

Wish list for treebanks

Fei Xia

Univ. of Washington

06/19/2011

Penn-style treebanks

- Nice properties:
 - Guidelines are very detailed
 - Functional tags: e.g., -SBJ, -TMP
 - Empty categories and coindexation
 - Annotation is good (enough?) for POS tagging and parsing, at least in some domains.
- Wish list:
 - More genres and more data
 - Head/predicate/argument information
 - Multi-representational: Dependency structure, phrase structure, PropBank
 - Bio-NLP applications:
 - Original document (e.g., indentation) and structure of the document (e.g., sections, subsections, table)
 - Scope of quantifiers, negation, and adverbs
 - Attributes of events: possible, conditional, suggestion

Head/predicate/argument information

- Head:
 - Head percolation table
- Argument vs. adjunct
 - He loaded hay onto the truck.
 - He loaded the truck with hay.
- Predicates:
 - non-verbal predicate: He was **upset**; I am **out**.
 - verb-particle: John **picked up** Mary
 - light verb: John **did a survey** on X last week
 - fixed expressions: Please **keep an eye on** him
 - raising predicates: He **seems** to be happy
 - adverbs: He **apparently** missed the bus

Multi-representational, multi-layered treebanks

- Parsing: phrase structure (PS) or dependency structure (DS)
- There are treebanks for PS or DS, but not both.
- Treebanks and PropBanks are developed at different times.
- Hindi-Urdu treebank which includes PS, DS, and PropBank.

Some BioNLP applications

- 2009 i2b2 challenge: find medication information (e.g., name, dosage, frequency, reason) in hospital discharge summaries
- Find recommendations in radiology reports
- Find phenotypes (e.g., pneumonia) in ICU reports

1 Chief concern
2 cough/vomiting
3 History of present illness
4 28 yo man with extensive PMH
5 including DM, bipolar, and
6 ...
7 Past medical history
8 - has been given diagnosis of bipolar
9 - Seizure disorder
10 Allergies
11 sulfa drugs
12 ...
13 Home medications
14 divalproex sodium 1000 mg PO BID
15 ativan 0.50mg IV pm anxiety
16 ...
17 Social history
18 tobacco 1ppd, denies drugs
19 Family history
20 Father deceased from PNA

21 Physical exam
22 VS: Data Result Last MIN – MAX
23 04/30/10 09:59 Temp C: 37.2 37.2 – 37.4
24 04/30/10 09:59 HR: 128 128 – 12
25 GI: abdomen obese, nontender, +BS
26 Musk: strength 5/5 bilaterally
27 ...
28 Labs:
29 ABG inconsistent with PNA
30 ...
31 Assessment
32 28 yo man with extensive PMH presents with
33 complaints of cough, vomiting, ...
34 ...
35 Plan:
36 - PNA or flu?
37 - PNA: CAP or Aspirational?
38 ...

Task: Does the patient have PNA
at the time of hospital admission?

1 Chief concern
2 cough/vomiting
3 History of present illness
4 28 yo man with extensive PMH
5 including DM, bipolar, and
6 ...
7 Past medical history
8 - has been given diagnosis of bipolar
9 - Seizure disorder
10 Allergies
11 sulfa drugs
12 ...
13 Home medications
14 divalproex sodium 1000 mg PO BID
15 ativan 0.50mg IV pm anxiety
16 ...
17 Social history
18 tobacco 1ppd, denies drugs
19 Family history
20 Father deceased from PNA

21 Physical exam
22 VS: Data Result Last MIN – MAX
23 04/30/10 09:59 Temp C: 37.2 37.2 – 37.4
24 04/30/10 09:59 HR: 128 128 – 12
25 GI: abdomen obese, nontender, +BS
26 Musk: strength 5/5 bilaterally
27 ...
28 Labs:
29 ABG inconsistent with PNA
30 ...
31 Assessment
32 28 yo man with extensive PMH presents with
33 complaints of cough, vomiting, ...
34 ...
35 Plan:
36 - PNA or flu?
37 - PNA: CAP or Aspirational?
38 ...

➔ Knowing document structure is very helpful.

1 Chief concern
2 cough/vomiting
3 History of present illness
4 28 yo man with extensive PMH
5 including DM, bipolar, and
6 ...
7 Past medical history
8 - has been given diagnosis of bipolar
9 - Seizure disorder
10 Allergies
11 sulfa drugs
12 ...
13 Home medications
14 divalproex sodium 1000 mg PO BID
15 ativan 0.50mg IV pm anxiety
16 ...
17 Social history
18 tobacco 1ppd, denies drugs
19 Family history
20 Father deceased from PNA

21 Physical exam
22 VS: Data Result Last MIN – MAX
23 04/30/10 09:59 Temp C: 37.2 37.2 – 37.4
24 04/30/10 09:59 HR: 128 128 – 12
25 GI: abdomen obese, nontender, +BS
26 Musk: strength 5/5 bilaterally
27 ...
28 Labs:
29 ABG inconsistent with PNA
30 ...
31 Assessment
32 28 yo man with extensive PMH presents with
33 complaints of cough, vomiting, ...
34 ...
35 Plan:
36 - PNA or flu?
37 - PNA: CAP or Aspirational?
38 ...

1 Chief concern
2 cough/vomiting
3 History of present illness
4 28 yo man with extensive PMH
5 including DM, bipolar, and
6 ...
7 Past medical history
8 - has been given diagnosis of bipolar
9 - Seizure disorder
10 Allergies
11 sulfa drugs
12 ...
13 Home medications
14 divalproex sodium 1000 mg PO BID
15 ativan 0.50mg IV pm anxiety
16 ...
17 Social history
18 tobacco 1ppd, denies drugs
19 Family history
20 Father deceased from PNA

21 Physical exam
22 VS: Data Result Last MIN – MAX
23 04/30/10 09:59 Temp C: 37.2 37.2 – 37.4
24 04/30/10 09:59 HR: 128 128 – 12
25 GI: abdomen obese, nontender, +BS
26 Musk: strength 5/5 bilaterally
27 ...
28 Labs:
29 ABG inconsistent with PNA
30 ...
31 Assessment
32 28 yo man with extensive PMH presents with
33 complaints of cough, vomiting, ...
34 ...
35 Plan:
36 - PNA or flu?
37 - PNA: CAP or Aspirational?
38 ...

➔ Word ngrams only are not sufficient

Summary

- Wish list:
 - More genres and more data
 - Head/predicate/argument information → use functional tags and other markups
 - Multi-representational treebanks → Hindi/Urdu Project
 - Bio-NLP applications:
 - Original document (e.g., indentation) and structure of the document (e.g., sections, subsections, table) → use html markup
 - Scope of quantifiers, negation, and adverbs → mark the scope??
 - Attributes of events: possible, conditional, suggestion → mark attributes somehow?