

의미주석 표준

이기용^o
고려대학교
ikiyong@gmail.com

Standards on Semantic Annotation

Kiyong Lee^o
Korea University

요 약

최근에 ISO/TC 37/SC 4 산하의 Working Group 2 Semantic annotation에서 자연언어의 의미주석에 관한 4 개의 국제표준을 완성하여 출판하였다. 그 중에서 2 개의 국제표준 ISO 24617-1 SemAF-Time (ISO-TimeML)[1]과 24617-7 ISOspace[2]를 간략히 소개하는 것이 이 발표의 목적이다. 자연언어 텍스트에서 전자는 사건과 관련된 시간 정보를 주석처리하고, 후자는 사건(event), 특히 운동(motion)과 관련된 공간 정보를 주석 처리하는 주석체계(annotation scheme)들을 구축, 기술하는 명세언어(specification language)이다. 이 표준들은 또한 ISO 24612:2012 LAF (Linguistic annotation framework)[3]의 제약조건들을 준수하며 언어 주석체계를 구축하였다. 오늘의 발표는 이들 두 개의 국제표준에 준한 주석체계들 *ASisoTime*과 *ASisoSpace*가 LAF를 따라 어떻게 구축되었는지 그 추상통사구조(abstract syntax)를 명시하고, 의미주석체계로서의 이들 주석체계의 타당성을 보이기 위하여 주석기반의 의미형식(semantic form)들을 체계적으로 도출하는 과정을 또한 보이도록 한다.

주제어: 주석 (annotation), 주석체계(annotation scheme), 추상통사구조(abstract syntax), 의미형식(semantic form)

1. 서론

이 발표는 의미주석에 관한 ISO 국제표준을 소개하는 것이 그 목적이다. 다만, 오늘의 발표는 자연언어의 텍스트에서 사건과 관련된 시간 정보와 공간 정보를 주석 처리하는 SemAF-Time(ISO-TimeML)¹⁾[1]과 ISOspace[2]에 초점을 둔다.

발표 내용의 전개 순서는 다음과 같다. [2절]에서는 이 두 표준들의 기초가 되는 기존의 연구 노력들을 간단히 소개하고, [3절]에서는 두 표준을 따라 구축된 주석 체계(annotation scheme)들인 *ASisoTime* 과 *ASisoSpace*의 추상통사구조(abstract syntax)를 명시한다. [4절]에서는 조그마한 데이터셋(dataset)을 주석 대상의 예제로 삼고, 이 데이터셋 그 안에 어떤 표현들이 표시가능 표현(markable)²⁾이 되는지 보이도록 한다. [5절]에서는 ISO-TimeML과 ISOspace를 이용하여 이 데이터셋이 실제로 어떻게 주석 처리되는지 보이겠다. [6절]에서는 이러한 주석체계들이 의미주석으로서 과연 타당한지 입증하기 위하여 의미주석 처리된 자료에서 의미형식(semantic form)들을 체계적으로 도출하는 과정을 보이겠다(Lee[4] 참조). [7절]에서는 의미역(semantic role)의 주석을 다루는 SemAF-SR[5]과 비교하면서 이들 국제표준들의 일관성과 중복 여부를 간단히 논한다. [8절]은 결론이다.

1) SemAF-Time은 ISO 24617:1 국제표준 자체를 가리키는 약어인 반면에, ISO-TimeML은 이 국제표준을 따르면서 XML을 표상언어로 하는 구체적인 주석처리 방법을 명시해 주는 명세언어(specification language)다.

2) markable은 markable expression을 줄인 말이다.

2. 관련 연구

ISO-TimeML[1]에 대한 개관은 Pustejovsky et al. (2010)[6]에 실려 있다. TimeML[7]이 그 기반이 된다. 이 둘이 크게 다른 것은 LAF의 요구조건대로 전자는 추상통사구조(abstract syntax)에 기반을 두고 형식의미론을 통한 의미주석 체계로서의 타당성을 검증했다는 점이다. 그 한 예로, TimeML은 사건유형(event type)와 사건사례(event instance)를 구별하는데, ISO-TimeML은 사건 하나만을 설정하여 이것을 하나의 유형(type)으로도 해석하고 맥락에 따라 구체화된 하나의 사건사례(token, instance)로도 해석한다는 점이다.

이 둘은 주석 내용을 표시하는 표상언어(representation language)로 XML을 택한다. TimeML은 텍스트 안에 주석 내용을 표시하는 방법(inline)을 택하는 데에 반하여, ISO-TimeML은 LAF[3]의 요구조건을 수용하여 텍스트와 분리된 주석 표시 방법(standoff)을 택한다.

ISOspace[2]는 MITRE의 SpatialML[8]을 시발점으로 삼는다. 전자와 후자가 다른 점은 근본적으로 두 가지다. 첫째, 전자는 운동(motion)이란 개체를 주석 대상으로 도입하여, 운동과 같은 사건들을 처소(place)나 경로(path)와 같은 위치(location) 개체들과 연관시키는 데에 반하여, 후자는 오직 처소들 간의 관계인 질적 공간관계(qualitative spatial relation)나 상대적 방향관계(relative, orientational relation)만을 다룬다. 둘째, 주석 내용을 표시하는 표상언어로 후자는 SpatialML이라는 이름처럼 XML에 의존하지만, 전자는 술어논리 형식과 같은 언어도 표상언어로 허용함으로써 주석체계의 추상

통사구조(abstract syntax)가 구상통사구조(concrete syntax)와 별개의 기술층위(descriptive level)를 이룸을 보여준다. 이는 ISOspace가 LAF[3]의 주장을 따른 것이다. ISOspace은 Brandeis 대학의 연구팀을 중심으로 처음 개발되었다.³⁾

SemAF-SR[4]은 의미역에 대한 기존의 여러 명세언어(specification language)를 통합하려는 노력에서 이루어졌다. FrameNet[11], PropBank[12], VerbNet[13], LIRICS[14] 등이 그 기초가 되었다.

3. 주석체계로서의 추상통사구조

3.1 추상통사구조

LAF[3]는 주석체계(annotation scheme)가 추상통사구조(abstract syntax)와 구상통사구조(concrete syntax)의 2개 층위로 기술된다고 주장한다. 추상통사구조는 주석체계를 일반화하여 기술하는 반면, 구상통사구조는 XML과 같은 표상언어를 이용하여 주석 내용을 어떻게 표시할 것인가를 명시하여 준다.

Lee[15]는 주석체계의 추상통사구조를 공식화하여 다음과 같이 정의하였다.

- (1) 어떤 언어의 데이터셋(dataset)들의 집합 C 가 주어졌을 때, 주석체계 AS 의 추상통사구조는 $\langle M, E, R, @ \rangle$ 4개의 원소로 구성되며 각 구성소는 각각 다음과 같이 정의된다.
 - M 은 표시가능 표현(markable)이라고 일컬어지는 C 에 속한 언어 분절(segment)들로 구성된 유한비공집합(non-empty finite set)으로서 그 유형은 E 에 의해 명시된다.
 - E 는 기본개체(basic entity)라고 일컬어지는 원소들의 유한 비공집합(non-empty finite set)으로 원자개체와 복합개체로 나뉜다.
 - R 은 E 의 요소들 간의 (주로 2항) 관계들의 유한 비공집합이다.
 - $@$ 는 E 와 R 의 각 원소들의 속성(attribute)들과 가능한 속성치(value)들의 목록을 명시하여 준다.

3.2 ISO-TimeML에 기반한 주석체계

ISO-TimeML의 기반이 에 기반 주석체계 $AS_{isoTime}$ 은 다음과 같이 명시된다.

- (2) 어떤 언어의 데이터셋들의 집합 C 가 주어졌을 때, 주석체계 $AS_{isoTime}$ 의 추상통사구조는 $\langle M, E, R, @ \rangle$ 4개의 원소로 구성되며 그 구성소 각각은 다음과 같이 정의된다.
 - M 은 C 의 데이터셋의 분절(segment: 단어나 구 또는 형태소)이며 그 유형은 다음의 E 에 의해 명시된다.
 - E 는 (1) 시간(time)⁴⁾, (2) 사건(event), (3) 신호(signal)의 3개 유형의 기본(원자)개체(basic

atomic entity)들로 구성된다.

- R 에는 (1) 시간관계, (2) 종속관계, (3) 상관관계, (4) 측량관계 등 4개의 관계가 있다.
- $@$ 는 E 와 R 의 각 원소에 대하여 명시되는 속성들과 가능한 속성치들을 명시하여 주는데 구체적인 목록은 SemAF-Time(ISO-TimeML)에 명시되어 있다.

이를 기초로 하여 주석내용을 표시해 주는 $AS_{isoTime}$ 의 표상언어는 XML이다. 위의 E 와 R 의 원소들을 나타내 주는 XML 태그들은 구상통사구조에 의해 다음과 같이 명시된다.

(3) XML기반의 구상통사구조

기본개체: 시간 <timeX3>, 사건 <event>,

신호 <signal>

관계: 시간관계 <tLink>, 종속관계 <sLink>,

상관계 <aLink>, 측량관계 <mLink>.

3.3 ISOspace에 기반한 주석체계

ISOspace에 기반한 주석체계 $AS_{isoSpace}$ 는 다음과 같이 명시된다.⁵⁾

- (4) 언어자료 데이터셋들의 집합 C 가 주어졌을 때, 주석체계 $AS_{isoSpace}$ 의 추상통사구조는 $\langle M, E, R, @ \rangle$ 4개의 원소로 구성되며 그 구성소 각각은 다음과 같이 정의된다.
 - M 은 C 의 데이터셋의 분절(segment: 단어나 구 또는 형태소)이며 그 유형은 다음의 E 에 의해 명시된다.
 - E 는 (1) 처소(<place>), (2) 경로(<path>), (3) 운동(<motion>), (4) 비운동사건(<event>), (5) 공간개체(<spatialEntity>), (6) 공간신호(<spatialSignal>), (7) 운동신호(<motionSignal>), (8) 측량신호(<measureSignal>) 등 8개 유형의 기본개체(basic entity)들로 구성된다.⁶⁾
 - R 에는 (1) 질적 공간관계(<qsLink>), (2) 방향관계(<oLink>), (3) 운동관계(<moveLink>), (4) 측량관계(<mLink>) 등 4개의 관계가 있다.
 - $@$ 는 E 와 R 의 각 원소에 대하여 명시되는 속성들과 가능한 속성 값들인데 이는 ISOspace에 의해 명시되어 있다.

5. 원 데이터(raw data)와 표시가능 표현

$AS_{isoTime}$ 과 $AS_{isoSpace}$ 가 어떻게 언어 데이터를 주

3) [9]와 [10]을 참조할 것.

4) 시간은 시간(time), 이외에 날짜(date), 기간(duration) 및 빈도(frequency)로 하위 유형으로 나뉜다.

5) 편의상 구체적 통사구조에 의해 명시되는 XML tag를 함께 소개한다.

6) 사건경로(<eventPath>)를 기본개체 유형으로 다룰 것인가는 더 논의해야 할 문제다. 그리고 경로(path)는 시작점과 종점을 언급할 때에 다른 기초개체를 언급하기 때문에 복합기본개체(complex basic entity)로 Lee[15]는 다루고 있다.

석 처리하는지 보이기 위한 자료로 다음 데이터셋 (dataset)을 쓰기로 한다.

(5) Dataset01

```
<text xml:id="dataset01">
  In the summer of 1987 we drove down from
  Palo Alto to San Diego, mostly through
  the scenic drive California Highway One.
  During the trip we dropped by to see Carmel and
  the Hearst Castle besides going to the Disney
  Land. </text>
```

이미주석 처리를 하기 전에, 주어진 데이터셋을 단어나 형태소로 먼저 분절하고, 그 다음에 표시가능 표현들을 선정한다. 아래에 그 예를 편의상 inline 주석으로 보이도록 한다.

(6) 부분적으로 주석된 Dataset01a

```
<text xml:id="dataset01a">
<s xml:id="s1" from="w1" to="w22">7)In_sig1
the [summer of 1987]_t1 we_sel drove_m1 down
from_ms1[Palo Alto]_p1 to_ms2 [San Diego]
_pl2, mostly through_ms3 the scenic drive_p1
[California Highway One]_p12.</s>
<s2 xml:id="s2" from="w23" to="w41">
During the trip_m2 we dropped_m3 by to see_el
Carmel_p13 and the [Hearst Castle]_p14
besides going_m4 to_ms the [Disney
Land]_p15./>
<s3 xml:id="s3" from="w42" to="w48"> We
spent_e2 [almost a week]_t2 traveling_m5
around.</s>
</text>
```

4. 주석 처리의 실제

이 데이터셋에 보면 *ASisoTime*의 대상이 되는 markable은 몇 안 되지만 *ASisoSpace*의 markable은 꽤 많다. 이들 중 몇몇은 주석처리하면 다음과 같다.

```
(7) <semAnn xml:id="sem01" target="#01a:s1">
  <isoTimeML xml:id="time01">
    <timeX3 xml:id="t1" target="#range(w3,w5,3)"
      type="date" value="1987-SU"/>8)
    <event xml:id="m1" target="#w7"
```

7) 단어를 계산할 때 구두점은 무시하였다.

8) 날짜와 시간 표시는 ISO의 국제표준[16]을 따른다. 절기 (season) 표시에 대해선 Extended Date/Time Format (EDTF) 1.0, Draft Submission, 13 January 2012에서는 봄을 21, 여름을 22 등으로 표시했지만, 여기서는 계절의 영어 이름 앞 글자 2 개를 따서 봄을 SP, 여름을 SU, 가을을 AU, 겨울을 WI로 정했다.

```
  pred="DRIVE" tense="past")
  <tLink xml:id="tL1" eventID="#m1"
    relatedTotime="#t1" relType="isIncluded"
  </isoTimeML>
</semAnn>
```

데이터셋01a, 문장 s1 안에는 주석대상 표현으로 시간과 사건을 표시해 주는 표현 2개가 있다.

(7)은 이들 기본개체표현을 먼저 주석처리하고 이것들 사이의 관계를 <tLink>로 연결해 주고 있다.

다음은 동일한 자료, 데이터셋01a, 문장 s1에 대한 공간 주석처리 과정을 보여준다.

```
(8) <semAnn xml:id="sem02" target="#01a:s1">
  <isoSpace xml:id="space01">
    <spatialEntity xml:id="se1" markable="#w6"
      type="human" signature="set"/>
    <motion xml:id="m1" target="#w7"
      tense="past" semanticType="path"
      semanticClass="moveOut"/>
    <place xml:id="pl1"
      markable="#range(w10,w11,2)" type="ppl"
      ctv="city" state="CA" country="US"/>
    <place xml:id="pl2"
      markable="#range(w13,w14,2)" type="ppl"
      ctv="city" state="CA" country="US"/>
    <path xml:id="p1" markable="#w19"/>
    <path xml:id="p2" markable="#w20,#w21"/>
    <path xml:id="p0" markable=""
      endPoint1="#pl1" endPoint2="#pl2"/>
    <moveLink xml:id="mvL1" figure="#m1"
      ground="#p0" mover="#se1"
      relType="traverse"/>9)
  </isoSpace>
</semAnn>10)
```

운동관계(<moveLink>)에서 경로(path)가 운동이 실행된

9) <moveLink>의 속성 목록에서 속성 @figure의 값은운동 (motion)을 허용하고, 속성 @relType를 도입하고 그 값으로 traverse(통과)를 도입하였다.

10) signal들은 주석하지 않았다.

터전(ground)이 된다. 그런데 (8)에서 p1과 p2는 동일한 경로인데 위의 운동 경로의 일부밖에 되지 않는다. 따라서, 비섭취태그(non-consuming tag)라는 실제로는 주석 대상표현이 없는 요소(element)를 도입하여, 운동 m1의 경로가 되도록 설정한다. 그 다음에, 최근에 국제표준안으로 제출한 ISOref[17]와 Lee[18]의 제안대로 이러한 관계를 다음과 같이 주석처리한다.

```
(9) <semAnn xml:id="sem03" target="#01a:s1">
  <isoRef xml:id="ref01">
    <refLink xml:id="rL1" figure="#p1"
      ground="#p2" relType="identity"/>
    <refLink xml:id="rL2" figure="#p0"
      ground="#p1" relType="partOf"/>
    <refLink xml:id="rL3" figure="#p0"
      ground="#p2" relType="partOf"/>
  </isoRef> </semAnn>
```

위의 주석은 p1과 2가 동일한 경로임을 보여주며, 동시에 이들 경로가 운전 m1의 전체 경로 p0의 일부임을 보여준다.

6. 의미형식의 도출

의미주석의 타당성(validity)를 보여주는 주된 기준은 의미주석에서 타당한 의미형식(semantic form)을 체계적으로 도출할 수 있는냐는 점이다. Bunt[19]와 Lee[20] 등이 주석기반 형식의미론에 체계화에 노력해 오고 있다.

이 절에서는 앞에 든 예를 기초하여 의미형식을 도출하는 절차를 보이도록 한다.

```
(10) sem01의 의미형식
 $\sigma(t_1) := [\text{year}(t_1)=1987 \ \& \ \text{season}(t_1)=\text{summer}]$ 
 $\sigma(m_1) := [\text{drive}(m_1) \ \& \ \text{past}(m_1)]$ 
 $\sigma(tL_1) := \text{isIncluded}(\tau(m_1), t_1)$ 
 $:= [\tau(m_1) \subseteq t_1]$ 
 $\sigma(\text{sem01}) := \sigma(t_1) \oplus \sigma(m_1) \oplus \sigma(tL_1)$ 
 $:= [\text{year}(t_1)=1987 \ \& \ \text{season}(t_1)=\text{summer}$ 
 $\ \& \ \text{drive}(m_1) \ \& \ \text{past}(m_1) \ \& \ \tau(m_1) \subseteq t_1]$ 
```

여기서, $\sigma(i)$ 는 i 의 확인자(identifier)를 가진 표현의 의미형식을 말하며, $\tau(e)$ 는 사건 e 가 성립하는 시간 t 를 말한다. 그러면, (10)의 의미형식 $\sigma(\text{sem01})$ 은 차를 몰았던 과거의 사건이 일어난 시간이 1987년 여름이었음을 말하는 것으로 해석된다.

sem02의 의미형식은 그 도출 과정을 거꾸로 보이던 다음과 같다.

```
(11) sem02의 의미형식
 $\sigma(\text{mvL1}) := [\text{drive}(m_1) \ \& \ \text{past}(m_1) \ \&$ 
 $\ \text{path}(p_0) \ \& \ \text{traverse}(m_1, p_0) \ \&$ 
 $\ \text{agent}(se_1, m_1)]$ 
```

```
 $\sigma(p_0) := [\text{path}(p_0) \ \& \ \text{endPoint1}(p_{11})$ 
 $\ \& \ \text{endPoint2}(p_{12})]$ 
 $:= [\text{path}(p_0) \ \& \ \text{endPoint1}(p_{11})$ 
 $\ \& \ \text{endPoint2}(p_{12}) \ \& \ \text{PaloAlto}(p_{11})$ 
 $\ \& \ \text{SanDiego}(p_{12})]$ 
```

위의 의미형식들을 합치면 sem02의 의미형식이 된다.

7. 과제: 일관성과 중복

국제표준을 만드는 과정에서 항상 부딪치는 문제가 표준들과의 상호연산성(interoperability)과 중복의 문제다. Lee[]는 ISO-TimeML과 ISOspace가 상호연산성을 유지하기 위한 수정제안을 했는데 추상통사구조상의 문제보다는 구체적으로 주석처리 내용을 어떻게 표상하느냐의 문제이므로 쉽게 수정할 수 있음을 지적하였다.

ISOspace와 SemAF-SR이 중복되는 데가 많음을 Bunt[]가 지적하였다. 그 중 하나가 ISOspace에서 공간개체(spatial entity)를 도입하는 데서 생긴다. 그 예로 앞에 든 Dataset01a 문장 s3의 주석을 들겠다.

```
(12) <s3 xml:id="s3" from="w42" to="w48">We_se2
  spent_e2 [almost a week]_t2 traveling_m4
  around.</s>
```

```
(13) <semAnn xml:id="sem03" target="#01a:s3">
  <isoSpace xml:id="space03">
    <spatialEntityxml:id="se2" markable="#w42"
      type="human" signature="set"/>
    <motion xml:id="m4" target="#w47"
      tense="past" semanticType="compound"
      semanticClass="move"/>
    <moveLink xml:id="mvL1" figure="#m1"
      mover="#se2" relType="traverse"/>
  </isoSpace></semAnn>
```

문제는 (13)의 <moveLink>에서 속성 @mover라든가 @path와 같은 것들이 SemAF-SR에서 semantic role로 다루어지므로 ISOspace가 SemAF-SR와 중복되는 점들이 있다는 것이다. 다음은 그 한 예다.

```
(14) <semAnn xml:id="sem04" target="#01a:s3">
  <semAFsr xml:id="sr03">
    <entity xml:id="x2" markable="#w42"
      type="human" signature="set"/>
    <event xml:id="m4" target="#w47"
      eventFrame="travel.01"/>
    <srLink xml:id="srL1" event="#m4"
      participant="#x2" semRole="agent"/>
  </semAFsr></semAnn>
```

8. 맺음말

ISO-timeML이나 ISOspace는 주식체계를 만들기 위한 표준화된 명세언어(specification language)에 불과하다. 그리고 이 명세언어들을 기반으로 구축한 *ASisoTime* 과 *ASisoSpace*도 일차적으로는 사람인 주석자를 위한 것이다. 그런데 커다란 크기의 말뭉치, 다양한 소스에서 다양한 데이터 형식의 빅데이터를 주석 처리하려면 기계의 도움이 필요하다.

따라서 여기서 소개한 의미주석 프레임워크들은 인간과 기계의 협업을 전제로 한다. 이를 위해서는 자연언어 주석에 대한 기계 학습을 필요로 한다. 최근의 좋은 참고 자료로 Pustejovsky와 Stubbs (2013)의 *Natural Language Annotation for Machine Learning*[1]을 추천한다.

참고문헌

- [1] ISO, 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, 2012. Project leader: Kiyong Lee, Editorial committee: James Pustejovsky (chair), Branimir Boguraev, Harry Bunt, Nancy Ide, and Kiyong Lee.
- [2] ISO, ISO 24617-7:2014(E) Language resource management - Semantic annotation framework - Part 7: Spatial information (ISOspace), The International Organization for Standardization, Geneva, 2012.
- [3] ISO, ISO 24612:2012(E) Language resource management - Linguistic annotation framework (LAF). Project leader: Nancy Ide.
- [4] K. Lee, Formal semantics for interpreting temporal annotation. In P. van Sterkenurg (ed.), *Unity and Diversity of Languages: Special Lectures for the 18th International Conference of Linguists*, Amsterdam, pp. 97-108. Benjamins, 2008.
- [5] ISO, ISO 24617-4:2014(E) Language resource management - Semantic annotation framework - Part 4: Semantic roles (SemAF-SR), The International Organization for Standardization, Geneva, 2014. Project leader: Martha Palmer.
- [6] J. Pustejovsky, K. Lee, H. Bunt, and L. Romary, ISO-TimeML: an international standard for semantic annotation, *Proceedings of LREC 2010*, ELDA, La Valette, Malta, 2010.
- [7] J. Pustejovsky, R. Ingria, R. Sauri, J. Castano, J. Littman, R. Gaizauskas, A. Setzer, G. Katz, and I. Mani, The specification language TimeML, in I. Mani, J. Pustejovsky, and R. Gaizauskas (eds.), *The Language of Time: a Reader*, pp. 545-557, Oxford University Press, Cambridge, 2005.
- [8] MITRE, *SpatialML: Annotation Scheme for Marking Spatial Expressions in Natural Language*, Version 3.1., The MITRE Corporation.
- [9] J. Pustejovsky, J. Moszkowics, and M. Verhagen, 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)*, pp. 1-6, Oxford, in conjunction with the Ninth International Conference on Computational Semantics IWCS 2011, January 12-14, 2011.
- [10] J. Pustejovsky, J. Moszkowics, and M. Verhagen, 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, I2MRT Workshop on Multimodal Resources and Tools*, a satellite workshop held in conjunction with LREC 2012, pp. 23-30, Istanbul, Turkey, 2012.
- [11] C.J. Fillmore, C.F. Baker, and H. Sato, FrameNet as a "Net", *Proceedings of LREC 2004*, pp. 1091-1094, 2004.
- [12] J.D. Choi, C. Bonial, and M. Palmer, Propbank instance annotation guidelines using a dedicated editor, Jubilee, *Proceedings of the 7th International Conference on Language Resources Management and Evaluation (LREC)*, Valletta, Malta, 2010.
- [13] K. Schuler, VerbNet: a broad-coverage, comprehensive verb lexicon, Ph.D. thesis, University of Pennsylvania, 2005.
- [14] H. Bunt, V. Petukhova, and A. Schiffrin, LIRICS Deliverable D.4.4., Multilingual test suites for semantically annotated data, <http://lirics.loria.fr>, 2007.
- [15] K. Lee, 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, 2012.
- [16] ISO, ISO 8601:2004(E), Data elements and interchange formats - Information interchange - Representation of dates and times,

- International Organization for Standardization, Geneva, 2004.
- [17] ISO, ISO WD 24617-9 Language resource management - Part 9: Reference and coreference (ISOref), 2005. Project leaders: Laurent Romary and Kiyong Lee.
- [18] K. Lee, Semantic annotation of anaphoric links in language, in H. Bunt (ed.), *Proceedings of the 10th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (isa-10)*, a satellite workshop of LREC 2014, pp.29-38, ELDA, 2014.
- [19] K. Lee, A compositional interval semantics for temporal annotation, in E. Lee and A. Yoon (eds.), *Recent Trends in Language and Knowledge Processing*, pp. 122-156, Hankookmunhawsa, Seoul, 2011.
- [20] H. Bunt, 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*, pp. 157-204. Hankookmunhawsa, Seoul, 2011.
- [21] ISO, ISO DIS 24617-6 Language resource management - Semantic annotation framework - Part 6: Principles (SemAF-principles), International Organization for Standardization, Geneva, 2005. Project leader: Harry Bunt.
- [22] J. Pustejovsky and A. Stubbs, *Natural Language Annotation for Machine Learning*, O'Reilly, Sebastopol, CA, 2013.