Making Sense of Word Sense Variation

Rebecca J. Passonneau*, Ansaf Salleb-Aouissi*,
Nancy Ide†
*Columbia University
†Vassar College

Outline

- Related work: human WSD
- MASC word sense annotation task
- Observation: varying IA, depending on word
- Explanatory factors
- Quantifying sense similarity
- Mining for systematic variation: assoc. rules
- Future work

Related Work: Human WSD

- Interannotator agreement using external sense inventory
  - Senseval (e.g., Ng 1999; Pedersen, 2002; Palmer et al. 2005)
  - Ontosem/Mikrokosmos (Passonneau et al. 2006; Dorr et al. forthcoming)
- Psycholinguistics: reaction time (Klein & Murphy 2002; Ide & Wilks 2006)

MASC

- MASC
  - OANC subset (OANC: 14 million words, 1990+, heterogeneous corpora)
  - Hand validation, e.g., NE, NP & VP Chunking
  - Hand annotation: word senses
- Goals
  - Freely available, heterogeneous WSD corpus
  - Align WordNet & FrameNet senses
  - Study WSD agreement issues

MASC WSD Task

- Collaborative selection of 10 words (ANC, WordNet, FrameNet)
  - Relatively polysemous (5-10 senses)
  - Balanced for part-of-speech
- Materials
  - Annotation tool linked to WordNet
  - Six trained annotators
  - **Small, multiply annotated set (50 examples/word)**
  - Large set with one annotator (1000 examples/word)

Ten Polysemous Words

Word	POS	No. senses	Freq
fair	Adj	10	463
long	Adj	9	2,706
quiet	Adj	6	244
land	Noun	11	1,288
time	Noun	10	21,790
work	Noun	7	5,780
know	Verb	11	10,334
say	Verb	11	20,372
show	Verb	12	11,877
tell	Verb	8	4,799

Annotation Interface

Interannotator Agreement

- Agreement coefficients (kappa, alpha, scott's pi; see Artstein & Poesio 2008)
- Corrects for chance agreement

$$\frac{p(A_O) - p(A_E)}{1 - p(A_E)}$$

- Ranges from 0 for chance to 1 or -1
- Calculate $p(A_E)$ on basis of observed proportion of each coding value *k* in the *i* by *j* matrix of *i* annotators, *j* annotation units

Variations in IA

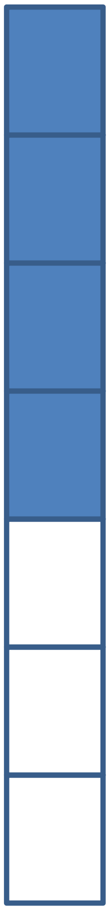
- 50 examples per word, 6 annotators
- For each pair of words, one has much better agreement
  - Adj > Noun > Verb
  - POS does not account for all the variation

POS	Word	Alpha, 6-way	No. Senses	No. Used
Adj	long	0.67	9	8
	fair	0.35	10	5
Noun	work	0.54	7	7
	land	0.26	11	8
Verb	tell	0.42	8	8
	show	0.26	12	11

Negative vs. Positive Factors

- Negative factors
  - Same 6 annotators for all tasks, with same experience, same tool, same training
  - Few consistent differences across annotators (one has lower $\overline{IA}_2$): $\overline{IA}_2$ = 0.39; sd=0.04)
- Positive factors
  - Concrete versus abstract senses
  - More versus less specific contexts
  - Productive polysemy
  - Similarity of senses in sense inventory

Concreteness: *Long* (highest IA)



Objective vs. metaphorical measurement

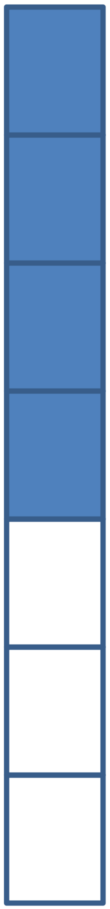
For 18 long months Michael could not find a job.

- Dimension = time; unit = months; 6 annotators

my editor . . . suggested a number of cuts to streamline what was already a long and involved chapter . . .

- Dimension = (*metaphor)* spatial extent; unit=words; 3 annotators
- Dimension = time; no unit of measurement; 2 annotators
- Dimension = unspecified (WN sense 9: *more than normal or necessary*); 1 annotator

Concreteness: *Work* (2nd highest IA)



Concrete: 2. product of effort; 3. occupation one is paid for; 5. unit of force; 6. place where one works; 7. artist's/writer's total output

. . . military service is so different from civilian work that most wage comparisons are extremely suspect.

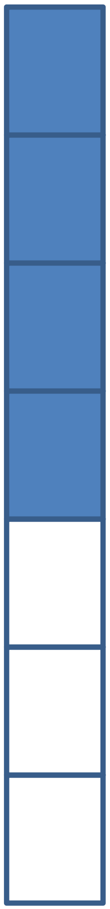
Sense 3: 6 annotators

versus abstract: 1. activity directed toward doing something; 4. applying the mind to learning a subject

The attorneys can search . . . based on criteria including the location where the work needs to be done and the field of law

Sense 1: 4 annotators; Sense 4: 1 annotator; No Selection: 1

Specific Contexts



Named Entities; descriptive NPs

Land (lower of middle pair): *India is exhilarating exhausting and infuriating, a land where you'll find the practicalities of daily life overlay the mysteries that popular myth attaches to India.*
Sense 5: territory occupied by a nation, 6 annotators

Show (higher of last pair): *. . . it should be pointed out that gene-regulatory regions, which . . . have been shown to underlie some QTLs, can be studied in the same way.*
Sense 2: prove, establish the validity of; 6 annotators

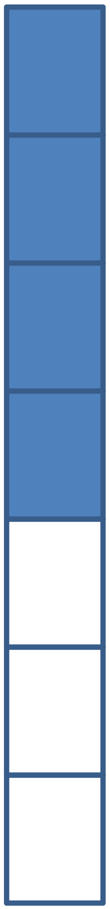
Productive Polysemy

Activity sense versus product of activity

A close friend is a plastic surgeon who did some minor OK semi-major facial work on me in the past.

Sense 1, activity directed toward making sth, 4 annotators;
Sense 2, product of effort, 2 annotators

Sense Similarity



Land: 11 senses

- Two property senses (1 and 7)
- Two soil/earth senses (2 and 4)
- Two geopolitical senses (3 and 5)

and uh we have five acres of land up against a hillside where I grew up and so we did have a garden about a one a half acre garden

Sense 4, solid part of the earth's surface, 1 annotator
Sense 1, location of real estate, 2 annotators
Sense 7, extensive landed property, 3 annotators

Measuring Sense Similarity

- Inter-sense Similarity Measures (ISM, Ide, 2006)
  - for all pairs of a words's *n* senses
  - lesk similarity (best of 6 measures in WordNet::Similarity package)
- Confusion Threshold

$$CT = \mu_A + 2\sigma_A$$

ISM Statistics

- IA inversely correlates with number of ISMs > CT, R= -0.78
- IA inversely correlates with std. dev. ISMs, R=-0.74

POS	Word	Max ISM	Mean ISM	Std. Dev.	> CT	IA
Adj	long	0.71	0.28	0.18	0	0.67
	fair	1.25	0.28	0.34	5	0.35
Noun	work	0.63	0.22	0.16	0	0.54
	land	1.44	0.17	0.29	3	0.26
Verb	tell	1.22	0.15	0.25	1	0.42
	show	1.38	0.18	0.27	3	0.26

Association Rules

Agrawal et al., 1993: mining frequent attributes (here--words, annotators, context descriptions) in a dataset D; where C expresses conditions on attributes of objects in D, find $C_1 \Rightarrow C_2$

$$Support(C) = \frac{freq(C)}{|D|}$$

$$Support(C_1 \Rightarrow C_2) = Support(C_1 \wedge C_2)$$

$$Confidence(C_1 \Rightarrow C_2) = \frac{Support(C_1 \wedge C_2)}{Support(C_1)}$$

Association Rules and Annotations

- Mine for association rules relating annotators
  - "Flatten" data to produce 2D tables
  - Look for high support rules
  - Bidirectional association rules, i.e., where there is support for a conjunct of attributes, and confidence for the LHS and for the RHS

$$A_i\_Word_m : Sense_x \Leftrightarrow A_j\_Word_m : Sense_y$$

$$A_i\_show.S2 \Leftrightarrow A_j\_show.S3$$

- Sense 2: establish the validity of something, as by an example, explanation, experiment
- Sense 3: provide evidence for

i	j	Supp. (%)	Conf. –i (%)	Conf.–j (%)
A1	A3	32	84.2	69.6
A5	A3	24	63.2	80.0
A4	A3	22	91.7	57.9
A4	A6	14	58.3	47.6
A4	A2	12	60.0	50.0
A5	A2	12	60.0	40.0

Conclusions

- Quantitative results showing that words differ regarding the ability of annotators to agree on word sense disambiguation in context
- Correlation between annotators' agreement on a set of words, and a measure of the sense similarity of the words' senses
- Qualitative results suggesting that words differ regarding their characteristic contexts of use
- Association rules indicate systematic patterns of agreement within the disagreement on senses

Future Work

- Represent relations among a word's senses
  - A list of senses is too restrictive
  - Alternative to *flattening* the sense inventory
- Relate measures of IA to association rules
- Use association rules to find support for the observations regarding specificity
  - Develop features to represent context specificity
  - Search for rules that relate senses to subsets of instances that are more specific