Classifying Processes and Basic Formal Ontology

Mustafa Jarrar\textsuperscript{1,2} and Werner Ceusters\textsuperscript{1}
\textsuperscript{1} Department of Biomedical Informatics, University at Buffalo, 77 Goodell Street, Buffalo, USA
\textsuperscript{2} Department of Computer Science, Birzeit University, Ramallah, Palestine

\textbf{ABSTRACT}

Unlike what is the case for physical entities and other types of continuants, few process ontologies exist. This is not only because processes received less attention in the research community, but also because classifying them is challenging. Moreover, upper level categories or classification criteria to help in modelling and integrating lower level process ontologies have thus far not been developed or widely adopted. This paper proposes a basis for further classifying processes in the Basic Formal Ontology. The work is inspired by the aspectual characteristics of verbs such as homeomericity, cumulativity, telicity, atomicity, instantaneity and durativity. But whereas these characteristics have been proposed by linguists and philosophers of language from a linguistic perspective with a focus on how matters are described, our focus is on what is the case in reality thus providing an ontological perspective. This was achieved by first investigating the applicability of these characteristics to the top-level processes in the Gene Ontology, and then, where possible, deriving from the linguistic perspective relationships that are faithful to the ontological principles adhered to by the Basic Formal Ontology.

\section{INTRODUCTION}

Living, growing, learning, purchasing, producing, sleeping, and mating are examples of different types of processes, i.e. entities with temporal parts that depend on other entities to occur.

The importance of process ontologies is rapidly increasing in several domains such as in event discovery (Nevatia et al 2004, Li et al 2017), industry and engineering (Morbach at 2007), software engineering (Ruy et al 2015), affective computing (Li et al 2016), among others. In biomedicine, there are several ontologies that have been recently developed or extended to cover a wide spectrum of types of processes, such as the process components of the Gene Ontology, the Emotion Ontology, and the Mental Functioning Ontology, to name a few. The Gene ontology (GO 2001) groups all processes under \textit{biological process}, which is informally defined as a collection of molecular events, specifically pertinent to the functioning of living beings and with defined temporal beginning and end. There are about 26k biological processes in the Gene ontology classified under 27 top-level processes, such as growth, development, behavior, response to stimulus, metabolic and immune system processes. These top-level processes were developed and are being extended without a formal or foundational framework. The Mental Functioning Ontology (Hastings et al. 2012) was developed as a framework to enable developing other ontologies related to mental health and diseases. It distinguishes between the aspects of mental functionings that are occurrents and those that are continuants. For example, intelligence and personality are dispositions while behaviors and mental processes are \textit{bodily processes}. The \textit{memory image} I have about my dad now (a cognitive representation) is related to my \textit{remembering} him now (a mental process). There are about 80 mental processes (e.g., learning, thinking, wanting, arousal, and perception) and about 500 behavioral processes (e.g., cognitive, rhythmic and social behaviors). A related ontology, the Emotion Ontology (Hastings et al., 2011) distinguishes between three notions related to specifying emotions: emotional processes, emotional dispositions, and mental representations. About 170 types of processes have been covered in this ontology including emotion processes, mood processes, and emotional behaviors.

In BFO (Arp et. al. 2015, Smith et. al. 2012), processes and process boundaries are defined under occurrents, but they are not elaborated further. Processes are those entities that occur, happen, unfold, or develop in time, have temporal proper parts, and depend on some continuant entity to happen. Process boundaries are other types of occurrents that occupy zero-dimensional temporal regions, thus they do not have temporal parts and are not processes themselves. Examples of process boundaries are the moment of a person’s birth, and that what is described by terms such as ‘midnight’, ‘departure’, and ‘arrival’.

In DOLCE (Masolo 2002), the corresponding entities are called \textit{perdurants}, which are those entities that happen in time, by accumulating different temporal parts: thus at any time \( t \) at which they exist, only their temporal parts at \( t \) are present. DOLCE was inspired for its classification by the lexical semantics literature, thereby leaning on properties such as homeomericity and cumulativity, which were used to classify perdurants into \textit{states} and \textit{processes} if they are accumulative, or into \textit{accomplishments} and \textit{achievements} if they are non-accumulative. States are distinguished from processes if they are homeomeric, and achievements are distinguished from accomplishments if they are atomic.

Although DOLCE and BFO are based on distinct perspectives, it is not unreasonable to view processes and accomplishments in DOLCE as what are processes in BFO. However, how achievements and states are to be interpreted by BFO remains unclear. If achievements are instantaneous happenings with zero-time duration then they correspond to

\textsuperscript{1} To whom correspondence should be addressed: mjarrar@birzeit.edu
BFO’s process boundaries. If, however, instantaneity is meant to be a short period of time, then they are BFO processes. Additionally, states in DOLCE cannot be BFO processes as they do not involve change.

Early research in philosophy and linguistics provided different accounts of the so-called events, suggesting various views and criteria to distinguish between actions, activities, accomplishments, achievements, processes, performances states, mental and physical events, bodily movements, and others (c.f., Casati et al 2015). Most of this research was originally devoted to the semantics of verbal phrases, the so-called ‘lexical aspect’ of verbs, and how they are structured in relation to time. Figure 1 reflects our understanding of how some of these accounts (e.g., Moens et al 1988, Bhatt 2005, Levin 2009) may be combined. (Vendler 1957) classified events into: activities, states, accomplishments, and achievements. This proposal was later revised and extended (e.g., Mourelatos 1978, Bach 1986, Krifka 1998, Caudal et al. 2005, Trypuz et al 2007). Krifka made it very clear that caution is required: these classifications are about predicates, i.e. descriptions, denoting entities such as processes in reality, and are not classifications about processes themselves: the same process can be described by distinct predicates each one of which can be classified differently. Moreover, processes and events might be lexicalized in different ways in natural languages. For example, the exact same event instance of John’s speaking to Mary yesterday, can be said in different ways, such as ‘I heard John speaking with Mary’, ‘John speaks with Mary’, or ‘John spoke with Mary’. The difference between those English phrases is ontologically irrelevant as they all refer to the same event. Furthermore, there are verbs that do not denote or refer to events, such as ‘It costs 205’, ‘It weights 20kg’, or ‘It looks easy’, which do not imply changes, as discussed earlier. Similarly, there are events that cannot be typically lexicalized using verbs, such as ‘war’ and ‘conference’.

In this paper, we examine the most commonly discussed criteria in the literature to classify events and verbal predicates: homeomerity, cumulativity, atomicity, telicity, durativity and instantaneity. Other aspehtical notions, such as incrementality and structure (Caudal 2005), distributivity and collectivity (Champollion 2014), and quantization (Krifka 1998), are here not dealt with.

2 GENE ONTOLOGY PROCESSES

To test the applicability of these notions, we attempted to use them to analyze and annotate the top level processes in the Gene Ontology. Table 1 illustrates a sample of these GO processes. The full annotations (and documentation of choices) can be accessed online1. We selected the 35 most top-level processes, that top all other (~45k) processes in the Gene Ontology.

It is worth noting that the top-level processes in the Gene ontology include not only process universals, but also defined classes, such as Single-Organism Process and Multi-Organism Process. Such process types cannot be classified according to criteria such as whether they are e.g., homeomorphic or telic because they are collections of different types of processes, defined based on the number of the organisms involved. Therefore, instead of selecting these processes, we selected and annotated their subtypes.

This initial effort turned out to be a challenging task for two reasons: (1) shortcomings in the literature about what precisely is to be understood by these notions, and (2) their focus on how matters are described (the linguistic perspective), rather than on what is the case (the ontological perspective).

Table 1. Sample of the top-level GO processes annotated.

In the next sections, we examine the most commonly used aspehtical criteria and discuss whether and how they might be used for further classifying processes in the spirit of BFO.

3 DEFINITIONS

Central in our approach is the distinction between temporal parts of processes, such as the first year of a toddler’s life, and mere occurrent parts of processes, as each of the eating processes that are part of a toddler’s life. 2

\[ p \text{ occurrent-part-of } q \]  \hspace{1cm} (R1)

a primitive relation of parthood holding independently of time between two process instances when one is a sub-process of the other (Arp et al 2015:135).

---

2. From here on forward, we follow the standard typographical conventions for particulars (lowercase, italics in definitions), universals (upper case) and relationships (bold for particular-level relationships, italics for universal-level relationships) (Arp et al 2015)
be it half the duration of temporal ('bpm' variations in the number of heart beats per minute) of a pianist's heart or of the number of bars in a piece of music produced by that pianist. There is some particular occurrent part which is defined by an instance of \( P \) and \( p \) temporal-part-of \( q \) (Arp et al 2015:139).

\[ p \text{ temporal-part-of } q = \text{def.} \]

\[ p \text{ temporal-part-of } q \]

& for some temporal region \( r \) \( p \) spans \( r \)

& for all occurrents \( c, r' \)

if \( c \text{ spans } r' \) & \( r' \) occurrence-part-of \( r \)

then \( c \text{ occurrence-part-of } p \) iff \( c \text{ occurrence-part-of } q \) (Smith 2012, corrected).

The relation ‘\( p \) spans \( r \)’ in R3 is shorthand for ‘\( p \) occupies spatiotemporal region \( str \) which occupies temporal region \( r \)’. R3 allows us to define the corresponding universal-level relation:

\[ P \text{ temporal-part-of } Q = \text{def.} \]

for every particular occurrent \( p \), if \( p \) instance-of \( P \), then there is some particular occurrent \( q \) such that \( q \) instance-of \( Q \) and \( p \) temporal-part-of \( q \).

Figure 2 depicts an instance of PROCESS PROFILE, a subtype of PROCESS (Smith 2012). Displayed is a representation of an instance of a rhythm profile, \( p1 \), which could be an occurrence-part-of the beating of some person’s heart or of the performance of a piece of music by a pianist and which in these cases would reflect the variations in the number of heart beats produced by that person’s heart, resp. musical bars played by the pianist (‘bpm’ = beats/bars per minute). The figure also depicts two processes of equal duration, \( p2 \) and \( p3 \) which both are temporal-part of \( p1 \), as well as the further temporal parts \( p4, p5, p7 \) and \( p8 \) which also are each of the same duration, be it half the duration of \( p2 \) and \( p3 \). Further depicted are the various defined classes instantiated by these processes.

### 3.1 Homeomericy

DOLCE considers an occurrence to be homeomeric iff all of its temporal parts, in BFO sense, are described by the very expression used for the whole occurrence, a view which is close to the notion of homogeneity as defined in (Dowty 1977:60). An example of a homeomeric process in this DOLCE sense would be what is described by ‘sitting’. Defined in this way, homeomericy is a property of universals, not of instances. Although BFO is a classification of instances and not of universals, we can work around this by recognizing several flavors of this principle at the level of instances:

(1) \( p \) isotypic-part-of \( q \) =def.

\[ p \text{ temporal-part-of } q \]

& \( p \) instance-of all types instantiated by \( q \).

Example (figure 2): \( p6 \) isotypic-part-of \( p5 \).

(2) \( p \) weakly-homeomeric-in \( P \) =def.

\[ p \text{ temporal-part-of } q \]

& \( p \) instance-of all types which are not process boundaries are instances of \( P \).

Examples (figure 2): \( p2 \) weakly-homeomeric-in DC1; \( p3 \) weakly-homeomeric-in DC1.

From R5 and R6 it follows that all parts of a process that is weakly-homeomeric-in some type are isotypic-part-of that process.

(3) \( p \) strongly-homeomeric-in \( P \) =def.

\[ P \text{ temporal-part-of } q \]

all temporal parts of \( p \) which are not process boundaries are instances of \( P \) and there is no such part of \( p \) that instantiates a subtype of \( P \).

Example (figure 2): \( p3 \) strongly-homeomeric-in DC9, \( p8 \) strongly-homeomeric-in DC11.

From R5 and R7 it follows that all parts of a process that is strongly-homeomeric-in some universal are isotypic-part-of that process.

Further to be investigated is whether it is worthwhile to introduce the DC ‘weakly-homeomeric-process’ as the DC formed by all processes which are – or have an occurrent part which is – weakly-homeomeric-in some process type. Also whether it would pay off to implement the notion of ‘strongly-homeomeric-process’ as the DC formed by all processes which are– or have an occurrent part which is – strongly-homeomeric-in some process type or whether it would be sufficient to have an axiom directly at the level of the types to which it applies. If section 2.1 would hold for DC8, 9, 10 and 11. It is also worth noticing that whether any of the relations R5…7 holds, depends on whether (1) a pure metaphysical stance is taken or (2) reality is looked at through what an ontology allows us to see and what is observable at the level of instances. If a restricted ontological commitment would not allow us to observe or describe instances \( p_e…p_s \), then, under that commitment, e.g., \( p_1 \) would be strongly-homeomeric-in DC5 and DC6.
3.2 Cumulativity

Cumulativity was extensively discussed in the lexical semantics literature (see e.g., Krifka 1989, and Champollion 2014) in an attempt to describe verbs – not processes! – a distinction similar to what exists for mass nouns on the one hand and count nouns on the other hand: whereas two portions of water together make one (bigger) portion of water, two bottles of water together do not make one (bigger) bottle. This view was later adopted in DOLCE for classifying perdurants: a perdurant is cumulative if the mereological sum of two instances of a type of perdurant is also an instance of the same perdurant type; for example, ‘the sum of two sittings is still a sitting’ (Masolo 2002:24). Nevertheless, both this definition and the example are rather unclear.

Looking at it from the BFO perspective, it leaves open the question what sort of mereological sum is intended. Does it include occurrent parthood (see R1), as when my sitting (p1) simultaneously with your sitting (p2) is summed to form p3 which is the sitting of the mereological sum of me and you? Would certain process aggregates such as an orchestra playing a symphony being the mereological sum of the playing of the individual musicians plus the directing of the conductor? Or is it exclusively temporal parthood (R3) such as my sitting during the first 30 minutes of my total sitting and my sitting during the last 30 minutes?

The DOLCE documentation about cumulativity as well as related proposals advanced in the linguistic community are not at all clear whether the sort of cumulativity they have in mind is a property of processes or of how a process is described. (Champollion 2014, Galton 2016), for instance, contend that cumulativity has more to do with the level of detail at which a process is described, rather than what it is ontologically: if a process p is described as ‘flying’ then cumulativity holds, as the sum of any two parts of such a flying is also a flying. However, if the very same p is described as ‘flying from London to New York’ then cumulativity, under their view, does not hold. From the BFO perspective, this meaning of cumulativity does not make sense: no entity becomes different because it is described differently. As with homeomericty, cumulativity as defined in DOLCE can be regarded as a property of some universals which requires some work-arounds in BFO, f.i.:

\[ p \text{ cumulative-with } q = \text{def.} \quad (R8) \]

all process types instantiated by p and all process types instantiated by q are instantiated by p, q and \( p + q \).

Example (figure 2): \( p_1 \text{ cumulative-with } p_2 \); DC1, DC5 and DC6 are the all and only types that are instantiated by both \( p_2 \) and \( p_3 \), as well as \( p_1 \) (the mereological sum of \( p_2 \) and \( p_3 \)).

\[ p \text{ cumulative-in } Q = \text{def.} \quad (R9) \]

\[ P \text{ isa } Q \]

& for all \( p_1, p_2 \) instance-of \( P : (p_1 + p_2) \text{ instance-of } Q \).

Example (figure 2): DC5 cumulative-in DC1.

Clearly, if some process \( p \) is at least weakly-homeomeric-in \( P \) then it is also instance-of a type which is cumulative-in \( P \), but not vice versa. One could assume, for example, that instances of Growth as defined in the Gene ontology would be weakly-homeomeric-in Growth and that all subtypes of Growth would be cumulative-in Growth. Subtypes of Cell Aggregation, however, could be assumed to be cumulative-in Cell Aggregation, but for sure no instance of it would be strongly-homeomeric-in Cell Aggregation: not every temporal part of a cell aggregation process is of the same type (e.g. its sub-process ‘cells coming close to each other’ is not itself an aggregation). However, the mereological sum of two cell aggregations would be a cell aggregation.

3.3 Telicity

From a lexical semantics perspective, an action predicate is telic if it refers in one way or another to a terminal point for the action described as tending ‘towards a goal envisaged as realized in a perfective tense but contingent in an imperfective tense’ (Garey 1957:106). An action predicate is atelic if it does not mention any goal, purpose or endpoint, but can be used to denote an action as soon as it begins. For example, the predicate running is from such perspective considered aletic because it does not mention any goal or termination point. Furthermore, if somebody is running a marathon (process \( p \)), then the predicate running can be used to describe \( p \) immediately after its start, and independent of the outcome, i.e. whether or not the runner finishes. The predicate running a mile is however considered telic (Krifka 1989:9): it mentions an endpoint and it can only safely be applied to \( p \) after completion. The question we seek to answer is whether this notion of telicity can in one or other form be applied to processes themselves, rather than to predicates under which these processes are described. In BFO terms: is telicity a notion that applies only to representational units, or can it also be applied to that what the representational units are about? If the latter is the case, then telic processes would be those that have some terminal point or goal – whether or not these processes are described as such – and aletic ones those that do not.

There are for sure processes that are telic under a specific interpretation of ‘terminal point’, i.e. all those which are finished! This is of course not what we have in mind here. What we do mean is that for a process to be telic, there must be something ‘in it’ what we informally can describe as ‘a change in the process’. However, this is just a matter of speaking as under the perspective of BFO, processes do not change because they are changes!

Consider the current laws of physics on Earth, one being that objects such as a rock without support fall down to
Earth. The rock cannot keep falling down forever, it naturally stops when reaching a surface. Therefore, any such ‘falling under natural Earth conditions’ is a telic process. This ‘falling’ terminates with (is followed by) a process that can be roughly described as ‘the coming to a stop’ and which can take the form of bouncing a bit (on a hard surface), or penetrating a soft surface (sand, for instance); it is this coming to a stop which is that what informally can be described as ‘the change in the process’. As another example, imagine that by looking through the representational units in figure 2 we do not just see the process profile \(p_1\) and the parts \(p_2, p_3, \ldots\) thereof, but the actual beating of somebody’s heart \(b_1\) within that timeframe, and the process parts thereof – process parts being processes in their own right – \(b_2, b_3, \ldots\), such that \(p_1\) is an occurrence-part of \(b_1, p_2\) occurrence-part of \(b_2, \ldots\), and so forth. Then \(b_2\) did not ‘change’ when \(b_2\) terminated and \(b_3\) started, and neither did \(b_1, b_2\) is the change! Remember also that from the BFO perspective it does not make sense to refer to processes ‘at a time’. We believe that telicity is once more not some unary property of an individual process, but rather something in relation to a universal. So we define:

\[
p \text{ is-telic-in } R \stackrel{\text{def.}}{=} \text{def.} \\
& \quad \text{there exists some process } q \text{ instance-of } Q \text{ and} \\
& \quad \text{some process } r \text{ instance-of } R, \text{ such that} \\
& \quad (1) q \text{ not instance-of } P, (2) p \text{ not instance-of } Q, \\
& \quad (3) p \text{ precedes } q, \text{ and} \\
& \quad (4) p \text{ and } q \text{ are temporal-part-of } r.
\]

Example (figure 2): each of \(p_4, p_5, p_7\) is-telic-in DC1.

\[
P \text{ is-telic-in } R \stackrel{\text{def.}}{=} \text{def.} \\
& \quad \text{for all process } p \text{ instance-of } P, p \text{ is-telic-in } R.
\]

For example: falling (i.e. under normal earth conditions) is-telic-in moving (\(p\) is the falling proper, \(q\) is the coming to a stop (bouncing, penetrating, \(\ldots\)), \(r+p+q\) is the entire motion process). Note that in general language, both \(p\) and \(r\) might be denoted by ‘falling’ Another example: ‘walking leg swing’ is-telic-in ‘walking’.

We do not follow (Hennig 2008:262-4) according to which ‘a telic process can be interrupted any time but it can be only complete when its goal is reached’. Thus, if a person on a cross road intends to cross over and lifts up his left leg to start walking but gets immediately run over by a car, the movement that was initiated was a ‘leg swing’, though not a ‘walking leg swing’. Our view is however compatible with (Krifka 1998) in that if a process is telic (in \(P\)), then there is no temporal part of it that is telic (in \(P\)).

### 3.4 Instantaneity

Some researchers distinguish between processes that are instantaneous and others that are durative (e.g., Garey 1957, Mourelatos 1978, Krifka 1998). They consider an event to be durative if it takes place over extended intervals, or instantaneous (also called punctual) if it takes just a moment, i.e., a complete action with no explicit internal temporal structure, such as arriving and departure. We argue that all processes are durative, and that there are no processes that happen instantly in zero-time. We distinguish between e.g., ‘the arriving of a train’ as a process extending over time, and ‘the arrival of a train’ as a process boundary.

Nevertheless, the instantaneous events discussed in the linguistics literature are likely not meant to be time points or process boundaries, but rather, a sort of processes. Thus, we revise the definition of instantaneous processes to be those types of processes that are the peak moments of some longer processes. Processes like, knocking, hitting, departure, arrival, and dying are examples of processes that extend over time, but they have temporal parts that we call peak moments. For example, knocking the door starts when one moves his hand towards the door, then continues by touching and hitting the door for a short moment (the peak moment), and then releasing the hand. Similarly, the arriving of the train may start when the first trolley of the train starts entering the station, and decreases its speed into zero (the peak moment), wait a bit, and then open its doors.

Notice that peak moments are critical parts. For example, if a process started with its agent intending it to be a knocking process but is terminated before its peak moment (i.e., before touching the door) then it is not a knocking, but if it is terminated directly after its peak then it is a knocking. In fact, this shows that instantaneous processes cannot be strongly-homeomeretic in some universal because their temporal parts before and after their peak are not of the same type. One may even slice a process into a higher granularity and ask at what time point a process is considered done, but we believe that this is irrelevant since the boundaries of instantaneous processes are typically flat boundaries. For example, the exact starting and ending of knocking, and similarly a train’s arrival, are typically a matter of flat determination. Further to be investigated is whether an entity (continuant or occurrent) is always created at a peak moment, e.g. a sound when knocking, a new individual when giving birth, a hole when drilling a wall, and so on. For this reason, no formalization for instantaneous processes has yet been decided upon.

### 3.5 Atomicity

Atomicity was adopted by DOLCE to distinguish between accomplishments (non-atomic) and achievements (atomic) (Masolo 2002:24) but its definition is unclear. A common understanding of it in the event semantic literature is that it is a one-step change-of-state, i.e., lacking any internal sub events (c.f., Caudal et al. 2005, Krifka 1998). But also this understanding of atomicity is problematic. Since every part of the temporal region occupied by a
process is occupied by a process which is part of the larger process, atomicity depends on the granularity level and is subject to one’s perspective. As pointed out in (Krifka 1989:80): “The notion of different granularities can be applied in cases where an entity appears as atomic under one description and as composed of different entities under another description. For example, an arrival of a train at a station may be considered as atomic event or as an event which is composed of subevents... atomicity depends on the position we take towards the minimal part problem”. Therefore, we do not find it opportune (yet) to propose a formalization compatible with BFO.

4 CONCLUSION AND FUTURE DIRECTIONS

We have overviewed some known aspectual notions used to classify verbal phrases and examined their reuse to build process ontologies under BFO. We provided BFO-compatible interpretations of homeomericity, cumulativity and telicity, discussed instantaneity, and explained that atomicity does not seem to be beneficial in classifying processes. We illustrated the use of these notions to analyze and annotate the top levels of the Gene Ontology processes. We plan to extend this work in several directions, including the full ontological analyses of the Gene Ontology processes, in order to detect possible modeling inconsistencies, as, for example, would be the case when a non-homeomeric process is declared a subtype of a homeomeric process, extending BFO process categories, as well as examining the use of these notions at a larger scale, such as restructuring of the process types found in WordNet.

ACKNOWLEDGEMENTS

Part of this research was conducted during Prof. Jarrar’s visit to the Ontology Research Center at UB, which was funded by the Fulbright Visiting Scholars Program. This work was supported in part by CTSA NIH 1 UL 1 TR001412-01 from the National Institutes of Health.

REFERENCES