Aspectual Coercion and Lexical Semantics Part 1: Using Selection to Describe the Interaction between Construction and Verb Meaning

Laura A. Michaelis
Department of Linguistics, University of Colorado Boulder, Boulder, Colorado, USA
laura.michaelis@colorado.edu

Abstract

Coercion is an inferential strategy used to resolve conflict between an operator and its argument. Such conflicts are resolved in favor of the semantic requirements of the operator (Talmy, 2000). Jackendoff (1997) and De Swart (1998), among others, represent coercion through type-shifting operators that intervene between an aspectual operator and its situation-type argument, ensuring that the argument is of the appropriate type for the operator. This framework has a mapping problem: the rules that it uses to represent aspectual-type shifts simply replace one aspectual type (the input) with another (the output), so it does not explain how the input representation constrains the output representation. This article offers a solution to the mapping problem: treating aspectual type shifts as operations on the decomposed semantic representations of verbs. I will show that two such operations can capture both implicit and explicit aspectual type-shifts in English, involving both tense constructions and aspectual constructions.

Keywords
aspect – tense – coercion – lexical semantics – argument structure – Construction Grammar
1 Introduction

In compositional accounts of sentence meaning, a phrase has the internal composition and meaning that it does because of the meaning and combinatoric requirements of its head word. These frameworks have refined our understanding of the syntax-semantics interface, but they make certain incorrect predictions. For example, while state verbs do not license frequency adverbials, the sentence *She lived there twice* is readily interpreted, as describing two distinct episodes of residing in a given place. How are the ‘extra’ components of meaning that characterize such contextually enriched interpretations, which lack reflexes in morphosyntax, ‘read in’ during sentence interpretation? This question has generated a rich analytic tradition, with diverse applications—from computational semantics to typology to cognitive neuroscience of language. This tradition encompasses works by Verkuyl (1972, 1993), Moens and Steedman (1988), Jackendoff (1990, 1997), Pustejovsky (1995), Pustejovsky and Bouillon (1995), Kamp and Reyle (1993), Koontz-Garboden (2007), Kuperberg et al. (2010) and Piñango and Deo (2016). The inferential process in question is known in the literature by several names, including coercion (Moens and Steedman, 1988; De Swart, 1998; Pustejovsky, 1995; Michaelis, 1998, 2004, 2011), implicit conversion (Talmy, 2000) and external override (Smith, 1997: 53 et passim). The first term has gained general currency, so I will use it here. Coercion, according to De Swart (1998: 360), is “syntactically and morphologically invisible: it is governed by implicit contextual reinterpretation mechanisms triggered by the need to resolve [semantic] conflicts”. Although often considered as a formal semantic problem, coercion belongs to the study of grammar and cognition. For Talmy (2000), coercion is a consequence of a fundamental design feature of language—its partitioning into grammatical and the lexical subsystems. As he observes, some grammatical constructions effect lexical type shifts because they are designed to do so (e.g., partitive and progressive constructions) and some—called coercion triggers—simply happen to perform that function (when language users exploit the lexical selectivity of the construction):

Some of the grammatical forms in a language function specifically to perform a particular conversion operation. Others simply make structural specifications that can come into conflict with the specification of a neighboring lexical item. In the latter case, as discussed in the preceding section, the basic pattern is that the grammatical form’s specification takes precedence and triggers a kind of operation, a “shift”, in the lexical item’s referent that brings it into accord.
Coercion phenomena are difficult to reconcile with a head-driven model of semantic composition because they suggest that complements, e.g., the adverbial expression *twice*, are semantic selectors in the very same way that heads are. However, coercion phenomena find a natural home in frameworks based on representational modularity, e.g., Jackendoff (1990, 1997, 2010), Jackendoff and Audring (2020). In the modular framework, constraints on argument-operator combinations do not refer to syntactic categories like head. Jackendoff (1997: 49) proposes that syntactically transparent composition, as represented by head-driven semantic composition, is a default within a wider array of combinatorial options, which he refers to collectively as enriched composition.

Under enriched composition

[...]he conceptual structure of a sentence may contain, in addition to the conceptual content of its LCSs [lexical-conceptual structures], other material that is not expressed lexically, but that must be present in conceptual structure [...] in order to achieve well-formedness in the composition of the LCSs into conceptual structure.

Jackendoff, 1997: 49

When an entity in conceptual structure is not a suitable argument for a functor F, the process of composition interpolates a “coercing function” G to create instead the structure F(G(X)), where X is a suitable argument for G, and G(X) is a suitable argument for F (Jackendoff, 1997: 53). For example, in order to account for the fact that the NP *a beer* denotes a portion or variety of beer, we assume a coercing function that derives a count type from a mass type, making *beer* a suitable argument for the operator expressed by the indefinite article. The interpolated-function model successfully extricates two head properties—that of being a syntactic head, i.e., determining the distributional properties of the phrasal projection, and that of being a semantic head, i.e. calling for an argument of a particular type (Zwicky, 1985; Croft, 1996). In other words, the indefinite article, while it does not determine the syntactic category of its phrasal projection, can nevertheless be said to select a nominal sister denoting a bounded entity (Van Eynde, 2020). The type-shifting functions that Jackendoff describes are not specific to nominal syntax, or even to coercion: the types figure in the semantics of both entities and events and the functions are used to describe both implicit type shifts and type shifts effected by morphosyntactic constructions dedicated to specific type-shifting functions (e.g., the English partitive and plural). De Swart (1998), henceforth DS, uses this same general framework to describe the interaction of tense, grammatical aspect and Aktionsart during semantic composition. DS (1998: 348) represents sentential semantic structure as in (1):
As shown in (1), the model has a nested structure: tense operators scope aspectual operators, which in turn scope eventuality descriptions. Eventuality descriptions, or situation radicals, are predicate-argument combinations that can be assigned to specific Aktionsart classes. For example, the situation radical *She wins the race* belongs to the class of achievements (Vendler, 1957; Dowty, 1979; Van Valin and LaPolla, 1997). DS proposes a domain of eventualities comprising three basic types: states, processes and events. Cross-cutting categories are used to capture the fact that processes may be aligned grammatically either with states, on the basis of the subinterval property (Dowty, 1986), or with events, on the basis of dynamicity.1

Aspectual operators map eventuality descriptions onto other eventuality descriptions (see also Herweg, 1991). DS (1998: 368–369) distinguishes two types of aspectual operators: type-shifting operators, called shift constructions by Michaelis (1998, 2004, 2011), and type-sensitive operators, which Michaelis (1998, 2004, 2011) refers to as concord constructions. Type-sensitive operators are identity functions; they are used to represent grammatical concord relations, e.g., the relationship between the indefinite article *a* and a singular count term in English. Frame adverbials like *in ten minutes* are type-sensitive aspectual operators: they take telic event radicals as input, and output an event of this same type. Type-shifting operators are used to represent the meanings of inflectional and derivational constructions that ‘output’ a type distinct from that of the input lexical item, as in Sign-Based Construction Grammar (SBCG; Sag, 2012; Kay and Michaelis, 2019; Michaelis, 2012). For example, the English Plural construction in SBCG licenses a unary-branching local tree representing an inflectional rule that maps a bounded type to an unbounded type. An example of a type-shifting aspectual operator is the English Progressive construction: as shown by stativity diagnostics discussed in Section 2, Progressive sentences in English denote states. The ‘input type’, denoted by the past-participle complement, is by contrast a dynamic eventuality. Because both type-sensitive and type-shifting aspectual constructions call for verbs belonging to specific Aktionsart classes, and because such requirements create the potential for semantic conflict between operator and argument, the DS model correctly predicts that both type-sensitive and type-shifting operators are

---

1 For DS (p. 351), the category of events contains only those dynamic situations with an intrinsic endpoint, i.e., telic situations. Following Herweg (1991), I will use the term *event* more broadly to designate all dynamic situations, including processes (i.e., activities), using the compound term *telic event* to designate events that have intrinsic endpoints.
potential coercion triggers. For example, as shown by (2), the presence of the type-sensitive frame adverbial in ten minutes induces an inchoative (event) reading of the stative situation radical They be-bored. Example (3) shows that the type-shifting Progressive operator coerces a dynamic reading of the stative situation radical I live-on Pearl Street (DS, 1998: 363):

(2) They were bored in ten minutes.
(3) I am living on Pearl Street.

While aspectual operators locate the situation type relative to reference time, tense operators locate reference time relative to speech time. Accordingly, aspectual theorists, e.g., Herweg (1991), have assumed that tense operators apply regardless of aspectual class of the situation-radical argument. However, as DS demonstrates, certain tense operators, in particular the French imperfective past (Imparfait) and perfective past (Passé Simple or Passé Composé), display aspectual sensitivity: while the imperfective operator calls for a stative situation radical, the perfective operator calls for a perfective situation radical (DS, 1998: 368–375). As aspectual-type selectors, tenses are also potential coercion triggers. For example, the stative verb savoir (‘know’) denotes a state in the ‘concord condition’ in which it receives imperfective past-tense inflection (as in Je le savais ‘I knew it’), but an inchoative or episodic event in the ‘conflict condition’ in which its inflectional realization is perfective past: Je l’ai su (‘I understood it’).

Coercion operators perform type shifts analogous to those performed by aspectual constructions like the English Progressive, but differ from aspectual operators in two respects: (a) they lack reflexes in linguistic structure and (b) they are ‘macro-operators’, each of which subsumes a number of context-dependent semantic transitions (DS, 1998: 361). For example, the macro-operator C_{he} which maps homogeneous situations onto events, corresponds to two contextually invoked transitions: from state to bounded state, on the one hand, and from state to inchoative event, on the other. Similarly, the macro-operator C_{ds} which maps dynamic situations onto states, subsumes several stativizing functions, including iteration and habituality (DS, 1998: 383). By using macro-operators to represent coercion effects, DS captures the difference between semantic conflict and the various contextual effects that speakers achieve by inducing semantic conflict. The ultimate goal of the exercise is to ensure that the appropriate situation types enter the discourse model—while preserving the selectivity of operators that is crucial to their semantic definitions. It is in this sense that DS claims to provide a “fully compositional analysis of aspect shift in French and English” (DS, 1998: 373).
The aspectual mappings proposed by DS are defined over sets of eventualities in a standard set-theoretic way: situation types are expressed as \( n \)-place predicates whose extensions are the individuals that the predicate characterizes; type shifts are then described as implicational statements of the form “a certain individual belongs to the extension of a certain situation-denoting predicate if and only if that individual also belongs to the extension of some other situation-denoting predicate” (DS, 1998: 381–382). While formally precise, this framework does not describe the internal constitution of the eventuality descriptions that the mappings relate. For this reason, the model cannot be used as a framework for describing verb semantics. This is a problem because in a construction-based framework for grammar description like SBCG, operators capture a lexical relationship—that between a verb lexeme, defined in frame-semantic terms as a relation, and the word that realizes that lexeme in a particular morpho-syntactic context, e.g., as a present-tense verb.

To understand the nature of the problem, let us look at mappings that produce states in the DS model. One such mapping involves the iteration operator (\( \text{iter} \)), which is said to derive a state that “describes an unbounded number of eventualities of the type described by the predicate” (DS, 1998: 383). In what respect does an iterated accomplishment radical, e.g., *She greet-the customers*, qualify as a state, i.e., an unbounded and internally homogeneous eventuality? In Aktionsart-based classification, a situation consisting of a series of type-identical subevents, e.g., bouncing a ball or jumping up and down, qualifies as a dynamic situation—an activity in the Dowty-Vendler framework. Similarly, the pairing of an event radical with a frequency adverbial like *many times* yields not a state but an iterated event. Why then should the operator \( \text{iter} \) stativize, if in fact it does? The same question arises when we look at the habitual operator \( \text{hab} \), which DS describes as “mapping eventuality descriptions onto state descriptions” (DS, 1998: 383). This analysis makes sense in light of the stative properties that habitual predications display (Carlson, 2012). For example, habitual sentences denote situations which, like states, can extend to the present. This is shown by the fact that the conjoined past- and present-tense assertions in (4) are compatible, whereas those in (5) are not:

\[
\begin{align*}
(4) & \quad \text{She smoked back then and I think she still does.} \\
(5) & \quad \text{*They moved away and I think they still do.}
\end{align*}
\]

Habitual activity is also the default inference available to interpreters when they must reconcile perfective verbal Aktionsart with a grammatical context calling for an imperfective construal, as in (6) and (7):

\[
\begin{align*}
(6) & \quad \text{Michaelis}\quad \text{Michaelis} \\
(7) & \quad \text{Cognitive Semantics} \ 8 \ (2022) \ 383–408
\end{align*}
\]
(6) She smokes.
(7) She smoked when I met her.

In (6) and (7) the perfective situation radical *She smoke*- is expressed by a present-tense predication—a grammatical option not ordinarily available to event predications, as shown by the ill-formedness of (8):

(8) *Look! She smokes.

The *when*-construction in (7) entails that the denoted situation obtained prior to a past reference time (the time at which I met her)—again a hallmark of stative predications (Vlach, 1981; Herweg, 1991; Altshuler et al., 2019). If habitual events are states, as claimed by DS and others (e.g., Langacker, 1996, 1997), then the facts in (6) and (7) fall out. However, this view of habituality leaves open two basic questions. First, why should habitual situations be stative? They do not qualify as such on the basis of their internal composition, which is isomorphic to that of iterated events. If habituality does not entail stativity, we must determine what the semantic link between habitual situations and stative ones actually is. Second, what aspectual operators trigger the stative type shifts which lead, either directly or indirectly, to habitual readings? The only plausible sources of aspectual information in (6) and (7) are the tenses, and yet the English tenses are said to be aspectually neutral, apparently doing no more than locating reference time with respect to encoding time.

Certainly, typological studies attest to the semantic sympathy that exists between present tense and habitual aspect (see, e.g., Dahl (1995) and Bybee, Perkins and Pagliuca (1994: 151–153)), but if habituality is an implication, aspectual or otherwise, of the present tense, past-tense habitual predications remain unexplained. It seems circular to analyze past-tense habitual predications like (6) as *ipso facto* perfective, and factually incorrect too, since, as (7) shows, they denote extensible situations akin to states. Far from entailing perfective meaning, the past tense in fact appears to trigger certain stative type shifts. For example, the combination of a past-tense event verb with a state-selecting temporal adverb, e.g., *already*, leads to a ‘present perfect’ interpretation, as in *I already ate* (Michaelis, 1996). And yet it is difficult to reconcile this apparent stativizing behavior with the fact that the past tense also plays a role in perfective coercion, yielding inchoative readings of state verbs in temporal discourse (Dowty, 1986). If tenses trigger aspectual coercions, then tenses must be aspectually sensitive operators. It remains to be determined, however, what aspectual requirements tense operators have. The behavior of the English
past tense in particular is paradoxical, suggesting two antithetical patterns of aspectual concord. Puzzles like this can be solved only once we address two basic questions about verb meaning. First, how does a verb that denotes one kind of situation come to denote a completely different type of situation? Second, how does the original meaning of the verb constrain what a type shift can cause it to mean? In the set-theoretic model of aspectual transitions, as described above, there is no obvious way to address these questions, because verbs are not given semantic representations; instead, a verb (or verb phrase) refers to a class of eventualities, and via transition rules, it can pick out one kind of eventuality as easily as another. The problem being raised here can perhaps best be understood by analogy to theories of lexical relations in general. If a lexical paradigm (say the case forms of a given noun) were based entirely on suppletive relations, it would contain associations but no morphological generalizations; morphological generalizations rely on shared word-internal structure (Bybee, 1995, 2001). By the same token, if aspectual type-shifts were not constrained by verb-internal semantic structure they would be anomalous among semantic operations that affect the realization of predicate-argument structure. The prevalent view of verbal argument structure is based on the idea (due to Jackendoff, 1972) that an argument’s syntactic expression is determined by the location of that argument in the verb’s event-structure representation (Altshuler et al., 2019; Van Valin and LaPolla, 1997; Pinker, 1989; Levin and Rappaport Hovav, 1995; Wunderlich, 1997). Accordingly, when a verb has multiple subcategorization frames, this variability tends to be captured by operations that alter components of a verb’s decomposed semantic structure. Rappaport Hovav and Levin (1998, 2005) have proposed an incremental theory of such operations, in which new verb meanings are derived through the expansion of simple event-structure templates into more complex ones. In the spirit of this work, I will outline an analytic framework for aspectual type-shifting in which aspectual mappings permute Aktionsart representations of verbs, and can do so only when there is structure shared between input and output types. Aktionsart classes are interpretable as sets of entities, as described by DS (1998: 381–382), but our focus here will be on their behavior as lexical-semantic representations with internal structure rather than as names of sets.2

2 The framework proposed here resembles the transition networks of Moens and Steedman (1988), with several important difference. First, the present framework treats tenses as aspectual-class selectors, while Moens and Steedman do not. Second, the present framework treats state representations as potential inputs to type-shifting operations, while states serve only as outputs in the Moens and Steedman model. Third, while Moens and Steedman appear to identify iteration with stativity, we follow Boneh and Doron (2010: 340)
The major claim of this article is that, by describing aspectual type shifts as operations on Aktionsart structure, we can explain: (1) the relationship between input and output types in aspectual mappings effected through verbal morphosyntax and (2) constraints on possible aspectual type shifts. This article will appear in two parts. The present article is Part 1. It lays out the selection-based descriptive framework, as well as its conceptual underpinnings; Part 2, forthcoming in a later issue, demonstrates the application of the selection-based framework to both grammatical aspect and tense. The remainder of the present article, Part 1, is divided into two sections. In the following section, Section 2, I will outline a two-tier model of verbal Aktionsart and describe the operations on Aktionsart structure that underlie aspectual type shifts. In Section 3, I offer concluding remarks, as well as a preview of the applications to be demonstrated in Part 2, forthcoming.

2 Aspect, Aktionsart and Aspectual Shift

2.1 Aspectual Ontology

How, and in what form, is aspectual information made available to morphosyntax? The mechanism is invocation (Zwicky, 1989, 1994); the categories invoked by constructions are aspectual types. While constructions that index aspectual categories may be aspectual constructions, they need not be: tense and evidential constructions, and even complementation constructions like the accusative-infinitive pattern, place constraints on aspectual type. The aspectual types invoked by morphosyntactic constructions are identical to those denoted by verbs and their projections. As a consequence, aspectual information is represented in a uniform way throughout the grammar. This is not a traditional view. In the aspectual literature, it is generally assumed that while verbs denote states and various event types (e.g., processes and externally caused state changes), the grammatical aspects imperfective and perfective reflect instead the narrator’s ‘attention to endpoints’. On this style of account, of which Smith (1997) and Comrie (1976) are representative, perfective marking is used to present a situation as having begun and ended within the relevant interval. Imperfective marking, by contrast, “presents part of a situation, with no information about its endpoints” (Smith, 1997: 73). This view is based on a visual metaphor, in which the grammatical aspects are lenses of various powers through which speakers view the event schemas denoted by

in assuming that “although imperfectivity is most naturally associated with habituality, the two notions are not reducible to one another.”
verbs. While this basic metaphor is intuitive, it obscures the fact that aspectual presentation is a form of categorization. This point can be clarified by analogy to the domain of entities. While we could say, for example, that the speaker who pairs a mass noun with an indefinite article (e.g., *a thick muck*) is ‘attending to the boundaries of the substance’, such an account would fail to capture a generalization: this speaker is presenting a mass as an individuated entity by using the syntactic frame characteristic of count nouns. By the same token, the speaker who combines an event verb with the morphosyntactic trappings of a state verb (e.g., present-tense inflection) is presenting the event so denoted as an instance of the state category, just as combining a state verb with perfective morphosyntax (e.g., an iterative adverbial expression) conveys that the state so presented is a type of event. If aspectual encoding is ad hoc categorization, we can conclude the event-state distinction underlies semantic representation at both the level of the lexeme (i.e., lexical entry) and the level of the word: the lexeme when realized as the daughter node of a derivational or inflectional construction.

What is the semantic basis of the event-state distinction? According to Langacker (1987: 258), this distinction has a “primal character”, because it is linked to a basic cognitive capacity: the ability to perceive change (or stasis) over time. It is generally agreed that while events contain distinct subevents and are bounded in time, states lack internal structure and are not bounded in time. In Langacker’s words, “the covariant properties of change and bounding can be regarded as two sides of the same coin (as can their opposites, constancy and open-endedness)” (1987: 261). Although this characterization is generally valid, it does not obviously cover those events that have both imperfective and perfective properties. These events, which are generally referred to as either activities or processes, include running and reading. Like other event types, activities in English cannot be reported as ongoing at speech time by means of the simple present tense: *Look! Sue reads*. While activities like reading contain distinct subevents (e.g., page scanning and page turning), their endpoints are arbitrary. Unlike so-called telic events, they can be protracted indefinitely through the iteration of their subevents; no subevent represents a logical stopping point, since activities do not culminate in any resultant state. While there are activities that have episodic construals, e.g., sleeping, sitting in a chair and holding something in one’s hand, these activities lack subevents; they are simply periods of stasis. Following Langacker (1987), I will refer to such activities as homogeneous activities, to distinguish them from those activities which, like running and singing songs, have heterogeneous internal part-structure when parsed into sufficiently small subintervals. Because some activities are in principle unbounded while others lack subevents, the two properties
of change and boundedness, while jointly defining the class of telic events (accomplishments and achievements), are sufficient conditions on eventhood but not necessary ones.

The property that unifies all event types is epistemological (Brisard, 2002; De Wit, 2016): events are those situations whose existence cannot be verified on the basis of a momentaneous ‘sample’. Let us illustrate this criterion by application to the least prototypical class of events—activities. Verification of a heterogeneous event, e.g., running, requires several frames. Since running consists of successive leaps using alternating legs, witnessing a single leap is insufficient to verify an event of running. In the case of homogeneous activities like holding a broom, standing in a corner or sleeping, verification requires access to points of inception and termination, as well as several contiguous frames between those endpoints. Sleeping is distinct both from being comatose and from nodding off for a second, and staying at your sister’s house is distinct both from dropping in on your sister and living with her. While states like being tall have duration in the same way that the events of sleeping and standing in a corner do, states do not take time, since any subinterval of a state counts as an instance of that same state. The existence of a state can thus be confirmed on the basis of an atemporal sample. The same cannot be said of a state phase, e.g., *She was sick for three days* or *She was short as a child*, because a state phase, like an activity, takes time.3

The epistemic criterion described here is highly compatible with the picture of the event-state distinction that emerges from viewpoint-based models of grammatical aspect discussed above: perfective aspect involves ‘endpoint focus’ because the assertion that an event exists entails confirmation that this event has begun or ceased, or both. Under the assumption that grammatical aspect and Aktionsart have uniform semantic representations, we expect that categories at the two levels will have such isomorphic characterizations. The epistemic characterization of the event-state distinction also comports well with what we know about the differential behavior of events and states in temporal discourse (Partee, 1984; Dowty, 1986; Herweg, 1991): while events are included within the reference intervals for which they are asserted, states include those reference times. For this reason, a speaker who makes a stative assertion, e.g., *She was there at 3:00*, is not presumed to know whether the denoted situation also went on at a superinterval that includes the topical

---

3 The category of state phase differs from that of stage-level predications, as described by Partee (1991), *inter alia*. State-level predications denote temporary states like being on sale, on fire or angry. Stage-level predications, unlike state phases, have stative syntactic and interpretive behaviors, e.g., they are reportable by means of the simple Present in English.
interval. By contrast, speakers who make perfective assertions, e.g., *She ran at lunchtime*, thereby signal that no such superinterval exists. Because this implication is an entailment and not merely a quantity-based implicature, it cannot be suspended, as shown in (9):

(9) *She ran at lunchtime yesterday. In fact, she ran prior to that.*

Figure 1 gives a hierarchical classification of the Aktionsart classes, as well as the conceptual bases of the distinctions:

![Hierarchical structure for the Aktionsart classes](image)

In Figure 1, situations are divided into those that take place over time (events) and those that hold at a given point in time, states. The class of events is partitioned into those that culminate in a specific resultant state (directed events) and those that do not (episodic events). Directed events are further partitioned into events that come about or are brought about over time (accomplishments) and events that are simply changes made manifest within some period of time (achievements). In the class of episodic events, we distinguish between *energeiai*, or, alternatively, *processes*, and state phases. The label *energeiai* is used, following Aristotle, to refer to the class of actions that take time but do not culminate (Binnick, 1991: 142–143). The category of *energeiai*
includes activities, both homogeneous and heterogeneous. State phases are states that begin and end within the reference interval and to which an explicit duration may be assigned, e.g., *I was depressed when I lived there* (Herweg, 1991). In contrast to states, state phases have perfective behavioral properties. For example, they can be enumerated (*Iona was ill for two weeks twice*) and they cannot be reported by means of the simple present tense (*Iona is ill for two hours*). Like states, however, state phases require no energy expenditure for their maintenance.

### 2.2 The Two-Tier Model of Aktionsart Representation

Situation types are both topological structures and gestalts. They are topological structures because they occupy intervals in characteristic ways irrespective of the size of the interval. They are gestalts because each one indexes an idealized causative event. This idealized event, which has been described as a causative prototype (Lakoff and Johnson, 1980: 69–71), a contingency-based event structure (Moens and Steedman, 1988: 18) and a causal chain (Smith, 1997: 21–22; Croft, 1998, 2012), involves direct manipulation of an entity by an agent, who brings about a perceptible change of state in that entity. The situation types are characterized by the ‘span’ of the causal chain that is lexically expressed in each case. For example, activities prototypically represent motor programs executed by agents while states prototypically represent resultant states. On this model, some events contain other events, and contiguous events share participants.

#### 2.2.1 Causal Representation

Rappaport Hovav and Levin (1998), henceforth RHL, capture the distinction between aspectual and frame-specific features of verb meaning by proposing a set of fixed event-structure templates with which verbs can combine. Verbs ‘fill in’ information represented by constants; the type of the constant determines the information that the verb will be required to provide. Table 1 presents an adaptation of RHL’s inventory of event-structure templates. In these templates, operators (shown in small caps) represent subevent connectives in the Jackendoff-Dowty-Vendler tradition, while variables represent participant roles. Constants are represented by the italicized material in angled brackets. I have expanded the RHL inventory of event templates in order to represent Aktionsart classes and event properties which, while having no direct relevance to verbal argument structure, figure prominently in aspectual type shifts. The class of state phases has been added and the class of processes split into two classes: homogeneous and heterogeneous activities. The state-phase template, as shown, contains the operator HOLD. This operator combines with
A stative situation type to yield a state which begins and ends. The homogeneous-activity template, as shown, also contains the operator \textit{hold}. In this template, however, \textit{hold} takes two arguments, a state radical and an effector. The effector argument is also an argument of this state predication; this notation reflects the fact that the subject-denotatum, although nonagentive, is responsible for the maintenance of the denoted state. The template for heterogeneous activities contains the operator \textit{repeat}. This operator has the same valence and 'control' properties which \textit{hold} has in the homogeneous-activity template. The \textit{repeat} operator captures the observation that heterogeneous activities, e.g., \textit{skip}, consist of iterated type-identical events. Since a heterogeneous activity is itself an event, a heterogeneous activity may 'fill in' the event variable in the heterogeneous-activity template. The resulting event is an event chain, or, equivalently, a heterogeneous activity. As in \textsc{rhl}'s original model, the achievement template properly includes the state template, while the accomplishment template contains the templates for activities, achievements and states, respectively.

The \textsc{rhl} model is not unique in using Aktionsart class to predict the syntactic behavior of verbs, but it provides new insights into a traditionally vexing question: what are the constraints on semantic derivations? Models that employ operations on event structure, e.g., Gropen et al. (1991) and Wunderlich (1997), permute representations relatively freely in order to capture the semantic implications of argument-structure alternations. \textsc{rhl} propose instead a single mechanism of semantic derivation, \texttt{template augmentation}: "Event structure templates may be freely augmented up to other possible templates in the basic inventory of event structure templates" (1998: 111). The added structures are the subevents represented by operators, e.g., \texttt{become}. Template augmentation involves the unification of Aktionsart representations. Unification

<table>
<thead>
<tr>
<th>Aktionsart Class</th>
<th>Causal Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>State</td>
<td>([x \text{ &lt;STATE&gt;}]) e.g., \textit{seem}</td>
</tr>
<tr>
<td>State phase</td>
<td>([\text{HOLD} [x \text{ &lt;STATE&gt;}]]) e.g., \textit{be sick for two days}</td>
</tr>
<tr>
<td>Homogeneous activity</td>
<td>([x \text{ HOLD} [x \text{ &lt;STATE&gt;}]]) e.g., \textit{sleep}</td>
</tr>
<tr>
<td>Heterogeneous activity</td>
<td>([x \text{ REPEAT} [x \text{ &lt;EVENT&gt;}]]) e.g., \textit{skip}</td>
</tr>
<tr>
<td>Achievement</td>
<td>([\text{BECOME} [x \text{ &lt;STATE&gt;}]]) e.g., \textit{sink}</td>
</tr>
<tr>
<td>Accomplishment</td>
<td>([[[x \text{ REPEAT} [x \text{ &lt;EVENT&gt;}]]) \text{ CAUSE} [\text{BECOME} [y \text{ &lt;STATE&gt;}]])]) e.g., \textit{build}</td>
</tr>
</tbody>
</table>
can be described metaphorically as the stacking of transparencies upon which strings of characters are written. (The term *transparency* refers here to a clear plastic sheet of the kind once used to display slides on an overhead projector.) The transparencies can be stacked on top of one another (in any order) as long as all of the symbols on each slide show through. If two transparencies contain identical strings, e.g., *ab*, then information is neither lost nor gained by superimposition. Through template augmentation, an event-structure template, e.g., the heterogeneous-activity template, projects an event-structure representation by which it is entailed—the accomplishment template. Template augmentation thereby drives verbal valence augmentation at the syntactic level. For example, the verb *sweep* has both a monovalent activity pattern (*She swept for hours*) and a trivalent accomplishment pattern, in which it denotes causation of motion (*She swept the dust off the steps*);

It is not surprising that an Aktionsart-based model of argument projection also provides a model of aspectual type shifts. However, the two types of models target distinct aspects of the syntax-semantics interface, and accordingly require some divergent mechanisms. While models of argument projection represent the effect of verbal semantics on syntax, models of aspectual type-shifting represent the effect of syntax on verbal semantics. Since the latter type of model does not build syntactic structure, it is as likely to pare down as to expand event-structure representation. In fact, as we will see, both implicit and explicit type shifts involve such ‘paring’ operations, here called selection operations. Is template reduction incremental in the same way that template augmentation is? To answer this question, we must first broaden our idea of what constitutes an ‘increment’ of Aktionsart representation. The next section will demonstrate that subevents are not the only meaning units visible to Aktionsart-based operations. As we will see, this set also includes the units of temporal representation.

### 2.2.2 Temporal Representation

While causal representation describes relations among entities, the entities of temporal representation are the situation types themselves. Temporal representation captures the patterns of stasis and change which are characteristic of each situation type. Temporal representations do not, for example, represent causal links between contiguous situations or agentive implications attaching to certain participants. Table 2 gives temporal representations for each of the six Aktionsart classes discussed above. These representations utilize the three situation-type components described in Section 1. They are: states (*ϕ*), transitions (*τ*), and event chains (*κ*). States are internally homogeneous situations which include no transitions (i.e., temporal boundaries). For this
reason, we say that states include the intervals at which they hold (Partee, 1984; Herweg, 1991). Transitions are state-change events, and as such are isomorphic to achievements. However, the category of transitions is not limited to those inchoative events which are lexicalized as achievement verbs, since it also includes the events of inception and cessation, which jointly define the endpoints of a situation. For example, the endpoints of sleeping, a homogeneous activity, are, respectively, the events of falling asleep and waking up.

A transition is necessarily defined relative to a prior or subsequent state. For this reason, I will assume that all intervals that are contiguous to a transition event, including those which precede an onset transition and those which follow an offset transition, are states. Such states, which I will refer to as rests, are available for selection by the permutation operation, resulting in a stative output type. Because the selection operation can target any rest—whether initial, final or intermediate—it finds states within Aktionsart representations where none have been presumed to exist. For this reason, I will show, the selection operation provides a compositional account of a wide variety of stative type shifts, including those that create, respectively, prospective, progressive, perfect, and habitual/generic construals.

Unlike states, transitions cannot stand alone, nor can they be iterated without the mediation of a state; accordingly, the representations *[τ] and *[τ τ] are ill formed (Bickel, 1997: 126). By contrast, the representation [τ Φ τ] is well formed; it corresponds to both a state phase and a homogeneous activity (recall that agentive properties are invisible to temporal representation). When the representation [τ Φ τ] is iterated, it corresponds to an event chain or heterogeneous activity (κ). The representation corresponding to heterogeneous activities contains the notation [τ Φ]+, denoting one or more instances of particular state change, e.g., that of crossing the room in an episode of pacing back and forth. While both heterogeneous and homogeneous activities can

<table>
<thead>
<tr>
<th>Aktionsart Class</th>
<th>Temporal Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>State</td>
<td>Φ</td>
</tr>
<tr>
<td>State phase</td>
<td>τ Φ τ</td>
</tr>
<tr>
<td>Homogeneous activity</td>
<td>τ Φ τ</td>
</tr>
<tr>
<td>Heterogeneous activity</td>
<td>τ Φ [τ Φ]+ τ</td>
</tr>
<tr>
<td>Achievement</td>
<td>τ Φ</td>
</tr>
<tr>
<td>Accomplishment</td>
<td>x τ Φ</td>
</tr>
</tbody>
</table>

Table 2: Temporal representation (based on Bickel, 1997)
be protracted indefinitely, the mechanisms are different in each case. In the former case, expansion entails concatenation, while in the latter case expansion simply entails lack of change. Notice, however, that in neither case does expansion have any effect on bounding: the initial and final transitions are present whatever intervenes between them. When a heterogeneous activity is embedded in an accomplishment representation, shown in Table 2 as \([x \tau \phi]\), its offset transition is superimposed upon the initial transition of the embedded achievement, \([\tau \phi]\). This reflects the observation that, for example, in an event of walking home, the threshold-crossing transition is also the final step of the walk.

The constraint that rules out sequences of the form \([\tau]\) and \([\tau \tau]\) need not be stipulated, since one cannot logically conceive of an inchoative event that is unaccompanied by a resultant state. Notice, however, that in the temporal representations given in Table 2, resultant states are not consistently indicated. In particular, states that follow events of termination are missing from the representations. These states are not indicated because they can be ‘read in’ on the assumption that transitions are isomorphic to achievements. Notice, however, that antecedent states are equally crucial to the definition of transition, and our temporal representations lack these as well. Let us assume, therefore, that antecedent states, like consequent states, can be subsumed, along with periods of stasis between chained events, under the rubric of rests. The term rest is meant to be construed as it is in rhythmic representation: a pause between ‘beats’ or transitions. While in the foregoing remarks I have distinguished intermediate states from antecedent and consequent states, this distinction is not particularly meaningful: because events are located with respect to one another on a time line, all events potentially qualify as chained events and all states can be construed as intermediate states. This point will become particularly relevant when we consider chained events that represent habitual situations, or ‘gnomic imperfectives’ (Boneh and Doron, 2010; Carlson, 2012).

2.3 Aspectual Type Shifts as Operations on Aktionsart Structure

In this section, we will look first at operations which permute Aktionsart structure and then at the concatenation operation, which maps an event radical to a series of type-identical events, i.e., a heterogeneous activity. We will find that while all aspectual mappings involve the unification of two Aktionsart structures, as per the Principle of Aktionsart Preservation, given in (10) below:

(10) *The Principle of Aktionsart Preservation.* In an aspectual mapping, the event-structure representations of input and output types must unify.
What (10) means is that the input and output representations must be capable of being superimposed on one another with no loss of information. In the course of this exploration, we will encounter a number of examples that appear to violate Aktionsart Preservation, but which are in fact indirect type shifts, in which the input and output types are related by two mappings, each of which conserves Aktionsart structure.

2.3.1 Permutation

Permutation operations add or select a single component of the input Aktionsart representation. The definition of component differs according to whether we are using causal or temporal representation. In causal representation, a component corresponds to an operator, e.g., HOLD, and the arguments it projects. In temporal representation, a component corresponds to a state, transition or event chain. As an example of addition, consider the transition from state to achievement. This type shift occurs implicitly when, for example, a frame adverbial is combined with a state radical, as in (11):

(11) They were bored in a few seconds.

This type shift involves the addition of the operator BECOME, or, equivalently, a transition, to the causal or temporal representation of the state. A further example of addition is the shift from achievement to accomplishment. This type shift underlies the coercion effect in (12):

(12) He finished dying and got carried off the stage.

In (12), an achievement radical, He die-, is construed as an accomplishment radical in order to satisfy the semantic requirements of the verb finish, which selects for complements denoting effected results. This type shift involves the addition of the operator REPEAT and its arguments, or, equivalently, an event chain κ, to the representation of the achievement. As an example of selection consider the explicit type shift performed by the copular resultative construction in English, exemplified in (13):

(13) a. The truck is loaded.
    b. The soup is cooled.

The resultant-state predications in (13) denote states, or more specifically those states that are embedded in the Aktionsart representations of their participial complements. These states are, respectively, that of the truck being full and that of the soup being cool. The stative type shift performed by the resultative...
construction involves selection of the state component in the causal or temporal representation of the lexical verb. Since both the accomplishment verb *load* and the achievement verb *cool* entail a resultant state, the application of selection conforms to Aktionsart Preservation. Notice, however, that the type shift exemplified in (13a) is not incremental: states and accomplishments differ by more than a single component of Aktionsart representation, since the accomplishment entails two subevents which the state does not.

In fact, the accomplishment-state mapping is one that selection frequently performs. Consider, for example, rest selection, as illustrated in (14):

(14)  

(a) She's about to load the truck.  
(b) She will load the truck.

Both (14a) and (14b) exemplify type-shifting constructions which, as is typical, express the two types that they mediate in their periphrastic forms: while the auxiliary head denotes a state, the infinitival complement denotes an event. Although both of these constructions are traditionally regarded as exponents of ‘future tense’, this label seems unrevealing, since it fails to capture the stative properties which both constructions exhibit, e.g., both constructions are used to express present-tense reports. While the stative character of these and other future markers has traditionally resisted a compositional explanation, in the present framework their stativity is a straightforward product of selection. The selected component is an antecedent rest—a state preceding the initial transition of the event chain embedded in the temporal representation of the input event radical.

Occasionally, however, permutation operations appear to violate Aktionsart Preservation. These violations are only apparent, since the relevant mappings are in fact sequenced mappings, as described by Moens and Steedman (1988: 21–22): ordered pairs of mappings, the first of which feeds the second. I will refer to these sequenced mappings as INDIRECT TYPE SHIFTS since they involve the mediation of a third aspectual category. Indirect type shifts exist because semantic transitions, as equivalence relations, are transitive; that is, if A=B and B=C then it follows that A=C. Indirect type shifting will be invoked below in the analysis of the Progressive.

2.3.2 Concatenation
The concatenation operation is similar to the repetition operations assumed by a number of aspectual theorists (Jackendoff, 1997: 51–52; Bickel, 1997: 117–119; DS, 1998: 361–362), but there are crucial differences between concatenation, as envisaged here, and these antecedent notions. Concatenation applies to an event type (i.e., dynamic situation radicals), and outputs a series of events that
are identical both to one another and to the input event. In addition, like other iteration operations, concatenation is used to represent both implicit and explicit type shifting, e.g., coerced readings triggered by frequency adverbials. The difference between concatenation and its predecessor notions lies in the nature of the output type. While repetition operations are typically assumed to output state types, concatenation instead outputs an event chain, which qualifies as a heterogeneous activity rather than a state. The identification of event chains with heterogeneous activities makes sense because, as has been widely observed, telic verbs with multiplex complement denotata receive activity readings. Note, for example, the contrast between the sentence She ate mushrooms, which reports an activity, and the sentence She ate a mushroom, which reports an accomplishment. Further, by rejecting the assumption that repeated events are ipso facto stative, we avoid the logical paradox alluded to in Section 1: situations that consist of multiple type-identical subevents, e.g., pacing, qualify as dynamic situations rather than states; it is not obvious therefore why event radicals that otherwise qualify as unique events receive coerced repeated-event interpretations in morphosyntactic contexts which call for state radicals. Two such contexts are illustrated in (6-7), repeated here as (15a-b):

(15)  a. She smokes.
      b. She smoked when I met her.

As discussed in Section 1, these examples illustrate morphosyntactic contexts which in some as yet undefined way call for stative situation types, and the inference of repeated/habitual action links input and output types. And yet if iteration entails stativity, it should always be possible to iterate one's way from an event to a state, we see above that it is not. Further, as Smith observes (1997: 51), the syntactic behavior of habitual predications suggests that they are perfective: they can appear in imperatives, with agent-oriented adverbials like deliberately, and in pseudo-cleft constructions. The syntactic constructions in question do not in general appear capable of coercing perfective readings of stative predications: sentences like??What she did was prefer white wine and??I persuaded her to prefer white wine are awkward at best.

A possible solution to the paradox is suggested by Langacker (1996, 1997): while iteration is sufficient for a stative construal, it does not entail a stative construal, since repeated events may be also be construed perfectively. But Langacker does not explain precisely why habitual predications invite stative construals. It cannot be, for example, that habitual predications, like states, necessarily denote unbounded situations, since, as Langacker (1996: 292) observes,
habitual and generic predications can denote situations which hold “for either a bounded or an unbounded span of time, i.e., their validity has a temporal scope” [emphasis in original]. If habitual predications can be either perfective or imperfective, what then is the basis for distinguishing between iterated-event sentences and habitual-event sentences? According to Langacker, iterated events and habitual events have distinct implications for our theories of the world. He describes these implications using Goldsmith and Woisetschlaeger’s (1982) distinction between structural and phenomenal knowledge. Phenomenal knowledge is akin to awareness. Iterated-event predications, like other kinds of episodic predications, express actual events — those which impinge on consciousness. Structural knowledge is akin to pattern recognition. Habitual sentences express structural events — those which one can predict to recur on the basis of world knowledge. Structural events are also conveyed by gnomic sentences, e.g., Oil floats on water or A periodontist treats gum disease, and many aspectual theorists, including Krifka et al. (1995), conflate habitual and gnomic sentences under the general rubric of generic sentences. In accordance with Krifka et al. (1995) and Bybee, Perkins and Pagliuca (1994: 152), we will assume that the differences between habitual sentences (which Krifka et al. refer to as characterizing sentences) and gnomic sentences (which Krifka et al. refer to as reference to types) can be traced to characteristic properties of nominal reference. Nominals in gnomic sentences have attributive reference, leading to contingency readings. For example, one can paraphrase the sentence Oil floats on water by means of a conditional sentence: if there is something that counts as oil, it will float on whatever substance qualifies as water. Habitual sentences do not have contingency readings, since such sentences attribute properties to specific entities. However, both gnomic and habitual sentences express nonincidental facts. The question before us is whether the structural-actual (or, equivalently, generic-episodic) distinction is relevant for aspectual coding. There is evidence to suggest that it is not.

In a typological survey of the generic-episodic distinction, Dahl (1995: 425) reaches the conclusion that although languages use grammatical markers to distinguish between generic and episodic sentences, no language uses dedicated morphosyntax to express this distinction. Dahl assumes that there is a single marker of genericity in each of the languages in his study, taking the present tense to be the ‘generic marker’ for English. As we know, however, generic statements are compatible with a number of other tense-aspect combinations, including the simple past and past progressive: Dogs chased cars in those days, During that summer parents were keeping their children indoors. Generic sentences appear to be recognized as such only on the basis of a mismatch between perfective verbal Aktionsart and the syntactic context in which
that verb appears. For example, Bybee et al. (1994: 152), in motivating a grammatical category of present habitual sentences, observe that “the difference between habitual and present stative resides entirely in the lexical meaning of the predicate: the present habitual reading of dynamic predicates covers many different instances of the same situation, while the present stative covers one continuous situation”. It therefore appears appropriate to conclude that generic meaning is a specific type of coercion effect, achieved by combining an event-chain radical with a state-sensitive operator, whether aspect or tense.

The connection between habituality/genericity and stativity is an inferential one: an iterated, temporally stable situation which is also construed as including reference time (whether past or present) will invariably be construed as gnomic/habitual. From this correlation, however, we cannot conclude that genericity entails stativity, since perfective sentences can also express structural events. Rather, this correlation suggests that genericity is a contextual inference, and one which is based upon a semantic prototype. The generic-episodic distinction is a contextual one in part because it hinges on inferences about the size of the relevant time scales. If the intervals separating the events are judged to be small, as in the case of The light flashed, the predication will be judged as episodic; if the relevant events are judged to be widely dispersed through time, as in The Romans laid siege to Gallic cities, the predication will be judged generic. The time-scale issue recalls a point about grammatical meaning made by Talmy in his classic paper ‘The Relation of Grammar to Cognition.’ Based on topological properties of deictic expressions, Talmy argues that the meanings of grammatical (closed-class) expressions are magnitude neutral, with respect to both time and space (Talmy, 2000: Chapter 1). If we accept that conclusion, magnitude-dependent semantic distinctions, like the generic-episodic distinction, are much more like pragmatic inferences than grammatical categories.

In light of the foregoing considerations, I propose to treat habitual-event radicals and iterated-event radicals as indistinguishable at the level of Aktionsart structure: both qualify as heterogeneous activities. Accordingly, the concatenation operation takes us only part of the way toward a stative interpretation: it yields a heterogeneous activity. It is at this juncture that perfective and habitual meanings are compatible. The permutation operation of selection provides the ultimate bridge to stative meaning: since iterated events contain intermediate rests, and since such rests qualify as states, those type shifts which require stative input types (whether implicit or explicit) are free to select intermediate rests.

While it is generally said that present-tense sentences report situations ongoing at speech time, the event radical denoted by a habitual predication need not overlap speech time. For example, a speaker can truthfully assert (15a) whether or not the person described happens to be smoking at encoding time.
Under the present analysis of present-tense habitual sentences, this interpretative fact is explained: the situation that is treated as ongoing at speech time is not that denoted by the verb and its arguments; rather, it is a state which lies between any two occurrences of the type denoted by the lexical verb.

3 Conclusion

On the Aktionsart-based approach to type shifting outlined here, aspectual information is both grammatically integrated and uniformly represented. It is uniformly represented because a single type of lexically encoded structure—Aktionsart structure—drives both a verb’s external syntax (the array of complements with which it combines) and its morphosyntactic realization (its inflectional and derivational properties). It is grammatically integrated because Aktionsart representations are indexed by constructions that are neither strictly aspectual nor even strictly temporal: we find Aktionsart-based concord constraints not only in tense constructions like the English present tense, but also in evidential constructions like the Turkish inferential past tense (Slobin and Aksu, 1982), as well as non-finite clausal complements, e.g., I believe the plan to fail/have failed (Katz, 2003). The distinction between grammatical aspect and lexical aspect is here recast as the distinction between aspectually sensitive constructions and the verb classes that those constructions select for. In Part 2 of this two-article series, “Using Selection to Describe Implicit and Explicit Type-Shifting Constructions”, we will use the Aktionsart-based model of aspectual mapping developed in this article to describe implicit and explicit types shifts performed by the past and present tenses of English, as well as two stativizing constructions of English, the Progressive and Perfect.

References


Piñango, Maria and Ashwini Deo. 2016. Reanalyzing the complement coercion effect through a generalized lexical semantics for aspectual verbs. *Journal of Semantics* 33: 359–408.


