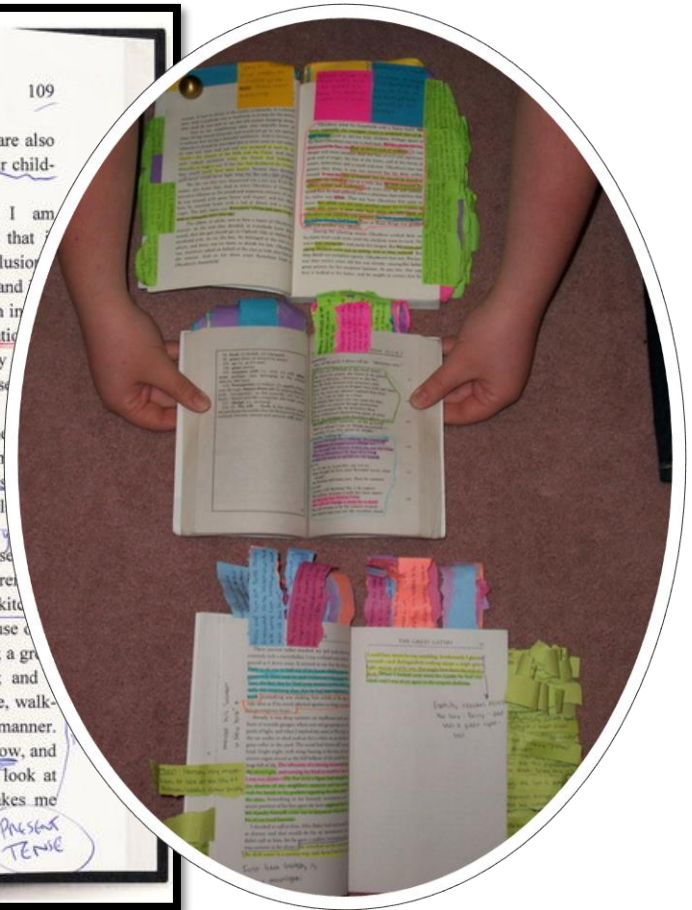
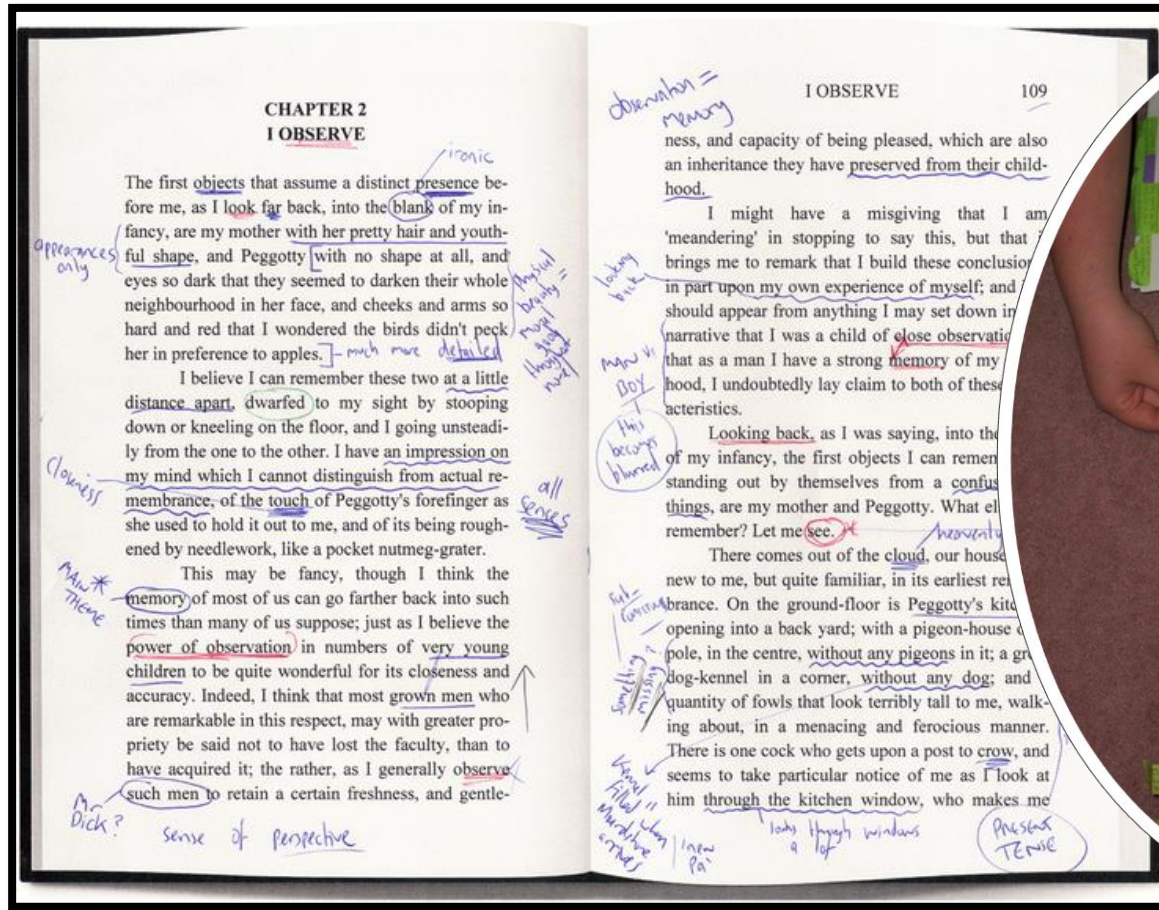


Practices in Text Annotation

Keigh Rim

1/31/2017

Good Old Text Annotation



Modern World Annotation

Themes

Abortion is wrong
- destroy potential (another Jesus?) (a gift from God)
- they were rich - had no reason apart from it was inconvenient

STRUCTURE - continuous monologue of 35 lines

Turning Point

Good reputation but does evil things

It removed an inconvenience

PURPOSE - to make reader sympathetic with narrator of foetus and therefore disagree with abortion

ORGANISATION - 35 line continuous monologue

EMOTIVE TONE - 1-16 happy 17-35 angry and bitter

TECHNIQUES - metaphor repetition alliteration

INDIVIDUAL WORDS - I was I was! Lived on trust 'Had no say' 'dropped'

CONTRAST - Joy of new life which is destroyed. Contrast is in narrator's feelings before/after turning point.

Narrator - the unborn child

VOICE 9.6 "For into us a child is born."

unfinished quotation "Made love or had sex?"

Repetition for emphasis

Alliterative (Joy of being conceived)

safety Security (Metaphor)

Feels love already

unable to fight back

Bitter at being aborted (betrayed)

no legal redress. A 'thing' not a 'human'

She had no more right to call herself a woman than a drug queen.

Somehow at sometime
They committed themselves to me
And so, I was!
Small, but I WAS!
Tiny in shape
Lusting to live
Thung in my pulsing cave
Seen they knew of me
My mother - my father
I had never on my being
I was there!
And love
The 'I couldn't think
Each part of me was saying
A silent, "Wait for me
I will bring you love!"
I was taken.
Blank school of fish
By the hand of one
Whose good name
Was graven on a brass plate
In Wimpole Street,
And dropped on the sterile floor
Of a foot-operated plastic waste bucket.
There was no Queens Counsel
To take my brief.
The cot I might have warmed
Stood in Harrods shop window.
When my passing was told,
My father smiled.
No grief filled my empty space
My skin was celebrated
With two tickets to see Danny La Rue
Who was pretending to be a woman
Like my mother was.

File Edit View Journal Tools Options Help

designed to be as a collaborative space where people can interact" (Lanningham 2006).

Social Media is a term used by some to describe the methods of communication and interaction that enable users to participate within the Web. There are five basic categories of social media, all of which fall under the banner of participatory culture. Each of the five can be divided into subgroups. An example accompanies each subgroup.

1. Collaboration: wikis -- ~~Wikipedia~~, open content -- ~~Creative Commons~~, social bookmarking -- Delicious, and social news -- Digg.

2. Communication: blogs -- ~~Wordpress~~, Micro-blogging -- ~~Twitter~~, social networking -- ~~Facebook~~.

3. Multimedia: photo sharing -- ~~Flickr~~, video sharing -- ~~YouTube~~, Livecasting -- ~~Livevideo.com~~.

Notes View Go Tools Bookmarks Window Help

2006_Rutter_calcium_microdomains_insulin.pdf (13 pages)

39
39
40
42
42
43
45
46
47
47
47

boxes and circles

sticky notes, like the paper equivalent (appear in the flat on the right; text searchable)

Fig. 3. Rapid imaging of Ca^{2+} in single MIN6 β -cells. Cells induced with adenovirus expressing green fluorescent protein were imaged in KRER medium (22 °C) with a 43x oil immersion objective using a spinning disc confocal microscope (Olea System with Yokogawa head controlled with Depixion Technology™ software) with acquisition of a complete 1024 pixel x 1024 image through the full axial depth (10 μ m, 10.7 μ m spacing) every 0.183 s using a Chroma EM charge-coupled device camera. Acetylcholine (100 nM) and KCl (10 mM) were added as indicated at the arrows. After acquisition background correction, pseudocolored rate images were generated off-line (A). The dashed line indicates the point of contact between two individual cells as identified morphologically under bright field illumination. (B) Average changes in Ca^{2+} in the three example shown are plotted, traces 2 and 1 for the same separate represent regions in the cell soma (intracellular) and periphery, respectively. (C) Average changes in Ca^{2+} in the three example shown are plotted, traces 2 and 1 for the same separate represent regions in the cell soma (intracellular) and periphery, respectively.

also in line with earlier studies on INS-1 cell populations expressing the much less pH sensitive probe acquerin, targeted to the ER lumen [26] and with studies in which the fluorescence of ER-targeted sensors was assessed rapidly after the permeabilisation of β -cells pre-exposed to varying glucose concentrations [29]. Nevertheless, none of the above data are alone sufficient to delineate the mitochondrially-targeted pericam [126,127], respectively (Fig. 4). By activating intramitochondrial dehydrogenases [131] these increases in intramitochondrial Ca^{2+} may then indicate repeated oscillations in both mitochondrial NAD(P)H [43,132] and cytosolic ATP [11] (Fig. 4). Interestingly, however, the slow oscillations in the two compartments occurred with a clear phase shift, with the increases

Annotation as Language Resource

- Language Resources for NLP/ML should be **Machine-readable**
 - Using common character set(s)
 - Serializable
 - Parse-compatible
 - Interoperable
- See Linguistic Annotation Framework (ISO standard #24612) literatures
 - Ide, N., & Romary, L. (2006). Representing Linguistic Corpora and Their Annotations.
 - Ide, N., & Suderman, K. (2007). GrAF: A Graph-based Format for Linguistic Annotations.

Scopes of Annotation Tasks

- Segmentation task
 - Entity identification
 - Find anchors, boundaries, or units
 - Also called unitization
 - *Can be non-consuming!*
- Linking task
 - Mark-up relations between entities/units
 - Only practicable once segmentation is given
- Labeling task
 - Giving names and attributes to segment and links

Column based data format

- MUC, CoNLL

	A	B
634	American	B-ORG
635	Depository	I-ORG
636	Receipts	I-ORG
637	(O
638	ADR's	B-ORG
639)	O
640	.	O
641		
642	Sólo	O
643	falta	O
644	el	O
645	visto	O
646	bueno	O
647	de	O
648	las	O
649	autoridades	O
650	brasileñas	O
651	para	O
652	hacer	O
653	efectiva	O
654	la	O
655	operación	O
656	de	O
657	compra	O
658	en	O
659	junio	O
660	próximo	O
661	.	O
662		
663	O	O
664	,	O
665	afirmó	O
666	.	O
667		
668	Por	O
669	Juan	B-PER
670	Miguel	I-PER
671	Núñez	I-PER

Development of XML

- Early XML serialization

In Washington <TIMEX3 tid="t1" TYPE="DATE" VAL="PRESENT_REF" temporalFunction="true" valueFromFunction="tf1" anchorTimeID="t0">today</TIMEX3>, the Federal Aviation Administration <EVENT eid="e1" class="OCCURRENCE">released</EVENT> air traffic control tapes from the night the TWA Flight eight hundred <EVENT eid="e2" class="OCCURRENCE">went</EVENT> down. There's nothing to show on why the plane <EVENT eid="e3" class="OCCURRENCE">exploded</EVENT>, but you <EVENT eid="e4" class="OCCURRENCE">cannot</EVENT> <EVENT eid="e5" class="OCCURRENCE">miss</EVENT> the moment. ABC's Lisa Stark <EVENT eid="e6" class="OCCURRENCE">has</EVENT> more.

<MAKEINSTANCE eventID="e1" pos="VERB" eiid="ei1" tense="PAST" aspect="NONE" />
<MAKEINSTANCE eventID="e2" pos="VERB" eiid="ei2" tense="PAST" aspect="NONE" />
<MAKEINSTANCE eventID="e3" pos="VERB" eiid="ei3" tense="PAST" aspect="NONE" />
<MAKEINSTANCE eventID="e4" pos="VERB" eiid="ei4" tense="PRESENT" aspect="NONE" />
<MAKEINSTANCE eventID="e5" pos="VERB" eiid="ei5" tense="INFINITIVE" aspect="NONE" />
<MAKEINSTANCE eventID="e6" pos="VERB" eiid="ei6" tense="PRESENT" aspect="NONE" />

<TLINK eventInstanceID="ei1" relatedToEventInstance="t1" relType="IS_INCLUDED" rule="2-1" />
<TLINK eventInstanceID="ei2" relatedToEventInstance="t1" relType="IS_INCLUDED" rule="2-1" />
<TLINK eventInstanceID="ei1" relatedToEventInstance="ei3" relType="BEFORE" rule="3-19" />
<TLINK eventInstanceID="ei3" relatedToEventInstance="ei4" relType="BEFORE" rule="6-1" />
<TLINK eventInstanceID="ei3" relatedToEventInstance="ei6" relType="BEFORE" rule="3-23" />

Stand-off Annotation

- Modern XML serialization in MAE

```
▼<NounVerbTask>
  ▼<TEXT>
    ▼<![CDATA[
      JABBERWOCKY By Lewis Carroll 'Twas brillig, and the slithy toves Did gyre and gimble in the wabe; All
      mimsy were the borogoves, And the mome raths outgrabe. 'Beware the Jabberwock, my son! The jaws that
      bite, the claws that catch! Beware the Jubjub bird, and shun The frumious Bandersnatch!' He took his
      vorpal sword in hand: Long time the manxome foe he sought-- So rested he by the Tumtum tree, And stood
      awhile in thought. And as in uffish thought he stood, The Jabberwock, with eyes of flame, Came whiffling
      through the tulgey wood, And burbled as it came! One, two! One, two! And through and through The vorpal
      blade went snicker-snack! He left it dead, and with its head He went galumphing back. 'And hast thou
      slain the Jabberwock? Come to my arms, my beamish boy! O frabjous day! Callooh! Callay!' He chortled in
      his joy. 'Twas brillig, and the slithy toves Did gyre and gimble in the wabe; All mimsy were the
      borogoves, And the mome raths outgrabe.
    ]]>
  </TEXT>
  ▼<TAGS>
    <NOUN id="N0" start="1" end="12" text="JABBERWOCKY" type="thing" comment=""/>
    <NOUN id="N1" start="61" end="66" text="toves" type="thing" comment="default value"/>
    <NOUN id="N2" start="94" end="98" text="wabe" type="place" comment="default value"/>
    <NOUN id="N3" start="119" end="128" text="borogoves" type="thing" comment="default value"/>
    <VERB id="V0" start="71" end="75" text="gyre" tense="past" aspect="simple"/>
    <VERB id="V1" start="80" end="86" text="gimble" tense="past" aspect=""/>
    <VERB id="V2" start="956" end="964" text="outgrabe" tense="" aspect="perfect progressive"/>
    <ADJ_ADV id="A2" start="37" end="44" text="brillig" type=""/>
    <ADJ_ADV id="A3" start="54" end="60" text="slithy" type=""/>
    <ADJ_ADV id="A4" start="104" end="109" text="mimsy" type=""/>
    <ACTION id="A0" fromID="V1" fromText="gimble" toID="N3" toText="borogoves" relationship="performed_by"/>
    <ACTION id="A1" fromID="N0" fromText="JABBERWOCKY" toID="V0" toText="gyre" relationship="performs"/>
    <DESCRIPTION id="D0" fromID="A3" fromText="slithy" toID="N1" toText="toves" relationship="describes"/>
    <DESCRIPTION id="D1" fromID="A4" fromText="mimsy" toID="N3" toText="borogoves" relationship=""/>
  </TAGS>
</NounVerbTask>
```


Interoperable Annotation

- Gate XML(Gate)
- UIMA XML (Dkpro)
- LIF JSON-LD(Lappsgrid)
- RDF & OWL (Semantic web community & Europe)
- ...

From model to specification

- A model is
 - Elements (terms)
 - Their relations
 - Interpretation of them
- A model is formalized into a specification
 - Technical descriptions of the model
- An annotator does their job based on guideline
 - Verbal descriptions of the model

Document Type Definition

- DTD is used to define XML scheme
 - Writing a DTD is re-writing your annotation scheme (model) for an XML based output
 - In MAE, we use a bit simplified DTD
 - For full details, see http://www.w3schools.com/xml/xml_dtd_intro.asp

DTD for MAE

- DTDs have 3 parts
 1. Task Name
 2. Elements
 - Extent Tags: unitization task
 - Link Tags: linking task
 3. Element Attributes: detailed labeling task

Task Name

- You can define your task name with a line such as the following:

```
<!ENTITY name "myTask">
```

Elements

- Extent Tags are defined with:

```
<!ELEMENT TagName ( #PCDATA ) >
```

- Link Tags are defined with:

```
<!ELEMENT LinkName EMPTY >
```

Attributes, pre-defined

- Extent Tags are automatically given the following attributes:
 - id
 - spans
 - text
- Link Tags are automatically given the following attributes
 - id
 - to-argument
 - from-argument

Defining an Attribute

- Full syntax for defining an attribute

```
<!ATTLIST
```

```
    __TAG_NAME__
```

```
    __ATT_NAME__
```

```
    __VALUE_TYPE__
```

```
    prefix
```

```
    __REQUIRE__
```

```
    __DEF_VALUE__>
```


Defining an Attribute

- `_TAG_NAME_` : the name of tag an attribute is associated
- `_ATT_NAME_` : the name of the attribute
- `_VALUE_TYPE_` : one of ID, CDATA, IDREF, or a list
(X | Y | ...)
- `prefix` : only used to give argument name for a link tag
- `_REQUIRE_` : #REQUIRE or #IMPLIED
- `_DEF_VALUE_` : the default value, double-quoted

More on predefined attributes

- id Attribute

- The id attribute is generated automatically
- Values are also automatically assigned by MAE
- An optional ID prefix can be specified as follows:
 - Unless you want to give a specific prefix, you don't technically need to put this line in your DTD

```
<!ATTLIST TagName id ID prefix="TN" #REQUIRED >
```

- spans Attribute

- Only when you want to allow an extent tag to be *non-consuming*, you make spans optional

```
<!ATTLIST Tag1 spans #IMPLIED >
```

More on predefined attributes

- argN Attribute

- A link tag grants from and to as its default argument
- To define custom argument structure, use argN

```
<!ELEMENT ARGUMENTS EMPTY >  
<!ATTLIST ARGUMENTS arg0 IDREF prefix="agent" #REQUIRED>  
<!ATTLIST ARGUMENTS arg1 IDREF prefix="patient" #REQUIRED>  
<!ATTLIST ARGUMENTS arg2 IDREF prefix="predicate" #REQUIRED>
```

- Use prefix and required-ness for finer definition
- All argN attributes automatically gets additional argNText attributes to keep text segment of the arguments