

The Annotation of Measure Expressions in ISO Standards

Kiyong Lee
Korea University, Seoul
ikiyong@gmail.com

Abstract

In annotating measure expressions such as *three days* and *about 123 km*, two recently published ISO standards, ISO-TimeML (ISO, 2012b) and ISOspace (ISO, 2014a), show some inconsistencies, as pointed out in ISO SemAF Principles (ISO, 2014c), a third ISO standards on semantic annotation to be published soon. Other than terminological or semantic inconsistencies introduced in ISO SemAF Principles, there are some formal inconsistencies between or within these standards. This paper attempts to resolve such inconsistencies by proposing some minimally possible modifications into the annotation schemes of those two standards. Despite these modifications, the interoperability between these standards is preserved, each retaining its own annotation scheme for either temporal or spatial information involving measures. An attempt is also made to partially merge ISO-TimeML and ISOspace as a step towards the integration of ISO SemAF standards into a modularly usable general annotation scheme for the semantic annotation of language.

1 Introduction

Measure expressions such as *three days* and *about 123km* are ubiquitous in language. Here is a short travel log which contains these measure expressions.¹

(1) Travel Log

*We flew to Toronto by Air Canada and drove to Niagara Falls **three days** before Christmas Day. Niagara Falls is **approximately 130 km (80 miles)** southwest of Toronto, **an average drive of one and a half hours** without traffic delays. According to Google maps, Niagara is **about 123km (76 miles)** from Pearson Airport and it takes **nearly 1 hour and a half** using highways having **speed limits of 100km/h**. We had estimated it would take **2 hours maximum** and hoped to get to Niagara before 6:00 pm, but arrived at the hotel in Niagara after 10:30 pm. We had to drive for **more than 6 hours** because of an unexpected heavy snow storm. We drove at **an average speed of around 20 kilometers per hour**. We moved **so slow**, consuming **so many hours** on the road, that Niagara seemed **very far**. We stopped for coffee after barely driving **50km** (a little over 30 miles) from Pearson Airport.*

The words or strings of words in boldface refer to quantities or amounts, called *measures*.² Some of them refer to time amounts and others to spatial measures of various dimensions such as distances. Two of them refer to speed limits that involve a spatio-temporal dimension. Some expressions (e.g., *2 hours*) are then quantitatively explicit and others (e.g., *so many hours*) are not.

For the purpose of language resource management, an ISO Working Group³ on semantic annotation recently published two ISO international standards, ISO-TimeML (ISO, 2012b)⁴ and ISOspace (ISO, 2014a)⁵. Parts of these standards treat measure expressions, spatial and temporal, while specifying how

¹Copied from <<http://www.tripadvisor.com/Travel-g155019-c97995/Toronto:Ontario:Niagara.Falls.A.Side.Trip.From.Toronto.html>> and slightly modified to suit our needs.

²The temporal expressions *6:00 pm* and *10:30 pm* are not in boldface because they do not refer to durations or amounts of time, but points in time.

³ISO/TC 37/SC 4 (Language Resource Management)/WG 2 (Semantic Annotation)

⁴Based on TimeML developed by Branimir et al. (2005), Pustejovsky et al. (2005) with some modifications which were discussed in Pustejovsky et al. (2010).

⁵Based on MITRE (2009)'s SpatialML and the Spatial Annotation Scheme developed by the Brandeis Working Group headed by James Pustejovsky. See Pustejovsky et al. (2011) and Pustejovsky et al. (2012).

to relate events (motions), paths, and some other basic entities to these measure expressions. Meanwhile, ISO SemAF Principles (ISO, 2014c)⁶, a third ISO standard on semantic annotation soon to be published, has pointed out some inconsistencies between the treatments of measure expressions by these two published standards and their inadequacies as semantic annotation.

Lee (2012) had earlier argued for the merging of ISO-TimeML and ISOspace into a unified annotation scheme, especially based on functional similarities of spatial and temporal signals (e.g., various prepositions in English) that trigger the anchoring of events, motions, durations, and paths to times and locations. In this paper, we may still opt for a partial merging of these two standards by removing any inconsistencies, especially formal inconsistencies to be described in Section 3 but puts its focus on the interoperability rather than the over-all integration of the two annotation schemes, especially concerning spatial, temporal, and spatio-temporal measure expressions. We claim that only a few minor modifications need to be made to resolve any formal, but not necessarily terminological or semantic inconsistencies between the standards in annotating measure expressions, either spatial or temporal, while keeping their original overall annotation schemes almost intact.

The rest of the paper develops as follows: Section 2 Review of ISO-TimeML and ISOspace, Section 3 Formal Inconsistencies, Section 4 Partial Merging, Section 5 Informal Semantics, and Section 6 Concluding Remarks.

2 Review of ISO-TimeML and ISOspace

In this section we briefly introduce ISO-TimeML (ISO, 2012b) and ISOspace (ISO, 2014a) that specify how to annotate measure expressions, temporal and spatial, respectively. For illustrations, we focus on the two basic entity types of measure: duration and distance.

2.1 Overview: Duration and Distance

Duration and distance are two types of a basic entity, named *measure*, that share structurally common features. As measure expressions, they are both structured as a pair $\langle n, u \rangle$, consisting of a numeric standing for quantity and a unit, possibly with a modifier that is optional: e.g., (1) *three_n days_u* and (2) *nearly_{mod} 130_n km_u*.

Furthermore, they are also interpreted at times as involving a temporal or a spatial interval, delimited by two end points, as shown below:

- (2) a. *We drove [_{endPoint1} three days _{endPoint2}] before [Christmas Day]_{t2=endPoint2}.*
 b. *We barely drove [_{endPoint1} 50km _{endPoint2}] from [Pearson Airport]_{p1=endPoint1}.*

Here (a) is interpreted as an event of driving that occurred on December 22 (_{endPoint1}). Similarly, (b) is interpreted as a motion of driving with an event path which covered the distance of 50 kilometers (_{endPoint2}).⁷ In the ensuing two subsections 2.2 and 2.3, we illustrate how durations and distances are annotated by ISO-TimeML and ISOspace, respectively, and represented in XML.

2.2 Durations in ISO-TimeML

There are two sorts of temporal expressions in our dataset (1) that are both treated by ISO-TimeML as durations:⁸

- (3) a. *We had to drive for **more than 6 hours**_{t1}.*
`<TIMEX3 xml:id="t1" type="DURATION" value="P6H" mod="moreThan"/>`

⁶See also Bunt (2015) in this volume for the main ideas of ISO SemAF Principles.

⁷This event is a directed and terminated dynamic motion each movement of which can be described structurally as a pair $\langle l_i, t_j \rangle$ consisting of a location l_i and an associated time t_j that increases incrementally. It thus forms an event path which is again analyzed as a sequence of movements with at least two end points, initial and terminal: $\langle l_0, t_i \rangle, \langle l_1, t_{i+1} \rangle, \dots, \langle l_m, t_n \rangle$. Here, the distance of an event path is measured as a length between each pair of a location and a time in the sequence. See (Mani and Pustejovsky, 2012), pp. 90-107, for further details on directed motion and dynamic interval temporal logic (DITL).

⁸For the sake of illustrations, dataset fragments are inline annotated with their IDs in this paper, while the specification of the attribute @target or @markable is omitted from the annotation of basic entities in ISO-TimeML or ISOspace, respectively.

- b. *We drove to Niagara Falls [t₂₁ **three days** t₂₂] t₂ before Christmas Day t₃.*
`<TIMEX3 xml:id="t2" type="DURATION" value="P3D" beginPoint="#t21" endPoint="#t22"/>`
`<TIMEX3 xml:id="t3" type="DATE" value="XXXX-12-25"/>`
`<TIMEX3 xml:id="t21" target="" type="DATE" value="XXXX-12-22" temporalFunction="TRUE" anchorTimeID="#t3"/>`⁹

Both of the temporal expressions *[more than 6 hours]*_{t₁ and *[three days]*_{t₂ are measure expressions, each specified with a numeric (quantity) and a unit. They carry different sorts of information, however. The first expression t₁ refers to an amount of time, the amount of time consumed by the motion of driving. On the other hand, the latter expression t₂ refers to a time interval that identifies Christmas (t₃=XXXX-12-25) with its end point t₂₂ and another date (XXXX-12-22) with its beginning point t₂₁.}}

Then there are two different link relations in ISO-TimeML: (1) <MLINK> between the event_{e₁} of driving and the amount of time t₁ and (2) <TLINK> between that same event_{e₁} and the initial point of time t₂₁ of a time interval t₂, as shown below:

- (4) a. *We had to drive_{e₁} for [more than 6 hours] t₁.*
`<MLINK eventID="#e1" relatedToTime="#t1" relType="MEASURES"/>`
 b. *We drove_{e₁} to Niagara Falls [t₂₁ **three days** t₂₂] t₂ before Christmas Day.*
`<TLINK eventID="#e1" relatedToTime="#t21" relType="DURING"/>`

Here the event (motion) of driving_{e₁} in Example (a) is linked to the amount of time_{t₁} that it consumed, while that same event_{e₁} in Example (b) is anchored to the date_{t₂₁} (December 22) during which it occurred.

2.3 Distances in ISOspace

There are three quantitatively explicit spatial measure expressions in our dataset Travel Log (1):

- (5) a. *Niagara Falls is approximately 130km (80 miles) southwest of Toronto.*
 b. *Niagara is about 123km (76 miles) from Pearson Airport.*
 c. *We stopped for coffee after barely driving 50km (a little over 30 miles) from Pearson Airport.*

According to ISOspace, these measure expressions are annotated as below:

- (6) a. `<measure xml:id="mes1" markable="approximately 130km" value="130" unit="km" mod="approx"/>`
 b. `<measure xml:id="mes2" markable="about 123km" value="123" unit="km" mod="approx"/>`
 c. `<measure xml:id="mes3" markable="barely ...50km" value="50" unit="km" mod="equalOrLess"/>`

While the annotation of these basic entities (measures) is routine, their linking relations slightly differ from one another:

- (7) a. *[Niagara Falls]_{p11} is [approximately 130km]_{mes1} southwest_{ss1} of Toronto_{p12}*
`<oLink xml:id="o11" figure="#p11" ground="#p12" trigger="#ss1" relType="southwest" frameType="absolute" referencePt="southwest" projective="true"/>`
`<mLink xml:id="m11" relType="distance" figure="#p11" ground="#p12" trigger="#mes1" val="#mes1"/>`
 b. *Niagara_{p11} is [about 123km]_{mes2} from_{ss2} [Pearson Airport]_{p13}.*
`<mLink xml:id="m12" relType="distance" figure="#p11" ground="#p13" trigger="#mes2" val="#mes2"/>`

⁹<TIMEX3 xml:id="t21"/> may be treated as an element, called *non-consuming tag*, which has no associated markable expression in text, thus the value of its attribute @target being empty "". See ISOspace (ISO, 2014a), A.3.4 Special Section: Non-consuming tags.

- c. *We stopped for coffee after barely_{mes3} driving_{m1} [50km]_{mes3}.*
`<mLink xml:id="m13" relType="generalDimension" figure="#m1"
ground="#m1" trigger="#mes3" val="#mes3"/>`

Examples (a) and (b) both represent a distance type relation, while Example (a) carries additional information about the orientation expressed by the spatial signal *southwest*. On the other hand, example (c) is annotated as referring to a `generalDimension` type relation in ISOspace,¹⁰ but may also be annotated as referring to a relation of the distance type grounded to the event-path created by the motion *drive*, as will be discussed in the following Section 3.

3 Formal Inconsistencies

The specification of semantic annotation schemes can be inconsistent in three different ways. The first two are introduced as *terminological* and *semantic* inconsistencies in ISO SemAF Principles¹¹ to be briefly discussed in the following Subsection 3.1. The third kind of inconsistency that we name *formal inconsistency* is discussed in Subsection 3.2.

3.1 Terminological or Semantic Inconsistency

Terminological inconsistency arises if two different terms are used for one and the same concept. For example, ISOspace has an element named *measure* for the concept referring to a quantity, whereas ISO SemAF Principles (ISO, 2014c) proposes the name *amount* for the same concept. Hence, the use of these two terms (names) is terminologically inconsistent.

Semantic inconsistency is caused by the use of a term for two different concepts. In ISO-TimeML, the term *event* refers to an eventuality, whereas it refers to a non-motion event in ISOspace. Hence, the use of the term *event* in ISOspace is semantically inconsistent with its use in ISO-TimeML. The term *duration* in ISO-TimeML refers to an amount of time and also to an interval of time. The use of this term is again semantically inconsistent.

The use of the tag (name of an element) `<event>` can, however, be intrinsically consistent within ISOspace, for it explicitly specifies the tag `<event>` as standing for a non-motion event (e.g., love), while using the tag `<motion>` to annotate motion verbs such as *drive* or *run*. The use of the tag `<event>` in ISOspace becomes inconsistent only if ISOspace is integrated with ISO-TimeML to form a single annotation scheme, for the tag `<event>` in ISO-TimeML stands for eventuality.

3.2 Formal Inconsistency between or within Standards

3.2.1 Intrinsic vs Extrinsic Inconsistency

The term *formal inconsistency* is here used to refer to structural differences between or within standards in their specification of annotation schemes. Each annotation scheme has two levels of specification: one is the level of specification, called *abstract syntax*, and the other the level, called *concrete syntax*, (e.g., an XML serialization of an abstract syntax for temporal annotation). The concrete syntax of an annotation scheme specifies how to represent the annotations specified by the abstract syntax. Formal inconsistency may occur between an abstract syntax and its associated concrete syntax when the concrete syntax fails to properly represent the annotations based on the abstract syntax.¹² Such a case of formal inconsistency, as is described now, we call *intrinsic* inconsistency.

In contrast, there is another case of formal inconsistency which we call *extrinsic* inconsistency. Given at least a pair of markables which are of two different sorts, but which have isomorphic (similar) structures (e.g., *We drove for nearly 2 hours.* vs. *We drove nearly 50 miles.*), two annotation schemes are understood to be formally inconsistent, if and only if, they specify different sets of basic entities or link relations over them or associate different lists of attributes and possible values for some of the entities or links.

If an annotation scheme is intrinsically inconsistent, then it is a serious problem for the annotation scheme itself. A concrete syntax becomes useless. Extrinsic inconsistency causes no problem for the interoperability of two annotation schemes, unless they are merged into a single annotation scheme. In the rest of Subsection 3.2, we focus on possible cases of formal inconsistency between the two standards, ISO-TimeML and ISOspace, in the annotation of measure expressions (distances vs durations) that are considered isomorphic in Subsubsection

¹⁰See ISOspace (ISO, 2014a), A.6.5.2, Example (c)

¹¹See ISO DIS 24617-6 SemAF Principles ISO (2014c), Clause 9.3 Quantities and measures, pp. 18-19.

¹²See Ide and Romary (2004), ISO 24612 LAF (ISO, 2012a), Bunt (2010), Bunt (2011), and Bunt and Pustejovsky (2010) for the distinction between annotation and its representation, and also between an abstract syntax and a concrete syntax.

3.2.2 and their annotation of links that relate and other basic entities to measure in Subsubsection 3.2.3. We finally discuss the formal inconsistency of specifying optional attributes in ISOspace and ISO-TimeML in Subsubsection 3.2.4.

3.2.2 Annotation of Measure Expressions

As mentioned in Subclause 2.1, measure expressions are treated in abstract terms as consisting of a pair $\langle n, u \rangle$, where n is a numeric referring to some quantity and u a unit. The measure expressions *nearly 2 hours* and *about 123 km* are similar in structure. In representing their annotations, ISO-TimeML and ISOspace are different from each other or extrinsically inconsistent (to use our term), as shown in Example (8):

- (8) a. *nearly 2 hours*
 ISO-TimeML: `<TIMEX3 xml:id="t1" type="DURATION" value="P2H" mod="APPROX"/>`
 b. *about 123 km*
 ISOspace: `<measure xml:id="mes1" value="123" unit="km" mod="approx"/>`

First, the tags of the two elements are different: `<TIMEX3>` vs `<measure>`. Second, ISO-TimeML specifies the type "DURATION" of its element `<TIMEX3>`, while ISOspace specifies no type for its element `<measure>`. Third, the value for the measure is represented as one chunk "P2H" in ISO-TimeML, while ISOspace represents the value of a measure separately from its unit by introducing two attributes `@value` and `@unit`.

In specifying ways of assigning a value (e.g., P6H) to the attribute `@value` for temporal expressions, ISO-TimeML follows ISO 8601 (ISO, 2004). The value P6H stands for "a period (P) of 6 hours (H)" and the period (P) is understood to be a duration of time or an amount of time, thus allowing a proper interpretation of the value P6H as a duration. As is argued in ISO SemAF Principles, the specification of annotating amounts of time (e.g., *nearly two hours*) in ISO-TimeML is, however, intrinsically inconsistent, for the attribute-value specification `value="P6H"`, for one thing, fails to conform to the abstract specification of annotating measure expressions.

Unlike ISO-TimeML, ISOspace is found intrinsically consistent. Consider the following list of attributes and possible values for the element, tagged `<measure>`, in XML:

- (9) Attributes of `<measure>`¹³
attributes = identifier, markable, value, [unit], [mod], [comment];
value = "real" | CDATA;
unit = CDATA;
mod = CDATA;

Bracketed attributes are optional ones, while non-bracketed ones are required attributes. There are two alternative values for the attribute `@value`: either a real number with its unit specified or any CDATA such as *far* with no unit specified. To allow non-explicit measure expressions as markables, ISOspace treats the attribute `@unit` as an optional (implied) attribute.

Here are two illustrations, one for an explicit measure expression and another for a non-explicit measure expression:

- (10) a. `<measure xml:id="mes1" markable="about 123 kilometers" value="123" unit="km" mod="approx"/>`
 b. `<measure xml:id="mes2" markable="very far" value="far" mod="very"/>`¹⁴

Hence, the concrete representation in XML of annotations of measure expressions in ISOspace is shown to be intrinsically consistent with some of its abstract specifications or the abstract syntax in general.

ISOspace may be extended to annotate temporal durations simply by adding temporal units to the list of possible values of the attribute `@unit`. Here is an illustration:

- (11) a. *more than 6 hours*
 b. `<measure xml:id="mes1" markable="more than 6 hours" value="6" unit="hour" mod="moreThan"/>`

¹³The format of listing these attributes follows the representation language of ISO/IEC (1996)'s Extended BNF.

¹⁴The adjectival or adverbial intensifier *very* is here treated as a value of `@mod`. It is not explicitly listed among the possible values of `@mod`, but is allowed by CDATA.

ISO DIS 24617-6 SemAF Principles argues against the representation of quantity modifiers as attribute-value pairs (e.g., `mod="moreThan"`) of the element `<measure>`. Instead, it proposes that a quantity modifier should be treated as a relation between two amounts or measurements, as shown below:¹⁵

(12) *We had to drive for more than 6 hours*

```
<amount xml:id="a1" target="#range(token6,token9)>
<amount xml:id="a2" target="#token8,#token9" num="6" unit="hour"/>
<relation xml:id="r1" arg1="#a1" arg2="#a2" relType="greaterThan"/>
```

This representation of a quantity modification should be formally and intrinsically *consistent* with the abstract syntax that specifies the notion of a quantity modification.

3.2.3 Measure Links

ISOspace introduces the tag `<mLink>` to annotate and represent the linking of events (motions) or any other relevant entities to a measure such as a distance or other spatial dimensions. Here is an example:

(13) a. *We drove_{m1} [about 122 kilometers]_{mes1}.*

```
b. <motion xml:id="m1" motionType="manner"/>
<measure xml:id="mes1" value="122" unit="km" mod="approx"/>
<mLink xml:id="ml1" relType="distance" figure="#m1" ground="#m1"
trigger="#mes1" val="#mes1"/>
```

This representation is based on the following specification of ISOspace:

(14) Current List of Attributes for the Element `<mLink>`¹⁶

```
attributes = identifier, markable, [trigger], [figure], [ground],
relType, val, [endPoint1], [endPoint2], [comment];
```

This specification fails to be consistent with the abstract structure $\langle e_1, e_2, R \rangle$ of a link R that relates a basic entity e_1 to another basic entity e_2 , for there is no pair of required attributes in the current list (14) of attributes for the element `<mLink>` which refer to two related entities.¹⁷ Links are basically binary relations between two entities. All of the links, `<TLINK>`, `<ALINK>`, `<SLINK>` and `<MLINK>`, in ISO-TimeML (ISO, 2012b) are binary relations between two entities, each having a pair of required attributes that specify a pair of entities that are to be related. `<TLINK>`, for instance, relates an event to a time or another event, thus having two required attributes like `@eventID` and `relatedToTimeID` or `relatedToEventID` also with a third required attribute `@relType` specifying the type of their relation.

This problem can, however, be easily fixed by treating the attributes `@figure` and `ground` as well as the attribute `relType` in the list (14) as required attributes and then making the two attributes `@figure` and `ground` stand for the two basic entities e_1 and e_2 that are to be linked by the relation R specified by the required attribute `relType`. The attribute `@val` is no longer necessary, for it is replaced by the newly required attribute `ground` which is now understood as referring to the value of an element `<measure>`.

(15) Modified List of Attributes for the Element `<mLink>`

```
attributes = identifier, markable, figure, ground, relType,
[trigger], [endPoint1], [endPoint2], [comment];
```

With this modified specification (15), the measure link, tagged as `<mLink>`, in ISOspace is now understood as a binary relation from a motion, a location or some other spatial entity (`figure`) to a measure (`ground`), as shown below:

(16) a. *We drove_{m1} [about 122 kilometers]_{mes1}.*

```
b. Old: <mLink xml:id="ml1" relType="distance" figure="#m1"
ground="#m1" trigger="#mes1" val="#mes1"/>
```

¹⁵See SemAF Principles (ISO, 2014c), Clause 9.3 Quantifiers and measures, pp. 18-19, example (16).

¹⁶List A.13 in ISOspace (ISO, 2014a).

¹⁷The specification of `<oLink>` and `<moveLink>` also run into the same problem and should be the topic for discussion on another occasion.

- c. New: `<mLink xml:id="m1" relType="distance" figure="#m1" ground="#mes1"/>`

Now the two measure links in ISOSpace and ISO-TimeML structurally resemble each other, as shown below:

- (17) a. *We had to drive*_{e1/m1} [*about 122 kilometers*]_{mes1} *for* [*more than 6 hours*]_{t1}.
- b. ISOSpace
`<mLink relType="distance" figure="#m1" ground="#mes1"/>`
- c. ISO-TimeML
`<MLINK relType="MEASURES" eventID="#e1" relatedToTime="#t1">`

These two are formally consistent, although they are terminologically inconsistent.

Compare now this modified treatment of the measure link (`<mLink>`) in ISOSpace with the proposal of ISO SemAF Principles (ISO, 2014c) that the measure link both in ISO-TimeML and ISOSpace be replaced by `<srLink>` for semantic roles, introduced by SemAF-SR (ISO, 2014b). Here is an example:

- (18) a. *I would walk*_{m1/ev1} [*500 miles*]_{mes1/am1}.
- b. ISOSpace:
`<measure xml:id="mes1" value="500" unit="mile"/>`
`<mLink xml:id="m1" figure="#m1" ground="#mes1" relType="distance"/>`
- c. SemAF Principles/SemAF-SR:
`<timeAmount xml:id="am1" aNum="500" unit="mile"/>`
`<srLink xml:id="sr1" arg1="#ev1" arg2="#am1" semRole="distance"/>`

These two treatments are formally consistent, for they both conform to the abstract structure $\langle e1, e2, R \rangle$ of the binary link relation.

3.2.4 Specification of Optional Attributes

In ISO-TimeML, the annotation of information related to an interval with its `@beginPoint` and `@endPoint` is associated with the basic entity element `<TIMEX3 type="DURATION"/>`. In ISOSpace, on the other hand, the annotation of information related to a path with its `@endPoint1` and `@endPoint2` is associated with the link `<mLink>`. Here are examples:

- (19) a. ISOSpace:¹⁸
*The width of the office*_{pl3} *is* [*25 feet*]_{mes5} *from the bookcase*_{se3} *to the* [*white board*]_{se4}.
`<measure xml:id="mes5" value="25" unit="ft"/>`
`<mLink xml:id="m15" relType="distance" figure="#pl3" ground="#mes5" endPoint1="#se3" endPoint2="#se4"/>`
- b. ISO-TimeML:¹⁹
*We left*_{e6} [*two weeks*]_{t62} *from* [*June 7, 2003*]_{t7}
`<EVENT xml:id="e6" pred="LEAVE" tense="PAST"/>`
`<TIMEX3 xml:id="t6" type="DURATION" value="P2W" beginPoint="#t61" endPoint="#t62"/>`
`<TIMEX3 xml:id="t7" type="DATE" value="2003-06-07"/>`
`<TIMEX3 xml:id="t62" type="DATE" value="2003-06-21" temporalFunction="true" anchorTimeID="#t7"/>`
`<TLINK eventID="e1" relatedToTime="#t62" relType="DURING"/>`

ISO SemAF Principles freely allows the specification of optional attributes associated with basic entities or links. Hence, the variation shown above may not be considered as causing formal inconsistency. Nevertheless, they create a problem for the integration of ISO-TimeML and ISOSpace for the annotation of measure expressions and their links, as will be discussed in Section 4.

¹⁸Taken from ISOSpace (ISO, 2014a), Annex A.6.5.2 Example (d), p.48, with some modifications.

¹⁹Taken from ISO-TimeML (ISO, 2012b), Clause 7.3.4.2 `<TIMEX3>`, page 17, Example (13) with the addition of `<TLINK>`.

4 Partial Merging

Two annotation schemes are interoperable only if each of them is formally and intrinsically consistent. They can also refer to each other, as shown below:

- (20) a. *We had to drive_{m1} [about 123km]_{mes1} for [more than 6 hours]_{t1}.*
- b.

```
<semAF xml:id="sem01">
  <isoSpace xml:id="sAnn01">
    <motion xml:id="m1" type="drive"/>
    <measure xml:id="mes1" value="123" unit="km" mod="approx"/>
    <mLink xml:id="ml1" figure="#m1" ground="#mes1" relType="distance"/>
  </isoSpace>
  <isoTimeML xml:id="tAnn01">
    <TIMEX3 xml:id="t1" type="DURATION" value="P6H" mod="moreThan"/>
    <MLINK eventID="#m1" relatedToTime="#t1" relType="MEASURES"/>
  </isoTimeML>
</semAF>
```

Here, `<isoSpace>` shows how the spatial measure (distance) expression `mes1` is annotated, while `<isoTimeML>` shows the annotation of the temporal measure (duration) expression `t1`. Furthermore, `<isoTimeML>` allows its `<MLINK>` to refer to the element `<motion xml:id="m1"/>` in `<isoSpace>` for the value `#m1` of the attribute `@eventID`. Otherwise, the motion of *drive_{m1}* may not be understood as referring to one and the same event of driving.

Despite their intrinsic formal consistency, these two annotation schemes are extrinsically inconsistent. This inconsistency can easily be resolved by introducing a few modifications into `ISOspace` and then by merging the treatment of temporal measure expressions into it. To merge the annotation of temporal measure expressions such as *more than 6 hours* into `ISOspace`, as in Illustration (21), it is only necessary to extend the list of possible values for the attribute `@unit` for the element `measure` of `ISOspace` to include temporal units such as *hours*. This is done automatically because that list is an open list, consisting of any `CDATA`.

- (21)

```
<isoSpace xml:id="sAnn01">
  <motion xml:id="m1" markable="#token4" type="drive"/>
  <measure xml:id="mes1" value="123" unit="km" mod="approx"/>
  <measure xml:id="mes2" value="6" unit="hours" mod="moreThan"/>
  <mLink xml:id="ml1" figure="#m1" ground="#mes1" relType="distance"/>
  <mLink xml:id="ml1" figure="#m1" ground="#mes2" relType="duration"/>
</isoSpace>
```

Here, the list of values for the attribute `@relType` of the element `<mLink>` is also extended to `"duration"`.²⁰

Such merging, however, requires further modifications. Consider the following pair of examples:

- (22) a. *We left_{e1} [t₁₁ two weeks t₁₂]_{t1} from_{s1} [June 7, 2003]_{t2}.*
- b. *We drove_{m1} [about 123 km]_{mes1} from_{ms1} [Pearson Airport]_{p11}.*

These two sentences are syntactically the same except that (a) contains two temporal expressions, a duration (`t1`) and a date (`t2`), while (b) contains two spatial expressions, a distance measure (`mes1`) and a location (`p11`). Their annotations are thus expected to be structurally the same, but the current versions of the two annotation schemes, `ISO-TimeML` and `ISOspace`, however, present two different annotation structures.

- (23) a. `ISO-TimeML`
- ```
<EVENT xml:id="e1" pred="LEAVE" tense="past"/>
<TIMEX3 xml:id="t1" type="DURATION" value="P2W"
beingPoint="#t11" endPoint="#t12"/>
<SIGNAL xml:id="s1" pred="FROM"/>
<TIMEX3 xml:id="t2" type="date" value="2003-06-07"/>
<TIMEX3 xml:id="t12" target="" type="date" value="2003-06-21"
temporalFunction="true" anchorTimeID="#t2"/>
<TLINK eventID="#e1" relatedToTime="#t12" relType="DURING"/>
```

<sup>20</sup>Later, this value will be changed to `"runtime"`.



b. ISOSpace

```
<motion xml:id="m1" type="drive" tense="past"/>
<measure xml:id="mes1" value="123" unit="km" mod="approx"/>
<signal xml:id="ms1" markable="from"/>
<mLink xml:id="ml1" relType="distance" figure="#m1"
ground="#m1" val="#mes1" endPoint1="#p11" />
```

There are at least three possible ways to integrate ISO-TimeML and ISOSpace for the annotation of measure expressions. One way is to modify the part of ISO-TimeML which annotates durations and merge it into ISOSpace, as was shown in Example (21), another way is to take the opposite approach, and a third way to follow ISO SemAF Principles<sup>21</sup> and merge the two different annotation schemes of measure expressions, both spatial and temporal, into a new annotation scheme or ISO SemAF-SR (semantic roles) (ISO, 2014b). For now, we take the first approach and show how ISO-TimeML's annotation (23a) can be partially merged into ISOSpace by extending the current version of ISOSpace to accommodate parts of ISO-TimeML. Here is an illustration:

(24) a. *We drove<sub>m1</sub> to Niagara Falls [t1 three days t2]<sub>mes1</sub> before<sub>s1</sub> [Christmas Day]<sub>t3</sub>.*

```
b. <semAF xml:id="sem02">
 <isoSpace xml:id="sAnn02">
 <motion xml:id="m1" type="drive" tense="past"/>
 <measure xml:id="mes1" value="3" unit="day"/>
 <mLink xml:id="ml1" figure="#t1" ground="#mes1"
relType="beginPointOf"/>
 </isoSpace>
 <isoTimeML xml:id="tAnn02">
 <TIMEX3 xml:id="t1" target="" type="date" value="XXXX-12-22"
temporalFunction="true" anchorTimeID="#t3" mLinkID="#ml1"/>
 <SIGNAL xml:id="s1" pred="BEFORE"/>
 <TIMEX3 xml:id="t3" type="DATE" value="XXXX-12-25"/>
 <TLINK xml:id="t11" timeID="#mes1" relatedToTime="#t3"
relType="BEFORE"/>
 <TLINK xml:id="t12" eventID="#m1" relatedToTime="#t1"
relType="DURING"/>
 </isoTimeML>
</semAF>
```

Here, the link `<mLink xml:id="ml1">` is interpreted as stating that the date `t1` is the initial point of the time interval with its length being "three days" `mes1`. The calculation of the date (`date(t1) = XXXX-12-22`) is then triggered by `mLinkID="#ml1"` with its interval value `mes1 = [3, day]` and also anchored to the date `date(t3) = XXXX-12-25` of `anchorTimeID="#t3"` in `<isoTimeML>`.

## 5 Informal Semantics

For the semantic justification of the proposed annotations of measure expressions, we show in this section how some of them are interpreted. Consider the following dataset segments, taken from Travel Log (1):<sup>22</sup>

(25) Semi-annotated Dataset 2

*We<sub>se1</sub> ... drove<sub>m1</sub> from [Pearson Airport]<sub>p10</sub> to [Niagara Falls]<sub>p11</sub> [t11 three days t12]<sub>t1</sub> before [Christmas Day]<sub>t2</sub>. We<sub>se1</sub> drove<sub>m1</sub> to<sub>ms1</sub> Niagara<sub>p11</sub> [about 122 kilometers]<sub>mes1</sub> for<sub>s</sub> [more than 6 hours]<sub>mes2</sub> at [an average speed of 20 kilometers per hour]<sub>mes3</sub>.*

### 5.1 Interpreting Event Paths

Here we may or may not introduce a non-consuming tag  $\emptyset_{p1}$  for an event path from `[Pearson Airport]p10` to `Niagarap11`. The following are two versions of an expected logical form for Dataset 2, one for a case with no event path annotated and another for a case with an event path annotated:

<sup>21</sup>See SemAF Principles, Clause 8.2 Spatial and temporal relations as semantic roles, and other places.

<sup>22</sup>For simplicity's sake, the same IDs are assigned to coreferential expressions in this dataset.

(26) a. No event path annotated:

$$[drive(m_1) \wedge agent(se_1, m_1) \wedge goal(pl_1, m_1) \wedge distance(m_1) \approx [122, km] \\ \wedge runtime(m_1) \geq [6, hour] \wedge speed(m_1) =_{average} [20, km/h]]$$

b. An event path annotated:

$$[drive(m_1) \wedge agent(se_1, m_1) \wedge goal(pl_1, m_1) \wedge path(p_1, m_1, pl_0, pl_1) \wedge distance(p_1) \approx [122, km] \\ \wedge runtime(m_1) \geq [6, hour] \wedge speed(m_1) =_{average} [20, km/h]]$$

In (a),  $distance(m_1)$  is interpreted as the quantity (length) of a distance traversed by the motion of driving ( $m_1$ ). In (b), on the other hand,  $distance(p_1)$  is interpreted as the quantity (length) of a path traversed by the motion of driving ( $m_1$ ) from Pearson Airport ( $pl_0$ ) to Niagara ( $pl_1$ ). The second interpretation is more detailed than the first one, but they are practically equivalent and equally acceptable interpretations.

Here we simply focus on the parts of the interpretations that are related to measures. Consider:

$$(27) [drive(m_1) \wedge distance(m_1) \approx [122, km]]$$

This is based on the following annotation:

```
(28) <motion xml:id="m1" type="drive" motionType="path"/>
<measure xml:id="mes1" value="122" unit="km" mod="approx"/>
<mLink xml:id="ml1" figure="#m1" ground="#mes1" relType="distance"/>
```

Then we have:

(29) Interpretation 1:

- (i)  $\sigma(m_1) := drive(m_1)$
- (ii)  $\sigma(mes_1) := q_{measured}(mes_1) \approx [122, km]$
- (iii)  $\sigma(ml_1) := q_{distance}(m_1) = q_{measured}(mes_1)$
- (iv)  $\sigma(s_1) := [drive(m_1) \wedge q_{distance}(m_1) \approx [122, km]]$

Here  $q_{measured}(mes_1)$  is a quantity measured at a particular situation  $mes_1$ , while  $q_{distance}(m_1)$  is interpreted as the quantity (length) of a distance traversed by a motion  $m_1$ .

We can also have:

(30) Interpretation 2:

- (i)  $\sigma(m_1) := drive(m_1)$
- (ii)  $\sigma(p_1) := path(p_1, m_1, pl_0, pl_1)$
- (iii)  $\sigma(mes_1) := q_{measured}(mes_1) \approx [122, km]$
- (iv)  $\sigma(ml_1) := q_{distance}(p_1) = q_{measured}(mes_1)$
- (v)  $\sigma(s_1) := [drive(m_1) \wedge q_{distance}(p_1) \approx [122, km]]$

This is the interpretation when a path is introduced as a non-consuming tag into the annotation.

## 5.2 Interpreting Amount of Time

Here is another illustration:

(31) a. Fragment1:  $[We\ drove_{m_1}\ for_{ts_1}\ [more\ than\ 6\ hours]_{mes_2}.]_{s_2}$

b. Annotation:

```
<motion xml:id="m1" type="drive" motionType="manner"/>
<measure xml:id="mes2" value="6" unit="hour" mod="moreThan"/>
<mLink xml:id="ml2" relType="runtime" figure="#m1" ground="#mes2"
trigger="#ts1"/>
```

c. Interpretation:

- (i)  $\sigma(m_1) := drive(m_1)$
- (ii)  $\sigma(mes_2) := q_{measured}(mes_2) \geq [6, hour]$
- (iii)  $\sigma(ml_2) := q_{runtime}(m_1) = q_{measured}(mes_2)$
- (iv)  $\sigma(s_2) := [drive(m_1) \wedge q_{runtime}(m_1) \geq [6, hour]]$

### 5.3 Interpreting Intervals

The measure expression *three days* may be used with a date expression *Christmas Day* as in the following dataset fragment (32). The annotation of this fragment has been presented in Section 4, Example (24). Here the measure expression is understood as providing information on either the initial or the terminal boundary of an interval of time with its quantity measured to be the length of three days.

- (32) a. Fragment2:  $[We \dots drove_{m_1} \dots [t_1 \textit{three days}]_{mes_1} \textit{before} [Christmas Day]_{t_2}]_{sem02}$
- b. Annotation: based on Example (24).
- c. Interpretation:
- (i)  $\sigma(m_1) := drive(m_1)$
  - (ii)  $\sigma(mes_1) := q_{measured}(mes_1) = [3, day]$
  - (iii)  $\sigma(ml_1) := begins(t_1, \iota(mes_1)) \wedge length(\iota(mes_1)) = [3, day]$
  - (iv)  $\sigma(t_1) := [month(12, t_2) \wedge day(22, t_2)]$
  - (v)  $\sigma(tl_1) := before(\iota(mes_1), t_2)$
  - (vi)  $\sigma(t_3) := [month(12, t_2) \wedge day(25, t_2)]$
  - (vii)  $\sigma(tl_2) := [drive(m_1) \wedge during(m_1, t_1)]$
  - (viii)  $\sigma(sem02) := [drive(m_1) \wedge begins(t_1, \iota(mes_1)) \wedge length(\iota(mes_1)) = [3, day] \wedge month(12, t_1) \wedge day(22, t_1) \wedge during(m_1, t_1) \wedge before(\iota(mes_1), t_2) \wedge month(12, t_2) \wedge day(25, t_2)]$

The function  $\iota$  maps an amount of time to an interval with its length measured to be that amount.

## 6 Concluding Remarks

This paper has reviewed some cases of formal inconsistency between the two recently published ISO standards, ISO 24617-1 ISO-TimeML (ISO, 2012b) and ISO 24617-7 ISOspace (ISO, 2014a) in annotating measure expressions, temporal and spatial, respectively. With a focus on durations (amounts of time) and distances, it has shown how the part of ISO-TimeML that annotates durations can be merged into ISOspace but with some modifications to resolve formal inconsistencies between the two standards. Following ISO 24617-6 SemAF Principles (ISO, 2014c) and Bunt (2015), it has also briefly examined the possibility of further generalizing the annotation scheme for measure expressions or incorporating it into `<srLink>` of ISO (2014b), but left it as an open issue. Sometimes a domain-specific modular approach with some overlaps licensed and tasks distributed can be considered of more practical use with greater efficiency than a single unified and theoretically compact treatment. The ubiquity of measure expressions in language and their use over various domains may also require a separate annotation scheme that can generally apply to various parts of semantic annotation. This paper has done its best not to propose extensive revisions, but attempted to resolve any formal inconsistencies with minimal modifications.

## Acknowledgements

I owe thanks to David Lovisek for modifying the travel log, to Harry Bunt, Suk-Jin Chang, Jae-woong Choe, Thierry Declerck, Koiti Hasida, James Pustejovsky, and other reviewers for their constructive comments, and especially to Roland Hausser who helped produce the final version.

## References

- Branimir, B., J. Castaño, R. Gaizauskas, B. Ingria, G. Katz, B. Knippen, J. Littman, I. Mani, J. Pustejovsky, A. Sanfilippo, A. See, A. Setzer, R. Saurí, A. Stubbs, B. Sundheim, S. Symonenko, and M. Verhagen (2005). *TimeML 1.2.1: A Formal Specification Language for Events and Temporal Expressions*.
- Bunt, H. (2010). A methodology for designing semantic annotation languages exploring semantic- syntactic isomorphisms. In A. Fang, N. Ide, and J. Webster (Eds.), *Proceedings of the Second International Conference on Global Interoperability for Language Resources (ICGL 2010)*, Hong Kong, pp. 29–46. Department of Chinese, Translation and Linguistics, City Univesity of Hong Kong.
- Bunt, H. (2011). 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*, Seoul, pp. 157–204. Hankookmunhwasa.

- Bunt, H. (2015). On the principles of interoperable semantic annotation. In H. Bunt (Ed.), *Proceedings of the Eleventh Joint ISO-ACL/SIGSEM Workshop on Interoperable Semantic Annotation (ISA-11)*, Queen Mary University of London, pp. xx–yy. A satellite workshop of IWCS 2015.
- Bunt, H. and J. Pustejovsky (2010). Annotating temporal and event quantification. In H. Bunt (Ed.), *Proceedings of the Fifth Joint ISO-ACL/SIGSEM Workshop on Interoperable Semantic Annotation (ISA-5)*, Hong Kong, pp. 15–22. Department of Chinese, Translation and Linguistics, City University of Hong Kong.
- Ide, N. and L. Romary (2004). International standard for a linguistic annotation framework. *Natural Language Engineering*, 10, 211225.
- ISO (2004). *ISO 8601:2004 Data elements and interchange formats – Information interchange – Representation of dates and times*. Geneva: ISO.
- ISO (2012a). *ISO 24612:2012 Language resource management - Linguistic annotation framework (LAF)*. Geneva: ISO. ISO Working Group:TC 37/SC 4/WG 1, Convenor and Project leader: Nancy Ide.
- ISO (2012b). *ISO 24617-1:2012 Language resource management - Semantic annotation framework - Part 1: Time and events (SemAF-Time, ISO-TimeML)*. Geneva: ISO. ISO Working Group:TC 37/SC 4/WG 2, Editors: James Pustejovsky (chair), Harry Bunt, Kiyong Lee (convenor and project leader), Bran Boguraev, and Nancy Ide in cooperation with the TimeML Working Group, <http://www.timeml.org>.
- ISO (2014a). *ISO 24617-7:2014 Language resource management - Part 7: Spatial information (ISOspace)*. Geneva: ISO. ISO Working Group: TC 37/SC 4/WG 2, Project leaders: James Pustejovsky and Kiyong Lee, supported by the ISOspace Working Group headed by James Pustejovsky at Brandeis University, Waltham, MA, U.S.A. The following is the homepage for the ISO-Space project <<https://sites.google.com/site/wikiisospace/>>.
- ISO (2014b). *ISO 24617-7:2014 Language resource management - Semantic annotation framework - Part 4: Semantic roles (SemAF-SR)*. Geneva: ISO. ISO Working Group: TC 37/SC 4/WG 2, Project leader: Martha Palmer.
- ISO (2014c). *ISO DIS 24617-6 Language resource management - Semantic annotation framework - Part 6: Basic principles (SemAF-Basics)*. Geneva: ISO. ISO Working Group: TC 37/SC 4/WG 2, Project leader: Harry Bunt.
- ISO/IEC (1996). *ISO/IEC 14977:1996(E), Information technology – Syntactic metalanguage – Extended BNF*. Geneva: ISO.
- Lee, K. (2012). Interoperable spatial and temporal annotation schemes. In H. Bunt (Ed.), *Proceedings of the Joint ISA-7 Workshop on Interoperable Semantic Annotation, SRSL3 Workshop on Semantic Representation for Spoken Language, and I2MRT Workshop on Multimodal Resources and Tools*, Istanbul, Turkey, pp. 61–68. A satellite workshop of LREC 2012, Istanbul, Turkey.
- Mani, I. and J. Pustejovsky (2012). *Interpreting Motion: Grounded Representation for Spatial Language*. Oxford: Oxford University Press.
- MITRE (2009). *SpatialML: Annotation Scheme for Marking Spatial Expressions in Natural Language*. The MITRE Corporation. Version 3.1, October 1, 2009, Contact: [cdoran@mitre.org](mailto:cdoran@mitre.org).
- Pustejovsky, J., R. Ingria, R. Saurí, J. Castaño, J. Littman, R. Gaizauskas, A. Setzer, G. Katz, and I. Mani (2005). The specification language TimeML. In I. Mani, J. Pustejovsky, and R. Gaizauskas (Eds.), *The Language of Time: a Reader*, pp. 545–557. Cambridge: Oxford University Press.
- Pustejovsky, J., K. Lee, H. Bunt, and L. Romary (2010). ISO-TimeML: an international standard for semantic annotation. In *Proceedings of LREC 2010*, La Valette, Malta.
- Pustejovsky, J., J. Moszkowics, and M. Verhagen (2011). ISOspace: the annotation of spatial information in language. In H. Bunt (Ed.), *Proceedings of the Sixth Joint ISO-ACL/SIGSEM Workshop on Interoperable Semantic Annotation (ISA-6)*, Oxford, England, pp. 1–9.
- Pustejovsky, J., J. Moszkowics, and M. Verhagen (2012). The current status of ISO-Space. In H. Bunt (Ed.), *Proceedings of the Joint ISA-7 Workshop on Interoperable Semantic Annotation, SRSL3 Workshop on Semantic Representation for Spoken Language, and I2MRT Workshop on Multimodal Resources and Tools*, Istanbul, Turkey, pp. 23–30. a satellite workshop held in conjunction with LREC 2012, Istanbul, Turkey.