Language, Cognition and Neuroscience

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/plcp21

Decomposing unaccusativity: a statistical modelling approach

Songhee Kim, Jeffrey R. Binder, Colin Humphries & Lisa L. Conant

To cite this article: Songhee Kim, Jeffrey R. Binder, Colin Humphries & Lisa L. Conant (25 Jun 2024): Decomposing unaccusativity: a statistical modelling approach, Language, Cognition and Neuroscience, DOI: 10.1080/23273798.2024.2368119

To link to this article: https://doi.org/10.1080/23273798.2024.2368119

LANGUAGE, COGNITION AND NEUROSCIENCE
https://doi.org/10.1080/23273798.2024.2368119

REGULAR ARTICLE

Decomposing unaccusativity: a statistical modelling approach

Songhee Kim, Jeffrey R. Binder, Colin Humphries and Lisa L. Conant

Department of Neurology, Medical College of Wisconsin, Milwaukee, USA

ABSTRACT
While the two types of intransitive verbs, i.e. unergative and unaccusative, are hypothesised to be syntactically represented, many have proposed a semantic account where abstract properties related to agentivity and telicity, often conceptualised as binary properties, determine the classification. Here we explore the extent to which graded, embodied features rooted in neurobiological systems contribute to the distinction, representing verb meanings as continuous human ratings over various experiential dimensions. Unlike prior studies that classified verbs based on categorical intuition, we assessed the degree of unaccusativity by acceptability of the prenominal past participle construction, one of the unaccusativity diagnostics. Five models were constructed to explain these data: categorical syntactic/semantic, feature-based event-semantic, experiential, and distributional models. The experiential model best explained the diagnostic test data, suggesting that the unaccusative/unergative distinction may be an emergent phenomenon related to differences in underlying experiential content. The experiential model's advantages, including interpretability and scalability, are also discussed.

ARTICLE HISTORY
Received 14 November 2023
Accepted 5 June 2024

KEYWORDS
Unaccusativity; embodied cognition; experiential model; distributional semantics; statistical modelling

1. Introduction

1.1. The unergative/unaccusative distinction

This paper revisits two types of intransitive verbs in English, namely unergative and unaccusative verbs. As per the Unaccusative Hypothesis (Burzio, 1986; Perlmutter, 1978), intransitive verbs are not homogeneous but are divided into two classes, i.e. unergative, whose sole argument is an underlying subject, and unaccusative, whose sole argument is an underlying direct object. Numerous proposals have aimed to explain the unergative/unaccusative distinction using either syntactic or semantic accounts, or both. Here we explore this phenomenon under an embodied cognition framework, hypothesising that this distinction is rooted, at least partially, in the different ways the two types of intransitive events are *experienced*.

Although the unergative/unaccusative distinction was originally defined in syntactic terms, a semantic explanation of the phenomenon was suspected from the beginning. Perlmutter (1978) himself stated that whether a predicate is an unergative or an unaccusative can be predicted from semantic content. The large body of research that followed has varied in how much emphasis was placed on semantics vs. syntax to account for the phenomenon. A few researchers explained the distinction in syntactic terms (Froud, 1998; Rosen, 1984; van Hout, 2004), whereas others put forward a purely semantic account, not resorting to syntax at all (Kaufmann, 1995; Shannon, 1992; Van Valin, 1990; Zaenen, 1988). For instance, Van Valin (1990) accounts for the phenomenon solely on the grounds of aspectual and agentive characteristics of the predicates. On the other hand, some researchers deemed both syntax and semantics to be necessary to account for the phenomenon. Levin and Hovav (1994), for example, acknowledge the syntactic nature of the distinction, but emphasise the role of semantics as well, which they argue governs the mapping of the verb arguments onto the syntactic structure. Similarly, Randall and colleagues (2004) suggest that subtle differences in how verbs are assigned unergative vs. unaccusative status in German vs. Dutch is explained by a set of linking rules consisting of [± telic], [± actor], [± locomotion], and [± inherent] features.

Notably, some have abandoned a strictly binary classification scheme for the verbs in question, as the distinction between the two categories turns out not to be clear-cut in many cases. In particular, accumulated observations in the literature on the so-called unaccusativity diagnostics, such as auxiliary selection, *ne*-cliticization, *-er* suffixation, causative alternation, resultative construction, prenominal past participle, and so on, have made it increasingly clear that certain intransitive

CONTACT Songhee Kim songhee.kim@nyu.edu Department of Neurology, Medical College of Wisconsin, 8701 W Watertown Plank Rd, Milwaukee, 53226, WI, USA
Supplemental data for this article can be accessed online at https://doi.org/10.1080/23273798.2024.2368119.

verbs are neither strictly unergative nor strictly unaccusative, or show variable behaviour across languages (Rosen, 1984; Sorace, 2000; Zeyrek & Acartürk, 2014). With such observations, some researchers argued for a non-dichotomous classification of the intransitive verbs (Haegeman, 2020; Oh, 2005; Zaenen, 1988), which was also supported by a data-driven approach to the classification of intransitives (Surtani et al., 2011). Some researchers further argued that the unergative/unaccusative distinction has a *graded* nature, rather than being a dichotomy. For instance, Sorace (2000) proposes a "hierarchy" of intransitive verbs, from the categorically unaccusative ones, which tend to be telic and non-agentive, to those that are categorically unergative, which tend to be atelic and agentive. This continuum, according to Sorace (2000), not only accounts for why certain verbs show fuzzier patterns as to the unaccusative diagnostics within a given language, but also accounts for the patterns of auxiliary selection of intransitive verbs observed across several European languages (Table 1). A similar intuition is found in other literature as well. Baker (2019) implements the graded nature of the unaccusative/unergative distinction in English by postulating a functional hierarchy of semantic heads, i.e. [± control], [± initiation], [± state], [± change], and [± telic], the combination of which he argues accounts for the observed patterns of various unaccusativity diagnostics. The construction of a hierarchy based on formal properties is found in other works as well. Randall et al. (2004), for instance, explain the different classification schemes in German vs. Dutch by setting differential priorities for the linking rules they propose. As another example, Legendre (2007) proposes a set of constraints, under the framework of Optimality Theory (Prince & Smolensky, 2004), that governs the mapping from lexico-aspectual features to an auxiliary, whereby the different patterns of auxiliary selection across languages are derived from the different rankings of those constraints.

These accounts of the unergative/unaccusative distinction provide an in-depth analysis of the phenomenon but share some limitations. To begin with, most of these studies explain the hard-to-operationalise unergative/unaccusative distinction using constructs that are by no means easier to operationalise, including but not limited to: internal vs. external causation, initiation, control, state, change, directed change, agentivity, telicity, and so on. These notions are given formal definitions but are typically understood in the context of lexical examples that bear the relevant property in the clearest way, while avoiding more ambiguous cases. For example, telicity might often be introduced with a minimal pair such as *arrive* vs. *see*, where the former is a canonical telic verb and the latter a canonical atelic. Crucially, the usefulness of such a concept decreases rapidly as the analysis expands in scope from the most representative items to others that are more ambiguous (e.g. *remember*).

A second, related, common characteristic of previous accounts is that the theoretical distinctions on which they are based are often assumed to be binary, without much justification. For example, telicity is often considered a binary property, i.e. telic vs. atelic, although the variable nature of telicity, either at the word level (e.g. semelfactives) or at the phrasal level (e.g. *eat three sandwiches* vs. *eat sandwiches*), has been extensively discussed in the literature (Dowty, 1991; Engelberg, 1999; Krifka, 1989; Rothstein, 2008; Tenny, 1994). Taking agentivity as another example, a verb is often assumed to be either agentive or non-agentive, with no intermediate alternatives. Although such semantic constructs are typically assumed to be dichotomous, there is little empirical data to support such assumptions. Particularly given the less-than-categorical judgments observed around many such semantic variables (O'Bryan, 2003; Rissman et al., 2015; Van Hout, 1998), it may be useful to explore an alternative way of encoding semantic content as continuous variables, even though we do not yet have strong evidence that suggests their continuous nature. One aim of the current study is to explore the degree to which these and other semantic features manifest as dichotomous or continuous phenomena, using a coding scheme that allows for continuous representation. If those semantic properties *are* binary, their dichotomous nature should manifest even when using a continuous encoding method.

Another characteristic of some previous accounts, especially the theoretical ones, is the way the critical data are sourced. As mentioned earlier, unaccusativity diagnostics data play a central role in many theories on unaccusativity, in tandem with the semantic profiling of individual verbs, i.e. attribution of semantic properties to individual verbs. In many theoretical works, however, these data have been collected

Table 1. The auxiliary selection hierarchy (from Sorace (2000)).

Semantic Class	Example	Auxiliary Selection
Change of Location	*come*, *arrive*	Selects BE (least variation)
Change of State	*rise*, *wilt*	
Continuation of a Pre-Existing State	*remain*, *survive*	
Existence of State	*exist*, *seem*	
Uncontrolled Process	*waver*, *skid*	
Controlled Process (motional)	*run*, *walk*	
Controlled Process (nonmotional)	*work*, *play*	Selects HAVE (least variation)

primarily through introspection, i.e. either on the basis of the author's own judgment or on judgments anecdotally collected from native speakers, rather than using extrinsic methods such as behavioural experiments or population sample surveys (See Sprouse & Almeida, 2012 for a notable exception). This practice has to do, at least partially, with a long-standing assumption in linguistics about the categorical nature of linguistic judgments. That is, *if* linguistic judgments are binary (e.g. acceptable vs. unacceptable) *and* the judgments are shared between native speakers of the given language, then sampling a few responses from native speakers or introspection is sufficient to establish a claim. Crucially, however, whether judgments from unaccusativity diagnostics truly are dichotomous has not been confirmed on the basis of empirical data. There are some noteworthy exceptions that did employ extrinsic data collection methods. Zeyrek and Acartürk (2014), for example, evaluated Sorace (2000)'s hypothesis by collecting grammaticality judgments on two unaccusativity diagnostics, namely the prenominal participle construction and impersonal passives, for Turkish intransitive verbs, and a few others have also adopted population survey methods to collect diagnostics data (Allman, 2016; Baker, 2019; Huang, 2018). However, these studies have been limited by inclusion of only a small number of verbs (Allman, 2016; Huang, 2018; Zeyrek & Acartürk, 2014) or by survey methods that allowed only binary judgments (Baker, 2019). Aiming to overcoming these issues, we explore an alternative model of unaccusativity that is based on a set of operationalised semantic properties, which will be assessed via judgments from a larger, random sample of native speakers on a large number of intransitive verbs. Our approach also avoids the unverified assumption that either the unergative/unaccusative distinction or the semantic properties are necessarily binary in nature.

1.2. An experiential approach

Apart from the differences discussed so far, the most fundamental way in which our model differs from previous accounts is in the use of explicitly embodied semantic features to encode verb meaning. As noted above, prior accounts of the unergative/unaccusative distinction have appealed to properties such as agentivity and telicity that aim to capture abstract features of events. However, the seemingly universal status of the unergative vs. unaccusative distinction across languages might imply a deeper root for this division, namely a possibility that this distinction emerges from the universal ways our brains process basic sensory, motor, interoceptive, and affective phenomena, which we refer to here collectively as "experiential" phenomena. In fact, there is accumulated evidence on event processing in non-human animals showing that similar semantic role attributions and causality processing are found in animals as well (see Wilson et al., 2022 for review), suggesting that constructs such as agentivity and telicity, which are central to the unergative/unaccusative distinction, are prelinguistic and might be grounded in non-linguistic aspects of events. More interestingly, research on event processing by nonhuman primates and infants has suggested that recognition of agentivity is based on cues from so-called low-level features such as motion, direction of motion, animacy, spatiotemporal patterns or contingencies, and causation (Craighero et al., 2011; Frith & Frith, 2010; Hauser, 1998; Kourtzi et al., 2008; Phillips et al., 2009; Schlottmann et al., 2009; Surian & Caldi, 2010; Wilson et al., 2022).

Support for the so-called experiential approach to verb meaning also comes from human neuroscience research, which has demonstrated that event representation in the human brain is grounded in sensory-motor experiences (Aflalo et al., 2020; Aziz-Zadeh & Damasio, 2008; Barsalou, 2008; Gallese & Cuccio, 2018; Hauk et al., 2004; Kemmerer et al., 2008; Pulvermüller et al., 2001; Raposo et al., 2009; Sidhu et al., 2014; Tong et al., 2022). Many fMRI studies have found that processing action verbs that engage various effectors (e.g. face, arm, leg, etc), for example, activates the corresponding motor cortex (Aziz-Zadeh et al., 2006; Boulenger et al., 2009; Hauk et al., 2004; Kiefer et al., 2012; Tettamanti et al., 2005; Willems et al., 2010), which was also confirmed at the single neuronal level (Aflalo et al., 2020). Outside the motor domain, neuroimaging research has more broadly pointed out the importance of modality-specific experiential information in accounting for the semantic representations of concepts in the brain, which turn out to be widely distributed over the cortices, probably engaging the neural circuits that are involved with the actual execution, and the acquisition, of the concept in question (Anderson et al., 2017; Carota et al., 2017; Fernandino et al., 2016; Fernandino et al., 2022; Kemmerer et al., 2008; Thompson-Schill, 2003; Tong et al., 2022). Finally, language acquisition research has suggested that sensory-motor experiences play an important role in children's acquisition of language and event concepts (Pexman, 2019; Sloutsky, 2010; Wellsby & Pexman, 2014).

Considering these findings, a possibility emerges that the unergative/unaccusative distinction, a seemingly universal phenomenon in natural language, might be better explained by a model that encodes the ways events are experienced through our senses, i.e. sensory, motor, and other basic biological mechanisms, from which notions such as agentivity and other characteristic differences between unergative and unaccusative events might arise. As an illustrative example, change of state verbs (a canonical unaccusative class) and body movement verbs (a canonical unergative class) are typically differentiated by telicity and agentivity; that is, change of state verbs (e.g. *freeze*, *shatter*, *wilt*) tend to be telic and non-agentive, whereas the body movement verbs (e.g. *jog*, *crawl*, *swim*) tend to be atelic and agentive. However, telicity and agentivity are not the only traits that distinguish the two classes. Various sensory dimensions typical of inanimate objects, such as colour, texture, weight, and so on, might be salient to change of state events, whereas body movement events are characterised by the presence of visible and proprioceptive motion, body parts, spatial proximity, spatial navigation, and so on. Those covarying experiential characteristics might contribute to representation of the two syntactic categories, and to event-semantic properties that are more abstract.

To explore this hypothesis, we employ a semantic model proposed in Binder et al. (2016), which represents word meaning as a set of approximately 65 experiential attributes, each of which is believed to have a corresponding neurobiological representation. The idea of encoding word meaning as a collection of sensory and perceptual features is not new. Lynott and Connell (2009, 2013), for example, assessed the meaning of 423 adjectives and 400 nouns, respectively, by collecting human ratings on the involvement of the five major perceptual modalities (i.e. visual, haptic, auditory, olfactory, and gustatory) in experiencing those concepts. Gainotti and colleagues (Gainotti et al., 2009, 2013) added a motor dimension to the basic perceptual dimensions and collected ratings from Italian native speakers on 49 concepts from animal, plant, and artifact categories. Similarly, Hoffman and Lambon Ralph (2013) used 8 sensory-motor modalities, i.e. colour, visual form, observed motion, sound, taste, smell, tactile, and performed actions, as identified by Cree and McRae (2003), to rate 160 living and non-living object nouns on their association with each feature. Lynott et al. (2020) offer perhaps the largest existing norming dataset, collecting ratings for 39,707 English words and phrases across six perceptual modalities (touch, vision, hearing, smell, taste, and interoception) and five action effectors (mouth/throat, hand/arm, foot/leg, head excluding mouth/throat, and torso).

Notably, some researchers have expanded the set of features to represent word meaning in a more comprehensive manner. For example, Crutch et al. (2012) included the dimensions of time, space, quantity, emotion, polarity (i.e. valence), social interaction, morality, and thought, along with the sensory-motor ones, to investigate the meaning of 200 abstract and 200 concrete nouns. Finally, Binder et al. (2016), whose feature set was adopted in the current study, defined an even broader set of 65 features, encompassing sensory, motor, spatial, temporal, affective, social, and causal dimensions, rating 535 words in the categories of noun, adjective, and verb. What distinguishes this model from other experiential models is that, as will be detailed below, the individual features in this model are defined based on the presence of a distinctive neural process, or neural representation, as attested by neuroimaging or other neurophysiological evidence (thus referred to by the authors as a "brain-based" model).

We use the Binder et al. (2016) semantic model for the current study due to its comprehensiveness and suitability to describe verb meanings, which is our focus. While most prior studies focused on sensory-motor features to describe mainly object concepts, as reviewed above, Binder et al.'s (2016) system includes components such as spatial, temporal, affective, cognition, and causal components, which are essential aspects of event concepts. The features belong to one of the following domains of experience: visual, somatosensory, auditory, gustatory, olfactory, motor, spatial, temporal, causal, social, cognitive, emotion, drive, and attention. Most of these domains are further subdivided into more specific modes of experience for which there is evidence of distinct neural representation. For instance, in the vision category, attributes such as luminance, size, colour, shape, visual complexity, visual motion, biological motion, and speed of motion were included as individual features, because each of them is suggested to have a corresponding neural processor. See Binder et al. (2016) for further details on feature selection and supporting evidence. Some of these features, especially those in the vision and motor categories, also likely overlap with some of the so-called "low-level" features that have been shown to play a role in the perception of actual events, as discussed in the event perception literature (See Zacks et al. (2007) for review).

To Binder et al.'s (2016) original model, we added five new features that are designed to encode some of the traditional notions widely discussed in event semantics, namely, agentivity, telicity, transitivity, and stativity (Cruse, 1973; Dowty, 1979; Gruber, 1965, 1967; Hopper & Thompson, 1980; Vendler, 1957, 1967; Verkuyl, 1989), although they are by no means an exhaustive list for representing event meaning. With these features, one of our aims was to construct a model that contains only event-semantic features, reflecting prior linguistic accounts of the unergative/unaccusative distinction as closely as possible, and to compare this model against other models in terms of how well it accounts for the distinction. All of these features were given an operationalised definition (see Appendix B for the full list) and were assessed via human ratings using the same 7-point scale as for the other features. An important question arising regarding the event-semantic features is whether they should also be considered experiential features. Given the motivation of the Binder et al. (2016) model, where individual experiential dimensions are chosen based on their neurobiological distinctiveness, the same logic should be applied to the event semantic features. That is, whether they are experiential features or not depends on the neuroimaging evidence that suggests a distinct neural processor for each of them. As some studies have indeed suggested the existence of neural correlates for some of these variables (Dowty, 1979; Grewe et al., 2007; Meltzer-Asscher et al., 2015; Ntemou et al., 2021; Pulvermüller, 2018; Romagno et al., 2012; Spengler et al., 2009), along with the research discussed above that suggests a pre-linguistic representation of agentivity in non-human primates, we think it is not unreasonable to categorise them as experiential features. Having said that, it is clear that some experiential features are more complex than others, or emerge in the brain from combinations of more basic experiential representations. For example, as Binder et al. (2016) discuss, knowledge of agentivity can emerge from other kinds of more elemental experience, including perception of biological motion, perception of causality, affective and cognitive states, and so on.

A model of unaccusativity that includes simple experiential phenomena would have an advantage in dealing with the learnability problem (i.e. how do children learn whether a verb is an unergative or an unaccusative?) compared to purely syntactic and categorical semantic accounts, which, while proposing the linguistic properties that explain the distinction in question, do not tell us *how* those properties are learned. Given evidence that children seem to at least start distinguishing the two categories by the age of 2 (Costa & Friedmann, 2012; Friedmann, 2007; Wang et al., 2019), even in languages in which overt morphosyntactic markers of unaccusativity are sparse, influential theories of the acquisition of this distinction, such as semantic bootstrapping (Pinker, 1982) and the theta system (Reinhart, 2003), have highlighted the importance of semantic properties. Reinhart (2003, 2016), for example, derives the unergative/unaccusative distinction from two binary conceptual features, namely, perceived causation (±c) and mental state involvement (±m), the different combinations of which are argued to give rise to all thematic roles. Of note, due to the binary nature of the features, their combination would not explain the non-dichotomous pattern of the unergative/unaccusative distinction, which calls for a graded model rather than a categorical model. Given the growing literature on the development of infant constructs of animacy, intentionality, and causality over the first year of life based on perceptual-motor information or experiences (Di Giorgio et al., 2021; Kaduk et al., 2013; Kominsky et al., 2022; Muentener & Carey, 2010; Sommerville et al., 2005; Woodward, 1998), the experiential model could help to explain how young children learn this distinction.

1.3. Current study

The present study had two aims. The first was to create a dataset that demonstrates the performance of various intransitive verbs on one of the diagnostics of unaccusativity available in English, namely the prenominal past participle (PnPP) construction, which would allow an assessment of how categorical the distinction actually is. For 138 English intransitive verbs, data on the acceptability of the PnPP construction were collected via a crowdsource platform (see Section 2.2 for details). With these data in hand, the second aim was to assess the ability of various models to account for the observed diagnostics dataset. Starting with categorical models, we assessed whether, and to what extent, the acceptability judgments can be explained by *a priori* syntactic labels (i.e. unergative vs. unaccusative) or by *a priori* semantic class labels found in the literature. For the semantic class labels, we used the verb classification in Levin (1993) and the hierarchical verb classes in Sorace (2000), where the latter was specifically proposed to capture the graded behaviour of unaccusativity diagnostics from canonical unergative to canonical unaccusative. After evaluating these *a priori* linguistic models against the empirical data, we assessed the extent to which the judgments reflect experiential properties of these verb concepts. We used the brain-based experiential model of Binder et al. (2016), expanded with five additional event-semantic features, as discussed in the previous section. We first tested a model consisting only of event-semantic features to see how well these features can account for the data. Next, we tested the full experiential model, comprising the experiential feature set in Binder et al. (2016) and the newly added event-semantic features. These models were evaluated on how well they account for the diagnostics data relative to other models such as the *a priori* syntactic and semantic category models. Finally, the comparison was also made against a popular corpus-derived distributional word embedding model, GloVe (Pennington et al., 2014). Like other distributional models, GloVe represents both syntactic and semantic information about words based on patterns of co-occurrence with other words in a very large text corpus. Because it is derived directly from a language sample and has been shown to embed syntactic information, it may perform better than a purely semantic model in explaining the unergative/unaccusative distinction, which some authors consider a primarily syntactic phenomenon.

We note that a data-centred approach to the unergative/unaccusative distinction, using extrinsic data, is not entirely new. In the face of ever-increasing needs for various NLP tasks that require automated learning of various linguistic categories from data, computational linguists have attempted to infer the grammatical classes of intransitive verbs from externally available data, which would obviate the need for subjective linguistic analyses. Some researchers automatically classified intransitive verbs based on subcategorization patterns (Merlo & Stevenson, 2001; Schulte im Walde, 2006). Pross (2020) demonstrated that a model trained on word embeddings derived from co-occurrence statistics captured cues for unergativity and unaccusativity successfully, which is taken to support the validity of the distributional semantic model as a source for modelling the unergative/unaccusative distinction. In fact, this assumption, that the difference between unergative and unaccusative verbs is reflected in their co-occurrence statistics in large text data, is shared among all these data-cantered approaches. Our approach is unique in that we hypothesise that the difference between the two classes can be found in their experiential semantic content. By including the GloVe model in our study, we could investigate whether the naturalness judgements of the PnPP phrases are better explained by the experiential aspects or by word co-occurrence statistics.

2. Methods

2.1. Verb selection and syntactic classification

We chose 138 English intransitive verbs from VerbNet, 131 of which appear in Levin (1993). *A priori* syntactic classification, i.e. unergative or unaccusative, was assigned to the verbs according to the linguistics literature (Levin & Hovav, 1994; Perlmutter, 1978; Van Valin, 1990). If such classification was missing for a given verb, we manually classified it using PropBank (Palmer et al., 2005), which was used to extract the verb's argument specifications (e.g. whether an argument is an agent or a theme). In total, 60 verbs were classified as unergative, and 78 were classified as unaccusative (see Appendix A).

2.2. Prenominal past participle (PnPP) ratings collection

In tandem with the categorical syntactic classification, we assessed the unergativity/unaccusativity of the verbs using one of the unaccusativity diagnostics. From the few available diagnostics in English, which also include *out-* prefixation, *-er* suffixation, resultative construction, and causative alternation, we chose the Prenominal Past Participle construction, i.e. a noun phrase in which the noun is modified by a preceding past participle of a verb. This construction is reported to be well-formed with most unaccusative verbs (e.g. *a fallen leaf*) but not with most unergative verbs (e.g. **a sneezed boy*) (Haspelmath, 1994; Hoekstra, 2021; Levin & Rappaport, 1986; Mulder, 1992; Zaenen, 1993). We selected this diagnostic mainly because it is one of the few diagnostics of unaccusativity available in English, a language that lacks a grammatical marker of unaccusativity such as auxiliary selection, that can be applied to most verbs in a productive manner. That is, virtually every English verb has a past participle form, making the construction of the diagnostic feasible, whereas other diagnostics using a particular prefix or suffix might not work for some verbs due to lexical restriction inherent to the verbs or other idiosyncratic reasons. Also, the use of PnPP phrases as a diagnostic ensured a relatively easy-to-understand test compared to diagnostics based on the resultative construction or the causative/inchoative alternation.

That said, it must be emphasised that no existing diagnostic for unaccusativity gives rise to a classification that is entirely consistent with the intuition-based classification reported in linguistics, including the PnPP construction. For various idiosyncratic reasons, certain unaccusative verbs may not modify a noun as most unaccusative verbs do. For example, *die* is an unaccusative verb, but its past participle, *died*, is not used as a prenominal modifier (e.g. **a died soldier*), presumably due to the existence of the adjective *dead*. In addition, some researchers have observed that the PnPP construction works well for change of state verbs, but not as well for other types of unaccusative verbs, e.g. *the broken vase* vs.?*the remained girl* (Baker, 2019). Despite these cases where it seems to fail, we apply this diagnostic because our primary purpose is in *verifying* the unergative/unaccusative distinction with empirical data, quantifying the amount of difference both within and across the unergative and unaccusative verbs.

Three different PnPP phrases were created for each verb by a linguist. All three phrases took the form of *the* + [past participle] + [noun]. Among the three phrases, two had the noun in its singular form, and one used a plural form, e.g. *the frozen lake*, *the frozen ground*, *the frozen popsicles*. For a few verbs that rarely take a singular noun as a direct object, a plural noun was used for all three phrases, e.g. *the abounded rumors*, *the abounded opportunities*, *the abounded questions*.

Ratings on the naturalness of the PnPP phrases were collected via Amazon Mechanical Turk. Participants rated the naturalness on a scale from 1 to 5 (1 = "very unnatural", 5 = "very natural"). Each participant rated a total of 30 different phrases, in which no two phrases contained the same verb. Participants reported at the end of the survey whether they are a native speaker of English, and the responses from non-native speakers of English were excluded based on their answer. The final data were from 686 participants, and the number of responses per phrase ranged from 19 to 24, with 99.3% of the phrases having 20 or more ratings. Per verb, the number of responses ranged from 62 to 71 (mean = 68.0, SD = 1.8), which is a sufficiently large number to conduct a well-powered (e.g. 80%) study to investigate syntactic phenomena using a Likert scale, according to Sprouse and Almeida (2017)'s analysis. Before the analysis, the scores were averaged across participants and across the three phrases, yielding one score per verb.

2.3. Semantic classifications

We also assigned the verbs to *a priori* semantic classes using two resources, Levin (1993) and Sorace (2000), creating two different sets of semantic classifications. The rationale behind using Levin (1993) was as follows: (a) it is one of the most widely accepted verb classification systems; (b) it includes a large number of English verbs; and (c) its classification is based on argument alternation patterns, to which the unergative/unaccusative distinction is sensitive. It should be noted that verbs often belong to more than one class in Levin's classification, as they are often polysemous or take part in multiple argument alternation patterns. In such cases, we chose the class that seemed most representative of the core meaning of the verb. For example, the verb *fly* occurs in three verb classes: *verbs of sending and carrying*, *verbs of existence*, and *verbs of motion*. In this case, we selected *verbs of motion* as that is the most salient meaning when we think of this verb in isolation, without a context, even though the verb may participate in the various argument alternation patterns characteristic of verbs of existence and verbs of sending and carrying. Of the 138, 131 verbs were included in Levin (1993) and were assigned to one of 18 Levin verb classes using this method. Following the procedure in Section 3.2, we manually assigned the 7 remaining verbs to corresponding Levin categories. The results are in Table 2a. Note that classes varied substantially in size.

Table 2. Semantic classifications.

Verb Class	Count
a. Levin classification (1993)	
Change of State	30
Motion	24
Involving the Body	21
Existence	19
Communication	10
Emission	10
Dis/appearance and Occurrence	9
Aspectual	5
Body-Internal Motion	4
Combining and Attaching	2
Grooming and Bodily Care	2
Perception	2
Total	**138**
b. Sorace classification (2000)	
Change of Location	5
Change of State	45
Continuation of Condition	8
Existence of State	8
Uncontrolled Process	34
Controlled Motional Process	20
Controlled Nonmotional Process	18
Total	**138**

The second semantic classification we used is that of Sorace (2000), who postulated that verb classes lie on a continuum from the categorically unaccusative to categorically unergative (see Table 1). Since Sorace (2000) lists only a few example verbs under each class, our 138 verbs were manually classified into one of the Sorace classes by a consensus of three linguists. The resulting number of verbs in each category is presented in Table 2b. See Appendix A for the entire list of stimuli with their *a priori* syntactic and semantic classifications.

2.4. The experiential representation of verbs

The experiential model consisted of graded features representing domains or types of experience associated with each verb. As described above, our model expanded on the features in Binder et al. (2016), which spans across sensory (e.g. visual, auditory, tactile, olfactory, and gustatory), temporal, spatial, affective, and motor, causal, social, attentional, and evaluative features, with the addition of the five new features, i.e. Agentivity, Transitivity, Requires Energy Input (i.e. stativity), Dynamicity, and Telicity. The Complexity feature in Binder et al. was not included, as it is only relevant to noun concepts. The entire list of features, along with actual queries used to elicit ratings, is in Appendix B. In total, our model had 72 features, including 6 event-semantic features (i.e. the five event-semantic features and the feature "Caused" from the original set). Definitions of the event-semantic features are in Table 3.

The Dynamicity and Requires Energy Input features were intended to jointly capture the stativity of events. The Dynamicity feature is based on the classic definition of stativity in the linguistics literature, such that states are events that do not involve change (Dowty, 1979; Kearns, 2000; McClure, 1994), while the Requires Energy Input feature uses an operationalised definition of stativity, such that non-states, but not states, require energy input to maintain the event (Comrie, 1976). Telicity was defined as indicating whether an event has a logical (natural) endpoint or not, a conventional definition found in linguistic literature (Depraetere, 1995; Vendler, 1957).

The experiential ratings were collected through Amazon Mechanical Turk (www.mturk.com). Turkers were required to be a native speaker of English to participate. While they could participate in as many sessions as they wanted, they did not rate the same verb more than once. Compensation was $1 per session, with sessions taking on average 9.4 minutes to complete. At the beginning of each session, instruction appeared with concrete examples, familiarising participants with the task. Participants rated a single verb on all 72 features in a self-paced manner. On the screen, the verb for that session (e.g. *fly*) appeared at the top, followed by a query and two verb examples that would typically receive a high and a medium rating, respectively. For instance, for the feature "Loud," the following query was presented: *To what extent do you think of this verb as being associated with a loud sound*? This was then followed by verb examples that we predicted would receive a high rating (e.g. *erupt*) and a medium rating (e.g. *clang*). Each example was accompanied with a short explanation. Participants were asked to indicate the degree to which they think a particular feature is relevant to the meaning of the verb on a 7-point scale (0 = "not at all", 6 = "very much"). Notably, given that the intermediate points on the scale were not assigned any verbal description, the Likert-scale data were treated, in our analyses, as an interval data, not ordinal. The numerical interpretation of the Likert scales data, in fact, is not rare in experimental linguistics (See Schütze & Sprouse (2013) for more discussion), or in the fMRI studies that used word ratings to predict brain activation patterns (Anderson et al., 2017, 2019; Fernandino et al., 2022; Tong et al., 2022).

Table 3. The event-semantic features.

Feature	Description
Agentivity	The degree to which an event is actively or intentionally done.
Caused	The degree to which an event is about someone (or something) causing a change in something else.
Transitivity	The degree to which an event must be done to something or someone else.
Requires Energy Input	The degree to which physical or mental energy is required for the event.
Dynamicity	The degree to which an event is about a change or activity, mental or physical, as opposed to an unchanging state.
Telicity	The degree to which an event has a defined state of completion.

It should be noted that, unlike in Binder et al. (2016), where each word was presented with a sentential context, verbs in the current study were presented in isolation, without any sentence context that clarifies its meaning further. The decision not to provide a context was made because such context can bias judgments of event semantics of a verb toward those of the presented context (e.g. the telicity of a verb could be judged as high in the presence of an expression referring to an endpoint when that verb is atelic in the absence of such context). Finally, our study included five "catch trial" items to test whether participants were paying attention to the task. Catch trials were formulated similarly to other queries, but only one answer was correct, e.g. *to what extent do you think of this verb as being an object that is smaller than a shoe box and larger than a mountain*?

We collected 4464 responses from 1224 participants, who, on average, rated 3.4 verbs (SD = 2.7). The number of ratings per verb ranged from 29 to 40, with a mean of 32.3 and standard deviation of 1.3. After data cleaning, described below, the number of ratings per verb was 28.1, which was more than or similar to the number reported in similar word rating studies (Binder et al., 2016; Lynott et al., 2020).

The ratings were cleaned using the following criteria. First, sessions with 3 or more incorrect answers on the catch trials, or uniform answers for every feature, were excluded. In addition, responses from participants who reported themselves as non-native English speakers were also excluded. These criteria removed 31 of the 4464 responses. To assure the quality of data, we performed further data cleaning based on inter-subject agreement. For each verb, we calculated the mean score of each feature across all raters, then correlated each rater's 72 feature ratings with the vector of group mean scores for that verb. Responses with a Pearson's correlation coefficient less than 0.5 were excluded. This yielded a final dataset of 3879 responses, rejecting 13.1% of the total responses. The number of responses per verb in the final dataset ranged from 18 to 36, with 99% of the verbs having 20 or more responses, with a mean of 28.1 responses. As in Binder et al. (2016), three sets of features were collapsed into one (Heavy and Light -> Weight, Hot and Cold -> Temperature, Smooth and Rough -> Texture) resulting in a final set of 69 features.

The remaining data were analysed to find potential redundancy within the feature set. For all possible combinations of features, we calculated Mutual Information (MI), which is a measure of mutual dependence between two variables (See Figure 1). MI is equal to zero if and only if two random variables are independent. Overall, MI was very low, with a mean of 0.082 across the pairs of features. The highest MI was for the Biomotion – Body pair (MI 1.007), with the second highest being for Pleasant – Happy (MI 0.906).

2.5. Distributional model

In addition to the four models we constructed, we tested a distributional model of semantic content using GloVe word embeddings, which reflect word-word co-occurrence patterns from a corpus (Pennington et al., 2014). We used pre-trained GloVe word embeddings with 300 dimensions. A principal component analysis was performed to reduce the number of predictors, and a parallel analysis yielded 24 principal components that explained 55% of the variance in the original dimensions.

2.6. Plan of analysis

We investigated how much variance in the PnPP scores is accounted for by each of the five types of models (i.e. categorical syntactic, categorical semantic, event semantic, experiential, and distributional) by fitting a linear

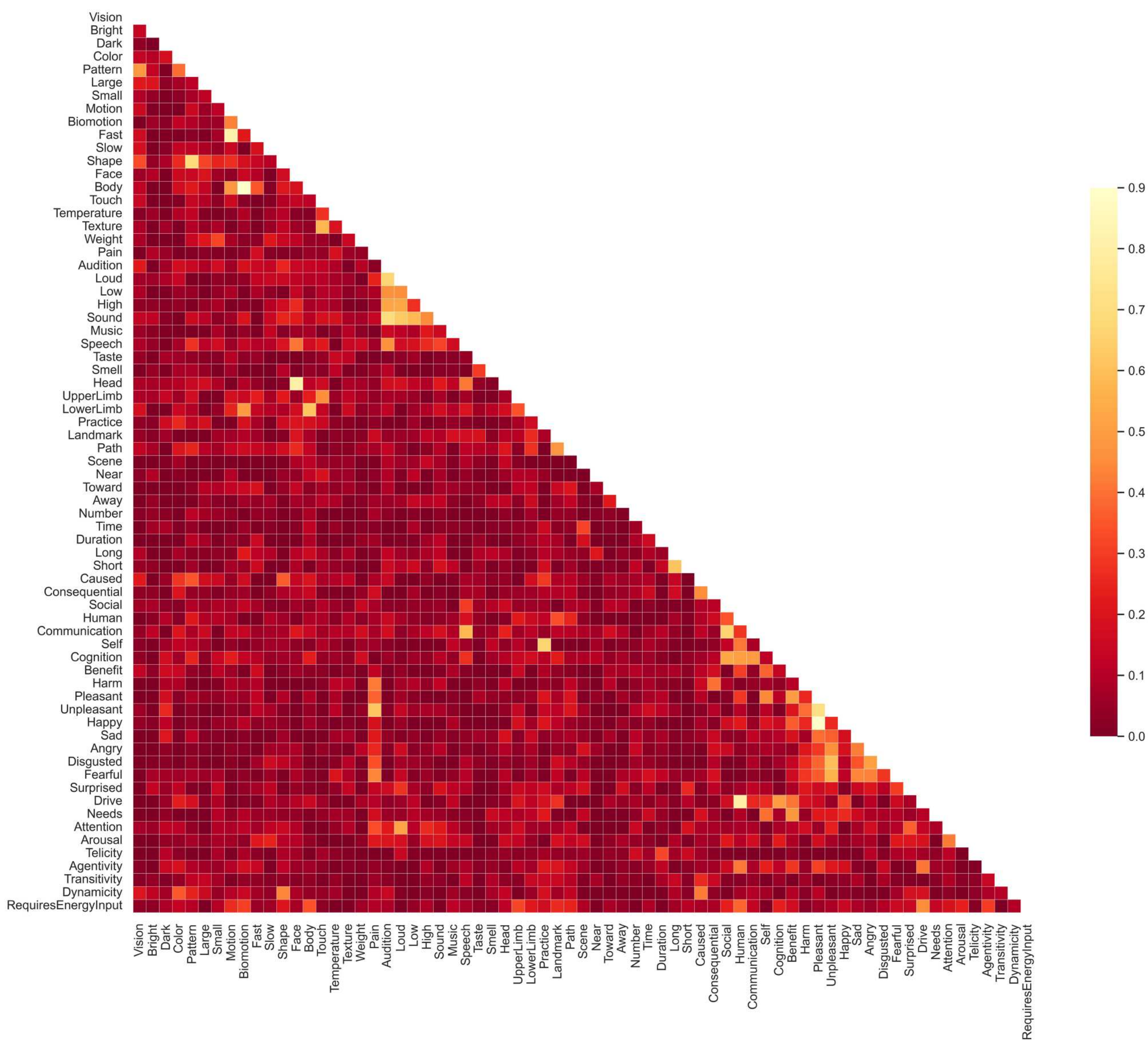

Figure 1. Mutual Information of 69 features. Mutual Information (MI) for all pairs of features, where 0 indicates complete independence of two features.

regression with either the categorical or graded variables in each model as predictors. With the experiential ratings dataset, two regression models were constructed, namely, a model with only event-semantic variables and a model with all experiential features (i.e. "full experiential model"). In all regression models, one predictor of no interest was included, namely, each verb's relative frequency of transitive usage (notated as FTU), calculated from VALEX (Korhonen et al., 2006). This variable was added to account for the variation in verbs' tendencies, sometimes for idiosyncratic reasons, to be used in the prenominal or transitive construction. By including this variable as a predictor of no interest, we effectively assessed how much variance in the PnPP scores is explained by the syntactic, semantic, or experiential predictors after the effect of their actual transitive usage is accounted for. Prior to model fitting, all gradable features were z-scored.

In addition to examining the amount of variance explained by each model, the generalizability of model performance to unseen data was evaluated using repeated 10-fold cross-validation (5 repetitions). The models were then evaluated by comparing the Root Mean Square Error (RMSE) of the prediction. The Akaike Information Criterion (AIC) measure was also calculated as a means of comparing various models' fit to the data, with the lowest AIC indicating the best fit to data. Adopting the general rule of thumb, a difference of more than 10 in AIC was considered lending no support for the model with the higher AIC (Burnham & Anderson, 2004). The *p*-values for the beta coefficients in each linear model were corrected for multiple comparisons via FDR.

3. Results

In the following we present the results of using each model to predict naturalness ratings on the PnPP diagnostic measure.

3.1. Syntactic category model

First, we examined how the PnPP scores differ by the two syntactic classes. As shown in Figure 2, the scores were indeed significantly different between the two categories ($t(134.27) = -7.43$, $p < 0.0001$), with the unaccusative verbs showing a higher mean score than the unergative verbs (mean for unaccusative = 3.39, mean for unergative = 2.55). However, the scores did not separate the two classes in a strictly categorical manner, and instead showed substantial overlap between classes. There was no significant difference in the magnitude of variance between the categories ($F(77,59) = 1.35$, $p = 0.22$).

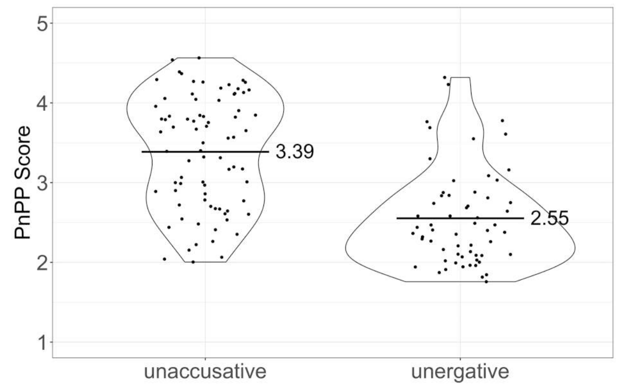

Figure 2. PnPP scores by syntactic category. Unaccusative verbs exhibited a higher mean score than unergative verbs, with the difference being statistically significant ($t(134.27) = -7.43$, $p < 0.0001$).

To examine how well the syntactic categories alone account for the data, we fit a linear regression model ("SynLM") with the syntactic category as a predictor and the naturalness score as the dependent variable. The regression was statistically significant (Adjusted $R^2 = 0.386$, $F(2, 135) = 44.12$, $p < 0.0001$). The regression coefficient for syntactic category was −0.68, meaning that forming a PnPP phrase with an unergative verb decreased the naturalness score by 0.68, on average. In other words, the PnPP scores were different across the two syntactic categories, with the unaccusative verbs being judged overall more acceptable in the PnPP construction than the unergatives, consistent with the literature. FTU was also significant, which is expected given the compatibility of PnPP constructions with transitive verbs. The model accounted for 38.6% of the variance in the judgment data, with only 17% uniquely accounted for by the syntactic class (see squared semi-partial correlation in Table 4). The AIC value of the model was 261.52, and RMSE of the cross-validation was 0.6059.

3.2. Levin semantic category model

We next examined how the PnPP scores varied across the Levin verb classes. While the verb classes in Levin (1993) were not devised to explain the unergative/unaccusative distinction, they are based mainly on argument alternation patterns, which purportedly reflect semantic properties that also underlie the unergative/unaccusative distinction.

Table 4. Linear regression result for syntactic model (SynLM).

	Regression coefficient	*P* value	Squared semi-partial correlation
Unergative	−0.677	< 0.0001	0.170
FTU	0.018	< 0.0001	0.114

Because 7 of the 138 verbs were missing in Levin's classification, we made a few amendments to the classification so that the Levin model can still be evaluated on a par with other models. First, the 7 verbs that do not appear in Levin (1993) (i.e. *circulate*, *converge*, *fail*, *meditate*, *persevere*, *retaliate*, *urinate*) were manually assigned a Levin class according to the class descriptions. Second, five Levin classes that had only one verb in them (i.e. Assuming a Position, Ingesting, Lingering and Rushing, Predicative Complements, and Social Interaction) were reassigned to the nearest category so as to make statistical analyses possible. A Kruskal–Wallis test, a non-parametric alternative to ANOVA, showed that the scores were significantly different across the semantic categories ($\chi^2(11) = 63.28$, $p < 0.0001$). Visual inspection of the data showed that the range of scores showed substantial overlap across semantic classes (Figure 3).

A linear regression with the Levin classes as predictors ("LevinLM") showed that 48.7% of the variance is explained by the model (Adjusted $R^2 = 0.487$, $F(12,125) = 11.90$, $p < 0.0001$). Among the classes, Change of State, Combining and Attaching, Existence, and Grooming and Bodily Care classes had a significant regression coefficient after FDR correction for multiple comparisons, all associated with an increase of the PnPP scores (See Table 5 for details). As before, FTU was also significant, uniquely accounting for 4.6% of the variance. Model comparison using AIC showed that the Levin model was better than the syntax model; AIC for SynLM = 261.52, AIC for LevinLM = 246.03. The RMSE of the cross validation was 0.5577.

3.3. Sorace classes

As a second semantic category model, we tested the Split Intransitivity Hierarchy (Sorace, 2000, 2011), which hypothesises that verb classes show a graded behaviour from core unaccusative to core unergative on morphosyntactic diagnostics of split intransitivity. Based on this theory, the PnPP naturalness ratings are expected to monotonically decrease from the core unaccusative, i.e. Change of Location, to the core unergative, i.e. Controlled Nonmotional Process (see Table 1).

As Sorace (2000) lists only a few verbs under each semantic class, most of our intransitive verbs were manually classified into one of the seven categories by a team of three linguists, based on the descriptions given in Sorace (2000) for each class. The result of the classification is in Table 2b (see Appendix A for the classification of all individual verbs). The PnPP scores for the Sorace categories are shown in Figure 4. As hypothesised, the naturalness of PnPP phrases generally decreases from the core unaccusative to the core unergative classes. The decrease, however, is not monotonic (e.g. Controlled Nonmotional Process has a higher mean than Controlled Motional Process), and the scores vary substantially within each class, showing considerable overlap between classes.

Using the Kruskal–Wallis test, we tested whether the differences in the PnPP scores across the Sorace categories were statistically significant. The scores were indeed significantly different across the semantic categories ($\chi^2(6) = 44.27$, $p < 0.0001$). We also fit a linear regression model ("SoraceLM") with the semantic classes as predictors. Here, the semantic classes were encoded as ordinal predictors, given that the dependent variable is hypothesised to monotonically decrease along the ordered predictors. Using the ordPens package in R (R version 4.3.1), the regression model penalised the sum of squared differences of adjacent dummy coefficients, effectively assuming the ordered nature of the semantic categories. A generalised additive model was fitted to quantify the association between the ordered predictors and the PnPP scores. The model was significant ($p < 0.0001$), with adjusted $R^2 = 0.413$. The Sorace model performed more poorly than the Levin model (i.e. AIC value of the Sorace model: 255.93, RMSE from the cross-validation: 0.6148). Compared to the syntactic model, it showed a larger RMSE.

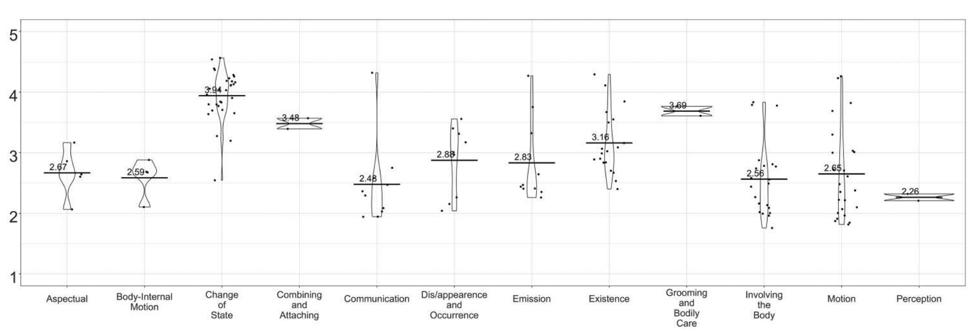

Figure 3. Naturalness scores by Levin classes. The PnPP scores varied significantly across the semantic categories, as demonstrated by a Kruskal-Wallis test ($\chi^2(11) = 63.28$, $p < 0.0001$).

Table 5. Linear regression result for Levin model (LevinLM).

	Regression coefficient	*p* val	Adj. *p* val	Squared semipartial correlation
Body-Internal Motion	0.236	0.543	0.626	0.001
Change of State	1.397	< 0.0001	< 0.0001	0.098
Combining and Attaching	1.119	0.020	0.049	0.021
Communication	0.125	0.697	0.697	0.001
Dis/appearance and Occurrence	0.757	0.032	0.065	0.018
Emission	0.440	0.167	0.287	0.007
Existence	0.875	0.004	0.018	0.031
Grooming and Bodily Care	1.171	0.014	0.043	0.023
Involving the Body	0.341	0.268	0.358	0.005
Motion	0.378	0.207	0.310	0.006
Perception	0.287	0.574	0.626	0.001
FTU	0.014	< 0.0001	0.004	0.046

In short, the unaccusativity diagnostics data shows an overall tendency that is consistent with the prediction of Sorace (2000, 2011). However, the linear model with Sorace categories as predictors did not perform much better than the categorical syntactic or Levin model, and it showed the greatest amount of error on unseen data among them.

3.4. The feature-based event-semantic model

We next investigated how the event-semantic features predicted the PnPP data. The following six features were included: Agentivity, Caused, Transitivity, Requires Energy Input, Dynamicity, and Telicity (see Table 3 for details). Among these, we predicted that Caused, Transitivity, and Telicity ratings would be positively associated with PnPP naturalness scores, and Agentivity would be inversely related to the scores. Requires Energy Input (coding stativity) and Dynamicity were not predicted to either increase or decrease the scores, as stative events are typically neither canonical unergatives nor canonical unaccusatives.

A multiple linear regression model ("EventLM") was built with these six features as predictors. The model accounted for significant variance (Adjusted $R^2 = 0.552$, $F(7,130) = 25.14$, $p < 0.0001$) (Table 6). While the polarity of the regression coefficients of the predictors was in the expected direction, only Caused was significant after FDR correction, being associated with an increase of 0.42 in PnPP naturalness. Agentivity moved the scores down, pushing the verbs toward the unergative side as predicted, but was not significant after correction for multiple comparisons. Telicity was not significant, either, contrary to prediction. Given that Agentivity and Telicity could interact when predicting the unaccusativity of the predicates (Baker, 2019; Levin & Hovav, 1994), we added an interaction term of Agentivity and Telicity to the model, which was also not significant ($p = 0.25$).

While only one predictor attained a corrected level of significance, the event-semantic model outperformed all the categorical models in terms of adjusted R^2, AIC, and prediction error on unseen data (AIC: 222.8, RMSE: 0.5173). Despite the overall good performance of

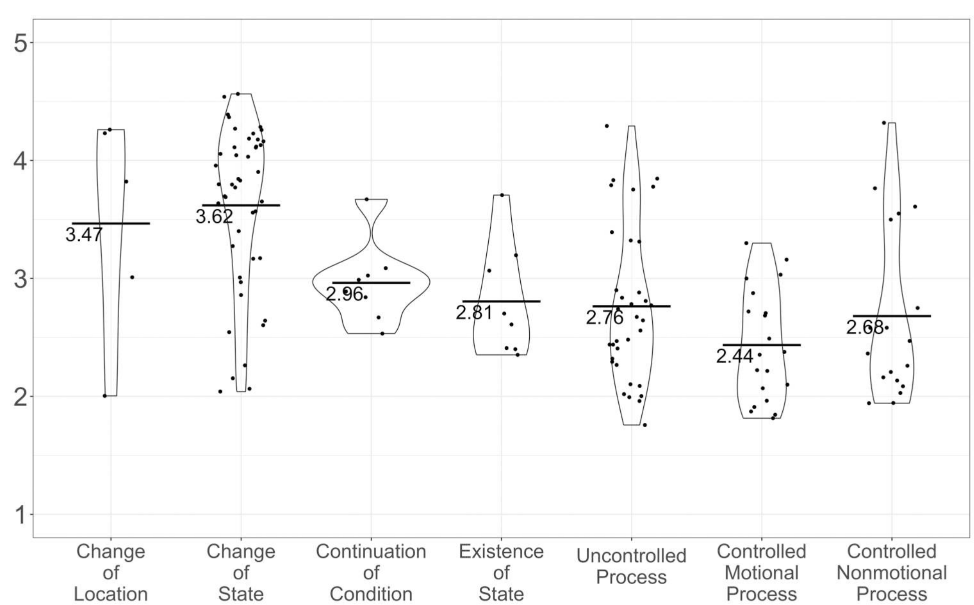

Figure 4. Naturalness scores by Sorace classes. The naturalness of PnPP phrases overall decreases from the core unaccusative to the core unergative classes as hypothesised by Sorace (2000), but the decrease is not monotonic. A Kruskal-Wallis test shows that the scores were significantly different across the categories ($\chi^2(6) = 44.27$, $p < 0.0001$).

Table 6. Linear regression result for event-semantic model (EventLM).

	Regression coefficient	*p* val	Adj. *p* val	squared semipartial correlation
Agentivity	−0.136	0.053	0.123	0.012
Caused	0.417	< 0.0001	< 0.0001	0.064
Transitivity	0.041	0.541	0.854	0.001
Requires Energy Input	−0.012	0.856	0.962	0
Dynamicity	−0.004	0.962	0.962	0
Telicity	0.026	0.610	0.854	0.001
FTU	0.178	0.001	0.003	0.038

the model, the variance attributable uniquely to each variable (i.e. the squared semipartial correlations) summed to only 11.6%, suggesting a large amount of variance jointly accounted for by the features. The feature-based event-semantic model shows that it is indeed possible to explain the unaccusativity diagnostics data based on graded, operationalised semantic features, with inputs from raters without expertise in linguistics.

3.5 The full experiential model

Given the large number of features in the full experiential model relative to the sample (stimulus set) size, along with the potential interdependence between some of the features, we first conducted a factor analysis to find latent dimensions (i.e. factors) in the experiential feature ratings data, which were then used as predictors in a linear regression.

A Kaiser-Meyer-Olkin test showed that the Measure of Sampling Adequacy was 0.75, validating a factor analysis for the data. Using principal axis factoring and the promax rotation method, 8 factors were extracted, as determined by a parallel analysis that kept factors with eigenvalues larger than the 95% percentile of the simulated data. The factors accounted for 71% of the variance in the ratings data (Figure 5).

Figure 5 shows the eight factors and each factor's five highest-loading features, along with ten verbs with the highest factor scores for that factor. As shown here, the affective factor accounted for the most variance in the experiential ratings, i.e. 17.17%. Among the eight factors, some factors (e.g. Sound, Change of State, and Movement) are reminiscent of the verb classes that have been implicated in the unaccusativity literature. Change of state verbs are canonically unaccusative, as already mentioned, and verbs involving sound or movement have been argued as subclassifying further, depending on various aspects of their lexicalised meaning (Baker, 2019; Levin & Hovav, 1994). For instance, verbs of sound emission (e.g. *jingle*) tend to be unaccusative, whereas verbs of communication (e.g. *speak*), which also often involve sound production, tend to be unergative (Levin & Hovav, 2005; Levin & Song, 1997; Sorace, 2000). With respect to movement, verbs that specify manner of motion (e.g. *run*) tend to be unergative, whereas those specifying direction of motion (e.g. *arrive*) tend to be unaccusative (Levin & Rappaport Hovav, 1992; Sorace, 2000).

Notably, the factor analysis reveals several dimensions that are largely overlooked in the semantics

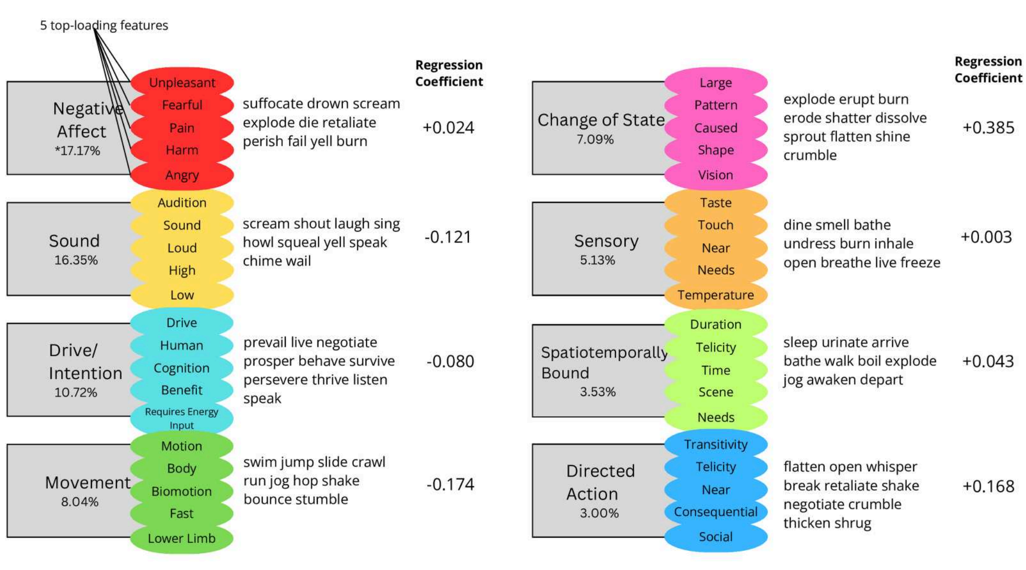

Figure 5. Factors in the experiential representation. Eight factors and their five highest-loading features, along with the ten verbs having the highest factor scores for each factor.

literature. For example, the factor that we call Spatiotemporally Bound seems to capture activities of daily living that typically happen in a specific setting (such as *bathroom* for *urinate* and *bathe*), have a typical duration, and meet basic needs. The factor we label Directed Action seems to indicate events with an inherently transitive quality, such as events involving communication and object-directed action, even if they syntactically realise as intransitives.

We fitted a linear regression model with these eight factors as predictors ("ExpFactorLM"). This model accounted for 59.71% of the variance in the data (Adjusted $R^2 = 0.597$, $F(9,128) = 23.56$, $p < 0.0001$) (Table 7). After FDR correction, Change of State, Movement, Directed Action, and Sound were statistically significant, with Change of State accounting for the greatest amount of variance. As shown in Table 7, Change of State and Directed Action were associated with an increase in the PnPP naturalness score, consistent with semantic theories that classify change of state verbs and alternating verbs (e.g. *open*) as unaccusative (Levin, 1993; Sorace, 2000). Sound was associated with a decrease in score, which might be an influence of communication verbs in our list, as demonstrated in the top ten verbs with the highest score for this factor. Similarly, Movement was associated with a decrease in score, which, again, might be an influence of verbs of motion that specify manner of motion. The overall result also clearly shows that not every dimension in the experiential semantic representation is relevant to the unergative/unaccusative distinction. For example, the affective dimension does not predict the PnPP scores, although it is the factor that accounts for the most variance in the ratings dataset. The factor-based linear model had an AIC value of 210.1, which is lower than the other models, and had RMSE of 0.482, which was also better than the other models.

While the factor analysis has an advantage of reducing the data into a few underlying factors, it is possible that, due to the breadth of these factors, features that have different effects on the PnPP scores could potentially load on the same factor. For example, the factor Movement integrates both Biomotion and Motion, where Biomotion tends to specify the manner of motion more than Motion, which is a more general feature. As discussed earlier, verbs referring to movement are known to subclassify differently depending on the specification of manner, direction, or external vs. internal cause of movement (Levin, 1993; Levin & Rappaport Hovav, 1992). Given this potential limitation, we performed a post hoc analysis with individual features as regressors. For this analysis, we performed feature selection prior to model fitting, as there are too many features given the size of the verb set. First, the features whose ratings show significant heteroscedasticity based on the Breusch–Pagan test were excluded, which were the following 19 features: Vision, Large, Shape, Face, Texture, Pain, Speech, Head, LowerLimb, Landmark, Duration, Consequential, Cognition, Pleasant, Unpleasant, Happy, Sad, Telicity, and Transitivity. In the remaining feature set, the following six features were excluded so that no two features are correlated at or above a Pearson's r of 0.9: Body, Fast, Human, Audition, Loud, and Social. Here, we excluded the feature with a poorer fit to the PnPP ratings in each pair of features. With these selection criteria, the final model included 44 features.

With these features, we fitted a linear regression model using backward stepwise regression. This model will not be compared to the others in terms of its performance relative to the PnPP scores because the method used is intended to optimise the fit of the model based on the obtained data, while the other models are based on theoretical justification. The following predictors were maintained in the model: Caused, Agentivity, Attention, Toward, Near, Taste, Weight, and FTU. The linear model ("ExpBackLM") accounted for 64.6% of the variance (Adjusted $R^2 = 0.646$, $F(8,129) = 32.28$, $p < .0001$). As shown in Table 8, Caused was the strongest predictor and was associated with an increase in the PnPP score of 0.46. Among other predictors, Agentivity, Attention, and Taste were associated with a decrease in the PnPP score, and Weight was associated with an increase. Here, the feature Weight measured the degree to which a particular event is associated with something light or heavy. It turns out that the verbs that have high ratings on this feature tend to refer to non-agentive events with a theme or patient event participant (e.g. *float*, *sprout*, *drop*, *evaporate*, *bounce*, *fly*, etc). Verbs with low ratings on this feature, on the other hand, tend to have an animate event participant (e.g. *mumble*, *cringe*, *behave*, *awaken, arrive, yell*, etc). In other words, it seems that the saliency of

Table 7. Linear regression result for experiential factor model (ExpFactorLM).

	Regression coefficient	*p* val	Adj. *p* val	squared semi-partial correlation
Negative Affect	0.024	0.620	0.698	0.001
Sound	−0.121	0.012	0.021	0.019
Drive/Intention	−0.080	0.077	0.116	0.009
Movement	−0.174	< 0.001	0.001	0.044
Change of State	0.385	< 0.001	< 0.001	0.188
Sensory	0.003	0.942	0.942	0.000
Spatiotemporally Bound	0.043	0.369	0.474	0.002
Directed Action	0.168	0.001	0.002	0.034
FTU	0.012	< 0.001	0.001	0.039

Table 8. Linear regression result for experiential individual feature model (ExpBackLM)

	Regression coefficient	*p* val	Adj. *p* val	squared semi-partial correlation
Weight	0.156	< 0.001	0.002	0.033
Taste	−0.094	0.035	0.047	0.012
Near	0.076	0.142	0.142	0.006
Toward	−0.084	0.070	0.080	0.009
Attention	−0.131	0.002	0.005	0.026
Agentivity	−0.126	0.008	0.013	0.019
Caused	0.459	< 0.001	< 0.0001	0.244
FTU	0.145	0.002	0.005	0.025

weight in an event meaning is negatively related to the agentivity of the event (see Section 4.2 for more discussion), although the Pearson correlation coefficient between Weight and Agentivity does not achieve significance ($r = -0.14$; $t(136) = -1.539$, $p = 0.11$). Along the same lines, the features Taste and Attention seem to be associated with the presence of an animate event participant who can sense the taste (e.g. *dine*, *burp*, *smell*) and perform the attention-grabbing action (e.g. *scream*, *yell*, *shout*), respectively. It should be noted, however, that these features also capture unique aspects of the unergative/unaccusative distinction, separate from agentivity or animacy of the events, given that each feature accounts for unique variance beyond that accounted for by the Agentivity feature (i.e. squared semi-partial correlation in Table 8).

3.6. A distributional semantic model – GloVe word embeddings

The final model we examined was a distributional model, in which word meanings are represented by their co-occurrence statistics derived from a large text corpus. We used pre-trained GloVe word embeddings with 300 dimensions. Principal component analysis yielded 24 components that were used as predictors of the PnPP data in a multiple linear regression.

The regression model ("GlovePCALM") was significant and explained 56.3% of the variance (Adjusted $R^2 = 0.563$, $F(25,112) = 8.04$, $p < .0001$). Results in Table 9 include only the first 9 principal components for the sake of brevity. While two principal components were statistically significant, they are inherently uninterpretable due to the nature of distributional models. The AIC value of the GloVe model was 235.05, which was larger than that of the full experiential model or the event-semantic model. The RMSE was 0.5696, which was also larger than that of the experiential models.

Table 10 shows a summary of the accuracy of each model in accounting for variance in the PnPP acceptability scores. The full experiential model explained the largest amount of variance and had the lowest AIC and RMSE of all models. The syntactic model and the Sorace categorical semantic model performed notably worse than other models.

Table 9. Linear regression result for distributional model (GlovePCALM).

	Regression coefficient	*p* value	Adj.*P*	squared semi-partial correlation
PC1	0.099	< 0.0001	< 0.0001	0.271
PC2	0.050	< 0.001	0.004	0.044
PC3	0.027	0.061	0.258	0.011
PC4	0.017	0.244	0.469	0.004
PC5	0.006	0.715	0.972	0.000
PC6	−0.028	0.076	0.262	0.010
PC7	−0.015	0.334	0.556	0.003
PC8	−0.001	0.972	0.972	0.000
PC9	0.027	0.128	0.290	0.008
FTU	0.193	0.003	0.025	0.029

Table 10. Summary of all models.

Model	Acronym	N predictor	Adj. R^2	RMSE	AIC
Categorical syntactic	SynLM	2	0.3863	0.6087	261.5
Categorical semantic	LevinLM	11	0.4875	0.5738	246.0
	SoraceLM	7	0.4130	0.6148	255.9
Event-semantic	EventLM	6	0.5522	0.5345	222.8
Experiential	ExpFactorLM	8	0.5971	0.5105	210.1
Distributional	GlovePCALM	24	0.5625	0.5696	235.1

4. Discussion

The primary aims of this study were to describe how the unergative/unaccusative distinction is manifested in a large dataset from a diagnostic test of this distinction, and to compare a range of models on their ability to account for the observations. Given the cross-linguistic attestation of the unergative/unaccusative split, combined with evidence alluding to the possibility of pre-linguistic event recognition, we hypothesised that the distinction can be better explained by a model that represents verb meanings with graded features over multiple dimensions of experience than by categorical syntactic or semantic models that are detached from such aspects of events. Our results show that the diagnostic judgments *were* different between the two syntactic categories, consistent with its use as a diagnostic for this distinction. However, neither the categorical syntactic classification nor the two categorical semantic classifications explained the data optimally. Instead, the full experiential model accounted for the most variance in the data, and it showed the least error when predicting unseen diagnostic ratings data. Overall, our results demonstrate the validity of an experiential semantic approach for understanding a primarily syntactic phenomenon. In what follows, we discuss how this new approach confirms existing theoretical findings and broadens the current understanding of verb semantics, along with its implications.

4.1. Implications for linguistic theory

The results of analyses using the experiential model align closely with prior findings in the literature. For example, change of state verbs are one of the most canonical members of the unaccusative class, and in our factor analysis (cf. ExpFactorLM), the factor related to this aspect of events (named Change of State) turned out to be the strongest predictor and was associated with an increased acceptability score on the PnPP diagnostic, a characteristic of the unaccusatives. The analysis with individual features (cf. ExpBackLM) also found that agentivity and causality features were significant. Agentivity was associated with a decrease in the diagnostic score, consistent with prior proposals linking this feature with unergative verbs, and causality was associated with an increase in the score, consistent with the fact that causation is a central element of transitive and unaccusative events. These two features were also found significant in the event-semantics model (i.e. "EventLM"), although only causality survived the FDR correction. In sum, the significance of causality and agentivity was confirmed consistently throughout the feature-based approach.

Notably, we did not find a significant effect of telicity in any of the models we tested. This was unexpected given that telicity, along with agentivity, has been posited as a key factor in determining the unaccusativity of verbs. As previously mentioned, this absence of effect might be because the telicity ratings in our study were made not on verb phrases but on isolated verbs, at which level telicity cannot be precisely determined and often remains underspecified (e.g. *run* vs. *run a mile* vs. *run a business*). In fact, our data show that raters varied substantially on the telicity judgments, with ratings on this feature exhibiting the second highest inter-subject variability among all features (SD = 1.83). In addition, the complexity of the concept itself might contribute to high inter-subject variability. For example, telicity judgment seems to be more difficult for events that do not have a logical endpoint but still come to an end due to physical or biological constraints (e.g. *sleep*, *urinate*).

Another plausible explanation is that the role of telicity in the unergative/unaccusative phenomenon may have been overestimated, especially if theoretical claims have been formed based on a rather small selection of verbs exhibiting canonical atelicity or telicity. Our ratings of telicity on a wider range of intransitive verbs, in fact, suggest that telicity may be far from a categorical property (Figure 6). While our findings provide preliminary insights, a comprehensive understanding of telicity, and potentially other linguistic variables, calls for additional systematic investigation.

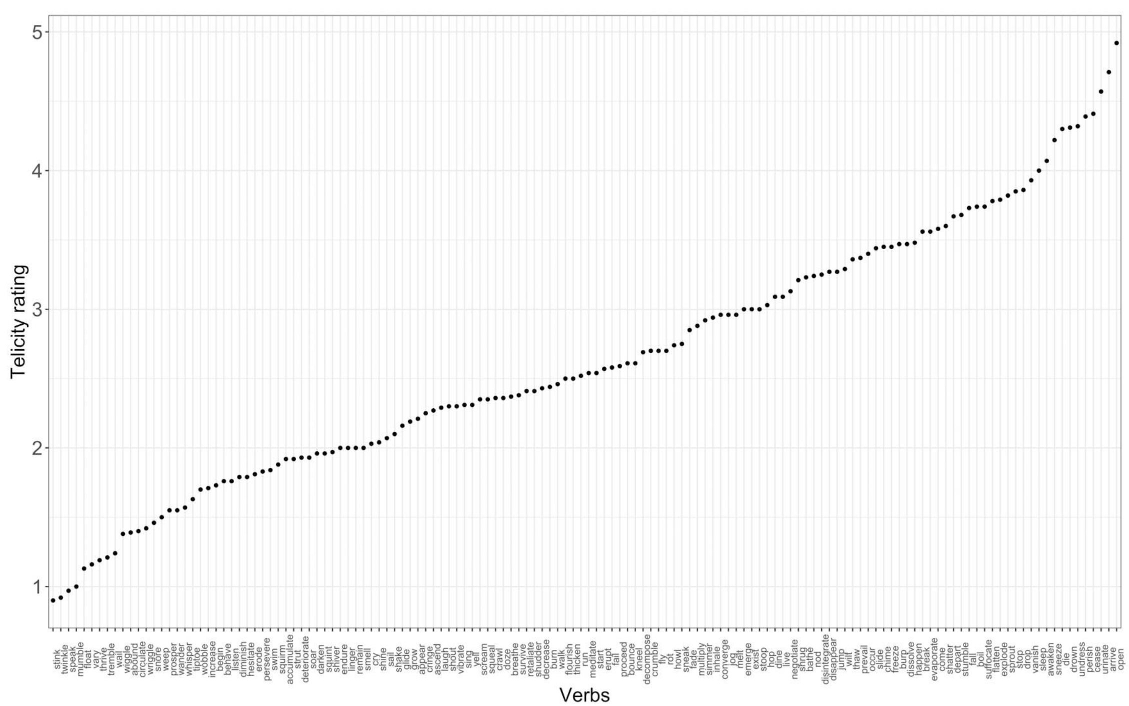

Figure 6. Ratings on Telicity

4.2. Novel findings from the experiential model

4.2.1. Experiential features relevant to the unergative/unaccusative distinction

Beyond the confirmation of existing theories, the experiential approach uncovers several novel dimensions of verb semantics that seem to underpin the unergative/unaccusative distinction. To begin, the individual feature analysis revealed that the saliency of taste and attention experiences are negatively correlated with PnPP acceptability, and that the saliency of weight experiences is positively correlated with PnPP acceptability. We propose that these types of experiences provide important cues as to the agentivity of the event subject. Taste-salient verbs typically indicate ingestion events (e.g. *eat*, *drink*), which entail a subject who actively ingests (and tastes). Similarly, attention-salient verbs typically refer to events in which an agent performs an action to attract attention (e.g. *scream*, *shout*), though clear exceptions exist (e.g. *explode*). In short, heightened saliency of taste or attention seems to be correlated with increased agentivity.

The link between weight salience and agentivity seems more complex. Weight salience measures the degree to which an event involves either very light or very heavy things. Given this definition, there are at least two ways weight salience serves as an indicator of agentivity. First, both extreme lightness and extreme heaviness tend to result in motion events induced by external forces, such as floating (induced by air currents) or falling (induced by gravity), wherein the event participant assumes a role of a theme argument, by definition. Second, from our ratings data, verbs with high ratings on Weight (e.g. *float*, *sprout*, *evaporate*, *flatten*, *thicken*, *accumulate*, *crumble*, *disintegrate*) seem to describe biological, physical, or other natural processes, which are often a function of weight or mass. In this case, the event participant undergoes a process, fitting the definition of a patient argument. In short, events with the weight dimension as a salient component of meaning are likely to involve a theme or patient argument, instead of an agent. This inference, however, may not apply to all events, particularly when the event can take multiple event participants (i.e. n-place predicates). For example, some transitive verbs encode an event where an agent interacts with a heavy or light object (e.g. *hoist*), in which case weight salience and agentivity are not mutually exclusive. This implies that the association between weight salience and agentivity might vary depending on the type of events or word categories.

Ultimately, the attribution of agentivity depends on a range of properties of the subject (e.g. mental capacity, mobility) that are largely inferred from observation, as proposed by Binder et al. (2016). Based on the current study's data, we propose that basic experiences involving taste, attention, and weight contribute to these inferences.

4.2.2. Latent dimensions of verb meaning

Aside from the individual features' association with the unergative/unaccusative distinction, our study also discovered a few latent dimensions in verb semantics. First, the Drive/Intention component from the factor analysis is characterised by high drive or motivation, human-like goals, cognitive thinking, benefit, and so on, according to the high-loading features for this factor. While this dimension has not been much discussed in the verb classification literature, it seems that many verbs that are represented by the Drive/Intention component (e.g. *prevail*, *prosper*, *survive*, *persevere, thrive*) belong to the stative verbs (Levin 1993; Sorace, 2000). Specifically, these verbs appear to describe a rather particular state, namely, a state of existence held by a cognitive being, that presumes some kind of effort to reach, or sustain, that state. While this subset of stative events has not been widely recognised, our experiential model provides a natural account of what constitutes this type of stative events, explaining how these stative events might be distinguished from more canonical members of the class, such as *exist*, *be*, or *remain*.

Some events turn out to be characterised by their typical spatiotemporal profile. The Spatiotemporally Bound component identifies a group of verbs with prominent temporal or locational scripts. This grouping primarily consists of: (i) daily routine verbs that typically have a predictable duration due to physical constraints (e.g. *sleep*, *urinate*, *bathe*, *jog*, *walk*, *awaken*) and (ii) verbs directly describing spatial displacement (e.g. *arrive*, *depart*). While these verbs have not traditionally been grouped together in prior literature, our findings suggest a novel experiential basis on which they might form a cohesive class.

Lastly, the Directed Action component seems to detect semantic (or conceptual) transitivity among intransitive verbs. This component seems to be robust in syntactically alternating verbs (e.g. *flatten*, *open*, *break*, *shake*, *crumble*, *thicken*) and social or communication verbs (e.g. *whisper*, *retaliate*, *negotiate*, *shrug*). In fact, these two groups of verbs appear to be quite dissimilar by syntactic and semantic grounds. However, the experiential approach finds a common ground on which they might share a certain characteristic, which we interpreted as semantic transitivity in a broader sense.

To summarise, the experiential semantic model reveals hidden facets of verb semantics, suggesting the possibility of an experience-based verb classification. Furthermore, it facilitates a detailed breakdown of verb meanings into experiential dimensions, which is not as readily possible with theories based on semantic or syntactic primitives. Determining whether these new dimensions should be considered semantic primitives will depend on their prevalence across a broader range of verbs than those examined in this study, a question that can be explored in future research.

4.3. The nature of event semantic properties

In this work, we represented event semantic properties as graded variables, departing from the common practice in linguistics, as discussed in the Introduction. Although we did not directly compare the graded and binary encodings of the linguistic variables, the event semantic model, comprising six graded features, outperformed both syntactic and semantic categorical models, as well as the distributional model. This suggests the validity of our new encoding scheme. Moreover, our data do not seem to support the binary status of the event-semantic variables. In fact, as per the ratings data, verbs might as well be argued to have varying *degrees* of an event semantic property. Taking agentivity as an example, an event of *mumble* (agentivity rating of 3.5) might be said to be *less* agentive than an event of *speak* (5.59), but *more* agentive than an event such as *snore* (0.42). Furthermore, the model with graded variables not only better fits the observed data, but also has a theoretical advantage in accommodating examples falling within intermediate ranges, which is a challenge for models assuming binary statuses for each property. For a more precise understanding of the nature of event semantic variables, future studies that directly compare models would be beneficial.

4.4. General implications

So far, we have discussed how the experiential model not only overlaps with linguistic theory but also introduces a new understanding of verb semantics where experiential dimensions play a central role. In the remainder of this paper, we revisit the cross-linguistic nature of the unergative/unaccusative distinction, reinterpret it in light of our new model, and discuss the overall implications of our work for future research.

The experiential approach in the current study was motivated in part by the cross-linguistic attestation of the unergative/unaccusative distinction. Overall, our results lend support to the idea that the two syntactic classes might indeed be grounded, at least partially, in differences in their experiential content. This is consistent with proposals that the theta system can be thematically driven (Reinhart, 2003), or that the unergative/unaccusative distinction might be based on so-called proto-roles, i.e. proto-agent and proto-patient, where each proto-role is explained as a set of distinct properties (Dowty, 1991). Given prior evidence on the contribution of embodied experience to development of language and concept processing (Baillargeon, 2004; Mandler, 1992; Wellsby & Pexman, 2014), it might indeed be the case that the acquisition of the unergative vs. unaccusative categories involves the experiential content of the events at least to some degree. Critically, however, our results do not entirely rule out the possibility that the experiential characteristics differentiating the unergative and unaccusative categories are an epiphenomenon, not the driving force behind category formation. We leave this question to future research.

Regarding the cross-linguistic status of the unergative/unaccusative distinction, our model offers a novel approach to understand their patterns. Firstly, the presence of unergative and unaccusative classes in many languages is naturally predicted, assuming the division is partly rooted in universal human experiences. In addition, the observed fuzziness of the distinction can be explained on the grounds that there are multiple components, not just one, that drive the distinction. That is, various experiential components, each with differing weights, collectively shape the cognitive representation of unaccusativity, positioning verbs along a continuum from canonical unaccusatives to canonical unergatives. Moreover, the language-specific patterns, i.e. the fact that different languages delineate the boundary between unergative and unaccusative at varying locations along the continuum, can potentially be explained in terms of the magnitude of individual weights (i.e. regression coefficients) assigned to experiential components in each language. In some languages, for example, the prominence of a movement component might lower the PnPP score more significantly than in other languages, leading movement-related verbs in that language to lean more towards the unergative territory compared to other languages. Such variations can be modelled by assigning weights of varying magnitude to the movement component.

Our work represents one of the few embodied approaches to a grammatical phenomenon. While there has been a substantial body of research in the embodied cognition literature (see Section 1.2), most of this research has focused on the semantic aspect of concepts, with little exploration of putatively syntactic phenomena (but see Brentari et al., 2015; Maffongelli et al., 2019; van Dam & Desai, 2016). Here we present a case where a linguistic phenomenon with a syntactic implication can be elucidated on the grounds of more elemental conceptual and semantic processes. A deeper understanding of these elemental processes might offer insights into other complex linguistic phenomena, particularly those observed cross-linguistically, which interacts with grammar but may have underlying semantic origins, as exemplified by the constraints on PnPP usage illustrated in the present study.

This work also enhances the current understanding of verb semantics. Particularly, our findings shed light on dimensions unique to events, which is made clearer when comparing our results with the findings of Binder et al. (2016), which encompassed nouns and adjectives along with verbs. For instance, in Binder et al. (2016), the factor accounting for the most variance was Vision/Touch, which is likely to reflect the significance of vision as a sensory modality. In contrast, the effect of the vision component was largely absent in our data, indicating a more limited role played by the visual system in verb meaning. However, a caution is required here, as the specific composition of factors can vary depending on the word stimuli included in a given study. For more meaningful investigations, future research should be conducted on a carefully selected, large-scale word set.

Lastly, the experiential componential approach has an advantage over categorical or distributed models from an engineering standpoint. That is, categorical models often draw heavily on linguistics expertise to assign syntactic/semantic properties to verbs, making it less feasible to extend the model to a novel word, and more costly to apply in NLP tasks. Conversely, the distributional model is extremely scalable but sacrifices any theoretical explanatory value, as the componential features are uninterpretable in a psychological or biological sense. The graded feature model lies between these two opposite ends; it is applicable to any word as long as ratings are available, and it is reasonably interpretable due to its operationalised definition of features. In short, the current model with graded, fully operationalised features is a viable model for future studies on linguistic phenomena at the intersection of syntax and semantics.

5. Conclusion

In this study, we have shown that an embodied model of semantics that integrates multiple experiential dimensions into a verb's semantic representation can be a useful model for exploring the unergative/unaccusative distinction. Compared to the categorical syntactic and semantic models described in previous linguistic studies, and to a distributional model of word meaning, the experiential model was better able to account for empirical unaccusativity diagnostics data, namely the acceptability of the prenominal past participle construction. The model provided by the experiential feature approach is also interpretable and readily scalable.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by National Institute on Deafness and Other Communication Disorders Grant R01-DC016622 to J.R.B. and Advancing a Healthier Wisconsin Project #5520462 to J.R.B.

Data availability statement

The data that support the findings of this study are openly available in OSF at http://doi.org/10.17605/OSF.IO/G5TB6.

References

Aflalo, T., Zhang, C. Y., Rosario, E. R., Pouratian, N., Orban, G. A., & Andersen, R. A. (2020). A shared neural substrate for action verbs and observed actions in human posterior parietal cortex. *Science Advances*, *6*(43), eabb3984. https://doi.org/10.1126/sciadv.abb3984

Allman, J. L. (2016). *Empirical Examination of Two Diagnostics of Korean Unaccusativity*. [PhD thesis]. The University of Texas at Arlington.

Anderson, A. J., Binder, J. R., Fernandino, L., Humphries, C. J., Conant, L. L., Aguilar, M., Wang, X., Doko, D., Raizada, R. D. S., & Sciences, C. (2017). Predicting neural activity patterns associated with sentences using a neurobiologically motivated model of semantic representation. *Cerebral Cortex*, *27*, 4379–4395. https://doi.org/10.1093/cercor/bhw240

Anderson, A. J., Binder, J. R., Fernandino, L., Humphries, C. J., Conant, L. L., Raizada, R. D. S., Lin, F., & Lalor, E. C. (2019). An integrated neural decoder of linguistic and experiential meaning. *The Journal of Neuroscience*, *39*(45), 8969–8987. https://doi.org/10.1523/JNEUROSCI.2575-18.2019

Aziz-Zadeh, L., & Damasio, A. (2008). Embodied semantics for actions: Findings from functional brain imaging. *Journal of Physiology Paris*, *102*(1–3), 35–39. https://doi.org/10.1016/j.jphysparis.2008.03.012

Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., & Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Current Biology*, *16*(18), 1818–1823. https://doi.org/10.1016/j.cub.2006.07.060

Baillargeon, R. (2004). Infants' reasoning about hidden objects: Evidence for event-general and event-specific expectations. *Developmental Science*, *7*(4), 391–414. https://doi.org/10.1111/j.1467-7687.2004.00357.x

Baker, J. (2019). Split intransitivity in English. *English Language and Linguistics*, *23*(3), 557–589. https://doi.org/10.1017/S1360674317000533

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Binder, J. R., Conant, L. L., Humphries, C. J., Fernandino, L., Simons, S. B., Aguilar, M., & Desai, R. H. (2016). Toward a brain-based componential semantic representation. *Cognitive Neuropsychology*, *33*(3–4), 130–174. https://doi.org/10.1080/02643294.2016.1147426

Boulenger, V., Hauk, O., & Pulvermüller, F. (2009). Grasping ideas with the motor system: Semantic somatotopy in idiom comprehension. *Cerebral Cortex*, *19*(8), 1905–1914. https://doi.org/10.1093/cercor/bhn217

Brentari, D., Renzo, A. D., Keane, J., & Volterra, V. (2015). Cognitive, cultural, and linguistic sources of a handshape distinction expressing agentivity. *Topics in Cognitive Science*, *7*(1), 95–123. https://doi.org/10.1111/tops.12123

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, *33*(2), 261–304. https://doi.org/10.1177/0049124104268644

Burzio, L. (1986). *Italian syntax: A government-binding approach* (Vol. 1). Springer Science and Business Media.

Carota, F., Kriegeskorte, N., Nili, H., & Pulvermuller, F. (2017). Representational similarity mapping of distributional semantics in left inferior frontal, middle temporal, and motor cortex. *Cerebral Cortex*, *27*, 294–309. https://doi.org/10.1093/cercor/bhw379

Comrie, B. (1976). *Aspect: An introduction to the study of verbal aspect and related problems (Vol. 2)*. Cambridge university press.

Costa, J., & Friedmann, N. (2012). Children acquire unaccusatives and A-movement very early. In Martin Everaert, Marijana Marelj, & Tal Siloni (Eds.), *The theta system: Argument structure at the interface* (pp. 354–378). Oxford University Press.

Craighero, L., Leo, I., Umiltà, C., & Simion, F. (2011). Newborns' preference for goal-directed actions. *Cognition*, *120*(1), 26–32. https://doi.org/10.1016/j.cognition.2011.02.011

Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General*, *132*(2), 163–201. https://doi.org/10.1037/0096-3445.132.2.163

Cruse, D. A. (1973). Some thoughts on agentivity. *Journal of Linguistics*, *9*(1), 11–23. https://doi.org/10.1017/S0022226700003509

Crutch, S. J., Williams, P., Ridgway, G. R., & Borgenicht, L. (2012). The role of polarity in antonym and synonym conceptual knowledge: Evidence from stroke aphasia and multidimensional ratings of abstract words. *Neuropsychologia*, *50*(11), 2636–2644. https://doi.org/10.1016/j.neuropsychologia.2012.07.015

Depraetere, I. (1995). On the necessity of distinguishing between (un)boundedness and (a)telicity. *Linguistics and Philosophy*, *18*(1), 1–19. https://doi.org/10.1007/BF00984959

Di Giorgio, E., Lunghi, M., Vallortigara, G., & Simion, F. (2021). Newborns' sensitivity to speed changes as a building block for animacy perception. *Scientific Reports*, *11*(1). Article 1. https://doi.org/10.1038/s41598-020-79139-8

Dowty, D. (1979). The semantics of aspectual classes of verbs in English. In Cleo Condoravdi, Orin Percus, & Zoltán Gendler Szabó (Eds.), *Word meaning and montague grammar: The semantics of verbs and times in generative semantics and in Montague's PTQ* (pp. 37–132). Springer.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, *67*(3), 547–619. https://doi.org/10.1353/lan.1991.0021

Engelberg, S. (1999). The magic of the moment: What it means to be a punctual verb. *Annual Meeting of the Berkeley Linguistics Society*, *25*(1), 109–121. https://doi.org/10.3765/bls.v25i1.3406

Fernandino, L., Binder, J. R., Desai, R. H., Pendl, S. L., Humphries, C. J., Gross, W. L., Conant, L. L., & Seidenberg, M. S. (2016). Concept representation reflects multimodal abstraction: A framework for embodied semantics. *Cerebral Cortex (New York, N.Y. : 1991)*, *26*(5), 2018–2034. https://doi.org/10.1093/cercor/bhv020

Fernandino, L., Tong, J.-Q., Conant, L. L., Humphries, C. J., & Binder, J. R. (2022). Decoding the information structure underlying the neural representation of concepts. *Proceedings of the National Academy of Sciences*, *119*(6), e2108091119. https://doi.org/10.1073/pnas.2108091119

Friedmann, N. (2007). Young children and A-chains: The acquisition of Hebrew Unaccusatives. *Language Acquisition*, *14*(4), 377–422. https://doi.org/10.1080/10489220701600523

Frith, U., & Frith, C. (2010). The social brain: Allowing humans to boldly go where no other species has been. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1537), 165–176. https://doi.org/10.1098/rstb.2009.0160

Froud. (1998). Aphasic evidence for the syntactic determination of unaccusativity. *UCL Working Papers in Linguistics*, 369–387.

Gainotti, G., Ciaraffa, F., Silveri, M. C., & Marra, C. (2009). Mental representation of normal subjects about the sources of knowledge in different semantic categories and unique entities. *Neuropsychology*, *23*(6), 803–812. https://doi.org/10.1037/a0016352

Gainotti, G., Spinelli, P., Scaricamazza, E., & Marra, C. (2013). The evaluation of sources of knowledge underlying different conceptual categories. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00040

Gallese, V., & Cuccio, V. (2018). The neural exploitation hypothesis and its implications for an embodied approach to language and cognition: Insights from the study of action verbs processing and motor disorders in Parkinson's disease. *Cortex*, *100*, 215–225. https://doi.org/10.1016/j.cortex.2018.01.010

Grewe, T., Bornkessel-Schlesewsky, I., Zysset, S., Wiese, R., Cramon, D. Y. v., & Schlesewsky, M. (2007). The role of the posterior superior temporal sulcus in the processing of unmarked transitivity. *NeuroImage*, *35*(1), 343–352. https://doi.org/10.1016/j.neuroimage.2006.11.045

Gruber, J. S. (1965). *Studies in lexical relations.* [PhD Thesis]. Massachusetts Institute of Technology].

Gruber, J. S. (1967). Look and see. *Language*, *43*(4), 937–947. https://doi.org/10.2307/411974

Haegeman, L. (2020). *Introduction to government and binding theory*. Wiley-Blackwell.

Haspelmath, M. (1994). Passive participles across languages. *Voice: Form and Function*, *27*.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*(2), 301–307. https://doi.org/10.1016/S0896-6273(03)00838-9

Hauser, M. D. (1998). A nonhuman primate's expectations about object motion and destination: The importance of self-propelled movement and animacy. *Developmental Science*, *1*(1), 31–37. https://doi.org/10.1111/1467-7687.00009

Hoekstra, T. (2021). *Transitivity: Grammatical relations in government-binding theory*. Walter de Gruyter GmbH & Co KG.

Hoffman, P., & Lambon Ralph, M. A. (2013). Shapes, scents and sounds: Quantifying the full multi-sensory basis of conceptual knowledge. *Neuropsychologia*, *51*(1), 14–25. https://doi.org/10.1016/j.neuropsychologia.2012.11.009

Hopper, P. J., & Thompson, S. A. (1980). Transitivity in grammar and discourse. *Language*, *56*(2), 251–299. https://doi.org/10.1353/lan.1980.0017

Huang, Y. (2018). *Linking form to meaning: Reevaluating the evidence for the unaccusative hypothesis*. [PhD Thesis]. Harvard University.

Kaduk, K., Elsner, B., & Reid, V. M. (2013). Discrimination of animate and inanimate motion in 9-month-old infants: An ERP study. *Developmental Cognitive Neuroscience*, *6*, 14–22. https://doi.org/10.1016/j.dcn.2013.05.003

Kaufmann, I. (1995). O– and D–predicates: A semantic approach to the Unaccusative–Unergative distinction. *Journal of Semantics*, *12*(4), 377–427. https://doi.org/10.1093/jos/12.4.377

Kearns, K. (2000). *Semantics*. St.

Kemmerer, D., Castillo, J. G., Talavage, T., Patterson, S., & Wiley, C. (2008). Neuroanatomical distribution of five semantic components of verbs: Evidence from fMRI. *Brain and Language*, *107*(1), 16–43. https://doi.org/10.1016/j.bandl.2007.09.003

Kiefer, M., Trumpp, N., Herrnberger, B., Sim, E.-J., Hoenig, K., & Pulvermüller, F. (2012). Dissociating the representation of action- and sound-related concepts in middle temporal cortex. *Brain and Language*, *122*(2), 120–125. https://doi.org/10.1016/j.bandl.2012.05.007

Kominsky, J. F., Li, Y., & Carey, S. (2022). Infants' attributions of insides and animacy in causal interactions. *Cognitive Science*, *46*(1), e13087. https://doi.org/10.1111/cogs.13087

Korhonen, A., Krymolowski, Y., & Briscoe, T. (2006). A large subcategorization lexicon for natural language processing applications. *LREC*, 1015–1020.

Kourtzi, Z., Krekelberg, B., & Van Wezel, R. J. A. (2008). Linking form and motion in the primate brain. *Trends in Cognitive Sciences*, *12*(6), 230–236. https://doi.org/10.1016/j.tics.2008.02.013

Krifka, M. (1989). Nominal reference, temporal constitution and quantification in event semantics. In R. Bartsch, J. van Benthem, & P. van Emde Boas (Eds.), *Semantics and contextual expression*. Foris Publications.

Legendre, G. (2007). On the typology of auxiliary selection. *Lingua*, *117*(9), 1522–1540. https://doi.org/10.1016/j.lingua.2006.06.005

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. University of Chicago press.

Levin, B., & Hovav, M. R. (1994). *Unaccusativity: At the syntax-lexical semantics interface (Vol. 26)*. MIT press.

Levin, B., & Hovav, M. R. (2005). *Argument realization (Vol. 10)*. university press.

Levin, B., & Rappaport, M. (1986). The formation of adjectival passives. *Linguistic Inquiry*, *17*, 623–661.

Levin, B., & Rappaport Hovav, M. (1992). The lexical semantics of verbs of motion: The perspective from unaccusativity. In I. M. Roca (Ed.), *Thematic structure* (pp. 247–270). De Gruyter Mouton. https://doi.org/10.1515/9783110872613.247

Levin, B., & Song, G. (1997). Making sense of corpus data: A case study of verbs of sound. *International Journal of Corpus Linguistics*, *2*(1), 23–64. https://doi.org/10.1075/ijcl.2.1.04lev

Lynott, D., & Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior Research Methods*, *41*(2), 558–564. https://doi.org/10.3758/BRM.41.2.558

Lynott, D., & Connell, L. (2013). Modality exclusivity norms for 400 nouns: The relationship between perceptual experience and surface word form. *Behavior Research Methods*, *45*(2), 516–526. https://doi.org/10.3758/s13428-012-0267-0

Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). The Lancaster sensorimotor norms: Multidimensional measures of perceptual and action strength for 40,000 English words. *Behavior Research Methods*, *52*(3), 1271–1291. https://doi.org/10.3758/s13428-019-01316-z

Maffongelli, L., D'Ausilio, A., Fadiga, L., Daum, M. M., Zwaan, R., & Fischer, M. (2019). The ontogenesis of action syntax. *Collabra: Psychology*, *5*(1), https://doi.org/10.1525/collabra.215

Mandler, J. M. (1992). How to build a baby: II. *Conceptual primitives. Psychological Review*, *99*(4), 587–604. https://doi.org/10.1037/0033-295X.99.4.587

McClure, W. T. (1994). *Syntactic projections of the semantics of aspect*. Cornell University.

Meltzer-Asscher, A., Mack, J. E., Barbieri, E., & Thompson, C. K. (2015). How the brain processes different dimensions of argument structure complexity: Evidence from fMRI. *Brain and Language*, *142*, 65–75. https://doi.org/10.1016/j.bandl.2014.12.005

Merlo, P., & Stevenson, S. (2001). Automatic verb classification based on statistical distributions of argument structure. *Computational Linguistics*, *27*(3), 373–408. https://doi.org/10.1162/089120101317066122

Muentener, P., & Carey, S. (2010). Infants' causal representations of state change events. *Cognitive Psychology*, *61*(2), 63–86. https://doi.org/10.1016/j.cogpsych.2010.02.001

Mulder, R. H. (1992). *The aspectual nature of syntactic complementation (Vol. 3)*. Holland Institute of Generative Linguistics.

Ntemou, E., Ohlerth, A.-K., Ille, S., Krieg, S. M., Bastiaanse, R., & Rofes, A. (2021). Mapping verb retrieval with nTMS: The role of transitivity. *Frontiers in Human Neuroscience*, *15*. https://www.frontiersin.org/articles/10.3389fnhum.2021.719461

O'Bryan, E. L. (2003). *Event structure in language comprehension*. [PhD Thesis]. The University of Arizona.

Oh, S. (2005). The case for split unaccusativity hypothesis. *Language Research*, *41*(1), 179–208.

Palmer, M., Gildea, D., & Kingsbury, P. (2005). The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, *31*(1), 71–106. https://doi.org/10.1162/0891201053630264

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language*

Processing (EMNLP), 1532–1543, https://doi.org/10.3115/v1/D14-1162

Perlmutter, D. M. (1978, September). Impersonal passives and the unaccusative hypothesis. In *Annual meeting of the Berkeley Linguistics Society* (pp. 157–190).

Pexman, P. M. (2019). The role of embodiment in conceptual development. *Language, Cognition and Neuroscience*, *34* (10), 1274–1283. https://doi.org/10.1080/23273798.2017.1303522

Phillips, W., Barnes, J. L., Mahajan, N., Yamaguchi, M., & Santos, L. R. (2009). Unwilling' versus 'unable': Capuchin monkeys' (Cebus apella) understanding of human intentional action. *Developmental Science*, *12*(6), 938–945. https://doi.org/10.1111/j.1467-7687.2009.00840.x

Pinker, S. (1982). A theory of the acquisition of lexical interpretive grammars. *The Mental Representation of Grammatical Relations*, 655–726.

Prince, A., & Smolensky, P. (2004). Optimality theory: Constraint interaction in generative grammar. *Optimality Theory in Phonology: A Reader*, 1–71.

Pross, T. (2020). Distributional semantics and the conceptual foundations of verb meaning: How neural word embeddings memorize the unaccusative hypothesis. *Empirical Issues in Syntax and Semantics*, *13*, 81–107.

Pulvermüller, F. (2018). The case of CAUSE: Neurobiological mechanisms for grounding an abstract concept. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1752), 20170129. https://doi.org/10.1098/rstb.2017.0129

Pulvermüller, F., Härle, M., & Hummel, F. (2001). Walking or talking?: Behavioral and neurophysiological correlates of action verb processing. *Brain and Language*, *78*(2), 143–168. https://doi.org/10.1006/brln.2000.2390

Randall, J., Van Hout, A., Weissenborn, J., & Baayen, H. (2004). Acquiring unaccusativity: A cross-linguistic look. In Artemis Alexiadou, Elena Anagnostopoulou, & Martin Everaert (Eds.), *The unaccusativity puzzle* (pp. 332–353). Oxford University Press.

Raposo, A., Moss, H. E., Stamatakis, E. A., & Tyler, L. K. (2009). Modulation of motor and premotor cortices by actions, action words and action sentences. *Neuropsychologia*, *47* (2), 388–396. https://doi.org/10.1016/j.neuropsychologia.2008.09.017

Reinhart, T. (2003). The theta system – an overview. *Theoretical Linguistics*, *28*(3), 229–290. https://doi.org/10.1515/thli.28.3.229

Reinhart, T. (2016). *Concepts, syntax, and their interface: The theta system (Vol. 72)*. The MIT Press. https://doi.org/10.7551/mitpress/8496.001.0001

Rissman, L., Rawlins, K., & Landau, B. (2015). Using instruments to understand argument structure: Evidence for gradient representation. *Cognition*, *142*, 266–290. https://doi.org/10.1016/j.cognition.2015.05.015

Romagno, D., Rota, G., Ricciardi, E., & Pietrini, P. (2012). Where the brain appreciates the final state of an event: The neural correlates of telicity. *Brain and Language*, *123*(1), 68–74. https://doi.org/10.1016/j.bandl.2012.06.003

Rosen, C. (1984). The interface between semantic roles and initial grammatical relations. *Studies in Relational Grammar*, *2*, 38–77.

Rothstein, S. D. (2008). Telicity, atomicity and the vendler classification of verbs. In S. Rothstein (Ed.), *Theoretical and crosslinguistic approaches to the semantics of aspect*. https://doi.org/10.1075/la.110.04rot

Schlottmann, A., Surian, L., & Ray, E. D. (2009). Causal perception of action-and-reaction sequences in 8-to 10-month-olds. *Journal of Experimental Child Psychology*, *103*(1), 87–107. https://doi.org/10.1016/j.jecp.2008.09.003

Schulte im Walde, S. (2006). Experiments on the automatic induction of German semantic verb classes. *Computational Linguistics*, *32*(2), 159–194. https://doi.org/10.1162/coli.2006.32.2.159

Schütze, C. T., & Sprouse, J. (2013). Judgment data. *In Research methods in linguistics*, 27–50.

Shannon, T. F. (1992). Split intransitivity in German and Dutch: Semantic and pragmatic parameters. In R. L. Lippi-Green (Ed.), *Recent developments in germanic linguistics (Vol. 93)* (pp. 97–113). John Benjamins Publishing.

Sidhu, D. M., Kwan, R., Pexman, P. M., & Siakaluk, P. D. (2014). Effects of relative embodiment in lexical and semantic processing of verbs. *Acta Psychologica*, *149*, 32–39. https://doi.org/10.1016/j.actpsy.2014.02.009

Sloutsky, V. M. (2010). From perceptual categories to concepts: What develops? *Cognitive Science*, *34*(7), 1244–1286. https://doi.org/10.1111/j.1551-6709.2010.01129.x

Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition*, *96*(1), B1–11. https://doi.org/10.1016/j.cognition.2004.07.004

Sorace, A. (2000). Gradients in auxiliary selection with intransitive verbs. *Language*, *76*(4), 859. https://doi.org/10.2307/417202

Sorace, A. (2011). Gradience in split intransitivity: The end of the unaccusative hypothesis? *Archivio Glottologico Italiano*, *XCVI*(1), 67–86.

Spengler, S., von Cramon, D. Y., & Brass, M. (2009). Was it me or was it you? How the sense of agency originates from ideomotor learning revealed by fMRI. *NeuroImage*, *46*(1), 290–298. https://doi.org/10.1016/j.neuroimage.2009.01.047

Sprouse, J., & Almeida, D. (2012). Assessing the reliability of textbook data in syntax: Adger's *Core Syntax*. *Journal of Linguistics*, *48*(3), 609–652. https://doi.org/10.1017/S0022226712000011

Sprouse, J., & Almeida, D. (2017). Design sensitivity and statistical power in acceptability judgment experiments. *Glossa: A Journal of General Linguistics*, *2*(1). https://doi.org/10.5334/gjgl.236

Surian, L., & Caldi, S. (2010). Infants' individuation of agents and inert objects: Individuation of agents and inert objects. *Developmental Science*, *13*(1), 143–150. https://doi.org/10.1111/j.1467-7687.2009.00873.x

Surtani, N., Jha, K., & Paul, S. (2011). Issues with the Unergative/Unaccusative classification of the intransitive verbs. *2011 International Conference on Asian Language Processing*, 55–58. https://doi.org/10.1109/ialp.2011.54

Tenny, C. (1994). *Aspectual roles and the syntax-semantics interface (1st ed., Vol. 52)*. Springer Dordrecht.

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S. F., & Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, *17* (2), 273–281. https://doi.org/10.1162/0898929053124965

Thompson-Schill, S. L. (2003). Neuroimaging studies of semantic memory: Inferring "how" from "where.". *Neuropsychologia*, *41*(3), 280–292. https://doi.org/10.1016/S0028-3932(02)00161-6

Tong, J., Binder, J. R., Humphries, C., Mazurchuk, S., Conant, L. L., & Fernandino, L. (2022). A distributed network for multimodal experiential representation of concepts. *Journal of Neuroscience*, *42*(37), 7121–7130. https://doi.org/10.1523/JNEUROSCI.1243-21.2022

van Dam, W. O., & Desai, R. H. (2016). The semantics of syntax: The grounding of transitive and intransitive constructions. *Journal of Cognitive Neuroscience*, *28*(5), 693–709. https://doi.org/10.1162/jocn_a_00926

Van Hout, A. (1998). On the role of direct objects and particles in learning telicity in Dutch and English. *Proceedings of the 22nd Annual Boston University Conference on Language Development*, *2*, 397–408.

van Hout, A. (2004). Unaccusativity as telicity checking. *The Unaccusativity Puzzle: Explorations of the Syntax-Lexicon Interface*, 60–83.

Van Valin, R. D. (1990). Semantic parameters of split intransitivity. *Language*, *66*(2), 221. https://doi.org/10.2307/414886

Vendler, Z. (1957). Verbs and times. *The Philosophical Review*, *66*(2), 143–160. https://doi.org/10.2307/2182371

Vendler, Z. (1967). *Linguistics in philosophy*. Cornell University Press. https://www.jstor.org/stable/10.7591j.ctvr7f2mf

Verkuyl, H. J. (1989). Aspectual classes and aspectual composition. *Linguistics and Philosophy*, 39–94. https://doi.org/10.1007/BF00627398

Wang, Z., Yang, X., & Shi, R. (2019). Mandarin-speaking toddlers' acquisition of unaccusativity. *Proceedings of the 43rd Boston University Conference on Language Development*, 705–715.

Wellsby, M., & Pexman, P. M. (2014). Developing embodied cognition: Insights from children's concepts and language processing. *Frontiers in Psychology*, *5*, 506. https://doi.org/10.3389/fpsyg.2014.00506

Willems, R. M., Hagoort, P., & Casasanto, D. (2010). Body-Specific representations of action verbs: Neural evidence from right- and left-handers. *Psychological Science*, *21*(1), 67–74. https://doi.org/10.1177/0956797609354072

Wilson, V. A. D., Zuberbühler, K., & Bickel, B. (2022). The evolutionary origins of syntax: Event cognition in nonhuman primates. *Science Advances*, *8*(25), eabn8464. https://doi.org/10.1126/sciadv.abn8464

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*(1), 1–34. https://doi.org/10.1016/S0010-0277(98)00058-4

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind/brain perspective. *Psychological Bulletin*, *133*(2), 273–293. https://doi.org/10.1037/0033-2909.133.2.273

Zaenen, A. (1988). *Unaccusative verbs in Dutch and the syntax-semantics interface (Vol. 123)*. CSLI/SRI International.

Zaenen, A. (1993). Unaccusativity in Dutch: Integrating syntax and lexical semantics. In James Pustejovsky (Ed.), *Semantics and the lexicon (Vol. 49)* (pp. 129–161). Springer.

Zeyrek, D., & Acartürk, C. (2014). The distinction between unaccusative and unergative verbs in Turkish: An offline and an eye tracking study of split intransitivity. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *36*(36), 1832–1837.