PARTS OF A WHOLE: DISTRIBUTIVITY AS A BRIDGE BETWEEN ASPECT AND MEASUREMENT

DISSERTATION SUMMARY

Introduction

This dissertation builds connections between three domains which are traditionally addressed by separate areas of research within mereological formal semantics: aspect, measurement, and distributivity. A number of natural language phenomena currently treated by separate theories turn out to be intimately related across these domains. I connect them through a bridge whose central pillar is a higher-order property, *stratified reference*. This concept is shown to be both general enough to connect and subsume several familiar notions, and formally precise enough to transfer insights across unrelated bodies of literature. Previous accounts of these notions either fail to generalize appropriately, or live on as limiting cases of a system presented here under the name of *strata theory*. This system is not a radical reorientation of the grammar. By subsuming and building on previous characterizations, strata theory retains much of what has been formerly gained, and provides a unified framework in which new correspondences are drawn between existing concepts.

The road to this claim starts with four semantic oppositions which are closely associated with the domains under consideration. These are the telic-atelic opposition, which is central to the study of aspect; the singular-plural opposition and the count-mass opposition, which are central to the study of plurality and measurement; and the collective-distributive opposition, which is central to the study of distributivity.

These oppositions can be formally related to one another. Intuitively, singular, telic, and collective predicates are delimited or bounded in ways that plural, mass, atelic, and distributive predicates are not. How to characterize the difference between boundedness and unboundedness – the *boundedness question* – is one of the central concerns of strata theory.

Answering this question amounts to specifying what it means for a predicate to be atelic, distributive, plural, or to have mass reference. It is not obvious that there should be a single property that is shared by all these predicates. As this work shows, however, it is indeed possible to isolate such a property. The identity of this property can be determined by analyzing a number of nominal and verbal constructions which are sensitive to the semantic oppositions listed above, and which are regarded as central to the study of aspect, measurement, and distributivity. These constructions are *for*-adverbials, which distinguish atelic from telic predicates (1); pseudopartitives, which distinguish plurals and mass nouns from singular count nouns (2); and adverbial *each*, which distinguishes distributive from collective predicates (3). I refer to these three constructions collectively as *distributive constructions*.

(1) a. John **ran** for five minutes. *atelic*
b. *John **ran to the store** for five minutes. **telic*

(2) a. thirty pounds of **books** *plural*
b. thirty liters of **water** *mass*
c. *thirty pounds of **book** **singular*

(3) a. The boys each **walked**. *distributive*
b. *The boys each **met**. **collective*

The novel angle of this dissertation consists in considering the constructions in (1) through (3) as parts of a whole. Previous work has produced separate theories to account for the phenomena that they exemplify. The resulting theories are often more limited in scope than they could be. For example, work on distributivity has focused on how best to formalize distributive readings, rather than on extending the notion of distributivity. Likewise, the study of aspect has concentrated entirely on temporal phenomena, and the study of measurement in natural language has focused largely on mass terms, partitives, and comparatives. This development has obscured the view on the common properties of these constructions. However, this problem is not inherent in the approaches encoded in these theories. Once the connection between them is made formally explicit, it is easy to connect them to each other, and to extend them to domains beyond the ones in which they have traditionally been applied. One can then combine the strengths of each account, and synthesize them to extend their empirical coverage. This is the motivation behind the present work.

This summary presents in broad terms the picture that is drawn in the dissertation. Due to space reasons, I ignore or gloss over many technical matters. Let me stress, however, that a central methodological goal of the dissertation consists in clarifying the foundational assumptions underlying the logic of language, specifically, mereological semantics. Chapter 2 of the dissertation presents classical extensional mereology, sets it in relation to compositional semantics, and discusses central assumptions made in both areas. The chapter is intended to be used as a reference point for future researchers and spells out the relevant background assumptions as explicitly as possible, especially in case of choice points where the literature has not yet reached consensus on a preferred analysis. Examples include the meaning of the plural morpheme, the question whether the meanings of verbs are inherently pluralized, the formal properties of thematic relations, the nature of group entities, etc. Choosing how to answer the controversial questions that relate to these concepts is not at the heart of the dissertation, but it is nevertheless a crucial prerequisite for the ability to formally implement the connections drawn below, and to carry out the proofs in the dissertation that explicate these connections.

The presence of distributive constructions in every one of the domains of interest makes it possible to place strata theory on a solid empirical foundation. Instead of asking abstractly what it is that atelic and distributive and mass and plural predicates have in common with each other, we can search for the property that the bold constituents in the grammatical examples in (1a), (2a), (2b) and (3a) have in common, to the exclusion of the ungrammatical examples in (1b), (2c), and (3b). I claim that these constituents are all subject to a constraint, a presupposition that is introduced into distributive constructions through certain words such as *for*, *of*, and *each*. This presupposition is formulated in terms of a higher-order property, which I call *stratified reference*. This property requires a predicate that holds of a certain entity or event to also hold of its parts along a certain dimension and down to a certain granularity. Dimension and granularity are understood as parameters which distributive constructions can set to different values.

The dimension parameter specifies the way in which the predicate in question is distributed. Different settings of this parameter allow one and the same predicate to be atelic but not distributive, or vice versa. When the dimension parameter is set to time, stratified reference applies to atelic predicates, as in (1). When it is set to a measure function like weight or volume, stratified reference applies to mass and plural predicates, as in (2). When it is set to a thematic role like agent, stratified reference applies to distributive predicates, as in (3).

The granularity parameter specifies that the parts in question must be either atomic or very small in size, as measured along the dimension. This parameter accounts for the differences between distributive constructions over discrete (count) domains, such as adverbial-*each* constructions, and those over domains involving continuous dimensions, such as *for*-adverbials and pseudopartitives.

Stratified reference can be seen as the result of generalizing the subinterval property, which has been largely adopted as a means to characterize atelic predicates. This property says that whenever a predicate holds at an interval t, it also holds at every subinterval of t, all the way down to instants. Figuratively speaking, the subinterval property requires that any event in the denotation of a predicate that has the subinterval property can be divided into infinitely thin "temporal layers" that are also in the denotation of this atelic predicate. For example, on this view, a predicate like *walk* is atelic because we can "zoom in" to any part of the runtime of a walking event to find another walking event:

(4) $\forall e[[\![\text{walk}]\!](e) \rightarrow \forall i[i < \text{runtime}(e) \rightarrow \exists e'[[\![\text{walk}]\!](e') \wedge e' < e \wedge i = \text{runtime}(e')]]]$

The subinterval property gives rise to the well-known "minimal-parts problem": it insists that an atelic predicate must be true at all subintervals, even infinitely short ones. But not all atelic predicates satisfy this stringent criterion. It seems sufficient for the predicate be true at length intervals that count as very small relative to the length of the bigger interval: Given that any waltzing event takes at least three steps to unfold, *John and Mary waltzed for* *an hour* does not entail that they waltzed within every single moment of the hour, only that they waltzed within every short subinterval of the hour. If the subinterval property is to have any viable chance, it must therefore be amended so that the event layers are constrained to be merely very thin, but do not have to be infinitely thin. I call these layers *strata*. This name is chosen to remind the reader of geological strata, the layers of rock which can be observed in geological formations in places such as the Grand Canyon. A geological stratum can be just a few inches thick (though not infinitely thin) and extend over hundreds of thousands of square miles. This aspect is mirrored in the theory, where strata are constrained to be very thin along one dimension, but may be arbitrarily large as measured in any other dimension.

Formally, this effect is achieved by adding a granularity parameter to the subinterval property and constraining this parameter to a low but nonatomic value. I assume that the model contains a predicate ε, such that $\varepsilon(K)(x)$ holds just in case x counts as very small as compared to the comparison class K. For example, $\varepsilon(\lambda t[\text{hours}(t) = 1])(t')$ is true just in case t' is very small with respect to one hour. We want to be able to say that whenever *waltz* holds of an event, there is a way of dividing this event into subevents with very small runtimes such that *waltz* also holds of each of these subevents. Formally, we can express this as follows:

(5) $$\forall e[\text{waltz}(e) \rightarrow e \in {}^*\lambda e' \left(\begin{array}{l} \text{waltz}(e') \wedge \\ \varepsilon(\lambda t[\text{hours}(t) = 1])(\text{runtime}(e')) \end{array} \right)]$$

This formula makes use of the star operator known from the literature on plural semantics. $x \in {}^*(\lambda y.P(y))$ means that x consists of one or more parts of which P holds. (Formally, $A \in {}^*P$ is defined as $\exists C[A = \bigoplus C \wedge C \subseteq P]$ (A is the sum of all the elements of a subset C of P).) Let us say that *waltz* has **stratified reference** with respect to the dimension *runtime* and the granularity $\varepsilon(\lambda t[\text{hours}(t) = 1])$ (true of any interval very short as compared to an hour), formally, $\text{SR}_{runtime,\ \varepsilon(\lambda t[hours(t) = 1])}(waltz)$, just in case (5) is true. By abstracting from this example, we arrive at the following definition:

(6) **Stratified reference (Definition)**

$$\text{SR}_{f,\varepsilon(K)}(P) \stackrel{\text{def}}{=} \forall x[P(x) \rightarrow x \in {}^*\lambda y \left(\begin{array}{l} P(y) \wedge \\ \varepsilon(K)(f(y)) \end{array} \right)]$$

As illustrated in Figure 1, this definition says that stratified reference applies to a predicate P just in case the following is true: whenever P holds of an entity or event x, there is a way to divide x into strata y_1, y_2, etc. such that each y_i is mapped by the function f to a value which counts as very small with respect to the comparison class K. This is illustrated in the following diagram, where the vertical axis represents the dimension f. In case f is instantiated with runtime, stratified reference approximately amounts to the subinterval property, except that the minimal-parts problem is avoided. We can now say that being atelic means having stratified reference with respect to time and a suitably instantiated granularity parameter.

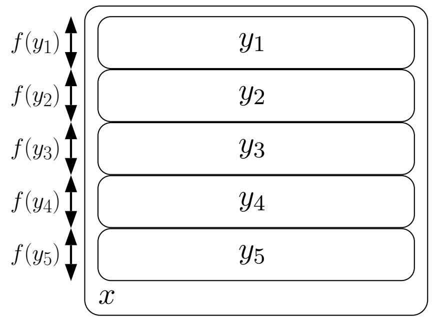


Figure 1. Illustration of stratified reference

In moving from the subinterval property to stratified reference, we have not only abstracted over the thickness of the strata but also over their orientation. I have introduced the concept of temporal strata as resulting from dividing an event along the temporal dimension, but we can also imagine spatial or "agental" strata – subevents that are constrained based on their spatial extent or based on their number of agents. Moreover, since stratified reference is untyped, we can also use it to express the fact that entities other than events, such as substances, can be divided along various dimensions. This allows us to generalize the insight encoded in the subinterval property and to carry it over from aspect to measurement and to distributivity. Stratified reference allows us to capture the parallelism between the telic-atelic, collective-distributive, singular-plural, and count-mass oppositions in a unified framework. Simply put, stratified reference describes the unbounded half of all these oppositions.

The rest of this document highlights four case studies which demonstrate the power of stratified reference, corresponding to Chapters 6 through 9 of the dissertation respectively.

Case Study 1: Spatial Aspect (Chapter 6)

Stratified reference allows us to treat time and space as different settings of a parameter. This makes it possible to extend theories of aspect from temporal to spatial phenomena. Specifically, we can account for differences in the distribution of temporal *for*-adverbials and their spatial counterparts. Spatial *for*-adverbials test for spatial atelicity but work analogously to temporal ones otherwise. For example, the predicate *meander* is spatially atelic because every part of a meandering road itself meanders, while the predicate *end* is spatially telic because if a road ends at a specific point, it does not end anywhere before that point:

(7) a. The road meanders for/*in a mile.
 b. *The road ends in/*for a mile.

The following minimal pair shows that spatial and temporal *for*-adverbials do not have the same distribution: (8a) is acceptable on an iterative interpretation, where John goes back and forth to the store many times, but (8b) is not acceptable on any interpretation.

(8) a. John pushed carts all the way to the store for fifty minutes. *temporally atelic*
b. #John pushed carts all the way to the store for fifty meters. *spatially telic*

The predicate *John pushed carts all the way to the store* involves reference to complex sum events, in some of which John may have gone back and forth and pushed carts little by little. Such an event is depicted in the space-time diagrams in Figure 2, which also shows the intuition behind the strata-theoretic explanation. The left-hand diagram corresponds to (8a) and the right-hand one to (8b).

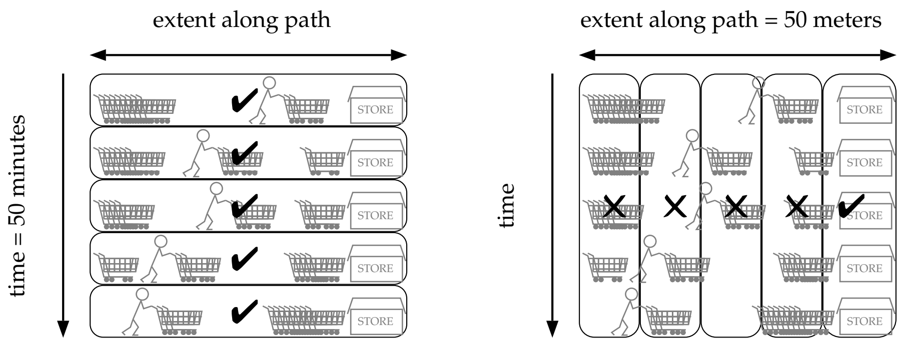


FIGURE 2. The predicate in (8) has the subinterval property in time but not in space

Both sentences require the predicate to hold at very small intervals that are parts of the interval they mention. In (8a), this interval is a fifty-minute time span. As long as each subinterval of this time span is the runtime of an event that satisfies the predicate *John pushed carts all the way to the store*, (8a) is predicted to be acceptable. The events marked with a checkmark satisfy this predicate in a scenario where John went back and forth and pushed carts all the way to the store little by little. In (8b), the interval is a fifty-meter long path. Most of the subintervals of this path do not contain the location of the store. That (8b) is unacceptable can then be explained by the assumption that an event whose location does not contain the store does not satisfy the predicate *John pushed carts all the way to the store*. For example, the events marked with an X on the right-hand diagram of Figure 2 do not satisfy this predicate. Thus, *push carts all the way to the store* has stratified reference with respect to the dimension *runtime* but not *spatial extent*:

(9) $\mathrm{SR}_{\text{runtime},[\![\text{50 minutes}]\!]}([\![\text{push carts all the way to the store}]\!])$
(Every event e in ⟦push carts all the way to the store⟧ can be divided into one or more parts, each of which is also in the denotation of ⟦push carts all the way to the store⟧ and has a very small runtime compared with fifty minutes.)

(10) $\neg\mathrm{SR}_{\mathrm{spatial\_extent},[\![50 \text{ meters}]\!]}([\![\text{push carts all the way to the store}]\!])$
(Not every event e in $[\![$push carts all the way to the store$]\!]$ can be divided into one or more parts, each of which is also in the denotation of $[\![$push carts all the way to the store$]\!]$ and has a very small spatial extent compared with fifty meters.)

We can now account for the contrast in (8) by the plausible assumption that *for 50 minutes* checks for the property negated in (9) and *for 50 meters* for the property in (10). More generally, the contrast between spatial and temporal *for*-adverbials supports a parametrized notion of the telic-atelic opposition, where the parameter is set either to time or to a spatial dimension. This parametrized nature of aspect has gone largely unnoticed in the past as researchers have focused on temporal aspect, but it is expected within the general picture of this work.

Case study 2: Measure Functions in Pseudopartitives (Chapter 7)

Strata theory makes it possible to draw a connection from the aspectual sensitivity of *for*-adverbials to the way in which pseudopartitives and other distributive constructions interact with formal properties of measure functions. A well known but poorly understood fact is that pseudopartitives accept certain measure functions and reject others. For example, weight, volume, and duration are admissible measure functions (*five pounds of books, five feet of snow, five hours of talks*), but speed and temperature are not (**five miles per hour of driving*, **five degrees Celsius of water*). This state of affairs raises two questions: First, how can we characterize the class of admissible measure functions in pseudopartitives? Second, why are not all measure functions admissible in the first place?

The following explanation is based on the idea that any account of the aspectual behavior of the *for*-adverbial *run for three hours* carries over to the pseudopartitive *three hours of running.* I view stratified reference as a presupposition that is imposed by both constructions on the constituent *run/running*:

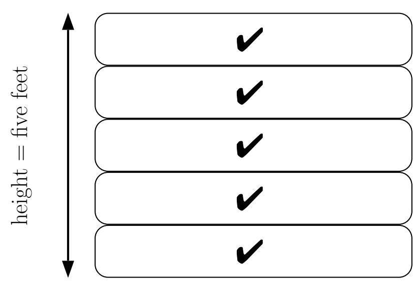
(11) run for three hours / three hours of running
Satisfied presupposition: $\mathrm{SR}_{\mathrm{runtime},\varepsilon([\![\text{three hours}]\!])}([\![\text{run}]\!])$
(Every running event consists of running subevents whose runtimes are very small compared to three hours.)

From this, we generalize to the assumption that the same presupposition that is found in *for*-adverbials is also found in pseudopartitives, just with other parameters. In substance-denoting pseudopartitives, we assume that the dimension parameter is the appropriate measure function. Mass nouns like *snow* are acceptable because they have divisive reference: whenever they apply to an entity, they also apply to all of its parts (again leaving aside very small parts).

(12) five feet of snow
Satisfied presupposition: $\mathrm{SR}_{\text{height},\varepsilon([\![\text{five feet}]\!])}([\![\text{snow}]\!])$
(Every snow amount consists of snow layers whose heights are very small compared to five feet.)

The idea behind the present account can again be understood via the visual metaphor. A plural or mass entity to which a pseudopartitive applies (for example some books, or some snow) is divided into strata which are very small as measured in the dimension determined by the pseudopartitive, but may extend arbitrarily in other dimensions. These strata are then required to be in the denotation of the noun. Singular count nouns always fail this test because the individuals in their denotation are atomic, and cannot be further subdivided into strata. Figure 3 illustrates at an intuitive level what (12) expresses formally: The measure function *height* is acceptable in the pseudopartitive *five feet of snow* because every amount of snow can be divided into (horizontal) layers of snow whose height is very small compared to five feet.

FIGURE 3. Accepting *five feet of snow*



A measure function like *temperature* is ruled out because smaller values are not guaranteed as you go from bigger to smaller amounts of substance:

(13) *thirty degrees Celsius of water
Failing presupposition: $\mathrm{SR}_{\text{temperature},\varepsilon([\![\text{thirty degrees Celsius}]\!])}([\![\text{water}]\!])$
(Every water amount consists of water parts whose temperatures are very small compared to thirty degrees Celsius.)

Thus, the measurement puzzle is answered in the following way. First, to be admissible in a pseudopartitive *Num N1 of N2*, a measure function μ has to be such that $\mathrm{SR}_{\mu,\varepsilon([\![\mathit{Num\ N1}]\!])}([\![N2]\!])$ is true. Second, the explanation for this phenomenon is that the constraint is independently attested in *for*-adverbials, and it also rejects telic predicates in *for*-adverbials and singular count nouns in pseudopartitives.

Case study 3: Connecting aspect, scope, and covers (Chapter 8)

Going beyond the case studies presented so far, the framework of strata theory offers new perspectives on a number of old problems. While most views on the minimal-parts problem agree that the level of granularity at which *for*-adverbials distribute must be less than maximally fine-grained, the present account makes it possible for the first time to relate this fact to cases of nonatomic granularity in other domains, such as distributivity over cells of a contextual cover. It is well known that VPs can only distribute over nonatomic entities (such as pairs) when there is supporting context that makes the relevant level of granularity salient. This is shown in the following contrast:

(14) a. These men weigh 250 pounds. *per pair reading unavailable*
b. These shoes cost fifty dollars. *per pair reading OK*

Within strata theory, this contrast can be related to the following fact: *For*-adverbials require quantifiers in their syntactic scope to be interpreted outside of their semantic scope out of the blue, but not when there is supporting context. Thus, (15a) is odd out of the blue, because *a flea* must take wide scope, resulting in an interpretation where one and the same flea is found over and over again. By contrast, (15b) is acceptable even when the pills covary with the subintervals of the month, in a context in which the patient's daily intake is salient, on a reading where *two pills* takes low scope under the *for*-adverbial.

(15) a. #John found a flea on his dog for a month.
b. The patient took two pills for a month and then went back to one pill.

The facts in (14) and (15) can be formally related in strata theory. To do so, I draw a connection between stratified reference and accounts which allow verb phrases to shift to a distributive reading under specific conditions. This operation plays a central role in standard theories of distributivity, where it is known as the D operator. In strata theory, applying the D operator to a verb phrase can be seen as a means to locally ensure that the stratified reference presupposition of a distributive item is satisfied.

Case Study 4: *All* as a Distributive Item (Chapter 9)

This dissertation takes the novel step of relating the atelic-telic opposition to the collective-distributive opposition, and expands the standard theory of collectivity to make it sensitive for certain distinctions within the class of collective predicates. Distributive predicates require any event in their denotation to be divisible into events that are atomic with respect to the appropriate thematic role. For example, any plural event in the denotation of the distributive predicate *smile* must be divisible into parts that have atomic agents and that belong to the denotation of the same predicate. This captures the distributive entailment from *John and Mary smiled* to *John smiled*. By contrast, collective predicates like *be numerous* or *be a* *motley crew* do not satisfy stratified reference on the thematic role of their subjects, because their subjects can be plural entities whose parts do not themselves satisfy the predicate. This view sheds new light on a set of problems concerning the interaction of *all* with cumulative quantification and dependent plurals. *All* normally cannot give rise to scopeless (cumulative) readings, in contrast to numeral quantifiers like *three*:

(16) a. Three safari participants saw thirty zebras.
Available scopeless reading: Three safari participants saw at least one zebra each, and thirty zebras were seen overall.
b. All the safari participants saw thirty zebras.
Unavailable scopeless reading: Each safari participant saw at least one zebra, and thirty zebras were seen overall.

However, if the VP contains a bare plural, the scopeless reading is available with *all*:

(17) a. Three safari participants saw zebras.
Available scopeless reading: Three safari participants saw at least one zebra each, and at least two zebras were seen overall.
b. All the safari participants saw zebras.
Available *scopeless reading:* Each safari participant saw at least one zebra, and at least two zebras were seen overall.

To explain these facts, the word *all* is claimed to presuppose that the VP has **stratified reference down to atomic agents**. This means that in a sentence of the form *All NP VP*, the agent of every VPing event is presupposed to consist of one or more atomic parts that are agents of VPing events. This presupposition rules out the scopeless reading of (16b), while the scopeless reading of (17b) is available because its presupposition is satisfied (18b):

(18) a. **Failing presupposition:** $\mathrm{SR}_{\text{agent, Atom}}([\![\text{see thirty zebras}]\!])$
(Every see-thirty-zebras event consists of subevents with atomic agents and in each of which thirty zebras are seen.)
b. **Satisfied presupposition:** $\mathrm{SR}_{\text{agent, Atom}}([\![\text{see zebras}]\!])$
(Every event in which at least one zebra is seen consists of subevents with atomic agents and in each of which at least one zebra is seen.)

Evidence for treating *all* as a distributive item comes from the fact that it rejects certain collective predicates. Thus it is unacceptable to say **All the men who run this country are politically homogeneous* or **All the soldiers in this bataillon sufficed to defeat the army*, even though these sentences are acceptable without *all*. That *all* is a distributive item raises the question why it behaves differently from other distributive items like *each* (*All the boys gathered* vs. **Each of the boys gathered*). I develop a novel theory of collectivity against whose background this difference can be traced to different settings of the granularity parameter.