Counting Time and Events

Kiyng Lee Korea University Deaprtment of Linguistics Ti Seoul, Korea ikiyong@gmail.com, klee@korea.ac.kr

Harry Bunt Tilburg University TiCC: Center for Cognition and Communication Tilburg, The Netherlands kr harry.bunt@uvt.nl

Abstract

Recurring events (e.g., John calls twice everyday) involve both temporal and event quantification. To annotate such events, there are two main approaches: one approach is represented by Pustejovsky et al. (2010a,b) and the other one, by Bunt and Pustejovsky (2010) and Bunt (2011a,b). In the framework of ISO-TimeML (2012), the first approach encodes information on quantification directly into both temporal and event entities, by introducing the attribute @quant into the element <EVENT> as well as the element <TIMEX3>. The second approach views quantification as a set of properties of the way a predicate applies to a set of arguments, or relates two sets of arguments, such as sets of events and their time of occurrence, and therefore annotates aspects of quantification as part of a relational link, such as <TLINK> or <TIME_ANCHORING>. In this paper, we discuss alternatives and explore possibilities to reach general consensus on the annotation of quantification over time and events and its extendibility to other entities.

1 Introduction

In January 2012, ISO-TimeML (2012) was published as an ISO's international standard: ISO 24617-1(E):2012 Language resource management - Semantic annotation framework - Part 1: Time and events (SemAF-Time, ISO-TimeML).¹ Although it was officially published as an international standard, this document still contains a couple of issues that remain to be resolved, especially those related to the annotation of recurring time and events that involves quantification, distributivity, and scopes.

These issues involve for example the annotation of the following sorts of expressions:

- (1) Sample Data for Recurring Time and Events a. Type 1: John called *twice*.
 - h. Tana 2. Jahr calle
 - b. Type 2: John calls *every day*.
 - c. Type 3: John calls *twice a day*.

In ISO-TimeML (2012), the predicate modifiers, italicized in (1), are all treated as referring to temporal entities, thus all are annotated into the element, named <TIMEX3>, and almost in the same manner.

Detailed analysis shows, however, that each of them should be annotated differently, involving temporal and event quantification and scopes. This difference requires minor or major modifications in the current version of ISO-TimeML (2012). Far before its publication, some issues on quantification had been known and much discussed. On these issues, we note at least two proposals besides the published standard itself: (1) Pustejovsky et al. (2010a,b) and (2) Bunt and Pustejovsky (2010) and Bunt (2011a,b).

The first proposal is a minimally modified version of ISO-TimeML (2012) with its representation scheme, which we tag <isoTimeML_m>. It differs from the representation scheme of ISO-TimeML (2012), <isoTimeML>, in two ways. First, this minimally modified version <isoTimeML_m> annotates event quantification by introducing the at-

¹SemAF-Time refers to the document as a whole, while *ISO-TimeML* refers to the XML-based annotation language specified in that document.

tribute @quant into the element <EVENT>. Second, it marks up the scopes of temporal and event quantification explicitly by introducing the attribute @scopes into both <TIMEX3> and <EVENT>.

The second proposal annotates quantification over events in a different way. Based on the analysis of quantification in Bunt (1985), it views quantification as a set of properties of the way a unary predicate applies to a set of arguments, or a binary predicate relates two sets of arguments. A predicate that relates a set of events to their times of occurrence, as expressed for instance by the preposition "at" in the sentence "John always calls at two o'clock" is annotated in ISO-TimeML with the relational link <TLINK>, and aspects of quantification are therefore annotated as part of this link. Since the *<*TLINK*>* tag is heavily overloaded in ISO-TimeML, as it is used for rather different purposes, hence for this specific use the tag <TIME_ANCHORING> is introduced. Annotation are marked up in a representation scheme, called the Ideal Concrete Syntax, which is designed according to the CASCADES methodology of designing annotation languages with an abstract syntax and a formal semantics (see Bunt, 2010; 2012) - this approach to the annotation of quantification is tagged <isoTimeML_ICSrep>.

While on the one hand quantification on the latter view is considered to arise when a predicate is applied to one or more sets of arguments (rather than to arguments which are single individuals), and it thus seems natural to annotate aspects of quantification as parts of relational link elements, it was noted in Bunt (1985) on the other hand that satisfactory semantic representations of sentences with quantifications can be obtained by considering aspects of quantification as parts of the compositional semantics of noun phrases. This is because NP representations can be defined in such a way that they anticipate on the use of the NP as an argument of a predicate, as already shown by Montague (1973).

The treatment of quantification proposed by Montague did not take the phenomenon of 'distributivity' into account, however, i.e. whether the members of an argument set are involved in the predication as individuals, in groups, or as a collectivity - see e.g. the example "Two men lifted the piano". Bunt (1985) showed that it is possible to construct semantic representations for noun phrases with different distributivities; interestingly, though, distributivity is often not expressed in a noun phrase, but by adverbials like "together" and "one by one", so it is not evident that this aspect of quantification would most conveniently be treated as part of NP semantics or in the semantics of combining an NP with a predicate.

The main purpose of this paper is to discuss and explore possibilities to annotate quantifications over time and events, for use in a future extended version of ISO-TimeML (2012), but also to contribute to the study of how to annotate quantification more generally, as explored in ISO project 24617-6, "Basic principles of semantic annotation", since in the end a treatment of quantification over time and events should be a special case of quantification more generally.

2 Two Annotation Schemes

Frequency is normally understood to be a number of occurrences of a repetitive event over a period of time; predicate modifiers such as "twice", "every day", and "twice a day" or "twice every day" are often treated as frequency expressions. In this section we discuss how these modifiers are annotated in two different representation schemes: <isoTimeML> and <isoTimeML_ICSrep>.

2.1 Annotation of Type 1 Modifier twice

ISO-TimeML (2012) annotates "twice" as a temporal entity expressing a frequency, encoding its information into the element <TIMEX3>.

(2) a. John called_{e1} twice_{t1}.

```
b.<isoTimeML xml:id="a1">
<EVENT xml:id="e1" pred="CALL"
tense="PAST"/>
<TIMEX3 xml:id="t1" freq="2X"/>
<TLINK eventID="#e1"
relatedToTime="#t1"
relType="DURING"/>
</isoTimeML>
```

This is interpreted as shown below:²

²The semantics of this interpretation was developed by Pratt-Hartmann as an extension of Pratt-Hartmann (2007). For the compositional process of deriving the two semantic representations, see Clause 8.4.3.3 Quantifying *<*TIMEX3*>* element in ISO-TimeML (2012), pp. 32-33.

(3) Interval_based First-Order Form: $\exists 2I_{e1}(R_{during}(I_{e1}, I_{t1}) \land p_{call}(I_{e1})$

This semantic form is understood as stating that, within a *contextually determined interval of time* I_{t1} , there were two instances of an event of type CALL-ING. Each interval I_{ei} is understood as an interval of time in which an event of type e is instantiated as e_i .³

Bunt and Pustejovsky (2010) argue that the modifier "twice" does not denote a temporal entity at all, but it is simply a counter, expressing how often a certain type of event occurred. <isoTimeML_ICSrep> thus annotates it as a part of the element <EVENT>.⁴ This element is then specified with two attributes @signature with a value SET and @cardinality for the cardinality of the (specified) set, as shown below:.

(4) a. John_{token1} called_{token2} twice_{token3}. b. <EVENT xml:id="e1" target="#range(token2,token3)" type="CALL" tense="PAST" signature="SET" cardinality="2"/>

This is interpreted as stating that, given a set E of two events, each event in E is of type CALL. This is also given a formal semantics in DRT (Discourse Representation Theory) of Kamp and Reyle (1993), as shown below:

(5) a. $\exists 2e[call(e)]$ b. $\exists S[|S|=2 \land \forall e[e \in S \rightarrow call(e)]]$

These two forms are equivalent.

2.2 Type 2 Modifier *every day* as a Temporal Quantifier

Both of the annotation schemes <isoTimeML> and <isoTimeML_ICSrep> treat type 2 modifiers as temporal quantifiers, but annotate them differently. The former annotates "every day" as part of the element <TIMEX3>.

This is interpreted as stating that, during each day, the event of John's calling occurred, as represented in two different forms:

(7) a. Interval-based: $\forall I_{day} R_{during}(I_{call}, I_{day})$ b. Event-based: $\forall t[day(t) \rightarrow call(e, t)].$

The latter scheme <ISO-TimeML_ICSrep>, however, divides the task of annotating "every day" over two elements: one is a new element, called <PERIOD>; the other is <TIME_ANCHORING>. The temporal noun "day" in "every day", for instance, is annotated into <PERIOD>, while the quantifier "every" is annotated into <TIME_ANCHORING> as a value of its attribute @timeQuant.⁵

(8) a. John_{token1} calls_{token2} every_{token3} day_{token4}. b. <isoTimeML_ICSrep> <EVENT xml:id="e1" type="CALL" target="#token2" signature="SET"/> <PERIOD xml:id="t1" type="DAY" target="#token4" signature="SET"/> <TIME_ANCHORING anchoredEvent="#e1"

 $^5 Lee~(2012)$ argues that there should be some formal constraints on the assignment of attriutes to links, on the basis of which we can, for instance, justify the validity of assigning such attributes as <code>@timeQuant</code> and <code>@eventDistr</code> to <TIME_ANCHORING>.

³In TimeML, from which ISO-TimeML was developed, <EVENT> was instantiated with an element <MAKEINSTANCE> and <TLINK> related these instances to each other or to a time. The ITL-based semantics of Pratt-Hartmann (2007) followed this version of TimeML and did not treat the semantics of "twice" or any quantifier expressions: it simply fails to treat quantification over events.

⁴The way of annotating "twice" directly into the element <EVENT> is exactly the same as that approach which annotates the negative expression "n't" or the tense "did", for instance, in "didn't call" into <EVENT> by providing it with information on its polarity and tense.

```
anchorTime="#t1"
tempRel="INCLUDED_IN"
eventDistr="INDIVIDUAL"
timeDistr="INDIVIDUAL"
timeQuant="EVERY"/>
</isoTimeML_ICSrep>
```

The semantics defined for the abstract syntax that underlies this representation yields the desired interpretation (see Bunt 2011a,b).

Note that the quantifier "every" in "every day" has ended up in the <TIME_ANCHORING> element relating the events and their times of occurrence, rather than in the <PERIOD> element that correspond to the word "day" rather than to the NP "every day". It may be considered a drawback of this approach that NPs as such are not treated as units, which may be more convenient for human annotators. It does however seem possible to modify the <ISO-TimeML_ICSrep> scheme such that the aspect of quantification expressed by quantifier words is moved from link elements to the elements annotating the linked arguments.

2.3 Annotation of Type 3 Modifier "Twice a Day"

The type 3 modifier "twice a day" is treated in <isoTimeML> as a one structural unit and annotated in a single element <TIMEX3>, as below:

(9) a. John calls_{e1} [twice a day]_{t1}. b. <isoTimeML> <EVENT xml:id="e1" target="#token2" pred="CALL"/> <TIMEX3 xml:id="t1" target="#token3 #token4 #token5" type="SET" value="DAY" quant="EVERY" freq="2X"/> <TLINK eventID="#e1" relatedToTime="#t1" relType="INCLUDED_IN"/> </isoTimeML>

This is interpreted as follows, again based on Interval Temporal Logic:

(10) $\forall J[[p_{day}(J) \land R_{during}(J, I_{t1})] \rightarrow \\ \exists_{2I_{e1}}(p_{call}(I_{e1}) \land R_{during}(I_{e1}, J)]]$

Here are two levels of restricted quantification: the range of the universal quantification $\forall J$ is restricted to the time interval I_{t1} , as expressed by $R_{during}(J, I_{t1})$, while that of the existential quantifier $\exists_{2I_{e1}}$ is restricted to the variable J for a set of days, again as expressed by $R_{during}(I_{e1}, J)$. There are at least two intervals of I_{e1} during which the event of calling holds.⁶

The <isoTimeML_ICSrep> approach, on the other hand, provides the following annotation:

```
(11) a. John calls<sub>tok2</sub> twice<sub>tok3</sub> a_{tok4} day<sub>tok5</sub>.
    b. <isoTimeML_ICSrep>
    <EVENT xml:id="e1"
    target="#token2" type="CALL"
    signature="SET"/>
    <PERIOD xml:id="t1"
    target="#token5" type="DAY"
    signature="SET"/>
    <TIME_ANCHORING
    anchoredEvent="#e1"
    anchorTime="#t1"
    tempRel="INCLUDED_IN"
    eventDistr="INDIVIDUAL"
    timeDistr="INDIVIDUAL"
    eventQuant="2"
    timeQuant="EVERY"/>
    </isoTimeML_ICSrep>
```

This yields the interpretation which says that "a set of call events is anchored time-wise in a set of days, such that the individual events are anchored at individual days, where every day includes a time anchor for two of these events."⁷

3 Quantification, Scopes, and Distributivity

In this section we first discuss event quantification and then a way to generalize quantification over other entities than time and events. We also discuss some issues concerning distributivity and some residual issues relating to set, scopes, and binding.

3.1 Event Quantification

ISO-TimeML (2012) annotates quantified temporal expressions, but has no provisions for anno-

⁶See for details ISO-TimeML (2012), p. 35.

⁷See the end of section 5, Bunt (2010b), example (50).

tating quantified events. Both Bunt and Pustejovsky (2010) and Pustejovsky et al. (2010a,b) extend the annotation of quantification to events, but in different ways. As was discussed in the previous section, Bunt and Pustejovsky (2010) annotate event quantification by introducing the attributes @signature="SET" and @eventQuant into the element <TIME_ANACHORING>.

Pustejovsky et al. (2010a,b), on the other hand, annotate event quantification by introducing the attributes @type="SET", @scopes and @quant with values such as EVERY into the element <EVENT>.

Here is an illustration:

(12) Event Quantification

```
a. Mary [read]<sub>e1</sub> during [every lecture]<sub>e2</sub>
b. <isoTimeML_m>
<EVENT xml:id="e1"
target="#token2" pred="READ"/>
<EVENT xml:id="e2"
target="#token4 #token5"
pred="LECTURE" type="SET"
quant="EVERY" scopes="#e1"/>
<TLINK eventID="#e1"
target="#token3"
relatedToEvent="#e2"/
relType="DURING"/>
</isoTimeML_mod><sup>8</sup>
```

Here, the element <EVENT xml:id="e2"> is specified with the attributes @type="SET" and @quant="EVERY", just as in the case of temporal quantification.

Each element in the annotation is then interpreted as below:

```
(13) a. <EVENT xml:id="e1"/>:

\exists e_1[read(e_1)]

b. <EVENT xml:id="e2" quant="EVERY"

pred="LECTURE" scopes="e1"/>:

\forall e_2[lecture(e_2)]

c. <TLINK>: \lambda y \lambda x [\tau(x) \subseteq y]
```

Note that the attribute @scopes is introduced to mark up the scopes of quantifiers explicitly. Note also that, to allow the interpretation (a) above, the event of reading (<EVENT xml:id="e1"/>) should be understood as having undergone existential quantification; in other words, the attribute @guant has the default value "SOME".

Given scope information, we can now combine each of the interpretations through the operation of conjunction and obtain the following overall interpretation:

(14) $\forall e_2 \exists e_1 [lecture(e_2) \rightarrow [read(e_1) \land \tau(e_1) \subseteq \tau(e_2)]]$

As is expected, this says that, during each lecture (e_2) , an event (e_1) of Mary's reading took place.

3.2 Generalizing Quantification

In natural language, almost any predication or relation can be quantified. Hence, the annotation of quantification over times and events should be viewed as a special case of quantification involving predicates about and relations between sets of any kinds of entity.

In an event-based semantics, quantification over events turns up in every sentence, not just for the relation between events and their time of occurrence, but also for the relation between events and their participants. Consider, for instance, the following example:

(15) Everybody will die.

This is an interesting example, cited in Bunt and Pustejovsky (2010), for the discussion of quantification, distributivity, and scopes, but it cannot be annotated just with temporal and event quantification only.

To extend quantification to non-temporal entities, one possibility is to introduce an element <EN-TITY>, and a linking tag <SRLINK> for annotating the relations between events and their participants. For illustration, the above example can be annotated as below:

(16) Annotation

a. Everybody_{x1} [will die]_{e1}.
b.<isoSEM xml:id="asr1">
<ENTITY xml:id="x1"
target="#token1" type="HUMAN"
signatue="SET" quant="EVERY"
scopes="#e1"/>

⁸This example is taken from Pustejovsky et al. (2010b), (28).

```
<EVENT xml:id="e1"
target="#token2 #token3"
type="DIE" tense="FUTURE"/>
<SRLINK event="#e1"
participant="#x1"
semRole="PATIENT"/>
</isoSEM>
```

The elements of this representation may be interpreted as follows:

(17) Interpretation a. $\sigma_{x1} := \lambda x_1[human(x_1)]$ b. $\sigma_{e1} := \lambda x[die(x, e_1)]$

Both of the elements are interpeted as denoting sets, a set of humans and a set of ones who die. Here are two notes. First, σ_{x1} is not bound by the universal quantifier, corresponding to the specification of @quant="EVERY". Second, the semantic role of the first argument of the predicate $die(x_1, e_1)$ can be spelled out to be $[die(e_1) \land Arg_1(patient, x_1, e_1).^9]$

3.3 Distributivity

Since the publication of Bunt (1985), the notion of distributivity has become an important issue as a property of quantification in formal semantics. Consider:

(18) The two men swallowed a beer and lifted the piano.

This sentence is interpreted as saying that each of the two men drank a beer and they together lifted the piano. To obtain such an interpretation, we need a formal mechanism of characterizing the so-called distributivity of events so that some are treated as *individual* events (e.g., "each drinking a beer") or *collective* events (e.g., "together lifting the piano").

To treat distributivity, one idea, originally proposed in Bunt (1985), is to bring in higher-order variables such as variables for sets. With these variables, we can have the following semantic form, where $\mathcal{P}_2(\text{MEN})$ denotes the set of all sets of two men.

(19)
$$\exists M[M \in \mathcal{P}_{2}(\text{MEN}) \land \forall x[x \in M \to [\\ \exists e_{1}[swallow_beer(e_{1}) \land agent(e_{1}, x)] \\ \exists e_{2}[lift_piano(e_{2}) \land agent(e_{2}, M)]]]]$$

Now the question is how to annotate sentences like the one given above, and how to derive such an interpretation. Again, we have several alternatives. One approach could be to encode distributivity for each relevant entity in the <ENTITY> element and use that information to trigger an appropriate link. Another way is to mark up that information on the attribute @eventDistr in the ICS representation.

The first approach runs into problems because the NP "The two men" are involved in *swallow*-events with individual distributivity and in a *lift*-event with collective distributivity. Trying to annotate this as different @distributivity values in the two <EVENT> elements makes no sense from a semantic point of view: it's not the elements of the sets of events that are involved individually or collectively, but only the participants in the agent role.

The second approach would give the following annotation (leaving out the parts that are not relevant for the present discussion), where the attribute @cardinality is used to represent the use of quantifier words for indicating the number of elements in argument set, as opposed to other uses that these words may have:

(20) a. [The two men]_{x1} [swallowed a beer]_{e1} and [lifted the piano] $_{e2}$. b.<isoTimeML_ICSRep xml:id="ad1"> <ENTITY xml:id="x1" target="#token1 #token2 #token3" type="MAN" signature="SET" cardinality="2" outscopes="#e1"/> <EVENT xml:id="e1" target="#token4" type="SWALLOW" signature="SET"/> <EVENT xml:id="e2" target="#token7" type="LIFT" signature="ELEMENT"/> <SRLINK xml:id="r1" event="#e1" participant="#x1" semRole="AGENT" participantDistr="INDIVIDUAL"/> <SRLINK xml:id="r2" event="#e2" participant="#x1" semRole="AGENT" participantDistr="COLLECTIVE"/> </isoTimeML_ICSRep>

⁹See Pustejovsky et al. (2007) for details on the annotation of event participants or argument role assignments <ArgLink>.

The DRT-based semantics for <isoTimeML_ ICSRep> given in Bunt (2011a,b), which does take the distributivity of quantifications into account, produces the semantic representation (19) for this annotation when extended with the treatment of <ENTITY> elements in the same way as the nonlinking elements for time and events, and with the treatment of the <SRLINK> linking element in the same way as the temporal linking elements.

4 Concluding Remarks

In this paper, we have reviewed two versions of <isoTimeML> in dealing with the annotation of temporal and event quantification.

We do not pretend to have presented a fully developed proposal for quantification over time and events that generalizes to quantification over other than temporal and 'eventual' entities, but we identified strengths and weaknesses of different proposals. We hope that this will contribute to the following tasks: (1) the revision and extension of ISO-TimeML (2012) and (2) the development of the new ISO project concerned with the annotation of spatial information ("ISO-Space"), where much the same issues relating to quantification arise as in ISO-TimeML (when the relation between a set of events and their place of occurrence is quantified, as in "Heavy thunderstorms are expected tomorrow all over the country"), and (3) the development of the ISO project concerning the basic principles of semantic annotation, ISO NP 24617-6, in which quantification has been identified as one of the burning issues to be dealt with, that cut across several attempts to define standards for semantic annotation.

Acknowledgments

We thank the five anonymous reviewers for their positive reviews and constructive comments and also Suk-Jin Chang, Jae-Woong Choe, Alex C. Fang for reading the pre-final version of this paper. The authors also wish to thank James Pustejovsky for many interesting useful discussions.

References

Bunt, H. 1985. *Mass Terms and Model-theoretic Semantics*. Cambridge University Press, Cambridge.

- Bunt, H. 2010. A methodology for designing semantic annotation languages exploiting syntactic-semantic iso-morphisms. In: A. Fang, N. Ide and J. Webster (eds.) Proceedings of ICGL 2010, the Second International Conference on Global Interoperability for Language Resources, Hong Kong City University, pp 29– 45
- Bunt, H. 2011a. Introducing abstract syntax + semantics in semantic annotation, and its consequences for the annotation of time and events. In E. Lee and A. Yoon (eds.), *Recent Trends in Language and Knowledge Processing*, 157-204. Hankukmunhwasa, Seoul.
- Bunt, H. 2011b. Abstract syntax and semantics in semantic annotation, applied to time and events. Revised version of Bunt (2011a). Unpublished.
- Bunt, H. 2012. CASCADES: A methodology for defining semantic annotations. Forthcoming in *Language Resources and Evaluation*.
- Bunt, H., and J. Pustejovsky. 2010. Annotating temporal and event quantification. In H. Bunt (ed.), *Proceedings* of ISA-5, the Fifth Joint ISO-ACL/SIGSEM Workshop on Interoperabile Semantic Annotation, 15-22.
- ISO/TC 37/SC 4/WG 2. 2012. ISO 24617-1:2012(E) Language resource management - Semantic annotation framework - Part 1: Time and events (SemAF-Time, ISO-TimeML). The International Organization for Standardization, Geneva.
- Kamp H and U. Reyle. 1993. *From Discourse to Logic*. Kluwer Academic Publishers, Dordrecht.
- Lee, K. 2012. Interoperable Spatial and Temporal Annotation Schemes. Proceedings of The Joint ISA-7, SRSL-3 and I2MRT Workshop on Interoperable Semantic Annotation, pp. 61-68. The Eighth International Conference on Language Resources and Evaluation (LREC 2012) Satellite Workshop, Istanbul.
- Montague, R. 1973. The proper treatment of quantification in ordinary English. In R. H. Thomason (ed.), *Formal Philosophy*, pp. 247-70. Yale University Press, New Haven, CT.
- Pratt-Hartmann, I. 2007. From TimeML to Interval Temporal Logic. In J. Geertzen, E. Thijsse, H. Bunt, and A. Schffrin (eds.), *Proceedings of The Seventh International Workshop on Computational Semantics*, pp. 166-180. Tilburg.
- Pustejovsky, J., J. Littman, and Roser Saur[i]. 2007. Arguments in TimeML: Events and Entities. In Frank Schilder, Graham Katz, and James Pustejovsky (eds.), *Annotating, Extracting and Reasoning about Time and Events*, pp. 107-127. Springer, Berlin.
- Pustejovsky, J., K. Lee, H. Bunt, and L. Romary. 2010a. ISO-TimeML: A Standard for Annotating Temporal Information in Language. in Proceedings of LREC 2010, the Seventh International Conference on Language Resources and Evaluation, 394-397. Malta.

- Pustejovsky, J., K. Lee, H. Bunt, and L. Romary. 2010b. Revised version of Pusejovsky et al. (2010a), a manuscript submitted to the journal *Language Resources and Evaluation*.
- Pustejovsky, J., J.L. Moszkowicz, and M. Verhange. 2012. Proceedings of the The Joint ISA-7, SRSL-3 and I2MRT Workshop on Interoperable Semantic Annotation, pp. 70-77. LREC 2012 Satellite Workshop, Istanbul.