Thirty Musts for Meaning Banking

Johan Bos & Lasha Abzianidze

1 Look at other meaning banks!

Groningen MEANING BANK

AMR

Parallel MEANING BANK

English Resource Grammar

U L F
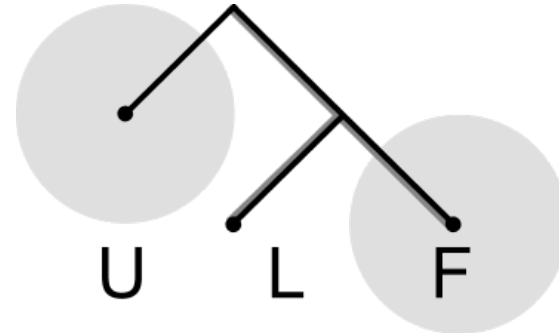
Damon showed me
his stamp album.

x1	x2	e1	t1

male.n.02(x1)
  Name(x1, damon)
time.n.08(t1)
  t1 < now
show.v.04(e1)
  Time(e1, t1)
  Recipient(e1, speaker)
  Theme(e1, x2)
  Agent(e1, x1)
stamp_album.n.01(x2)
  Owner(x2, x1)

2   Select public domain corpora!

3 Freeze the corpus before you start!

4 Work with raw texts in your corpus!

- words ≠ atoms of meaning
- tokenisation decisions can change
- never carry out annotation on tokenised text
- don't think "syntax" can fix word boundaries

New York-based companies …

5   Use stand-off annotation!

- keep annotations separate from original
- use character offsets of original (raw) text
- effective if various annotation layers are involved

Consider manual annotation!

- Many meaning banks are created with the help of a computational grammar (ERG, CCG)

- But there is an alternative: manually produce meanings (AMR)

```
(b/base-01
   :ARG0 (c/company)
   :Location (c/city
                 :Name (n/name
                           :Op1 "New"
                           :Op2 "York")))
```

7 Make a friendly annotation interface!

- Annotation can be fun – but also tedious

- Web-based, search facilities, statistics

- Sanity checker

8 Include an issue reporting system!

- Annotators will find bugs, raise questions, observe interesting phenomena

- Provide means to record this valuable information

- There are several existing issue reporting systems available (e.g., MANTIS)

9   Be careful with the crowd!

- **Gamification**:
slow and hard to recruit new players

- **Crowdsourcing**:
payment and dealing with cheaters

PHRASE·DETECTIVES

?
wordrobe
play what you mean

10   Profit from lexicalised grammars!

- A lexicalised grammar (e.g. CCG) gives an advantage in annotating syntactic structure
- Thematic roles can be directly coupled to lexical entries (words)

```
Amy  loves       Bob .
NP    (S\NP)/NP  NP  S\S
      [Stim,Exp]
```

29   Don't focus on just one language!

- Most meaning banks: one language (English)

- Risky!

- Meanings ought to be abstract (language neutral)

- Parallel corpora!

11   Try to use language-neutral tools!

- Tokenisers, taggers, parsers, boxers

- Resources tailored to a particular language …

- … prevent porting your method to other languages

12   Apply normalisation to symbols!

Expression	Meaning
2pm	14:00
half past eight	08:30
five to ten	09:55

Expression	Meaning
forty five	45
two hundred	200
367	367

13   Limit underspecification!

- Underspecification hinders inference
- Provide most plausible interpretation

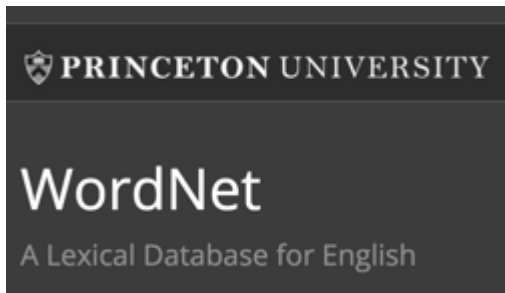*Tom got stuck in his sleeping bag.*
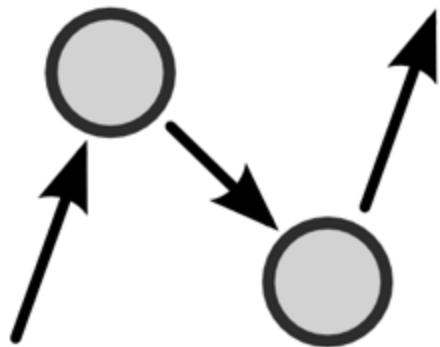
z Z Z Z

bag.n.01(x)
sleep.v.01(e)
Agent(e,x)

sleeping_bag.n.01(x)

14 Beware of annotation bias!

- Most plausible interpretation can also give unfair balance to stereotypes

- Examples:
  - gender
  - word sense

15   Use existing resources for word senses!

16   Apply symbol grounding!

- words → concepts
- named entities → wikification (AMR)
- toponyms → GPS coordinates
- visualisation of concepts or actions
- creating timelines

17 Adopt neo-Davidsonian events!

- *Davidson*   eat(e,x) eat(e,x,y)
- *Hobbs*   eat(e,x,y,z)
- *Parsons*   eat(e) Agent(e,x) Patient(e,y)

18 Use existing role-labelling inventories!

- PropBank: small set (6) of universal roles

- VerbNet: larger set (ca. 25) of universal roles

- FrameNet: specific roles for each frame

  Also needed: relations for
  prepositions, noun compounds, possessives

19 Treat role nouns differently!

- agent and recipient nouns are relational (*worker, employee, victim, dentist*)
- treating them as one-place predicates gives rise to contradictions
- AMR and PMB pursue this approach

(*Mia's husband is Vincent's boss*)

20   Beware of geopolitical entities!

- Names for GPEs can refer to
  locations, governments, sport squads, people

- Systematic polysemy:
  cities, states, countries, continents, etc.

- See ACE annotation guidelines

21 Give scope to negation!

- sentence meaning is about truth conditions
- *negation* plays a crucial role here
- proper treatment requires scope: ¬(…)

(concord, affixes, raising)

22   Pay attention to compound words!
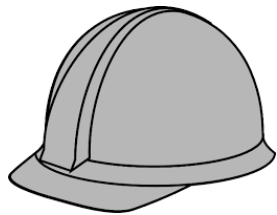
Expression	Meaning
hard hat	hard_hat.n.02(x)
peanut butter	peanut_butter.01(x)
old woman	old_woman.n.01(x) ???
secretary of state	secretary_of_state.n.01(x)
North and South Dakota	state.n.01(x) &Name(x,North_Dakota) & …
Jingle Bells	song.n.01(x) & Name(x,Jingle_Bells)

Lucy in the Sky with Diamonds

Several Species of Small Furry Animals Gathered Together in a Cave and Grooving With A Pict

23   Use inference tests in design!

- Amy is a smart woman $\xrightarrow{\text{yes}}$ Amy is a woman
  Amy is ten years old $\xrightarrow{\text{no}}$ Amy is old

23 Use inference tests in design!

- Amy is a smart woman --**yes**--> Amy is a woman
Amy is ten years old --**no**-> Amy is old

- Bob is unhappy   --**yes**-> Bob is not happy
Bob is not happy --**no**-> Bob is unhappy

23 Use inference tests in design!

- Amy is a smart woman —**yes**-> Amy is a woman
  Amy is ten years old —**no**-> Amy is old

- Bob is unhappy —**yes**-> Bob is not happy
  Bob is not happy —**no**-> Bob is unhappy

- Florence is the cultural capital of Italy —**no**->
  Florence is the capital of Italy

Divide and conquer!

- computing meaning is very hard
- many unsolved representation problems
  - aspect
  - information structure (focus)
  - presuppositions
  - generics
  - plurals

25 Put complex presuppositions on hold!

- some presuppositions triggers can be analysed as being anaphoric (definites, names)

- but more complex cases require copying or other sophisticated operations

(*the crowd applauded again*)

26 Respect elliptical expressions!

- Invisible, but omnipresent…
- Many different types of ellipsis

My hair is longer than Mary's ø.
Tom eats fish, but his brother doesn't ø.
ø Ho fame.

27 Think about generics!

- generics, habituals, kinds
- hard to model in first-order representations

A lion is strong.
Who invented the typewriter?
I play the piano.

Don't try to *be* clever!

The fork **is** on the table.
Mary **is** an engineer.
John **is** the teacher who graded my essay.

PRINCETON UNIVERSITY

WordNet
A Lexical Database for English

30 Measure meaning discrepancies!

- Need to be able to compare meaning representations

- Current approaches based on Allen et al. 2008 (Smatch for AMR, Counter for DRS)

- We need to make "matching" more semantical

The End

pmb.let.rug.nl