Constructions 15 (2023)
doi 10.24338/cons-571

Differentiating valence patterns: A quantitative analysis based on formal and semantic attributes

Kazuho Kambara[1], Hajime Nozawa[2], & Takeshi Takahashi[2]
[1] Ritsumeikan University [2] Kyoto University of Foreign Studies

Abstract

A verb can license multiple constructions. For instance, the verb *replace* licences (at least) three constructions: (i) *Alice replaced Bill with Charlotte*, (ii) *Alice replaced Bill*, (iii) *Bill replaced Charlotte*. Each construction realises different arrays of semantic roles (i.e., *Alice* as the responsible entity for the transition of *Bill* and/or *Charlotte*, *Bill* as the "new" entity, and *Charlotte* as the "old" entity) which are called valence patterns. Since each construction represents different meanings, a hearer (or reader) must identify the appropriate construction to understand a given utterance. Despite its theoretical importance, the cues used in identifying each valence pattern have yet to be explored. This paper aims to reveal the attributes used in identifying each pattern by annotating data with semantic attributes (i.e., Animacy of subjects and objects) and a formal attribute (i.e., the presence of *with*-phrase). The analysis based on conditional inference tree shows that active sentences with the verb *replace* can be classified by nearly 80% accuracy, which is fairly "good" (baseline: 35.4%). Moreover, the analysis of misclassifications reveals that a fine-grained semantic characterisation of participants is needed for a more accurate specification of a construction suggesting the effectiveness of event ontologies in constructional modelling.

1 Introduction

Construction grammars (Boas & Sag 2012; Croft 2001; Fillmore et al. 1988; Goldberg 1995, 2006; Hoffmann 2022; Kay & Fillmore 1999; Langacker 2008) provide a systematic approach to describe both regularity and idiomaticity in grammatical constructions. Since the primary unit of analysis in construction grammar is a **construction**, a pair of meaning and form, the scope of analysis includes the description of meaning, which imposes a significant challenge due to its "ungraspable" nature.

Despite the widely acknowledged difficulty, a few attempts for building a framework for semantic description have been made (Fellbaum 1998; Baker 2017). Two kinds of semantic annotation are currently available: (i) semantic class labelling and (ii) semantic role labelling. The former assigns the specific semantic class to a given text (e.g., assignment of Dog to the expression *"Nana"*) while the latter assigns the role

Corresponding author(s): Kazuho Kambara, Ritsumeikan University, kazy0324@pep-rg.jp

in a situation to a given text (e.g., assignment of ⟨Agent⟩ to the expression *"Nana"* in *"Nana hid a bone in the garden"*). These two criteria are independent but complementary of each other (Baker & Fellbaum 2009). The primary focus of this paper revolves around challenges in semantic role labelling.

Construction grammarians often employ **frame semantics** to represent the semantics of given expressions (Fillmore 1982, 1985, 2008; Fillmore & Atkins 1992; Fillmore et al. 2012). In frame semantics, frames are characterised as a set of finer-grained semantic roles, called **frame elements** and their relations among them. Frame semantics is well known for its flexible nature of describing different types of argument realisations. However, how a hearer (or reader) identifies the appropriate argument realisation pattern associated with a form is still open to discussion.

This paper explores semantic and formal features in identifying the appropriate argument realisation pattern by examining attested cases of the verb *replace*, and we argue that using ontologies is an effective method for the estimation of arrays of semantic roles. This paper is structured as follows. Section 2 introduces the basic assumptions of frame semantics and the problem of identifying the appropriate argument realisation pattern. Section 3 explains the method and the procedure of the corpus analysis employed in this study. Section 4 reports the quantitative and qualitative results based on the conditional inference tree. Section 5 argues the need for an ontology for finer-grained descriptions of frames. Section 6 briefly summarises the paper and discusses possible future developments.

2 Identifying semantic role realisation patterns

This section overviews the basic assumptions of frame semantics and introduces the current study's research question: how does a hearer identify the appropriate argument realisation pattern?

Identifying the appropriate argument realisations can be tricky when a verb licenses different constructions, as demonstrated in (1). As paraphrased as (1a) and (1b), the relation between "Alice" and "Bill" in (1) can vary depending on contexts.

(1) Alice replaced Bill.
  a. Alice took the (physical, psychological, or social) place that Bill used to have.
  b. Alice removed Bill from the (physical, psychological, or social) place.

The ambiguous readings of the form licensed by the same predicate (i.e., *replace*) impose a challenge to semantic theories of construction grammars. If a construction is a pairing of a form and a meaning, a hearer has to distinguish one reading from the other. Though this problem has been well-known in the literature of semantic role labelling, cues used in the identification processes have yet to be explored.

This section is structured as follows. Section 2.1 introduces the frame semantic approach to semantic role labelling, and Section 2.2 points out the limitations of the previous studies.

2.1 Frame semantic role labelling

This section overviews the characteristics of frame semantics (Fillmore 1982, 1985; Fillmore & Atkins 1992) and FrameNet (Baker et al. 1998, 2003; Baker 2017; Fillmore et al. 2003a; Fillmore & Baker 2015), which are approaches to semantic role labelling, and explains how frame semantics provides "deep" semantics with an example of a detailed analysis on the verb *replace* (Fillmore et al. 2001).

One of the goals in linguistic semantics is to identify and explain distributions of **semantic roles** in a given sentence (Levin & Rappaport Hovav 2005: Ch.5). Descriptions of semantic roles specify "Who did What to Whom, and How, When and Where?" in a given sentence (Palmer et al. 2010: 2). The sentences in (2) are annotated with some of the typical semantic roles: ⟨Agent⟩, ⟨Theme⟩, and ⟨Instrument⟩[1]. Assigning appropriate semantic roles to syntactic constituents is not necessarily straightforward since the manners of their distribution vary from context to context[2]. As exemplified in (2), the different semantic roles are assigned to the syntactically same constituents of the sentences with the verb *open*.

(2) a. [$_{\langle\text{Agent}\rangle}$ John] opened [$_{\langle\text{Theme}\rangle}$ the door].
b. [$_{\langle\text{Theme}\rangle}$ The door] was opened [$_{\langle\text{Agent}\rangle}$ by John].
c. [$_{\langle\text{Instrument}\rangle}$ The key] opened [$_{\langle\text{Theme}\rangle}$ the door].
d. [$_{\langle\text{Agent}\rangle}$ John] opened [$_{\langle\text{Theme}\rangle}$ the door] [$_{\langle\text{Instrument}\rangle}$ with the key].
(Fillmore 2003: 59)

Frame semantics (Fillmore 1982, 1985; Fillmore & Atkins 1992) aims to describe these various manners of semantic role distributions. In frame semantics, a lexical meaning is described relative to a conceptual structure called a **frame** (or **semantic frame**), which is a unit consisting of different frame elements (i.e., finer-grained semantic roles) structured with static, or dynamic relations[3]. One of the most known frames is `Commercial Transaction` (Fillmore 1977, 2003), which specifies the complex dynamic relations among the four participants, namely ⟨Buyer⟩, ⟨Seller⟩, ⟨Money⟩, and ⟨Goods⟩. Primary units of analysis in frame semantics are a frame, and a **lexical unit** (Cruse 1986: 49) defined as a pair of a word with a sense (Fillmore et al. 2003a: 235–236). When a sense of lexical unit *lu* is based on a frame f, *lu* evokes f (Fillmore et al. 2003a: 236). For example, the verbs such as *buy* and *sell* evoke `Commercial Transaction`.

Frames are essential resources for building "deep" semantics, which cannot be fully obtained from the syntactic (or, more generally, formal) properties of a sentence. Fillmore et al. (2001) demonstrate how frame semantics provide deep semantics through

[1] Roughly, ⟨Agent⟩ refers to "the actor" of the situation expressed by the predicate, ⟨Theme⟩ to "the influenced entity", and ⟨Instrument⟩ to "the entity used by the actor".

[2] Descriptions of semantic role distributions pose a challenge to semantic theories (Palmer et al. 2010: Ch.1) since semantic role labelling needs to map semantic roles onto syntactic constituents in complicated manners, as shown in (2).

[3] For instance, in contrast to a dynamic frame like `Opening`, `Human Body` is a static frame since relations between body parts (e.g., *hand*, *arm*) stay the same over a period of time, as in the way where *hand* is based on *arm* and profiles a part of it. For simplicity, this paper only deals with frames with dynamic relations.

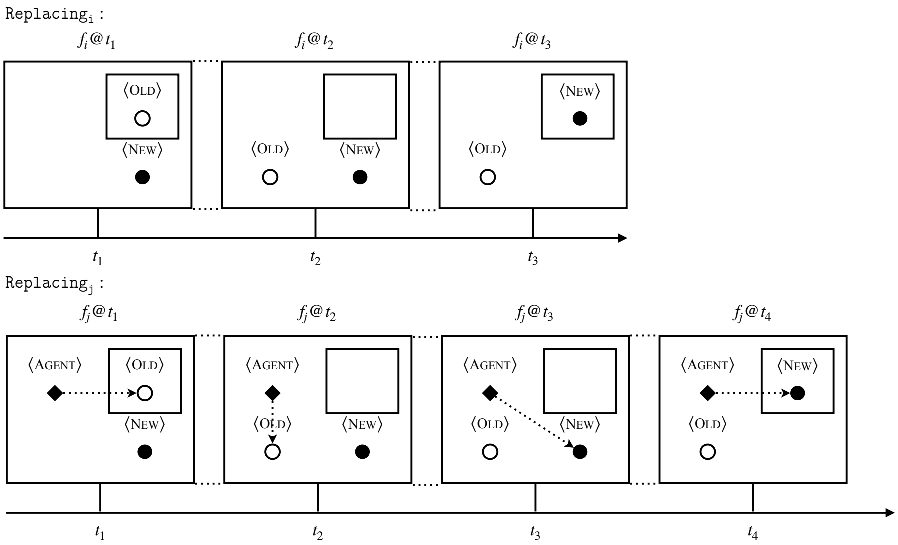

Figure 1: The structure of `Replacing` (A visualisation of Fillmore et al. (2001: 5–6))

a detailed analysis of `Replacing`, which consists of three frame elements (i.e., ⟨Agent⟩, ⟨New⟩, and ⟨Old⟩). The verb *replace* evokes a dynamic relation, which involves transitions of ⟨New⟩ and ⟨Old⟩ to or from a (physical, psychological, or social) place with (or without) an intervention of ⟨Agent⟩, as visualised in Figure 1. Fillmore classified the distribution of these frame elements into two patterns as shown in (3) (Fillmore et al. 2001: 4–7). (3a) exemplifies autonomous position changes of ⟨New⟩ and ⟨Old⟩ while the ⟨Agent⟩ causes transitions of ⟨New⟩ and/or ⟨Old⟩ in (3b)[4]. The distinction based on the presence of ⟨Agent⟩ is visualised as `Replacing`$_i$ and `Replacing`$_j$ in Figure 1.

(3) a. [$_{\langle\text{New}\rangle}$ Alice] replaced [$_{\langle\text{Old}\rangle}$ Bill] on the committee
b. [$_{\langle\text{Agent}\rangle}$ Charlotte] replaced [$_{\langle\text{Old}\rangle}$ Bill] [$_{\langle\text{New}\rangle}$ with Alice] on the committee

(Fillmore et al. 2001: 5–6)

The grammatical subjects in (3) do not share the same frame element, and each sentence denotes different subtypes of `Replacing`. Frame semantics provides deep semantics by employing a conceptual structure called a frame.

Frame semantics was later developed into an electronic lexicography project called

4 Fillmore et al. (2001: 5–7) also discuss that (3b) has two different kinds of events. One possible reading of (3b) is that (i) ⟨Agent⟩ is responsible for the transition of ⟨New⟩, and the other is that (ii) ⟨Agent⟩ is responsible for the transition of ⟨Old⟩ and ⟨New⟩. In the reading (i), Charlotte can replace Bill after he leaves the committee at his will. This reading suggests that ⟨Agent⟩ (i.e., Charlotte) has nothing to do with removing ⟨Old⟩ (i.e., Bill). In contrast, the reading (ii) exhibits how Charlotte is involved in removing Bill and adopting Alice. In explaining these two readings, Fillmore et al. (2001: 6) used a pair of parentheses around the grammatical object, "Bill" (i.e., "Charlotte replaced (Bill) with Alice"). Conventionally, the use of parentheses in a sentence signals the optionality of the constituent. However, removing the direct object "Bill" from the sentence is likely to be ungrammatical. We only used the one without parentheses to avoid unnecessary confusion because Fillmore's intention of parentheses was not clear.

FrameNet. FrameNet aims to provide a data set to construct a lexicon of English that is readable to both humans and machines, based on the theory of frame semantics and supported by annotating corpus examples of the lexical items (Baker 2017: 772)[5]. Because frame semantics succeeded in creating a foundation for consistent semantic role labelling, frame semantics and FrameNet are widely known in both linguistics circles and other neighbouring fields.

2.2 A constructionist account of argument realisation

Frames specify the event types and their available components (i.e., frame elements). However, the mechanism of identifying an array of frame elements is still open to discussion. As shown in (3), lexical units can realise their arguments in multiple ways. Following the conventions of FrameNet, we refer to such possible patterns of frame element realisation as **valence patterns** of a frame (Ruppenhofer et al. 2016: 8)[6], and FrameNet aims to describe the possible patterns (Baker 2017: 784). As visualised in Figure 1, `Replacing` evoked by the verb *replace*, has three frame elements (i.e., semantic roles)[7]: ⟨Agent⟩, ⟨New⟩, and ⟨Old⟩. Let `Replacing` be a set of {⟨Agent⟩, ⟨Old⟩, ⟨New⟩}. Since the size (cardinality) of `Replacing` is three, the total size of its power set is 8 ($= 2^3$)[8]. Despite the number of possibilities, only a handful of patterns can be realised. Langacker (1990) equates these patterns with how one construes the given event, implying that the event's human construal constrains possible realised valence patterns. We argue that examining such processes contributes to a finer-grained characterisation of constructions.

Identifying the appropriate valence pattern of a given sentence poses a significant challenge to linguistic semantics since the estimated realised pattern corresponds to

[5] For a more detailed overview of FrameNet and its goal and design, see Fillmore & Baker (2015); Baker (2017); Ruppenhofer et al. (2016).

[6] One of the anonymous reviewers pointed out that the phenomena under investigation are closely related to **mini-constructions** (Boas 2003), variants of constructions with the same verb. As shown in (i), the verb *beat* can license different subtypes of resultative constructions.

(i) a. They beat the olives out of the tree.
b. They beat the eggs creamy.
c. They beat the pebbles to a fine dust.
d. The mob beat them to death.

(Boas 2003: 353)

Each sentence represents a different subtype of event-frame associated with the verb *beat*. We decided to use the term "valence pattern" over "mini-construction" to clarify our framework. More importantly, our analysis aims to reveal the variation of the different manners of realising the same frame.

[7] In principle, it is possible to posit more frame elements such as ⟨Time⟩ for `Replacing`. However, as Baker (2017: 775) discusses, frame elements can be divided into core and non-core frame elements. The former corresponds to the roles essential to the definition of the given frame and usually occurs as arguments. This paper assumes the three frame elements (i.e., ⟨Agent⟩, ⟨Old⟩, ⟨New⟩) of `Replacing` are core frame elements.

[8] The power set of three frame elements (i.e., ⟨Agent⟩, ⟨Old⟩, ⟨New⟩) includes the following sets: {ϕ, {⟨Agent⟩}, {⟨Old⟩}, {⟨New⟩}, {⟨Agent⟩, ⟨Old⟩}, {⟨Agent⟩, ⟨New⟩}, {⟨Old⟩, ⟨New⟩}, {⟨Agent⟩, ⟨Old⟩, ⟨New⟩}}. Though it is highly debatable if we should include ϕ as a possible valence pattern, we do not discuss the possible implications of this assumption any further.

Table 1: Possible correspondence patterns

Example	Form	Meaning
Alice replaced Bill	[*X* REPLACE *Y*]	{⟨New⟩, ⟨Old⟩}
Charlotte replaced Bill	[*X* REPLACE *Y*]	{⟨Agent⟩, ⟨Old⟩}
Charlotte replaced Bill with Alice	[*X* REPLACE *Y with Z*]	{⟨Agent⟩, ⟨Old⟩, ⟨New⟩}

the semantics of a sentence. For instance, as demonstrated in (2), the verb *open* can have three valence patterns[9], namely: (i) {⟨Agent⟩, ⟨Theme⟩}, (ii) {⟨Instrument⟩, ⟨Theme⟩}, and (iii) {⟨Agent⟩, ⟨Theme⟩, ⟨Instrument⟩}. These patterns are obtained by conjoining the realised frame elements in a given sentence.

However, how a hearer (or reader) identifies the appropriate valence pattern is open to discussion. In construction grammar, a construction is treated as a part of form and meaning. A hearer must identify the appropriate valence pattern in understanding a sentence like (3). As shown in Table 1, sentences in (3) suggest that the verb *replace* in active form can license at least three constructions. The same logic can be applied to other lexical items (e.g., *open* in (2)). How a hearer understands that a participant (e.g., Alice) is more likely to be ⟨New⟩, rather than ⟨Agent⟩ is not entirely clear from previous studies, nor from the descriptions in FrameNet since FrameNet descriptions only list possible valence patterns of a given lexical item.

A usage-based construction grammar (Langacker 1999) assumes that every characteristic of a construction is a result of accumulated generalisations of actual uses, which suggests that some linguistic characteristics support these identification processes. It is possible to assume that such processes rely on pragmatics, world knowledge, and understanding of the context of occurrence. However, it is well known that the distinction between semantics and pragmatics is better seen as a gradient structure than a discrete one (Langacker 2008: 40–42). Since this assumption is also shared in construction grammar (Hoffmann 2022: 38–43), the mechanism of identifying the appropriate valence pattern should be analysed in more detail.

Analysts must consider the kinds of "cues" used in the processes before embarking on this path. As shown in Table 1, some constructions share the same form yet possess different semantics. Langacker (1990) discusses that realisations of participants are based on one's construals, suggesting that the semantics of participants play a critical role in the process. Since participants of a frame are realised as noun phrases, the semantics of nouns should be one of the strong candidates.

For these reasons, we assess how each semantic attribute of nouns can contribute to differentiating valence patterns along with some formal attributes. The kinds of features that contribute to identifying the appropriate valence pattern can help reveal finer-grained parings between form and meaning. To achieve this goal empirically, corpus-based methods can be applied (see Gries (2023) for a more recent and detailed discussion). The quantitative corpus method employing (semi-)manual annotation and statistical techniques can reveal the kinds of variables contributing to realised pattern identification, which cannot be accomplished only with intuition-based analyses

9 We ignore the order of realised frame elements for convenience.

(Geeraerts 2010).

3 Method

This section explains the methods and procedures. We performed quantitative corpus analysis using a conditional inference tree to reveal the features that contribute to identifying valence patterns. Section 3.1 describes the annotation strategies and examined data, and Section 3.2 introduces the statistical evaluation employed in this study.

3.1 Annotation strategies and the examined data

Following previous study (Fillmore et al. 2001), the verb *replace* was selected for a case study in which aimed to reveal the relationship between the valence patterns and other grammatical or semantic properties. Due to the high frequency of *replace* in most corpora and the labour-intensive nature of the analysis, BNC Baby[10] was employed as a data set for the manual annotation task, which yields 331 attested cases in total. Using Ruby, the first author extracted all instances of *replace* for the annotation task.

Manual annotation to corpus data was conducted employing the simplified version of `Replacing` as defined in (4) following Fillmore et al. (2001)[11]. Frame elements of the frame are defined in (5). In addition to the realised frame elements in (4)–(5), IsPassive, HasWithPhrase and Animacy were to observe the linguistic realisations of valence patterns as demonstrated by (Fillmore et al. 2001). These criteria are defined in (6)–(7).

(4) `Replacing`: ⟨New⟩ moves to a (physical, social, or psychological) place originally taken by ⟨Old⟩, with or without the intervention of ⟨Agent⟩.

(5) Frame elements of `Replacing`:
   a. ⟨Agent⟩: An entity that executes the act of replacing ⟨New⟩ or ⟨Old⟩.
   b. ⟨New⟩: An entity that takes place instead of what was available as a result of replacing (with or without the involvement of ⟨Agent⟩).
   c. ⟨Old⟩: An entity that loses its place due to availability of ⟨New⟩ as a result of replacing (with or without the involvement of ⟨Agent⟩).

(6) Grammatical features:
   a. IsPassive: TRUE iff the sentence in question is passive, FALSE otherwise (cf. *Bill was replaced*).

[10] Raw data of BNC Baby is available on: http://www.natcorp.ox.ac.uk/corpus/babyinfo.html

[11] Currently, FrameNet associates the two readings of *replace* with `Replacing` and `Take_place_of`. The description of `Replacing` in (4) can be seen as an integrated version of `Replacing` in FrameNet and `Take_place_of`. If we follow the descriptions in FrameNet, our research question revolves around the disambiguation of the verb *replace*, rather than differentiating valence patterns of the same verb. We decided to stick to the original characterisation given by Fillmore et al. (2001) to avoid discussing how to differentiate word senses.

b. HasWithPhrase: TRUE iff the predicate co-occurs with a *with*-phrase, FALSE otherwise (cf. *Charlotte replaced Bill with Alice*).

(7) Animacy: Whether each value of the grammatical subject or object of the predicate is animate or not (i.e., SbjIsAnimate, ObjIsAnimate).

Though it is possible in principle to annotate attested subjects and objects with categorical semantic categories (e.g., Human), creating a list of mutually exclusive semantic categories relevant to the distribution of valence patterns is extremely challenging. Moreover, the semantics of nouns is not well developed compared to other word classes such as verbs or adjectives (Murphy 2010: 149). For this reason, we only annotated the Animacy of the referents of grammatical subjects and objects, following the conventional procedures well employed in corpus linguistic literature (Gries 2010: 327).

In FrameNet, passivised sentences are said to exhibit null instantiation (Fillmore et al. 2003a: 245–246). For instance, the sentence in (8) does not realise ⟨Agent⟩ nor ⟨New⟩ as its grammatical subject though it evokes `Replacing`. Since the verb in question is passivised, the subject of the predicate cannot realise the same frame elements as ones in active voice. This pattern is caused by one of the grammatical constraints shared in the English-speaking communities.

(8) a. Alice replaced Bill
b. (i) Bill was replaced
(ii) Bill was replaced by Alice
(iii) ??Bill was replaced Alice

For this reason, we conducted separate analyses on *replace* based on the value of IsPassive.

3.2 Statistical analysis

As discussed in 2.2, factors in identifying valence patterns are not necessarily apparent from the previous studies. We conducted a tree-based analysis to confirm the kind of variables contributing to valence pattern identification (Levshina 2015, 2020; Gries 2020, 2021). Tree-based analysis is a group of statistical tools to classify the factors in classifying different entities. **Conditional inference trees** (CITs) and **conditional random forests** (CRFs) are popular methods to predict a linguistic behaviour with the given predictor variables.

To achieve a similar goal, linguists could employ **multinomial logistic regression analysis**. Regression analysis is a family of methods that aim to obtain a formula $y = a + bx$ (where y is a response variable, a an intercept, b a slope, x a predictor variable), which explains the distribution of the response variable y. In multinomial logistic regression analysis, the response variable y must be categorical with more than three values[12]. When the distributions of the predictor variable explain that of the

[12] When a categorical variable (e.g., IsDitransitive) has only two values (i.e., true or false; 1 or 0), it is called a nomial, or binary variable (Gries 2021: 17–18). Since our analysis does not deal with a nomial variable as a response variable, we only discuss the nature of categorical variables any

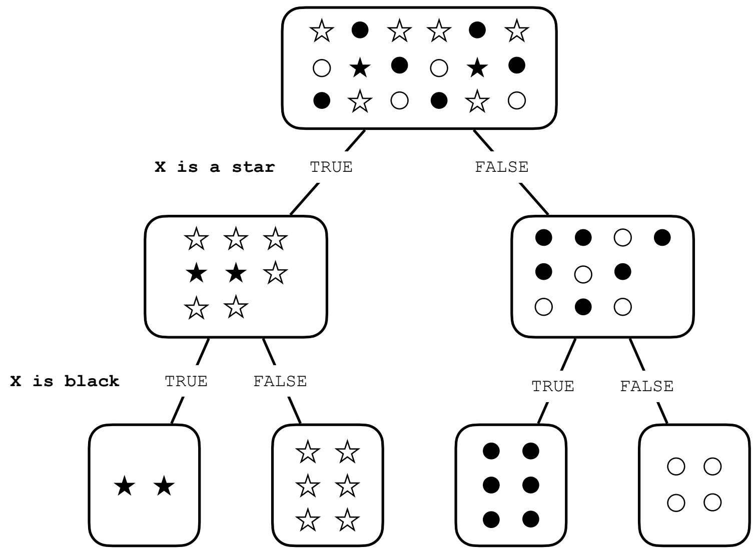

Figure 2: Binary partitioning of a data set (A modified version of Levshina (2020: 612))

response variable(s), it can be said that the distribution of the response variable(s) can be predicted. Unlike simple null-hypothesis significance tests, statistical modelling using regression analyses can deal with the data that was not included in the given observation. This is because the obtained model describes the distribution of given data and predicts the distribution of similar data.

The conditional inference tree is a family of classification and regression tree (CART) analyses. The goal of CART is to obtain maximum purity (or minimal impurity) in the terminal nodes. For instance, imagine a dataset that contains stars and circles with black or white colour. The goal of CART is to classify each object into a finite number of groups by using a given criterion. For instance, suppose an analyst decided to classify objects using two categorical variables, namely: `X is a star` and `X is black`. By using CART, analysts can obtain a tree in Figure 2. This analysis is achieved by the recursive procedures in (9). By combining the bootstrap method with a conditional inference tree, analysts can also identify the "important" variables as side effects. This approach is called (conditional) random forests. Random forests are useful when analysts use a large number of variables.

(9) **Step 1.** Select the predictor which helps best to distinguish between different values of the response variable, using some statistical criterion.
Step 2. Make a split in this variable, splitting the data in several data sets. Most algorithms use binary partitioning, although non-binary splits have also been implemented.
Step 3. Repeat Steps 1 and 2 recursively until no further splits can be made, based on certain pre-defined criteria.

(Levshina 2020: 612)

We employed CIT to reveal how each predictor contributes to realising each va-

further.

Table 2: Raw frequency of each valence pattern

	Active	Passive	Sum
Ag+Old	67	2	69
Ag+Old+New	60	8	68
New+Old	77	93	170
Old	9	13	22

lence pattern. The samples we extracted from the BNC Baby were expected to be relatively small. Though we could achieve the same goal by employing multinomial logistic regression analysis, a small sample-size can lead to unstable results with large confidence intervals. Though a small sample-size can also be challenging in CIT (Gries 2020), we employed it to reveal the kind of variables contributing to valence pattern identification since this study is still in preliminary stage. In addition, we decided not to use CRF since the number of predictors is small enough to track its all possible variations.

All computations were carried out using R (R Core Team 2022). Following Levshina (2020), we executed the CIT by using the partykit package (Hothorn & Zeileis 2015; Hothorn et al. 2023), and visualised the results using a family of ggplot2 packages (Wickham & Grolemund 2016).

4 Results

This section reports the results obtained from the procedures described in Section 3. In the following, we report quantitative and qualitative results of our studies of active voice and passive voice in Section 4.1 and Section 4.2, respectively.

In the following tables and figures, each valence pattern is presented in simplified versions. {⟨Agent⟩, ⟨Old⟩, ⟨New⟩} is simplified as Ag+Old+New, {⟨Agent⟩, ⟨Old⟩} as Ag+Old, {⟨New⟩, ⟨Old⟩} as New+Old, {⟨Old⟩} as Old.

The overall raw frequency of each valence pattern and voice is summarised as Table 2. Since the voice and HasWithPhrase in (6) are independent, we separated the overall frequencies depending on the value of voice.

Though we assumed the valence pattern identification of active voice is carried out by consulting the grammatical and semantic features of their subjects and objects, eight cases instantiating only ⟨Old⟩ were observed. We excluded these instances like (10) from our analysis since the number of observations was small.

(10) [$_{⟨\text{Old}⟩}$ The blade] needs **replacing** at the moment it is snatching at the grass. (dem/KCV)

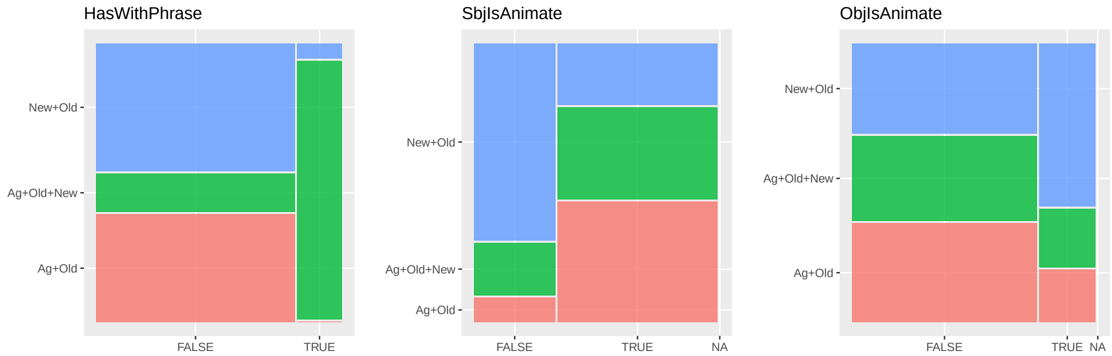

Figure 3: Mosaicplots of each variable in relation to valence patterns (Active Voice)

Table 3: Cross tabulation of all variables (Active voice)

	Freq	SbjIsAnimate	OjbIsAnimate	HasWithPhrase
Ag+Old	71	63	11	6
Ag+Old+New	61	48	6	49
New+Old	74	26	24	8
Old	8	0	0	0

4.1 Active voice

4.1.1 Quantitative analysis

The quantitative analysis of *replace* revealed that HasWithPhrase contributes to identifying Ag+Old+New, and the interaction of SbjIsAnimate and ObjIsAnimate contributes to identifying Ag+Old, and New+Old. The descriptive statistics of all possible variable combinations reveal that the matching values of SbjIsAnimate and ObjIsAnimate contribute to the realisation of New+Old. The same tendency was confirmed by the CIT, and the classification accuracy of the CIT was nearly 80%.

The descriptive statistics of sentences in active voice are summarised in Table 3, which is visualised as mosaicplots in Figure 3. In mosaicplots, the widths of the bars represent the proportional distribution of the variable on the x-axis and, within each of the (stacked) bars, the heights indicate the proportional distribution of the levels of the variable on the y-axis (Gries 2021: 123). The leftmost panel of mosaicplot shows the distribution of HasWithPhrase is dominant in the use of Ag+Old+New. In contrast, the distributions of SbjIsAnimate and ObjIsAnimate seem to correlate variables and each valence pattern.

All possible variable distributions of variables with raw frequencies are summarised as Table 4. AnSbj and AnObj correspond to animate subjects and objects, and InaniSbj and InaniObj to inanimate subjects and objects. Values conjoined with ":" represent the combination of each value (e.g., AnSbj:AnObj). The presence of a *with* phrase is shown as "+WithPhrase". Figure 4 visualises the observed proportional distributions of each valence pattern in Table 4. As suggested from the distributions of Figure 4, the matching values in subjects and objects contribute to identifying New+Old. Moreover, the presence of a *with* phrase is dominant in Ag+Old+New, which indicates that

Table 4: Every possible variable distribution of predictor variables

	Ag+Old	Ag+Old+New	New+Old
AnSbj:AnObj	9	5	28
AnSbj:AnObj+WithPhrase	0	5	0
AnSbj:InanObj	50	12	1
AnSbj:InanObj+WithPhrase	0	23	1
InanSbj:AnObj	0	0	0
InanSbj:AnObj+WithPhrase	0	0	0
InanSbj:InanObj	6	6	48
InanSbj:InanObj+WithPhrase	0	7	1

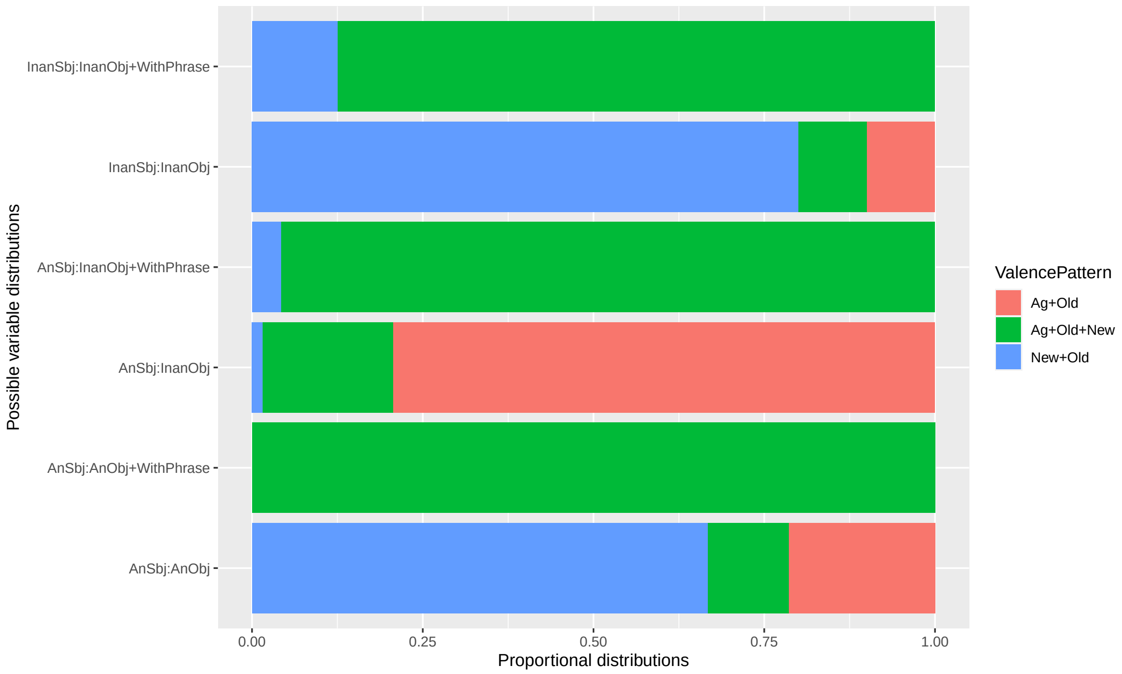

Figure 4: Proportions of valence patterns in relation to the possible distributions of each predictor

the ambiguity between Ag+Old and New+Old is trickier since it must be resolved by consulting the semantic features of grammatical subjects and objects.

The classification tree is visualised as Figure 5. The rectangles with node IDs show the names of the variables selected for the best split and the corresponding p-values (Levshina 2015: 294). The threshold of the split is set $p < 0.05$. The stacked bar plots at the bottom show the proportional distributions of each valence pattern and its raw frequencies. The plot shows that the presence of *with* phrases contribute to the identification of Ag+Old+New, while the matching values of SbjIsAnimate and ObjIsAnimate contribute to the identification of New+Old. If the combination of Animary in subjects and objects is met, New+Old is likely to be observed, Ag+Old otherwise.

CITs can also reveal the discrepancy between predicted and observed values. The cross-tabulation of each value is summarised in Table 5. The x-axis of the table corresponds to the predicted distribution of valence patterns, and the y-axis to the

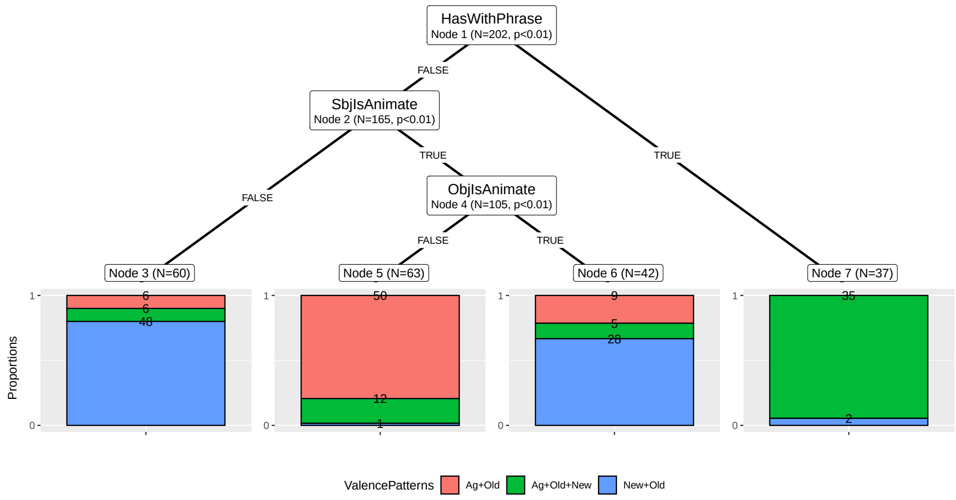

Figure 5: CIT of *replace* in active voice

Table 5: The result of classification by CIT (Accuracy: 79.7%)

	Ag+Old	Ag+Old+New	New+Old
Ag+Old	50	12	1
Ag+Old+New	0	35	2
New+Old	15	11	76

actual observed distribution of valence patterns. Since the total number of valence patterns is three, the baseline of the prediction is 0.333 ($\approx 1/3$), which means that if the algorithm assigned the valence patterns randomly, the correct prediction should be somewhere around 33% (Levshina 2015: 297). However, this baseline is based on the assumption that values are equally frequent (Gries 2021: 326). Instead, we computed the baseline as a mean of the proportion of the most frequent valence pattern (e.g., New+Old: 77 (37.7%)) and the conventional baseline (e.g., 0.33), which yields 35.4%. We employed this baseline to account for the dispersion of each valence pattern. The current model made 161 correct predictions ($= 50 + 35 + 76$) out of 202 cases. The model's accuracy is 0.797 ($= 163/206$), which is better than the baseline.

4.1.2 Qualitative analysis

Though the nearly 80% classification accuracy seems fairly "nice", we also analysed three types of misclassifications, namely: (i) Ag+Old misclassified as New+Old, (ii) Ag+Old+New misclassified as Ag+Old. The former case suggests that the kind of entity that participates in `Replacing` is crucial for an accurate understanding of valence pattern distribution, and the latter shows a (possibly) genre-specific use of *replace*.

Cases of Ag+Old misclassified as New+Old: The semantic subject of the verb *replace* in (11) is an enterprise "BR (British Rail)", which is annotated as an inanimate subject. However, since the organisation is a group of people with the potential of agency, its frame element should be ⟨Agent⟩. The semantics of nouns must be recognised to estimate the valence pattern of a sentence correctly.

(11) [$_{\langle\text{Agent}\rangle}$ BR] **replace** [$_{\langle\text{Old}\rangle}$ old trains on railway named after Lovejoy]. (news/E9S.xml)

Similarly, the subject of (12) is "Work", which denotes an activity, hence the inanimate object. It is not exactly fair to say that the subject of *replace* is "Work", but if we interpret the sentence as "Work repaired the weather vane ...", then *work* presupposes the existence of agents.

(12) [$_{\langle\text{Agent}\rangle}$ Work] began last July to repair the weather vane and to **replace** [$_{\langle\text{Old}\rangle}$ the original plaster], [...] (news/BM4.xml)

Furthermore, the social hierarchy of a participant contributes to the realisation of a valence pattern. For instance, the subject of (13) is "Clough", annotated as ⟨Agent⟩. Though the values of the subject and object are both animate, "Clough" must be annotated as ⟨Agent⟩ since the sentences preceding (13) include a description of "Clough" as the boss of "Forest" (a football team).

(13) [$_{\langle\text{Agent}\rangle}$ Clough] has been left kicking his heels in his search to **replace** [$_{\langle\text{Old}\rangle}$ England defender Des Walker], [...] (news/CH3.xml)

Similarly, the subject in (14) is "the shrewder restaurateurs", which is coded as animate. Though the values of animacy in the corresponding subject and object are animate, the relation between *restaurateurs* and *regulars* must be recognised to correctly estimate the valence pattern.

(14) [...] and [$_{\langle\text{Agent}\rangle}$ the shrewder restaurateurs] have set out to **replace** [$_{\langle\text{Old}\rangle}$ their lost business regulars] by constructing lunch menus that are brief, light and designed to show off the chef's talents rather than [...] (news/AHC.xml)

In addition, just because the lexical denotation of *it* is inherently "inanimate", it does not mean it cannot be interpreted as ⟨Agent⟩. In (15), since "it" refers to "the electorate" (underlined in (15)), the referent has a certain degree of agency.

(15) If <u>the electorate</u> disapproves of the policies or their outcomes, [$_{\langle\text{Agent}\rangle}$ it] has the opportunity to **replace** [$_{\langle\text{old}\rangle}$ the government] at the next election. (aca/J57.xml)

Cases of Ag+Old+New misclassified as Ag+Old: In mathematical discourses, the valence pattern Ag+Old+New is likely to be realised with the occurrence of *by* phrases, as demonstrated in (16). We refrain from claiming that the occurrence of *by* phrases in Ag+Old+New is genre-specific. However, it suggests the syntactic realisation of ⟨New⟩ can vary from one context to another.

(16) a. [$_{\langle\text{Agent}\rangle}$ They] have achieved this by **replacing** [$_{\langle\text{Old}\rangle}$ the usual initial-value problem [...]] by [$_{\langle\text{New}\rangle}$ an equivalent $2 \sqrt{\acute{o}}\, 2$ matrix [...]]. (aca/B2K.xml)

Table 6: Cross tabulation of all variables (Passive voice)

	FREQ	SBJISANIMATE	HASWITHPHRASE
Ag+Old	1	1	0
New+Old	101	10	8
Old	14	5	0

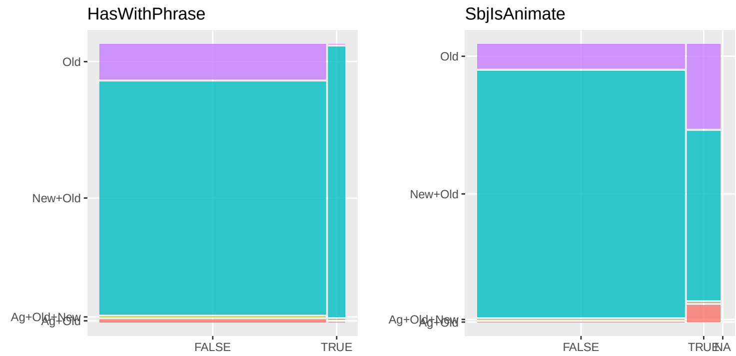

Figure 6: Mosaicplots of each variable in relation to valence patterns (Passive Voice)

b. [$_{\langle \text{AGENT} \rangle}$ We] obtain the corresponding solutions in regions II and III simply by **replacing** [$_{\langle \text{OLD} \rangle}$ g] by [$_{\langle \text{NEW} \rangle}$ $\neg\Omega$] and [$_{\langle \text{OLD} \rangle}$ f] by [$_{\langle \text{NEW} \rangle}$ $\neg\Omega$] alternately. (aca/B2K.xml)

4.2 Passive voice

4.2.1 Quantitative analysis

As discussed in Section 3, passive forms of *replace* behave differently from active forms. As summarised as cross-tabulation in Table 6, Ag+Old+New was not observed in passive forms of *replace*. As visualised as mosaicplots in Figure 6, the distribution of HASWITHPHRASE and SBJISANIMATE does not vary from one valence pattern to another.

All possible combinations of variable distributions are summarised as Table 7, which is visualised as stacked barplots as Figure 7. The majority of valence patterns are New+Old, and the realisation of Old and Ag+Old is quite rare.

We also conducted CIT to contrast the results with the results in active forms. The classification accuracy was 87.1% ($0.87 = (0 + 101 + 0)/116$), which is higher than that of active forms. However, as suggested, the majority of valence patterns in passive forms are New+Old (87.1%), which makes the classification much easier than that in active forms. For this reason, the accuracy of CIT in passive forms is not "better" than that in active forms. For this reason, we do not present the plot here.

4.2.2 Qualitative analysis

The valence pattern ambiguities also arise when passivisation is involved. Subjects of passive sentences need to be ⟨OLD⟩ since direct objects of active sentences are always

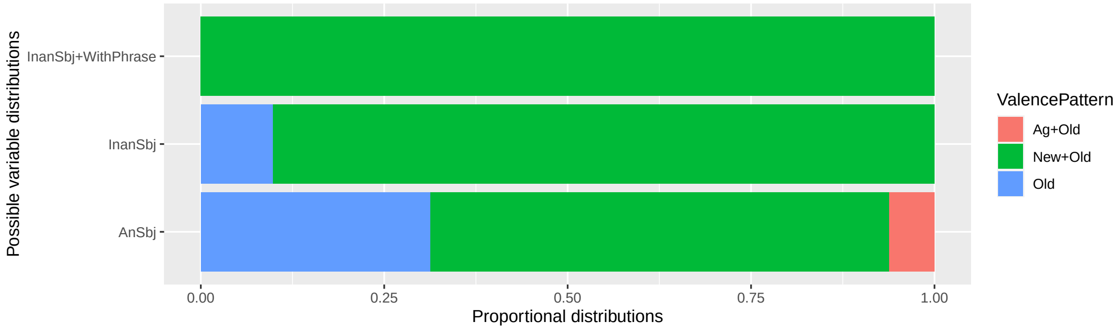

Figure 7: Proportions of valence patterns in relation to the possible distributions of each predictor

Table 7: Every possible variable distribution of predictor variables

	Ag+Old	New+Old	Old
AnSbj+WithPhrase	0	0	0
AnSbj	1	10	5
InanSbj+WithPhrase	0	8	0
InanSbj	0	83	9

realised as subjects. For this reason, it is difficult to estimate the kind of valence pattern from which a given sentence is passivised.

However, the ambiguity of this kind is sometimes resolved by different types of adjuncts. Since the value of *by* phrase represents the subject in a corresponding active sentence, one can infer that (17a) corresponds to New+Old (i.e., $\texttt{Replacing}_\texttt{i}$ in Figure 1), and (17b) to Ag+Old+New (i.e., i.e., $\texttt{Replacing}_\texttt{j}$ in Figure 1).

(17) a. [$_{\langle \text{OLD} \rangle}$ Vegeculture [...]], was **replaced** by [$_{\langle \text{NEW} \rangle}$ seed culture], [...] (aca/J18)

b. In a second, [$_{\langle \text{OLD} \rangle}$ the screen] was **replaced** with [$_{\langle \text{NEW} \rangle}$ a one-way mirror] so that [...] (news/A1M)

5 Discussion

This section discusses the theoretical implications of our analysis. Section 5.1 points out that examining semantic attributes that contribute to differentiating each valence pattern can reveal a construction's semantic specification. Section 5.2 argues for the effectiveness of examining the semantic attributes of participants, which can eventually lead to an ontological analysis of an event.

5.1 Semantic poles of a construction

Despite the challenging cases in valence pattern identifications, our analysis of *replace* revealed that the participants' semantic attribute (i.e., animacy) was effective enough (for a machine) to identify the various valence patterns by around 80% accuracy in active voice, and (although with some reservations) 87% in passive voice. However, it

Table 8: The result of classification by CIT (Accuracy: 87.1%)

	Ag+Old	New+Old	Old
Ag+Old	0	0	0
New+Old	1	101	14
Old	0	0	0

does not mean this attribute is effective in every situation. We point out that frame semantics can benefit from exploring semantic attributes that contribute to identifying the valence patterns of verbs.

The animacy played a crucial role in identifying valence patterns of `Replacing`. This result suggests that the lexical item *replace* evokes an event-frame associated with specific semantic attributes. Since our analysis showed that the matching values of animacy contributed to the differentiation of Ag+Old and New+Old, it suggests that semantic poles of constructions licensed by *replace* incorporate something similar to animacy.

However, this does not mean that animacy is an effective predictor in every valence pattern identification. Different frames impose different semantic constraints on their frame elements, which suggests that different semantic attributes should play an important role in different constructions. For instance, Fillmore et al. (2003b) demonstrated how an analyst conducts a frame semantic analysis with the example of the frame `Attaching`. This frame subsumes many attaching subevents (e.g., *glueing, welding*). In this family of events, one can easily predict that most of the attached entities are inanimate (e.g., *paper*, *metal*) because it is not likely to assume the situation where animate entities are attached without considering the metaphorical readings. Instead, it would be more productive to construct a finer-grained semantic class of object nouns participating in `Attaching`. This means that the semantic specification of a construction needs to incorporate varying degrees of semantic attributes in addition to frame elements.

It may seem obvious to assume that semantic attributes of collocating nouns are incorporated in the related valence patterns. Our analysis using quantitative corpus methods revealed that the kind of world knowledge contributes to the differentiation of valence patterns. Though it is highly challenging to assess the coverage of animacy in other lexical items, our analysis can lead to a more precise characterisation of constructions. If a construction grammarian aims at full coverage of the facts of any language without loss of linguistic generalisations (Kay 1997: 123), enriching frame structures with attributes of frame elements should be beneficial.

5.2 Describing frame participatability

Our analysis succeeded in revealing the kind of participants that are likely to participate in an event to some extent. As suggested from the analysis of misclassified instances, the animacy of collocating nouns should better be treated as a useful approximation. We suggest that ontological analysis of events (i.e., analysis of what an event is) can be effective for a more precise characterisation of a given construction's semantic pole.

Though being far from perfect, we succeeded in differentiating valence patterns of *replace* using the animacy of collocating nouns and the presence of a *with*-phrase. Our analysis showed the kind of entities that are likely to be realised in a given valence pattern. From Figure 5, we can draw the following two predictions regarding the differentiation of Ag+Old and New+Old.

(i) Given that the *with*-phrase is absent, if two animate entities or two inanimate entities occur as the subject and object of the verb *replace*, the valence pattern is likely to be New+Old.

(ii) Given that the *with*-phrase is absent, if an animate entity occurs as the subject and an inanimate entity as the object of the verb *replace*, the valence pattern is more likely to be Ag+Old.

However, these generalisations do not tell us why the interactions of animate or inanimate entities lead to one specific valence pattern. To answer this question, constructing a finer-grained event structure is needed. The misclassified cases in Section 4.1.2 suggest that world knowledge of how animate or inanimate things interact is critical in improving classification accuracy. Describing such knowledge is challenging and may not be linguistic (in a narrow sense). However, from the standpoint of frame semantics, incorporating such information to some extent is inevitable. Fillmore & Baker (2015: 791–797) discuss that frames can be divided into cognitive frames and semantic frames. The former corresponds to the non-linguistic world knowledge and the latter to the linguistic one. The distinction between these two frames is blurry. Our analysis shows that differentiation of valence patterns requires a finer-grained characterisation of frames, which can lead to describing a cognitive frame.

Detailed analyses of event structures can lead to the ontological analysis of events employing the ideas of ontological engineering. Roughly put, an **ontology** is a collection of categories of things (including events like `Replacing` and entities like `Human Body`) (Mizoguchi 2003, 2004a,b). Ontological engineering is an interdisciplinary field ranging from artificial intelligence to philosophy and attempts to provide a consistent framework to describe concepts[13]. Some linguists and scholars in neighbouring fields have integrated ontologies and semantic analyses of natural language to varying degrees (Paradis 2005; Hirst 2009; Murphy 2010; Moltmann 2022). FrameNet has already implemented the hierarchical relations among frames (Ruppenhofer et al. 2016: 9), which can be treated as an event ontology.

Building a more detailed frame using various attributes for linguistic analysis has been attempted (Kuroda & Isahara 2005; Kuroda et al. 2006), which can provide a precise correspondence between form and meaning. The caveat is that analysts must arrive at how "deep" they wish their analysis to be. For instance, if an analyst wishes to code the social power balance between two people, as observed in (13), they need to record the participants' social statuses (e.g., "Clough" being the boss of a football team). However, constructing such records is too trivial and probably does not attract linguists' interests. Instead, it would be more appealing to construct how likely a particular type of participant is to participate in a given event-frame. We call such

[13] See Mizoguchi (2004a) for an explanation of a practical application of ontologies.

tendencies as **participatability**.

Descriptions of participatability can vary from one noun to the other. For instance, the noun *restaurateur* as ⟨Agent⟩ of `Replacing` (cf. (14)), the noun *bird* is likely to be realised as ⟨Layer⟩ of `LayingAnEgg`, and the list goes on. Describing the participatability of nouns can enrich the frame structures provided by FrameNet. Since frames only specify the event-types and their relations among frames, we can arrive at describing the detailed and finer-grained correspondence between the event-types and constructions, as partially demonstrated in this paper. Though constructing a linguistically-relevant semantic attributes of participants can be daunting, systematic descriptions of participants should enrich the analyses of other related phenomena.

6 Conclusion

This article argued for the effectiveness of participants' semantic attributes (i.e., animacy of subject and object) and a formal property of a sentence (i.e., the presence of *with*-phrase) in differentiating valence patterns of *replace* (nearly 80% for active voice and 87% for passive voice). The results suggest that the semantic poles of constructions include the semantic attributes of participants. Moreover, we suggest that analysing the kinds of participants in a given frame can lead to ontological analyses of events.

Several issues remain unsolved. Although analyses of valence pattern identification suggest a lot to both (quantitative corpus) lexical analysis and ontology building, they are labour-intensive tasks. The demanding nature of semantic analysis poses a significant challenge to classifying valence patterns. Though the state-of-the-art automatic semantic classification could resolve this issue, its precision must be evaluated carefully. Secondly, as suggested in Section 5, a finer (or coarser) grained semantic classification of subjects and objects could lead to different results. A comparison of different classifications has to be conducted to determine the "right" degree of specificity. Lastly, we conducted different analyses of active and passive voices, which essentially put off the effects of passivisation in valence pattern identification. Incorporating such an aspect is significant to account for the relation between construals and linguistic forms.

Data availability

All codes and annotated data are available on OSF (Open Science Framework).

Acknowledgement

The authors thank the two anonymous reviewers of *Constructions* for their thorough comments and the editors, Stefan Hartmann and Lotte Sommerer, for their support. We also thank Atsushi Mizumoto, two anonymous reviewers of *English Corpus Studies*, and the audience at Kyoto Linguistics Colloquium (KLC). Usual disclaimers apply.

Authors' contributions

The order of authors reflects the amount of contribution. The first author initiated the project, proposed a preliminary annotation guideline, annotated the data, performed statistical analysis, and wrote and revised the substantial part of this paper. The second author critically assessed and revised the paper and annotation guideline, and annotated the data. The third author also participated in the critical assessment of the paper, and annotated the data.

References

Baker, Collin F. 2017. FrameNet: Frame semantic annotation in practice. In Nancy Ide & James Pustejovsky (eds.), *Handbook of linguistic annotation*, 771–811. New York: Springer.

Baker, Collin F. & Christiane Fellbaum. 2009. WordNet and FrameNet as complementary resources for annotation. In *Third linguistic annotation workshop*, 125–129.

Baker, Collin F., Charles J. Fillmore & Beau Cronin. 2003. The structure of FrameNet database. *International Journal of Lexicography* 16(3). 281–296.

Baker, Collin F., Charles J. Fillmore & John B. Lowe. 1998. The Berkeley FrameNet project. In *COLING-ACL 98*, 86–90.

Boas, Hans C. 2003. *A constructional approach to resultatives*. Stanford: CSLI Publications.

Boas, Hans C. & Ivan A. Sag (eds.). 2012. *Sign-based construction grammar*. Stanford: CSLI Publications.

Croft, William. 2001. *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.

Cruse, Alan D. 1986. *Lexical semantics*. Cambridge: Cambridge University Press.

Fellbaum, Christiane (ed.). 1998. *WordNet: An electronic lexical database*. Cambridge, Mass.: MIT Press.

Fillmore, Charles J. 1977. Topics in lexical semantics. In Roger W. Cole (ed.), *Current issues in linguistic theory*, 76–138. Indiana: Indiana University Press.

Fillmore, Charles J. 1982. Frame semantics. In The Linguistic Society of Korea (ed.), *Linguistics in morning calm*, 111–137. Seoul: Hanshin Publishing Company.

Fillmore, Charles J. 1985. Frames and the semantics of understanding. *Quaderni di Semantica* 6(2). 222–254.

Fillmore, Charles J. 2003. *Form and meaning in language, vol.1: Papers on semantic roles*. Stanford: CSLI Publications.

Fillmore, Charles J. 2008. Border conflicts: FrameNet meets construction grammar. In Elisenda Bernal & Janet DeCesaris (eds.), *Proceedings of the XIII EURALEX international congress*, 49–68.

Fillmore, Charles J. & Beryl T. Atkins. 1992. Toward a frame-based lexicon: The semantics of RISK and its neighbors. In Adrienne Lehrer & Eva Feder Kittay (eds.), *Frames, fields, and contrasts: New essays in semantic and lexical organization*, 75–102. London: Routledge.

Fillmore, Charles J. & Collin F. Baker. 2015. A frames approach to semantic analysis. In Bernd Hein & Heiko Narrog (eds.), *The Oxford handbook of linguistic analysis*, 791–

816. Oxford: Oxford University Press.

Fillmore, Charles J., Christopher R. Johnson & Miriam R. L. Petruck. 2003a. Background to FrameNet. *International Journal of Lexicography* 16(3). 235–250.

Fillmore, Charles J., Paul Kay & Mary Catherine O'Connor. 1988. Regularity and idiomaticitiy in grammatical constructions: The case of *let alone*. *Language* 64(3). 501–538.

Fillmore, Charles J., Russell R. Lee-Goldman & Russell Rhodes. 2012. The FrameNet constructicon. In Ivan A. Sag & Hans C. Boas (eds.), *Sign-based construction grammar*, 309–372. Stanford: CSLI Publications.

Fillmore, Charles J., Miriam R.L. Petruck, Josef Ruppenhofer & Abby Wright. 2003b. FrameNet in action: The case of Attaching. *International Journal of Lexicography* 13(3). 297–332.

Fillmore, Charles J., Charles Wooters & Collin F. Baker. 2001. Building a large lexical databank which provides deep semantics. In *Proceedings of the 15th Pacific Asia Conference on Language, Information and Computation*, 3–26.

Geeraerts, Dirk. 2010. The doctor and the semantician. In Dylan Glynn & Kerstin Fischer (eds.), *Quantitative methods in cognitive semantics: Corpus-driven approaches*, 63–78. Berlin: Mouton de Gruyter.

Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Gries, Stefan Th. 2010. Behavioral profiles: A fine-grained and quantitative approach in corpus-based lexical semantics. *The Mental Lexicon* 5(3). 323–346.

Gries, Stefan Th. 2020. On classification trees and random forests in corpus linguistics: Some words of caution and suggestions for improvement. *Corpus Linguistics and Linguistic Theory* 16(3). 617–647.

Gries, Stefan Th. 2021. *Statistics for linguistics with R: A practical introduction*. Berlin: Mouton de Gruyter 3rd edn.

Gries, Stefan Th. 2023. Quantitative corpus methods of cognitive semantics/linguistics. In Thomas Li (ed.), *Handbook of cognitive semantics*, 328–350. Leiden: Brill. https://stgries.info/research/ToApp_STG_QuantCorpCognSem_HdbkCognSem.pdf.

Hirst, Graeme. 2009. Ontology and the lexicon. In Steffen Staab & Rudi Studer (eds.), *Handbook on ontologies*, 269–292. Berlin: Springer.

Hoffmann, Thomas. 2022. *Construction grammar: The structure of English*. Cambridge: Cambridge University Press.

Hothorn, Torsten & Achim Zeileis. 2015. Partykit: A modular toolkit for recursive partytioning in R. *The Journal of Machine Learning Research* 16(1). 3905–3909.

Hothorn, Torsten, Achim Zeileis & Maintainer Torsten Hothorn. 2023. *Package 'partykit'*. https://partykit.r-forge.r-project.org/partykit/.

Kay, Paul. 1997. *Words and the grammar of context*. Stanford: CSLI Publications.

Kay, Paul & Charles J. Fillmore. 1999. Grammatical constructions and linguistic generalizations: The *what's x doing y?* construction. *Language* 75(1). 1–33.

Kuroda, Kow & Hitoshi Isahara. 2005. Proposing the multilayered semantic frame analysis of text: As an effective framework to reveal what you need to know before defining entries for a (generative) lexicon. In *Proceedings of the 3rd international workshop of generative approaches to the lexicon*, 124–133. http://clsl.hi.h.kyoto-u.ac.jp/~kkuroda/papers/msfa-gal05-paper.pdf.

Kuroda, Kow, Masao Utiyama & Hitoshi Isahara. 2006. Getting deeper semantics than Berkeley FrameNet with MSFA. In *Language resources and evaluation (LREC) 2006*, 2425–2430. http://clsl.hi.h.kyoto-u.ac.jp/~kkuroda/papers/msfa-lrec06.pdf.

Langacker, Ronald W. 1990. Settings, participants, and grammatical relations. In Savas L. Tsohatzidis (ed.), *Meanings and prototypes: Studies in linguistic categorization*, 213–238. London: Routledge.

Langacker, Ronald W. 1999. A dynamic usage-based model. In Michael Barlow & Suzanne Kemmer (eds.), *Usage-based models of language*, 1–63. Stanford: CSLI Publications.

Langacker, Ronald W. 2008. *Cognitive grammar: A basic introduction*. Chicago: University of Chicago Press.

Levin, Beth & Malka Rappaport Hovav. 2005. *Argument realization*. Cambridge: Cambridge University Press.

Levshina, Natalia. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam & Philadelphia: John Benjamins.

Levshina, Natalia. 2020. Conditional inference trees and random forests. In Magali Paquot & Stefan Th. Gries (eds.), *A practical handbook of corpus linguistics*, 611–643. Berlin: Springer.

Mizoguchi, Riichiro. 2003. Tutorial on ontological engineering, part 1: Introduction to ontological engineering. *New Generation Computing* 21(4). 365–384.

Mizoguchi, Riichiro. 2004a. Tutorial on ontological engineering, part 2: Ontology development, tools and languages. *New Generation Computing* 22(1). 61–96.

Mizoguchi, Riichiro. 2004b. Tutorial on ontological engineering, part 3: Advanced course of ontological engineering. *New Generation Computing* 22(2). 193–220.

Moltmann, Friederike. 2022. Natural language ontology. In Edward N. Zalta & Uri Nodelman (eds.), *The Stanford encyclopedia of philosophy*, Metaphysics Research Lab, Stanford University Winter 2022 edn. https://plato.stanford.edu/archives/win2022/entries/natural-language-ontology/.

Murphy, M. Lynne. 2010. *Lexical meaning*. Cambridge: Cambridge University Press.

Palmer, Martha, Daniel Gildea & Nianwen Xue. 2010. *Semantic role labeling*. California: Morgan & Claypool Publishers.

Paradis, Carita. 2005. Ontologies and construals in lexical semantics. *Axiomathes* 15(4). 541–573.

R Core Team. 2022. *R: A language and environment for statistical computing*. Vienna. https://www.R-project.org/.

Ruppenhofer, Josef, Michael Ellsworth, Miriam R. L. Petruck, Christopher R. Johnson, Collin F. Baker & Jan Scheffczyk. 2016. *FrameNet II: Extended theory and practice*. Berkeley: FrameNet. URL: https://framenet2.icsi.berkeley.edu/docs/r1.7/book.pdf.

Wickham, Hadley & Garrett Grolemund. 2016. *R for data science: Import, tidy, transform, visualize, and model data*. California: O'Reilly Media, Inc.