FrameNet Data Creation Nitty-Gritty

Michael Ellsworth
infinity@icsi.berkeley.edu

FrameNet Workshop
NAACL 2015

Outline (easy to hard)

- Lexicographic annotation
  - Corpus extraction
  - Sentence selection
  - Labels
  - Diversity checking
- Frame creation
  - Related frames
  - FEs
  - LUs
- Crowd sourcing and automation(?)

Background: What is FN?

- Map from form to natural meanings
    - Lemma, in context, to frames (via LU and anno)
    - Particular GFs and PTs to FEs (via annotation)
- Criteria for correct annotation and frame def.:
    - Same LU, same FE allows paraphrase
    - Common background assumptions across frame
- Must be checked in real, corpus exx.

Corpus Extraction

- Too many examples of LU

- Subselecting, keeping relevant diversity

- Strategies (e.g. finding a direct object):

  - WordSketch-based search: √ Obj

  - Rule-based search: T NP:Ground

  - Low-level grep-like search (cqp):
    @[lemma="in"&pos="IN"] []{0,4} [pos="N.*"];

Annotation

- Select a good sentence

- Apply FE/GF/PT

- Mark other contextual info: Supports, Relative clauses, Aspectual markers, Metaphor...

- Check that diversity is captured; iterate if not

  - Check against other LUs of the frame

  - Check against intuition

```
File  Edit  View  Search  Terminal  Help
101063074:   bncp=2802=36=25 {On} the little trolley
101063193:   bncp=2802=36=46 perfect {on} barbell wheels :
101063521:   bncp=2803=16=3 A woman , {on} the sandbank . "
101063576:   bncp=2803=18=2 {On} the sandbank . "
101063705:   bncp=2803=27=2 It stretched , broad and placid , to the horizon {on} both
sides .
101063711:   bncp=2803=27=3 Further {on} down , near to the city , a single felucca was
 gliding gracefully in towards the bank .
101063932:   bncp=2803=30=2 ` It was Lying right up {on} the shoal . "
101064240:   bncp=2803=41=5 I 'd expect them to be wasting their time {on} this foolish
 nonsense and not me .
101064339:   bncp=2803=44=3 Right {on} the spot , too !
101064425:   bncp=2803=46=3 He had a dozen men {on} the other side of town waiting for
him .
101064654:   bncp=2803=55=4 Nothing moved {on} the bank or out on the water .
101064659:   bncp=2803=55=4 Nothing moved on the bank or out {on} the water .
101064694:   bncp=2803=56=2 But why {on} earth &hellip; ? "
101064745:   bncp=2803=58=3 Normally it went about its business either {on} foot or in
an arabeah , the horse-drawn cab distinctive to the city .
101065035:   bncp=2803=65=1 The third person in the car was another Egyptian , definite
ly about thirty , slim and dressed , like the Prince , in a smart , European-style sui
t but with the usual pot-like tarboosh of the Egyptian professional {on} his head .
101065082:   bncp=2803=66=4 They had worked together often and got {on} well .
101065092:   bncp=2803=67=1 The Prince and McPhee had walked {on} to the top of the ban
k and were standing looking down at the river .
101065141:   bncp=2803=69=2 Someone else was going to do this one and then they suddenl
y switched me {on} to it . "
:
```

0   ` It was Lying right up ON the shoal . "

1   The friar then made his way back ON to the main highway , past the Priory of St Mary Overy and across London Bridge .

2   Only when the opposition are right back ON the defence or for free kicks .

3   This still leaves Britain way down ON the pasta–eating league .

4   He had n't seen an image in the corner of his eye , belatedly realised who it could be , and was making his way back ON foot to investigate ?

5   The Jerry observers were looking right down ON us .

6   Müncheberg and his Kette from 7/JG 26 were on their way out ON another hunting expedition .

7   they 'll fall right out ON the floor among the

8   We take great pride in our shows , we get right off ON playing live , always have

9   So the camera was right down ON the bottom .

10   who 're right down ON their luck .

- dist-dir-(15)

1. ` It was Lying right up ON the shoal . " Edit
2. So the camera was right down ON the bottom . Edit
3. It disgorged a crowd of bewildered passengers carrying buckets and spades and shrimping nets who pushed their way back ON again .DNI Edit
4. At night there were no longer any bonfires to be seen , either on the hill or way out ON the surrounding plain . Edit

Frame Creation

- Initial brainstorming of wordlists
  - Our own brains
  - Thesauruses
  - WordNet, PropBank, etc.
- Checking shared FEs and entailments for LUs
  - Imagined sentences
  - Initial corpus searches
- Not confirmed until annotation complete!

Example: Color

- Possible LUs:
  - What to include?
  - Wide net first, then find differences
- FEs
  - Example sentences?
  - FE hypotheses
  - Test with all LUs
  - Test against corpus for completeness

Color LU brainstorming

- Red, green, blue, black, white, yellow, purple, pink, brown, orange, gray

- Puce, chartreuse

- Burnt umber; cherry blossom pink?

- Midnight?, forest?; dark blue?, light green?, greenish blue?

- Reddened?, green-colored?

LU Answers 1

- Basic color terms are obviously in
  - Gray includes "grey" and "gray"
- Other dedicated color terms included as they occur in the corpus
  - Puce
  - Burnt umber is a dedicated color word, found in the corpus 22 times, enough to be useful
  - "Cherry blossom pink" isn't found in BNC, and is compositional, from "pink" and "cherry blossom"

LU Answers 2

- Midnight? One unambiguous BNC ex., + a few maybes:

    – The more I thought of that {midnight} face , the more intelligent and charming it became ; and it seemed too to have had a breeding , a fastidiousness , a delicacy , that attracted me as fatally as the local fishermen 's lamps attracted fish on moonless nights .

    – He turned slightly in the seat , the light from a street-lamp gilding the carved tanned planes of his face , turning the thickness of his hair to {midnight} silk , so that despite herself Fran felt a tiny ripple of appreciation for his male beauty .

    – Peter Scott provides the geese tastefully disappearing into a background of {midnight} green .

    – The rainbow tie showed up well against the {midnight} blue of the shirt .

LU Answers 3

- Compositional items not included
  - "Dark blue", etc., from "dark" + "blue"
  - "Fire engine red" from "fire engine" + "red"
- "reddened", etc., from *redden.v*, belonging to a different frame, not in FN yet
  - Different relationship to Time FE
  - Inherits from Event, not State
  - "reddened" ≠ "red"

FE brainstorming

- The entity that is colored: Entity

- The color: Color

- Comparand:

  – The *green* of a leaf; leaf *green*

- Degree: very, somewhat, etc.

- Color_qualifier: dark, bright, etc.

FEs found in annotation

- Attribute:
  - red in color
- Cause:
  - red with embarrassment
- Subregion:
  - red in the face
- Descriptor:
  - a beautiful red
- Type:
  - red 5

Can anno be done faster?

- Automatic Semantic Role Labeling
  - Several systems, most recently SEMAFOR from CMU
  - Reasonable, but variable, performance
  - Dependent on hand annotation
- Machine assisted
  - Tried pre-labeled annotation at FN, often worse than zero
- Crowd-sourcing
  - Most recently, collaboration btw FN and Google
  - Check it out at the LAW Workshop workshop!

Can frames be made faster?

- Automation
  - Many efforts (e.g., Borin et al. 2010)
  - Using other resources plus corpus (Green & Dorr 2004)
  - Using parallel corpora (Padó & Lapata 2005)
- Machine Assistance
  - Tool under development at Google under Abhijit Prahabal... Stay tuned!
- Crowd Sourcing?
  - No serious testing of this idea so far.

Conclusion

- FrameNet frames capture natural categories
- We spend a lot of time making that happen
  - Annotation is a vital part of accurately making frames
- There are possibilities for speeding things up
  - Crowdsourcing
  - Machine assisted creation
  - Machine created frames and annotation
- Still large qualitative difference between hand-curated data and other sources

Thanks!