2SG   one
     个   好   消息。   (ToRCH2014: L11)
     gè   *hǎo*   *xiāoxī.*
     CLF   good   news
     'Well, I also want to tell you some good news.'

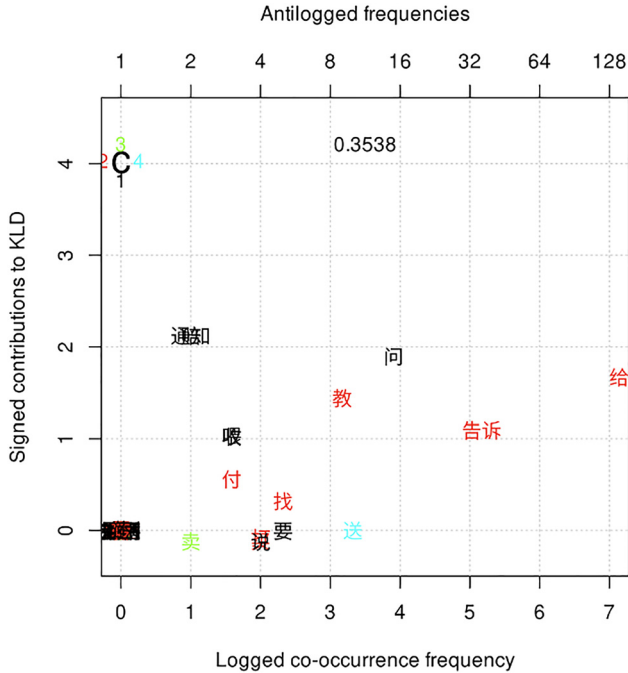7 The yuan is the basic unit of the renminbi, the official currency of the People's Republic of China.

Figure 3: Distinctive collexemes of variant c by frequency and association.

(6)	当时	我	爸	给-了	我	四万	多
	Dāngshí	*wǒ*	*bà*	*gěi-le*	*wǒ*	*sìwàn*	*duō*
	at.that.time	1SG	father	give-ASP	1SG	40,000	over
	块	钱,	结果	我	一	个	半
	kuài	*qián,*	*jiéguǒ*	*wǒ*	*yī*	*gè*	*bàn*
	yuan[8]	money	consequently	1SG	one	CLF	half
	月	就	全	花-了。	(ToRCH2014: F25)		
	yuè	*jiù*	*quán*	*huā-le.*			
	month	just	all	spend-ASP			

'At that time, my father gave me more than 40,000 yuan, but I spent it all in a month and a half.'

---

8 者 *zhě* '-er/-ist' is usually attached to a verb phrase (VP) or an adjective phrase (AdjP) and changes the VP/AdjP into a noun that denotes a person who has been conducting the action denoted by the VP or has the quality/trait denoted by the AdjP.

The remaining collexeme that does not belong to any of these categories, 找 *zhǎo* 'look.for', occurs in variant c with 麻烦 *máfan* 'trouble' as a partially filled idiomatic expression 找 <某人> 麻烦 *zhǎo <mǒurén> máfan* 'give <someone> a hard time'. See (7).

(7) B: | 法律 | 上 | 说 | 是 | 五 | 天, | 你 |
|---|---|---|---|---|---|---|
| *Fǎlǜ* | *shàng* | *shuō* | *shì* | *wǔ* | *tiān,* | *nǐ* |
| law | up | say | be | five | day | 2SG |
| 不是 | 五 | 天 | 我 | 就 | 找 | 你 |
| *búshì* | *wǔ* | *tiān* | *wǒ* | *jiù* | *zhǎo* | *nǐ* |
| NEG.be | five | day | 1SG | just | look.for | 2SG |

麻烦.   (CALLFRIEND Mandarin-Mainland: 4227)
máfan.
trouble

'Legally speaking, it should be five days. If not, I will give you a hard time.'

Zhu (1979) found that only few verbs that inherently carry the 'transfer' meaning can occur in variant e, and these verbs cannot occur with variant c. Meanwhile, there are a set of 'lexical dative' verbs that occur frequently with variant c but are hardly attested in variant e. His observations appear to be borne out by our analyses, which show that there is no overlap between the distinctive collexemes of variant e and that of variant c. For example, collexemes of verbal transfer/communication that are distinctively associated with variant e (e.g., 讲 *jiǎng* 'tell', 说 *shuō* 'say', and 介绍 *jièshào* 'introduce') occur less frequently than expected with variant c. Likewise, distinctive collexemes of variant c that express verbal transfer/communication (e.g., 问 *wèn* 'ask', 教 *jiāo* 'teach', and 告诉 *gàosu* 'tell') are rarely attested in variant e. Although variant e and variant c are quite different from each other, they have one thing in common, that is, among the five constructional variants, only variant e and variant c can express the movement of the theme from the source to agent; (4) and (8) are pertinent examples.

(8) | 光 | 你 | 爱人 | 就 | 收-了 | 他 |
|---|---|---|---|---|---|
| *Guāng* | *nǐ* | *àirén* | *jiù* | *shōu-le* | *tā* |
| alone | 2SG | husband/wife | just | receive-ASP | 3SG |

三百万。   (ToRCH2014: K27)
sānbǎiwàn.
three.million

'Your wife alone received three million yuan from him.'

Variant d attracts collexemes that denote physical transfer, specifically, short-distance or hand-to-hand/face-to-face transfer (e.g., 借 *jiè* 'borrow/lend', 租 *zū* 'rent', 还 *huán* 'return', 交 *jiāo* 'hand.over.to', 卖 *mài* 'sell', 传 *chuán* 'pass', 扔 *rēng* 'toss', 转 *zhuǎn* 'forward/transfer', 送 *sòng* 'give.as.a.gift', 留 *liú* 'leave', and 给 *gěi* 'give'). See (9). In addition, variant d also attracts 告诉 *gàosu* 'tell' and 介绍 *jièshào* 'introduce', which do not denote physical transfer but express verbal transfer/communication that involves a direct, face-to-face contact between the agent and recipient. See Figure 4. As mentioned earlier in this section, collexemes of verbal transfer/communication that are distinctively associated with variant c are repelled by variant e, and vice versa. For example, 告诉 *gàosu* 'tell' is distinctively associated with variant c but repelled by variant e, whereas 介绍 *jièshào* 'introduce' is distinctively attracted by variant e but rarely attested in variant c. Variant d is special in that it attracts both 告诉 *gàosu* 'tell' and 介绍 *jièshào* 'introduce'. In addition to verbs of short-distance transfer, variant d also attracts verbal
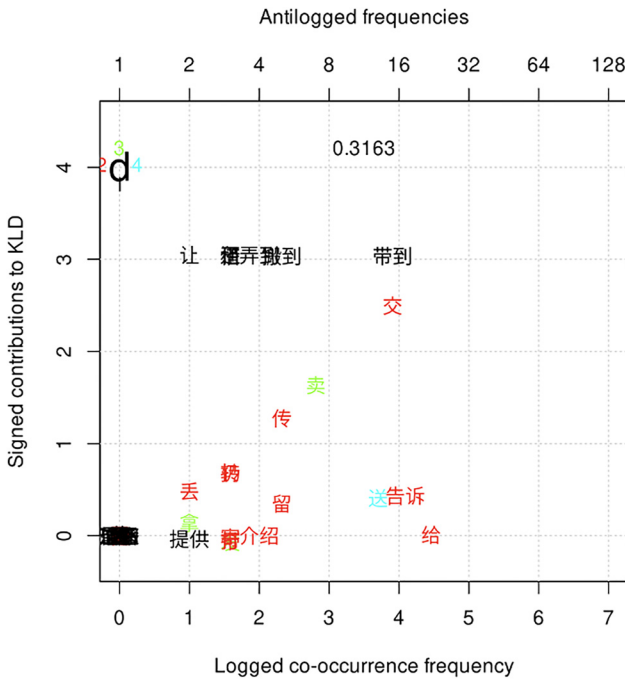

Figure 4: Distinctive collexemes of variant d by frequency and association.

compounds that imply a long-distance physical transfer, such as 带到 *dàidào* 'bring.arrive', 弄到 *nòngdào* 'get.arrive', and 搬到 *bāndào* 'move.arrive', which are exclusively associated with variant d, as indicated by the black color in Figure 4. (10) is one example.

(9) 我   把   录音   和   资料   都   交-给
    Wǒ   *bǎ*   *lùyīn*   *hé*   *zīliào*   *dōu*   *jiāo-gěi*
    1SG   BA   recording   and   material   all   hand.over.to-to
    律师   了，   中国   的   律师。   (ToRCH2014: M04)
    lǜshī   *le,*   *zhōngguó*   *de*   *lǜshī.*
    lawyer   PRT   China   GEN   lawyer
    'I gave all the recordings and materials to a lawyer, a Chinese lawyer.'

(10) 民警   赶来，   想   把   打人-者[9]
     Mínjǐng   *gǎnlái,* *xiǎng* *bǎ*   *dǎrén-zhě*
     police   arrive   intend   BA   beat.person-NMLZ
     带到   附近   的   定王台
     dàidào   *fùjìn*   *de*   *dìngwángtái*
     bring.arrive   nearby   GEN   PN
     派出所。   (ToRCH2014: B04A)
     pàichūsuǒ.
     local.police.station
     'The police arrived and intended to take the attacker to the nearby
     Dingwangtai police station.'

Variant b is characterized by collexemes that denote short-distance or hand-to-hand/ face-to-face transfer (e.g., 付 *fù* 'pay', 带 *dài* 'bring', 留 *liú* 'leave', 送 *sòng* 'give.-as.a.gift',[10] and 发 *fā* 'send.out'; 教 *jiāo* 'teach' does not denote physical transfer, but it encodes transfer of knowledge that involves a direct, face-to-face contact between the agent and recipient); see Figure 5. The constructional meaning of variant b can be clearly seen in the use of 发 *fā* 'send.out' in variant b compared to its use in variant a, as shown in (11) and (12). (11) describes the face-to-face/hand-to-hand transfer of the money between the agent and recipient, whereas (12) encodes the transfer of the email that implies some distance between the agent and recipient, which we will discuss later.

---

9 As mentioned earlier, 送 *sòng* 'give.as.a.gift' is distinctively associated with four variants (variants a, b, c, and d), but it yields its highest AM value with variant b.

10 Although 给 *gěi* 'to' is optional in variant d, it appears in nearly 80 % (74/93) of the uses of the 18 distinctive collexemes in variant d (not counting 给 *gěi* 'give', since it cannot occur with 给 *gěi* 'to' as a result of haplology [Tang 1978]). It is interesting that all 16 uses of 告诉 *gàosu* 'tell' in variant d do not occur with 给 *gěi* 'to', while 给 *gěi* 'to' occurs in all 3 uses of 介绍 *jièshào* 'introduce' in variant d.

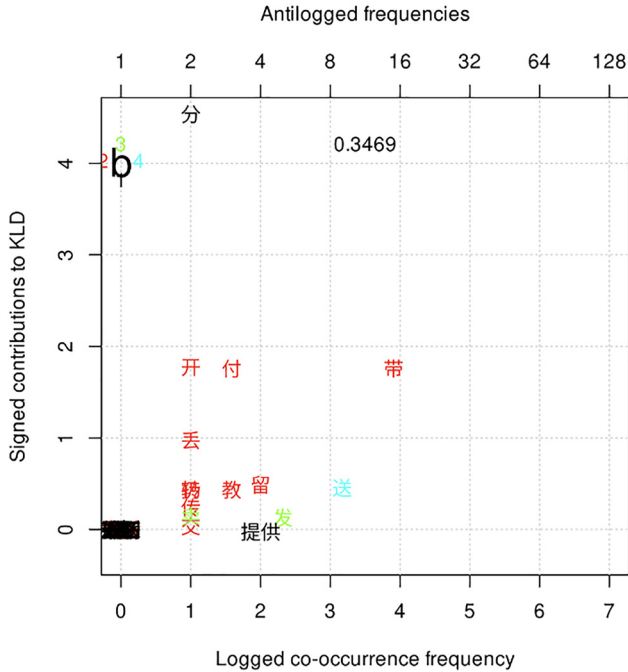**Figure 5:** Distinctive collexemes of variant b by frequency and association.

(11) D: 比如　　　　 说　　 大家　　　 要　　　　　　 去　　　 吃　　 中午饭

            Bǐrú　　　　 *shuō dàjiā*　　 *yào*　　　　　 *qù*　　 *chī*　　 *zhōngwǔfàn*

            for.example say　 everyone be.going.to go　　 eat　　 lunch

            了,　　　 他　 一　　 个　　　　 人　　　 发-给　　 你

            le,　　　 *tā*　 *yī*　　 *gè*　　　　 *rén*　　 *fā-gěi*　　 *nǐ*

            PRT　　　 3SG one　　 CLF　　　　 person send.out-to 2SG

            五　　 块　　 钱.　　 (CALLFRIEND Mandarin-Mainland: 5773)

            wǔ　　 *kuài qián.*

            five　　 yuan money

            'For example, if everyone is going to have lunch, he will give you five yuan each.'

(12)　 当　　　　 老师　　　 的　　　　　 生日　　　　 快　　　 到

       Dāng　　　 *lǎoshī*　　 *de*　　　　　 *shēngrì*　　 *kuài*　　 *dào*

       just.at　　 teacher　　 GEN　　　　 birthday　　 soon　　 arrive

       时,　　　 这　　　 位　　　　　 联络员　　　　 妈妈　　 就

       shí,　　　 *zhè*　　 *wèi*　　　　　 *liánluòyuán*　　 *māmā*　　 *jiù*

when	this	CLF	liaison	mother	just
会	发	电子邮件	给	所有	的
huì	*fā*	*diànzǐyóujiàn*	*gěi*	*suǒyǒu*	*de*
will	send.out	email	to	all	GEN
家长,	提议	大家	"凑	份子"	给
jiāzhǎng,	*tíyì*	*dàjiā*	*"còu*	*fènzi"*	*gěi*
parent	suggest	everyone	put.together	whip-round	to
老师	买	生日	礼物。	(ToRCH2014: B20B)	
lǎoshī	*mǎi*	*shēngrì*	*lǐwù.*		
teacher	buy	birthday	gift		

'When the teacher's birthday is coming soon, the mother who serves as a liaison between parents will send an email to all parents, suggesting that everyone contribute to buy a birthday gift for the teacher.'

One of the most distinctive collexemes of variant b, 带 *dài* 'bring', while seemingly denoting short-distance physical transfer, tends to occur with an abstract object in variant b, which is in accordance with Liu's (2006) observation; see (13).

(13)

我们	带-给	彼此	欢乐,	使
Wǒmen	*dài-gěi*	*bǐcǐ*	*huānlè,*	*shǐ*
1PL	bring-to	each.other	joy	make
对方	充满	自信、	对	未来
duìfāng	*chōngmǎn*	*zìxìn,*	*duì*	*wèilái*
the.opposite.side	be.full.of	confidence	for/to	the.future
满怀	希望。	(ToRCH2014: F07)		
mǎnhuái	*xīwàng.*			
be.full.of	hope			

'We bring joy to each other and make each other full of confidence and hope for the future.'

Variant b and variant d share a set of collexemes that denote short-distance physical transfer. Besides, among the five constructional variants, they are the only two variants in which 给 *gěi* 'to' can be attached to the verb;[11] see (9), (11), and (13).

Variant a is the least frequent constructional variant. Collexemes that are distinctively associated with variant a include 倒 *dào* 'pour', 发 *fā* 'send.out', 寄 *jì* 'send', 打 *dǎ* 'make.a.phone.call', 买 *mǎi* 'buy', and 送 *sòng* 'give.as.a.gift'. See Figure 6. Among these collexemes, the three most distinctive 'lexical dative' collexemes of variant a, 发 *fā* 'send.out', 寄 *jì* 'send', and 打 *dǎ* 'make.a.phone.call', all

---

11 王朔 *Wáng Shuò* is a famous contemporary writer in China whose writing is based on Beijing Mandarin.
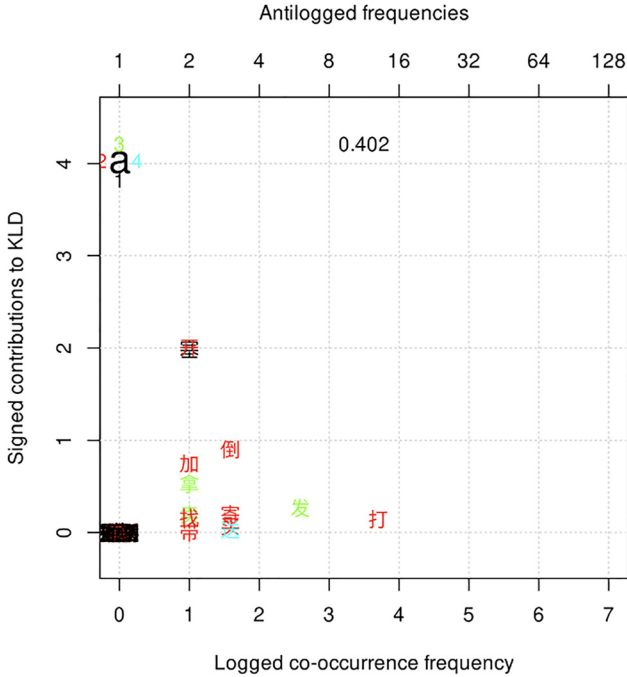
Figure 6: Distinctive collexemes of variant a by frequency and association.

involve some distance between the agent and recipient. 倒 *dào* 'pour' does not inherently carry the 'transfer' meaning and is thus an 'extended dative' verb. When 'extended dative' verbs like 倒 *dào* 'pour' occur in variant a, 给 *gěi* 'to' arguably contributes more to the 'transfer' reading than the verbs do. In such case, the construction describes an event that is composed of two sub-parts (Hsiao and Mahastuti 2020; Zhu 1979), as can be seen in (14): the first part in which the agent completed the action of pouring a cup of tea, immediately followed by the second part in which the action of 'giving' is completed.

(14)   他   倒-了   杯   茶   给   我。   (ToRCH2014: K14)
        Tā   *dào-le*   *bēi*   *chá*   *gěi*   *wǒ.*
        3SG   pour-ASP   CLF   tea   to   1SG
        'He poured me a cup of tea.'
        Lit. 'He poured a cup of tea and gave it to me.'

Five out of the six distinctive collexemes of variant a that reach the co-occurrence threshold (i.e., 倒 *dào* 'pour', 发 *fā* 'send.out', 寄 *jì* 'send', 打 *dǎ* 'make.a.phone.call', and 买 *mǎi* 'buy') are also distinctively associated with variant e. Variant a and

variant e are the only two constructional variants with which 'extended dative' verbs can occur. However, variant e differs from variant a in three aspects: (1) while in variant a 给 *gěi* 'to' and the recipient it marks follow the verb, in variant e the preposition and the recipient/source it marks precede the verb; (2) in addition to 给 *gěi* 'to', many other prepositions can mark the recipient/source in variant e, such as 向 *xiàng* 'towards' and 跟 *gēn* 'towards'; and (3) on the basis of (2), variant e can not only describe the movement of the theme from the agent to recipient, like variant a, but also the same type of 'transfer' action with the opposite direction, i.e., from the source to the agent, as in Example (4).

With regard to which constructional variant is used most frequently in Mandarin Chinese, previous empirical studies on dative alternation in Chinese give different results. For example, Chin (2009) calculates the frequency distribution of constructional variants a, b, c, and e in a self-built corpus of the writings of 王朔 *Wáng Shuò*; he finds that variant c is the most frequently used constructional variant, followed by variant e. Variant b and variant a are used far less frequently. Ma (2013) investigates the frequency distribution of all five constructional variants in a self-built corpus of Chinese newspapers and reports that variant e is the most frequently used constructional variant. Our results based on the data from the ToRCH2014 corpus (Xu 2017) and the CallFriend-Mainland Mandarin corpus (Canavan and Zipperlen 1996) suggest that variant e is the most frequently used constructional variant, followed by variants c and d. Similar to Chin's (2009) findings, variant b and variant a are used far less frequently.

As a side note, we observed several instances of idiomatic/routinized uses across the constructions in everyday language use. As previously mentioned, 找 *zhǎo* 'look.for' occurs in variant c with 麻烦 *máfan* 'trouble' as an partially filled idiomatic expression 找 <某人> 麻烦 *zhǎo <mǒurén> máfan* 'give <someone> a hard time' as in (7). Another routinized use involves the frequent co-occurrence of 打 *dǎ* 'make.-a.phone.call' and 电话 *diànhuà* 'phone.call' in variant a and variant e. All 12 uses of 打 *dǎ* 'make.a.phone.call' in variant a collocate with 电话 *diànhuà* 'phone.call', and 92.8 % (102/110) of the uses of 打 *dǎ* 'make.a.phone.call' in variant e collocate with 电话 *diànhuà* 'phone.call'.

4 Concluding remarks

4.1 Linguistic findings/implications

The results from the multiple distinctive collexeme analysis help us identify the similarities and differences between the five constructional variants, as summarized in Table 3.

Table 3: Semantic potential of constructional variants.

Variants	Semantic potential	Directionality of transfer
e	Short-distance physical transfer Short-distance verbal transfer/communication Long-distance physical transfer Long-distance verbal transfer/communication Extended dative verbs	Bidirectional
c	Short-distance verbal transfer/communication Short-distance physical transfer	Bidirectional
d	Short-distance physical transfer Short-distance verbal transfer/communication Long-distance physical transfer	Unidirectional
b	Short-distance physical transfer Short-distance verbal transfer/communication	Unidirectional
a	Long-distance verbal transfer/communication Long-distance physical transfer Extended dative verbs	Unidirectional

Table 3 shows that while there is significant overlap in semantic potential, the variants differ in terms of potential semantic scope, with some being much narrower than others. Variant e is the most frequent (in our data anyway), attracts the widest range and highest number of collexemes (i.e., verbs) and thus is the most inclusive semantically, and can express transfer bidirectionally – if we had to suggest which variant to teach first to second language learners of Chinese, our recommendation would be construction e. Construction c, the next most frequent variant, seems more focused on short-distance transfer events, and shares with construction e that it can be used to express transfer in either direction. As the reviewer pointed out, the ease of comprehension should also be considered when deciding which variant to teach first to L2 Chinese learners. The constructional variant in Chinese that has a parallel construction in the learner's L1 (structurally or/and semantically) is conceptually more accessible and easier for the learner to acquire (Teo 2023). For example, construction c (the ditransitive) in Chinese (我给了玛丽一本书 *Wǒ gěi-le mǎlì yī běn shū* 'I gave Mary a book') and the ditransitive construction in English (*I gave Mary a book*) are parallel constructions structurally in the sense that both constructions have the same syntactic ordering (i.e., SBJ$_{agent}$ V OBJ$_{recipient}$ OBJ$_{theme}$) and also semantically because both are more focused on short-distance/direct transfer events (c.f., Gries and Stefanowitsch 2004a). From that perspective, construction c would be the easiest construction for L1-English L2-Chinese learners to understand and therefore should be taught first. The remaining constructions are only used to express transfer

unidirectionally, with construction d being the one with the widest semantic range of transfer events of the three. Variant b is similar to c in its restriction to short-distance transfer scenarios, while variant a is specifically associated with long-distance transfer events.

The overall similarity of variants b and c poses an interesting question by re-raising the question of whether they should actually be considered the same construction or not, which in turn affords us the opportunity to highlight possible follow-up analyses our methods offer. In general, one would argue that two constructs instantiate two separate constructions if the constructs differ on the formal or the functional side (or of course both). In Section 1.2 above, the difference between the two examples is the presence of 给 *gěi* 'to' versus 了 *le* (glossed as an aspectual marker) and qualitative considerations do support the notion that b and c exhibit less than the high degree of similarity one would expect to see for conflation. Variants b and c differ along aspectual lines such that the former usually describes a completed action while the latter can be used to describe both a completed and a future action. Also, the presence of 给 *gěi* 'to' limits the direction of the transfer events that variant b can express, i.e., it makes b unidirectional. Some 'acquire' verbs, such as 收 *shōu* 'receive', can occur with variant c to denote the movement of the theme from the source to the agent (see [8]), but cannot occur with variant b because of 给 *gěi*'s presence. Another 'acquire' verb, 买 *mǎi* 'buy', denotes opposite directions when occurring with variant b (from agent to recipient) and variant c (from source to agent).

In addition to these qualitative considerations, we can approach the issue of the similarity of b and c using the quantitative data we have: If variants b and c were really just one construction, the distributional hypothesis underlying all of collostructions and nearly all of quantitative corpus linguistics would lead to the expectations that

1. their collexemes would behave very similarly, and that
2. the correlation between the distinctive collexeme strengths for b and the distinctive collexeme strengths for c would be high.

However, our results suggest both do not seem to be the case, as can be seen when one, first, plots the distinctive collexeme strengths of b and c against each other and when one, second, checks the correlation between their collexeme strengths, which is not strong at all (Spearman's *rho* = 0.026 with a bootstrapped 95 % confidence interval of [−0.21, 0.273]). In fact, the correlations and an exploratory cluster analysis on them suggest that b is more similar to a than to anything else, which corroborates Liu's (2006) observation. Thus, if we take the distributional hypothesis seriously, as one has to once one has committed to any kind of collocational/collostructional analysis, it does not seem as if b and c should be grouped (as opposed to Hsiao and Mahastuti [2020] and Zhang and Xu [2019]).

4.2 Methodological findings/implications

This study is one of a small number of MDCAs, small at least compared to collexeme and distinctive collexeme analyses. It is one of an even smaller number of such studies following recent discussion by using more than one association measure value (like p_{FYE} or an exact binomial test), namely a frequency value together with an association measure that is much less correlated with, or tainted by, co-occurrence frequency (Gries 2019, 2022b, 2023, 2024), and it is among the first to do so with a directional association measure while also considering the dispersion of each verb over the constructions. In bringing together this variety of innovations, this study, on top of the linguistic findings, was also a methodological proof of concept. It is, however, also essential to consider which MDCA approach is most useful under what circumstances. In general, we feel that the following conclusions and recommendations are warranted:

– Of the two measures that are highly correlated with co-occurrence frequency – the legacy exact binomial tests and the Pearson residuals – there is no good reason to use the former anymore. The results both measures produce are extremely highly correlated but the former runs into the -Inf problem of really small probabilities much more often than the latter, and taking steps to address that problem can sometimes amount to be computationally very resource-intensive and slow; computing all Pearson residuals for one analysis, by contrast, usually takes no time at all, which in turn allows for bootstrapping approaches etc. for confidence intervals of collexeme attraction values. However, downsides even for the Pearson residuals approach *are* the high correlation with co-occurrence frequency and the bidirectionality, which means its application should maybe be restricted to exploratory approaches.

– The two measures that are more independent of co-occurrence frequency – the multiple log odds ratio and the contributions to *KLD* – are both fast to compute; the main criterion differentiating between them is the ability of the latter to distinguish verb-to-construction and construction-to-verb associations. While we have focused only on one direction here, depending on one's goals, this can be a very attractive option, especially when coupled with confidence intervals (Olguín Martínez and Gries to appear).

That means, for any kind of study that has theoretical and/or explanatory goals (rather than just exploratory), the contributions to *KLD* approach is probably the best-suited one. It combines speed (for bootstrapping confidence intervals) with 'expressiveness' (capturing association largely independently of co-occurrence frequency) and directionality – it's hard to see why this should not be the method of choice for future studies.

4.3 Where to go from here

The results of this MDCA indicated a variety of semantic dimensions along which the five constructions differ. However, further exploration is required to capture the whole picture. For example, why do 打 *dǎ* 'make.a.phone.call' and 电话 *diànhuà* 'phone.call' co-occur much more frequently in variant e than in variant a (102:12)? While 75 % (9/12) of the routinized uses of 打 *dǎ* 'make.a.phone.call' and 电话 *diànhuà* 'phone.call' in variant a are found in the written data, 86.3 % (88/102) of the routinized uses of 打 *dǎ* 'make.a.phone.call' and 电话 *diànhuà* 'phone.call' in variant e are from the spoken data, indicating that mode/register might play a role in this biased distribution. This difference in distribution might also be influenced by other factors such as the pronominality, accessibility, and length difference between the recipient and the theme, etc. (e.g., Bresnan and Ford 2010; Zhang and Xu 2019, 2023). Thus, the next step could be to annotate our data for multiple variables and model the effects of these variables on speakers' choices between the five constructional variants, adopting methods such as random forests (e.g., Gries 2021; Levshina 2020); an alternative approach might be to use what is currently being developed as a kind of multivariate distinctive collexeme analysis, essentially an application of a hierarchical configural frequency analysis or association rules to multivariate collostructional data.

Data availability: The R code and data summaries are available at https://osf.io/ds7zk/.

References

Ambridge, Ben, Julian M. Pine, Caroline F. Rowland & Franklin Chang. 2012. The roles of verb semantics, entrenchment, and morphophonology in the retreat from dative argument-structure overgeneralization errors. *Language* 88(1). 45–81.

Anthony, Laurence. 2020. *AntConc (version 3.5.9)*. Tokyo: Waseda University. https://www.laurenceanthony.net/software.

Béchet, Christophe. 2015. The English substitutive complex prepositions and crosslinguistic constructional overlap. *Paper presented at the Conference "Language in Contrast (LiC): Diachronic, Variationist and Cross-linguistic Studies", Université Paris VIII, December*.

Bernaisch, Tobias, Stefan Th. Gries & Joybrato Mukherjee. 2014. The dative alternation in South Asian English(es): Modelling predictors and predicting prototypes. *English World-Wide* 35(1). 7–31.

Bresnan, Joan & Marilyn Ford. 2010. Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language* 86(1). 168–213.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina & R. Harald Baayen. 2007. Predicting the dative alternation. In Gerlof Bouma, Irene Krämer & Joost Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94. Amsterdam: Royal Netherlands Academy of Science.

Buysse, Manon. 2012. *The dative alternation: A corpus-based study of spoken British English*. Ghent: Ghent University MA thesis.

Canavan, Alexandra & George Zipperlen. 1996. *CALLFRIEND Mandarin Chinese-mainland dialect*. Philadelphia, PA: Linguistic Data Consortium.

Chang, Hui. 2014. Mǔyǔ wéi yīngyǔ hé fǎyǔ de xuéxízhě duì hànyǔ shuāngbīnjù jí qí yǔgé zhuǎnhuàn jiégòu de xídé yánjiū [The acquisition of Chinese double object constructions and dative alternation constructions by English and French native speakers]. *Yǔyán Wénzì Yìngyòng* [Applied Linguistics](2). 96–106.

Chin, Chi On. 2009. *The verb GIVE and the double-object construction in Cantonese in synchronic, diachronic and typological perspectives*. Seattle, WA: University of Washington dissertation.

Colleman, Timothy. 2009. Verb disposition in argument structure alternations: A corpus study of the dative alternation in Dutch. *Language Sciences* 31(5). 593–611.

Colleman, Timothy & Sarah Bernolet. 2012. Alternation biases in corpora vs. Picture description experiments: DO-biased and PD-biased verbs in the Dutch dative alternation. In Dagmar Divjak & Stefan Th. Gries (eds.), *Frequency effects in language representation (Trends in Linguistics. Studies and Monographs 244.2)*, 87–125. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110274073.87.

de Marneffe, Marie-Catherine, Scott Grimm, Inbal Arnon, Susannah Kirby & Joan Bresnan. 2012. A statistical model of the grammatical choices in child production of dative sentences. *Language and Cognitive Processes* 27(1). 25–61.

De Vaere, Hilde. 2020. *The ditransitive alternation in present-day German: A corpus-based analysis*. Ghent: Ghent University dissertation.

De Vaere, Hilde, Ludovic De Cuypere & Klaas Willems. 2021. Alternating constructions with ditransitive *geben* in present-day German. *Corpus Linguistics and Linguistic Theory* 17(1). 73–107.

Desagulier, Guillaume. 2014. Visualizing distances in a set of near-synonyms: *Rather, quite, fairly*, and *pretty*. In Dylan Glynn & Justyna A. Robinson (eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy (Human Cognitive Processing: Cognitive Foundations of Language Structure and Use 43)*, 145–178. Amsterdam: John Benjamins. https://doi.org/10.1075/hcp.43.06des.

Ellis, Nick C. 2006. Language acquisition as rational contingency learning. *Applied Linguistics* 27(1). 1–24.

Evert, Stefan. 2009. Corpora and collocations. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics: An international handbook, Volume 2 (Handbooks of Linguistics and Communication Science 29/2)*, 1212–1248. Berlin: Walter de Gruyter. https://doi.org/10.1515/9783110213881.2.1212.

Fonteyn, Lauren & Stefan Hartmann. 2016. Usage-based perspectives on diachronic morphology: A mixed-methods approach towards English *ing*-nominals. *Linguistics Vanguard* 2(1). 20160057.

Gilquin, Gaëtanelle. 2006. The verb slot in causative constructions: Finding the best fit. *Constructions* 1. https://doi.org/10.24338/cons-445.

Gilquin, Gaëtanelle. 2010. *Corpus, cognition and causative constructions (Studies in Corpus Linguistics 39)*. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.39.

Gilquin, Gaëtanelle. 2013. Making sense of collostructional analysis: On the interplay between verb senses and constructions. *Constructions and Frames* 5(2). 119–142.

Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure* (Cognitive Theory of Language and Culture). Chicago, IL: University of Chicago Press.

Gries, Stefan Th. 2003. Towards a corpus-based identification of prototypical instances of constructions. *Annual Review of Cognitive Linguistics* 1. 1–27.

Gries, Stefan Th. 2004. *Coll.analysis 3. A program for R for Windows 2.x.*

Gries, Stefan Th. 2005. Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research* 34(4). 365–399.

Gries, Stefan Th. 2013. 50-something years of work on collocations: What is or should be next …. *International Journal of Corpus Linguistics* 18(1). 137–166.

Gries, Stefan Th. 2019. 15 years of collostructions: Some long overdue additions/corrections (to/of actually all sorts of corpus-linguistics measures). *International Journal of Corpus Linguistics* 24(3). 385–412.

Gries, Stefan Th. 2021. *Statistics for linguistics with R: A practical introduction*, 3rd revised edn. (Mouton Textbook). Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110718256

Gries, Stefan Th. 2022a. *Coll.analysis 4.0. A script for R to compute perform collostructional analyses*. https://www.stgries.info/teaching/groningen/index.html.

Gries, Stefan Th. 2022b. What do (some of) our association measures measure (most)? Association? *Journal of Second Language Studies* 5(1). 1–33. https://www.stgries.info/teaching/groningen/index.html.

Gries, Stefan Th. 2023. Overhauling collostructional analysis: Towards more descriptive simplicity and more explanatory adequacy. *Cognitive Semantics* 9(3). 351–386.

Gries, Stefan Th. 2024. Frequency, dispersion, association, and keyness: Revising and tupleizing corpus-linguistic measures. *Studies in Corpus Linguistics*, vol. 115. Amsterdam: John Benjamins.

Gries, Stefan Th. & Anatol Stefanowitsch. 2004a. Extending collostructional analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics* 9(1). 97–129.

Gries, Stefan Th. & Anatol Stefanowitsch. 2004b. Co-varying collexemes in the *into*-causative. In Michel Achard & Suzanne Kemmer (eds.), *Language, culture, and mind* (Conceptual Structure, Discourse and Language), 225–236. Stanford, CA: CSLI Publications.

Gropen, Jess, Steven Pinker, Michelle Hollander, Richard Goldberg & Ronald Wilson. 1989. The learnability and acquisition of the dative alternation in English. *Language* 65(2). 203–257.

Hilpert, Martin. 2006. Distinctive collexeme analysis and diachrony. *Corpus Linguistics and Linguistic Theory* 2(2). 243–256.

Hilpert, Martin. 2008. *Germanic future constructions: A usage-based approach to language change (Constructional Approaches to Language 7)*. Amsterdam: John Benjamins. https://doi.org/10.1075/cal.7.

Hilpert, Martin. 2012a. Diachronic collostructional analysis: How to use it and how to deal with confounding factors. In Kathryn Allan & Justyna A. Robinson (eds.), *Current methods in historical semantics* (Topics in English Linguistics 73), 133–160. Berlin: De Gruyter Mouton.

Hilpert, Martin. 2012b. Diachronic collostructional analysis meets the noun phrase: Studying *many a noun* in COHA. In Terttu Nevalainen & Elizabeth Closs Traugott (eds.), *The Oxford handbook of the history of English* (Oxford Handbooks), 233–244. Oxford: Oxford University Press.

Hsiao, Huichen S. & Lestari Mahastuti. 2020. A collostructional analysis of ditransitive constructions in Mandarin. In Jia-Fei Hong, Yangsen Zhang & Pengyuan Liu (eds.), *Chinese lexical semantics. CLSW 2019 (Lecture Notes in Computer Science 11831)*, 37–51. Cham: Springer. https://doi.org/10.1007/978-3-030-38189-9_4.

Jenset, Gard B., Barbara McGillivray & Michael Rundell. 2018. The dative alternation revisited: Fresh insights from contemporary British spoken data. In Vaclav Brezina, Robbie Love & Karin Aijmer (eds.), *Corpus approaches to contemporary British speech: Sociolinguistic studies of the Spoken BNC2014 (Routledge Advances in Corpus Linguistics 21)*, 185–208. London: Routledge. https://doi.org/10.4324/9781315268323-10.

Kučera, Henry & W. Nelson Francis. 1967. *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Levshina, Natalia. 2020. Conditional inference trees and random forests. In Magali Paquot & Stefan Th. Gries (eds.), *A practical handbook of corpus linguistics*, 611–643. Cham: Springer. https://doi.org/10.1007/978-3-030-46216-1_25.

Liu, Feng-hsi. 2006. Dative constructions in Chinese. *Language and Linguistics* 7(4). 863–904.

Lorente Sánchez, Juan. 2018. "Give it him and then I'll give you money for it". The dative alternation in Contemporary British English. *Research in Corpus Linguistics* 6. 15–28.

Lu, Bingfu. 2007. *Ditransitive coding in Chinese. Paper presented at the Conference on Ditransitive Constructions*. Max Planck Institute for Evolutionary Anthropology.

Ma, Bei. 2013. *A corpus-based collostructional analysis of Mandarin ditransitive construction*. Shanghai: Fudan University MA thesis.

McEnery, Anthony, Zhonghua Xiao & Lili Mo. 2003. Aspect marking in English and Chinese: Using the Lancaster Corpus of Mandarin Chinese for contrastive language study. *Literary and Linguistic Computing* 18(4). 361–378.

Mukherjee, Joybrato & Stefan Th. Gries. 2009. Collostructional nativisation in New Englishes: Verb-construction associations in the International Corpus of English. *English World-Wide* 30(1). 27–51.

Olguín Martínez, Jesús Francisco & Stefan Th. Gries. to appear. Hypothetical manner constructions and filler-slot relations. *Constructions and Frames*.

Perek, Florent. 2015. *Argument structure in usage-based construction grammar: Experimental and corpus-based perspectives (Constructional Approaches to Language 17)*. Amsterdam: John Benjamins. https://doi.org/10.1075/cal.17.

Peyraube, Alain & Shanshan Lü. 2019. Guānyú hànyǔ jǐyǔ jiégòu lìshí yǎnbiàn de xīn jiànjiě – Jiān lùn hànyǔ fāngyán zhōng de qūbiéxìng yǔgé biāojì [New insights on the history of the dative constructions and on the differential dative marking in Mandarin and in some Chinese dialects]. *Lìshǐ Yǔyánxué Yánjiū [Research in Historical Linguistics]* 13. 218–242.

Rajeg, Gede Primahadi Wijaya. 2014. Metaphorical profiles of five Indonesian quasi-synonyms of ANGER: Multiple distinctive collexeme analysis. In *Proceedings of the International Congress of the Linguistic Society of Indonesia 2014 (KIMLI 2014)*, 165–170. Bandar Lampung: Masyarakat Linguistik Indonesia (MLI).

Rajeg, Gede Primahadi Wijaya. 2019. *Metaphorical profiles and near-synonyms: A corpus-based study of Indonesian words for HAPPINESS*. Melbourne: Monash University dissertation.

Rajeg, Gede Primahadi Wijaya. 2020. Linguistik korpus kuantitatif dan kajian semantik leksikal sinonim emosi bahasa Indonesia. *Linguistik Indonesia* 38(2). 123–150.

Reali, Florencia. 2017. Acceptability of dative argument structure in Spanish: Assessing semantic and usage-based factors. *Cognitive Science* 41(8). 2170–2190.

Rens, Dario. 2017. The semantics of the *aan*-construction in 16th-century Dutch: A semasiological and onomasiological approach. *Literator* 38(2). a1333.

Röthlisberger, Melanie. 2018. *Regional variation in probabilistic grammars: A multifactorial study of the English dative alternation*. Leuven: KU Leuven dissertation.

Röthlisberger, Melanie. 2021. Social constraints on syntactic variation: The role of gender in Jamaican English ditransitive constructions. In Tobias Bernaisch (ed.), *Gender in World Englishes (Studies in English Language)*, 147–175. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108696739.007.

Röthlisberger, Melanie, Jason Grafmiller & Benedikt Szmrecsanyi. 2017. Cognitive indigenization effects in the English dative alternation. *Cognitive Linguistics* 28(4). 673–710.

Schönefeld, Doris. 2013a. Go mad – come true – run dry: Metaphorical motion, semantic preference(s) and deixis. *Yearbook of the German Cognitive Linguistics Association* 1(1). 215–236.

Schönefeld, Doris. 2013b. *It is … quite common for theoretical predictions to go untested* (BNC_CMH). A register-specific analysis of the English *go un-V-en* construction. *Journal of Pragmatics* 52. 17–33.

Schönefeld, Doris. 2015. A constructional analysis of English *un*-participle constructions. *Cognitive Linguistics* 26(3). 423–466.

Stefanowitsch, Anatol. 2013. Collostructional analysis. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford handbook of Construction Grammar (Oxford Handbooks)*, 290–306. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195396683.013.0016.

Stefanowitsch, Anatol. 2014. Collostructional analysis: A case study of the English *into*-causative. In Thomas Herbst, Hans-Jörg Schmid & Susen Faulhaber (eds.), *Constructions collocations patterns (Trends in Linguistics. Studies and Monographs 282)*, 217–238. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110356854.217.

Stefanowitsch, Anatol. 2019. *Delivering a Brexit deal to the British people:* Theresa May as a reluctant populist. *Zeitschrift für Anglistik und Amerikanistik* 67(3). 231–263.

Stefanowitsch, Anatol. 2020. *Corpus linguistics: A guide to the methodology (Textbooks in Language Sciences 7)*. Berlin: Language Science Press. https://doi.org/10.5281/zenodo.3735822.

Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2). 209–243.

Stefanowitsch, Anatol & Stefan Th. Gries. 2005. Covarying collexemes. *Corpus Linguistics and Linguistic Theory* 1(1). 1–43.

Su, Hung-Kuan & Alvin Cheng-Hsien Chen. 2019. Conceptualization of CONTAINMENT in Chinese: A corpus-based study of the Chinese space particles *lǐ*, *nèi*, and *zhōng*. *Concentric* 45(2). 211–245.

Tang, Ting-chi. 1978. Double object constructions in Chinese. In Robert Liang-wei Cheng, Ying-che Li & Ting-chi Tang (eds.), *Proceedings of Symposium on Chinese Linguistics, 1977 Linguistic Institute of the Linguistic Society of America*, 67–96. Taipei: Student Book Co., Ltd.

Teo, Ming Chew. 2023. A parallel construction approach to teaching the Chinese "give" construction. In Fangyuan Yuan, Baozhang He & Wenze Hu (eds.), *Pedagogical grammar and grammar pedagogy in Chinese as a second language*, 56–73. London: Routledge. https://doi.org/10.4324/9781003161646-6.

Theijssen, Daphne. 2008. Using the ICE-GB Corpus to model the English dative alternation. In *Proceedings of the Postgraduate Conference in Corpus Linguistics*. Birmingham: Aston University.

Theijssen, Daphne. 2012. *Making choices: Modelling the English dative alternation*. Nijmegen: Radboud University dissertation.

Toutanova, Kristina, Dan Klein, Christopher D. Manning & Yoram Singer. 2003. Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, 252–259. Association for Computational Linguistics.

Tseng, Huihsin, Pi-Chuan Chang, Galen Andrew, Daniel Jurafsky & Christopher Manning. 2005. A conditional random field word segmenter for Sighan bakeoff 2005. In *Proceedings of the Fourth SIGHAN Workshop on Chinese Language Processing*, 168–171. Association for Computational Linguistics.

Valdeson, Fredrik. 2021. *Ditransitives in Swedish: A usage-based study of the double object construction and semantically equivalent prepositional object constructions 1800–2016*. Stockholm: Stockholm University dissertation.

Van Goethem, Kristel & Muriel Norde. 2020. Extravagant "fake" morphemes in Dutch. Morphological productivity, semantic profiles and categorical flexibility. *Corpus Linguistics and Linguistic Theory* 16(3). 425–458.

Wang, Linlin. 2013. Jīyú duō yǔtǐ yǔliào de yīngyǔ shuāngjíwù jiégòu yánjiū — Yǐ *give* wéilì [A corpus-based multi-genre study of English ditransitive construction: A case study of *give*]. *Wàiguó Yǔ* [Journal of Foreign Languages] 36(1). 45–54.

Wulff, Stefanie, Nick C. Ellis, Ute Römer, Kathleen Bardovi-Harlig & Chelsea J. Leblanc. 2009. The acquisition of tense-aspect: Converging evidence from corpora and telicity ratings. *The Modern Language Journal* 93(3). 354–369.

Xia, Fei. 2000. *The part-of-speech guidelines for the Penn Chinese Treebank (3.0)* (IRCS Technical Reports Series 00–07). Philadelphia, PA: University of Pennsylvania.

Xu, Jiajin. 2017. *ToRCH2014 corpus: Texts of Recent CHinese corpus 2014*. Beijing: National Research Centre for Foreign Language Education, Beijing Foreign Studies University.

Yao, Yao & Feng-hsi Liu. 2010. A working report on statistically modeling dative variation in Mandarin Chinese. In Chu-Ren Huang & Dan Jurafsky (eds.), *Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010)*, 1236–1244. COLING 2010 Organizing Committee.

Zhang, Dong & Jiajin Xu. 2019. Yīnghàn yǔgé jiāotì xiànxiàng de duō yīnsù yánjiū [A multifactorial study of dative alternation in English and Chinese]. *Wàiguó Yǔ* [Journal of Foreign Languages] 42(2). 24–33.

Zhang, Dong & Jiajin Xu. 2023. Dative alternation in Chinese: A mixed-effects logistic regression analysis. *International Journal of Corpus Linguistics* 28(4). 559–585.

Zhou, Jiangping. 2023. A corpus-based study of explicit objective modal expressions in English. *Studia Neophilologica* 95(1). 100–126.

Zhu, Dexi. 1979. Yǔ dòngcí "gěi" xiāngguān de jùfǎ wèntí [Syntactic issues related to the verb 给 *gěi* 'give']. *Fāngyán* [Dialect](2). 81–87.TnQ