Banks Meeting: Word Sense for Verbs and Nouns

2009-02-24

Martha Palmer and Ed Hovy

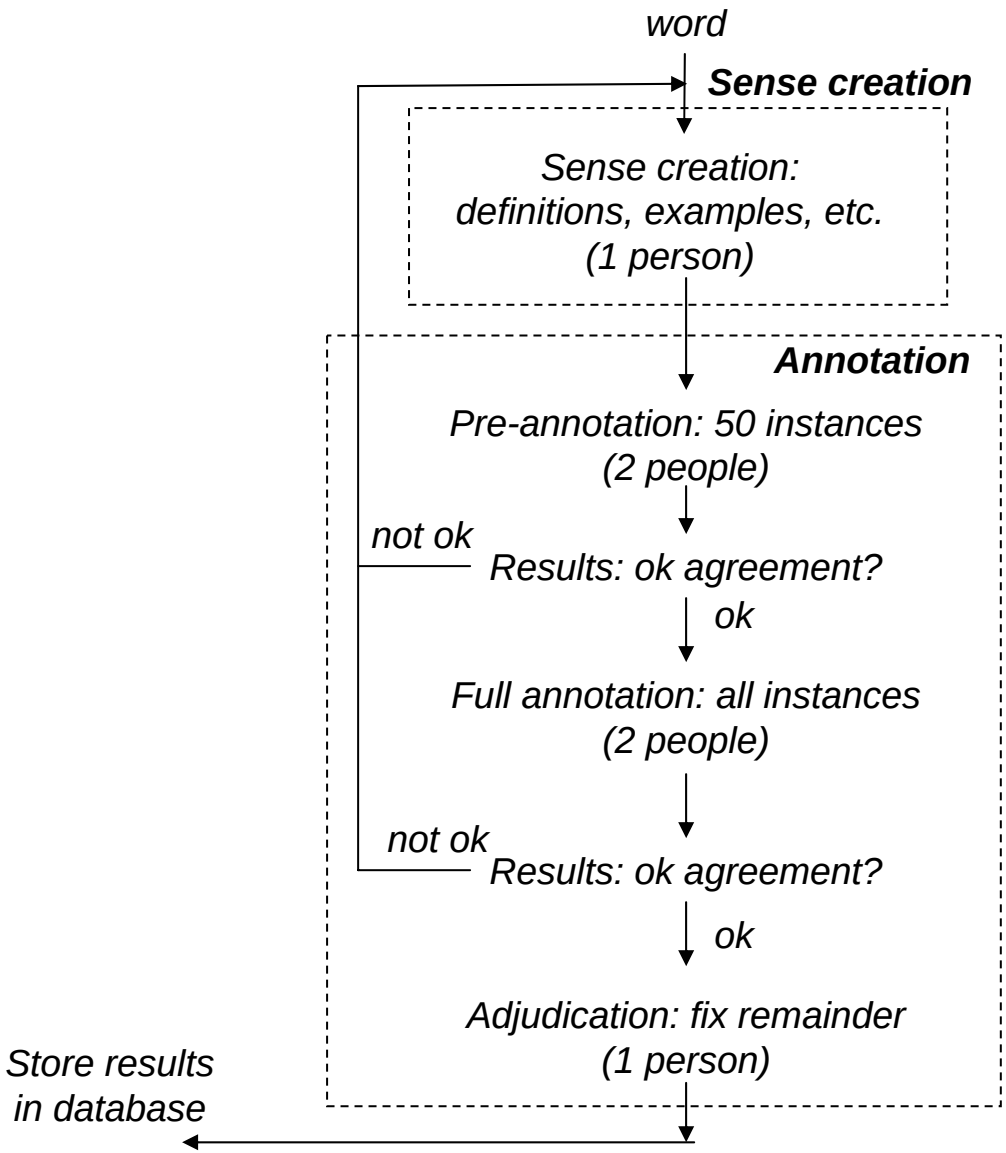
http://www.bbn.com/NLP/OntoNotes

Verb Sense Annotation

Sense Tagging: WordNet – *'press'*

- S: (v) press (exert pressure or force to or upon) *"He pressed down on the boards"; "press your thumb on this spot"*
- S: (v) press (place between two surfaces and apply weight or pressure) *"pressed flowers"*
- S: (v) compress, constrict, squeeze, compact, contract, press (squeeze or press together) *"she compressed her lips"; "the spasm contracted the muscle"*
- S: (v) press (crowd closely) *"The crowds pressed along the street"*
- S: (v) press (create by pressing) *"Press little holes into the soft clay"*

Sense Creation Process



Grouped senses for *'press'*

- **# 1: physically apply pressure or force on, over, or through something**
- **# 2: continue or move on; persevere on**
- **# 3: persuade or force someone into doing something**
- **# 4: (cause to) be a current burden or oppression**
- **# 5: bring petition, charges, or lawsuit for/against**
- **# 6: make a record or CD**
- **# 7: lift weights**
- **# 8: iron; make flat by applying pressure**
- **# 9: squeeze a fruit (usually grapes) for juice**

Grouped senses for *'press'*

- **# 1: apply pressure or force on, over, or through something**
- **# 2: continue or move on; persevere on**
- **# 3: persuade or force someone into doing something**
- **# 4: (cause to) be a current burden or oppression**
- **# 5: bring petition, charges, or lawsuit for/against (~10)**
  - **Both George W. Bush and Al Gore today continue [*-1] to press their cases literally and figuratively .**
- **# 6: make a record or CD (none)**
- **# 7: lift weights (none)**
- **# 8: iron; make flat by applying pressure (2)**
- **# 9: squeeze a fruit (usually grapes) for juice (none)**

Need more data

- New verbs
- New verbs with very few instances
- Old verbs with insufficient data for many senses, ex. *press*
- Old verbs with one clear predominant sense that is well represented, but which still need more data for other senses
- How to find the rare instances?
  - Random sampling?
  - Active learning *Chen, Schein, Ungar, Palmer, NAACL06, Hovy?*
  - Language Models? Dmitry Dligach
- See Data Selection Plan later today…

English Verb Sense Tagging Status

- **1700 lemmas have passed grouping and been delivered**
- **2172 have been grouped**

	WSJ	ECTB	EBN	EBC
Tokens	36331	16730	25631	22231
% coverage	88.7%	49.5%	87.2%	89.0%

WSD with English OntoNotes Verbs

- **Picked 217 sense group annotated verbs with 50+ instances each (*Chen, Dligach & Palmer, ICSC 2007)***
  - **35K instances total**
  - **WN polysemy range: 59 to 2;**
  - **Coarse polysemy (group) range: 16 to 2**
  - **Test: 5-fold cross-validation**
  - ***Automatic performance approaches human performance!***

WN Avg. Polysemy	Onto Avg. Polysemy	Baseline	ITA	MaxEnt	SVM
10.4	5.1	0.68	0.825	0.827	0.822*

**Now up to .83, Dligach & Palmer, ICSC08*

Chinese: How do we choose which verbs to sense-tag?

- Verbs that have
  - More than one frameset
  - More than 8 instances
- Sense file creators have access to
  - Frame files from the propbank
  - Definitions from CETA dictionary
  - Examples from the corpus
- Some words turn out to be monosemous

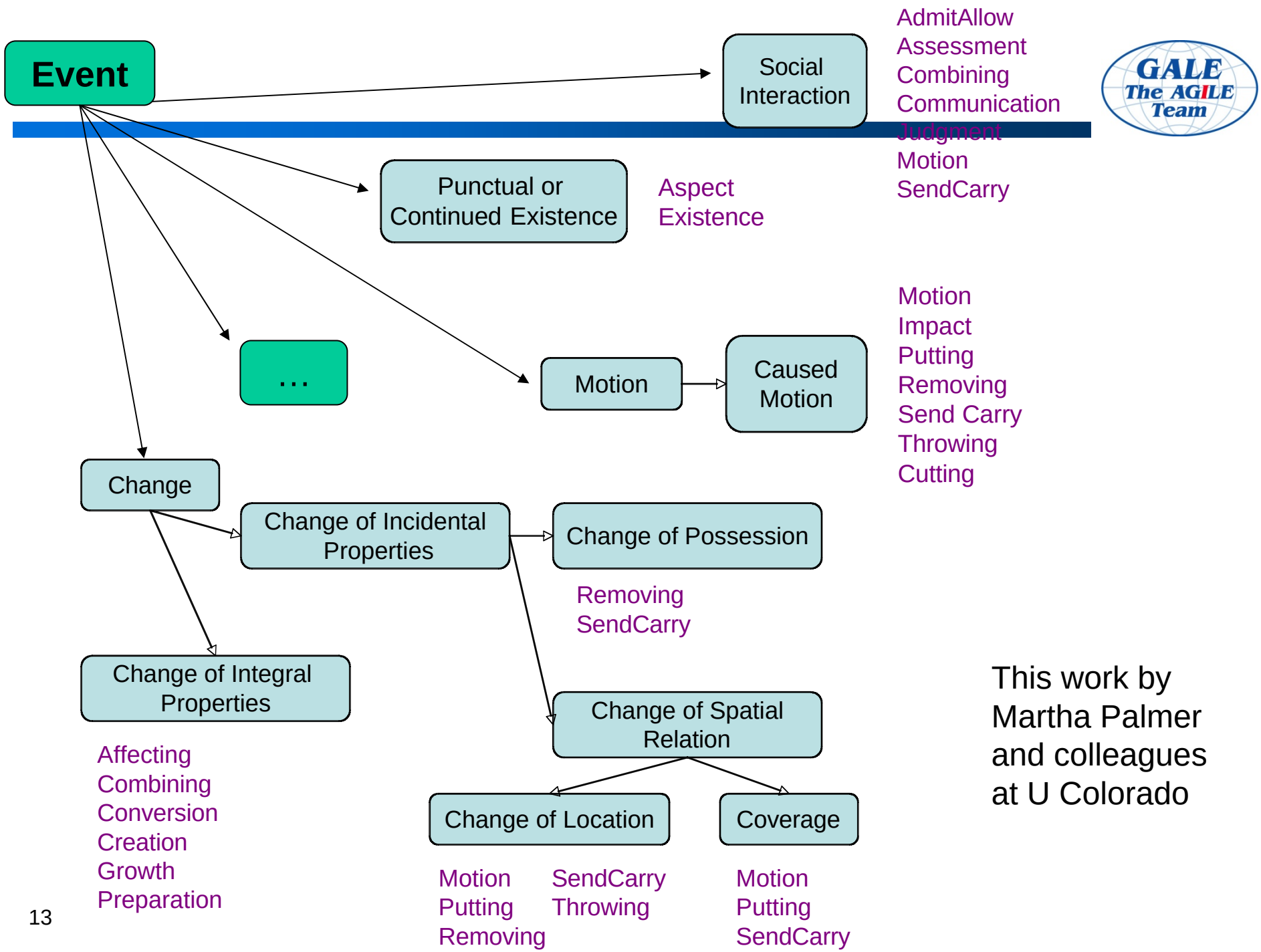
Chinese Verb sense tagging

- **OntoNotes 1.0:**
  - **301 verbs**
  - **24,727 instances (ctb data)**
- **OntoNotes 2.0**
  - **301 verbs**
  - **25,149 instances (bn data)**
- **OntoNotes 3.0***
  - **~111 verbs, 19888 instances (ctb + bn + bc)**
  - **301 verbs, 19839 instances (bc)**

Forward pointer to Ontologizing

A subset of the verbs we had grouped have been organized by semantic class using VerbNet.

- **Communication**
- **Predicative complements**
- **Creation**
- **Existence**
- **Send and Carry**
- **Change of Possession**
- **Motion**
- **Throw**
- **Put**
- **Miscellaneous**



This work by Martha Palmer and colleagues at U Colorado

OntoNotes – Arabic Status

- **Treebank: revised to improve consistency**
- **PropBank: 200K ATB-3, on hold pending revised trees**

- **Verb sense:**
- **Noun sense:**
- **Ontology:**

proceeding, planning to merge later with revised trees

- **YR1 – Propbank, 450 Framesets, token coverage, 80%**
- **YR2 – 100K sense tagged, 160 Framesets subdivided**
- **YR 3 – 200K sense tagged, same Framesets (48.4%)**
- **YR 4 – more sense tagging or PropBank revising?**
  - **Revise 200K, add 100K NW or 100K BN?**

Noun Sense Annotation

Annotation Coverage for Nouns (Eng,Ar)

Percent coverage

noun	english				arabic
	wsj	ectb	bn	bc	nw
y1	37.50%				
y2	44.80%		37.10%		
y2.9	56.60%	44.90%	46.70%	33.80%	
y3	65.40%	63.70%	62.40%	47.60%	10.1%

noun	english					arabic
total at end	wsj	ectb	bn	bc	total	nw
y1	25750				25750	
y2	30745		14542		45287	
y2.9	38861	29511	18010	9408	95790	
y3	45813	43016	24474	14588	127891	6583
Rest to 100%	70050	67529	39221	30647	207447	

rate for year	wsj	ectb	bn	bc	total	date
y1	25750				25750	Feb 07
y2	4995		14542		19537	Nov 07
y2.9	8116	29511	3468	9408	50503	Feb 09
y3	6952	13505	6464	5180	32101	Add. data

1. The y1, y2 and y2.9 numbers are only for high-quality annotations (>0.85 ITA) whereas the y3 numbers are using all the data.
2. The arabic numbers in the tables are using monosemous nouns. If we just use Arabic Wordnet monosemous nouns, then the coverage changes to 21%.

Feb 09 Noun Status — Internal Counts

- **Upload of English nouns, Feb 09:**

	Types	Instances
Double-annotated	1314	111628
Adjudicated	623	102847
Avg. agreement		.93

- **Upload of Arabic nouns, Feb 09:**

	Types	Instances
Double-annotated	98	10201
Adjudicated	61	9351
Avg. agreement		.89

- **Chinese will be uploaded by March 09**

Annotation Coverage for Nouns (Chi)

- **We are currently behind:**

Percent coverage

noun	chinese		
	ectb	bn	bc
y1	29.0%		
y2	31.9%	20.2%	
y2.9			
y3	20.7%	18.0%	15.0%

Raw counts

noun	chinese		
	ectb	bn	bc
y1	21091		
y2	23154	14779	
y2.9			
y3	14919	13199	4155

1. The y1, y2 and y2.9 numbers are only for high-quality annotations (>0.85 ITA) whereas the y3 numbers are using all the data.
2. The y3 Chinese noun numbers almost all account for monosemous cases in the corpora. The total raw annotations are around 4000 across all corpora
3. Part of the y1 and y2 Chinese noun data was annotated against erroneous sense files. This explains the decrease in coverage in y3
4. The Chinese sense data does not count the out of part-of-speech annotations

- **Recently recovered from y1 and y2 annotations:**

	Total annotated	Recovered	85%+	60%+
CTB	70K	35K		
BN	14K	7.5K		

Explanation for Chinese lag

- **Poor pinyin-to-chinese mapping:**
  - **Had to redo; all complete and conformant to Colorado scheme**
  - **Not all previously annotated words converted yet; some still**
  - **Need conversion**
  - **Need checking to see if senses are ok**
- **Poor sense creator:**
  - **Initially, used nouns whose senses had been created by Penn**
  - **During this time, our sense creator created new words' senses (to build up lead before annotation starts)**
  - **About 40% of these words turned out monosemous; so sense creator 'invented' senses**
  - **But we could only discover problem after annotation started: annotators could not reach agreement**
  - **Fired sense creator**
  - **Had to re-do sense creation and annotations**

Plan for Chinese

- **Annotate aggressively:**
  - **Need to annotate about 100K instances (bring total from 32K to 132K = 77% coverage of total 172K)**
  - **Experience shows individual annotation rate = 2.5 decisions / min**
  - **Need double-annotation, and usually about 20% re-done due to sense problems: about 1 double-decision / min**
  - **So need 100K mins = 1667 hours**
  - **With 10 annotators at 10hrs/week ➔ 17 weeks**
  - **Add 3 weeks' hiring and training time**
- **Current status:**
  - **4 annotators currently active**
  - **Hiring 12 new annotators now (tested about 35 candidates)**
- **Appoint dedicated Chinese annotation manager**
- **Monitor rate with BBN**