Semantics and Pragmatics of NLP
Lexical Semantics: Machine Learning

Alex Lascarides

School of Informatics
University of Edinburgh

Outline

1. What is Logical Metonymy?
2. Rule-Based Accounts
3. Some Shortcomings/Gaps
4. A probabilistic model for interpreting logical metonymies

What is Logical Metonymy?

Semantic type of a syntactic complement to a word differs from the semantic type of the argument in logical form:

(1) a. Mary finished the cigarette.
    b. Mary finished smoking the cigarette.

(2) a. Mary finished her beer.
    b. Mary finished drinking her beer.

(3) a. easy problem
    b. difficult language
    c. good cook

Things in Common

1. Additional meaning is predictable
   - The event that's finished/enjoyed/started is the *purpose* of the denotation of the noun
2. Interpretations can be rendered with a paraphrase

Major Challenges

- **Semi-productivity**
  ??*enjoy the tunnel*, ??*enjoy the door* etc.
- **Context-sensitivity**

  (4) My goat eats anything. He enjoyed your book

- **Ambiguity**
  fast scientist: publishes quickly, does experiments quickly
  researches quickly, persuades people quickly
  thinks quickly . . .
  are all highly plausible interpretations

We will tackle **ambiguity** and discuss **semi-productivity**.

Theoretical Accounts: Generative Lexicon

- Against sense enumeration;
- meaning of adjective/verb depends on noun;
- nouns have *qualia structures*:
  - This represents very simple world knowledge:
    what object is made up of;
    its purpose;
    how it was created.
- adjectives/verbs modify qualia for nouns.

Example (Simplified): enjoy the book

book: inherited qualia

$$\left[\begin{array}{l} \textbf{book} \\ \text{SEM} : \text{book}(y) \\ \text{QUALIA} : \left[\begin{array}{l} \text{CONST} : \textbf{pages} \\ \text{TELIC} : \textbf{read} \\ \text{AGENTIVE} : \textbf{write} \end{array}\right] \end{array}\right]$$

enjoy: inherited info; *begin*, *finish* etc.

$$\left[\begin{array}{l} \textbf{coercing} \\ \text{CAT SUBCAT} : \left\langle \left[\begin{array}{l} \textbf{np} \\ \text{SEM} : \boxed{\text{n}}\ [Q(y)] \\ \text{QUALIA TELIC} : \boxed{\text{P}} \end{array}\right] \right\rangle \\ \text{SEM} : [e][\text{enjoy}(e, x, e') \wedge \text{act-on-pred}/\boxed{\text{P}}(e', x, y) \wedge \boxed{\text{n}}\,] \end{array}\right]$$

enjoy the book:

$$\left[\begin{array}{l} \textbf{coercing} \\ \text{CAT SUBCAT} : \left\langle \left[\begin{array}{l} \textbf{np} \\ \text{SEM} : \boxed{\text{n}}\,\text{book}(y) \\ \text{QUALIA TELIC} : \boxed{\text{P}}\,\textbf{read} \end{array}\right] \right\rangle \\ \text{SEM} : [e][\text{enjoy}(e, x, e') \wedge /\boxed{\text{P}}\,\text{read}(e', x, y) \wedge \boxed{\text{n}}\,\text{book}(y)] \end{array}\right]$$

Gaps

- They assume noun classes have one (perhaps default) telic role.
- So don't investigate relative degree of ambiguity of various cases of metonymy (e.g., *fast scientist* vs. *fast programmer*)
- Or degree of variation
  - for an N with different verbs:
    begin the house (agentive) vs. *enjoy the house* (telic)
  - for verb with different Ns:
    begin the tunnel (agentive) vs. *begin the book* (telic)

Manually constructing a lexicon with very rich semantic information
so as to account for regular polysemy is impractical anyway.

An Alternative: Machine Learning

- Can the meanings of metonymies (and other forms of regular polysemy) be acquired automatically from corpora?
- Can we constrain the number of interpretations by providing a *ranking* on the set of meanings?

Finding Answers Empirically:

- Provide a probabilistic model
- Model parameters: exploit *meaning paraphrases*
  - co-occurrences of nouns, verbs and metonymic verbs/adjectives in the corpus
- evaluate results against human judgements

The Model: Metonymic Verbs **enjoy book**

Find *e* which maximises the probability
$P(e, book, enjoy)$ of seeing "enjoy e-ing book".

The Equations: *e*=event; *v*=metonymic verb; *n*=noun

(5) $$P(e, n, v) = P(e) \cdot P(v|e) \cdot P(n|e, v)$$

Estimating the probabilities:

$$P(e) = \frac{f(e)}{\sum_i f(e_i)}$$

$$P(v|e) = \frac{f(v,e)}{f(e)}$$

$$P(n|e, v) = \frac{f(n,e,v)}{f(e,v)}$$

Sparse Data for $f(n, e, v)$! BNC

enjoy movie:

(6) I've always enjoyed watching spy movies.

begin speech:

(7)
a. Churchill. . ., as he had begun to make public public speeches. . .
b. Liam sprang on to a table, raised a glass and began to declaim a speech.
c. The Prince. . .he began to make increasingly serious and significant speeches.
d. For the first time in ten years I'm gonna begin delivering a speech.

Grice predicts sparse data problem!

Solving the Estimation Problem

Assume that the likelihood of seeing n as object of e is independent of whether e is the complement of v

So:

$$P(n|e,v) \approx P(n|e)$$

$$P(n|e) = \frac{f(n,e)}{f(e)}$$

$$P(e,n,v) \approx \frac{f(v,e) \cdot f(n,e)}{\sum_i f(e_i) \cdot f(e)}$$

Example: *enjoy the film*

$f(enjoy, e)$		$f(film, e)$	
play	44	make	176
watch	42	be	154
work with	35	see	89
read	34	watch	65
make	27	show	42
see	24	produce	29
meet	23	have	24
go to	22	use	21
use	17	do	20
take	15	get	18

So events associated with enjoying films are:

- watching, making, seeing, using

Model is ignorant of context;
determines most dominant meanings in the corpus.

How We Estimated Parameters

Corpus: POS-tagged, lemmatised BNC (100 million words), parsed by Cass (Abney, 1996)

Verb-argument tuples: $f(e, n)$

- Can extract verb-SUBJ and verb-OBJ (need just verb-OBJ here)
- Errors make filtering necessary:

  e.g.: discard Vs that only occur once; particle Vs (*come off heroin*) retained only if particle is adjacent to N

Metonymic verb and its complement: $f(v, e)$

- Metonymic verb *v* followed by VBG (progressive) or To0 (infinitival)

Examples for $f(e, v)$

(8)
a. I am going to start writing a book
start write
b. I've really enjoyed working with you
enjoy work
c. The phones began ringing off the hook
begin ring

(9)
a. I had started to write a love-story
start write
b. She started to cook with simplicity
start cook
c. The suspect attempted to run off
attempt run off

Paraphrases from the Literature Verspoor 1997, Pustejovsky 1991, 1995

John began the book	$\longrightarrow$ reading/writing
John began the sandwich	$\longrightarrow$ eating/making
John began the beer	$\longrightarrow$ drinking
John began the cigarette	$\longrightarrow$ smoking
John began the coffee	$\longrightarrow$ drinking
John began the speech	$\longrightarrow$ writing
John began the lesson	$\longrightarrow$ writing/taking
John began the solo	$\longrightarrow$ playing
John began the song	$\longrightarrow$ singing
John began the story	$\longrightarrow$ telling
John enjoyed the symphony	$\longrightarrow$ listening to
John enjoyed the film	$\longrightarrow$ watching
Mary enjoyed the movie	$\longrightarrow$ watching
John quite enjoys his morning coffee	$\longrightarrow$ drinking
Bill enjoyed Steven King's last book	$\longrightarrow$ reading
Mary likes movies	$\longrightarrow$ to watch
Harry wants another cigarette	$\longrightarrow$ to smoke
John wants a beer	$\longrightarrow$ to drink
Mary wants a job	$\longrightarrow$ to have

Model-Derived Paraphrases (ranked by likelihood)

underline: value agrees with claims about meaning in the literature

$P(e, n, v)$	e_1	e_2	e_3	e_4	e_5
$P(e, book, begin)$	<u>read</u>	<u>write</u>	appear in	publish	leaf through
$P(e, book, enjoy)$	<u>read</u>	write	browse through	look through	publish
$P(e, sandwich, begin)$	bite into	<u>eat</u>	munch	unpack	<u>make</u>
$P(e, beer, begin)$	<u>drink</u>	pour	sip	crack	sell
$P(e, beer, want)$	<u>drink</u>	buy	sell	weep into	get
$P(e, cigarette, begin)$	<u>smoke</u>	roll	light	take	twitch
$P(e, cigarette, want)$	<u>smoke</u>	take	light	put	buy
$P(e, coffee, begin)$	pour	<u>drink</u>	sip	make	stir
$P(e, coffee, enjoy)$	browse through	<u>drink</u>	make	take	go for
$P(e, speech, begin)$	make	read	recite	disclaim	slur
$P(e, lesson, begin)$	learn	teach	<u>take</u>	read	recite
$P(e, solo, begin)$	<u>play</u>	sing	tun	hem	work through
$P(e, song, begin)$	<u>sing</u>	rehearse	write	hum	play
$P(e, story, begin)$	<u>tell</u>	write	read	re-tell	recount
$P(e, symphony, enjoy)$	play	<u>listen to</u>	write	hear	serve
$P(e, film, enjoy)$	<u>watch</u>	make	see	go to	work with
$P(e, movie, enjoy)$	<u>watch</u>	go to	make	see	eat in
$P(e, movie, like)$	see	go to	<u>watch</u>	make	film
$P(e, job, want)$	get	lose	take	make	create

Evaluation: Comparison Against Human Judgements

- Randomly select 12 metonymic verbs
  - *attempt*, *begin*, *enjoy*, *expect*, *postpone*, *prefer*, *resist*, *start*, *survive*, *try*, *want*
  - frequency in BNC between 10.9 per million and 905.3 per million
- Randomly select 5 nouns which are attested as objects to these verbs.
- Use model to derive meanings of the resulting 60 combinations.
- Divide set of generated meanings into three probability bands
  - High, Medium, Low (equal size)

Example Stimuli

60 V-N pairs $\times$ 3 bands = 180 stimuli

Michael attempted a smile	Michael attempted to give a smile
Michael attempted a smile	Michael attempted to rehearse a smile
Michael attempted a smile	Michael attempted to look at a smile
Jean enjoyed the concert	Jean enjoyed listening to the concert
Jean enjoyed the concert	Jean enjoyed throwing the concert
Jean enjoyed the concert	Jean enjoyed making the concert

The Experimental Procedure

Magnitude estimation of linguistic judgements (ME):

- Subjects see a modulus item and assign it an arbitrary number; other stimuli are rated proportional to the modulus;
- ME yields highly robust and maximally delicate judgement data;

Subjects

- experiment administered over the Web using WebExp (Keller *et al.* 2001);
- 60 subjects, each subject saw 90 stimuli;
- judge meaning paraphrases.

Results

- Performed analysis of variance (ANOVA) to test whether paraphrases with high probs are perceived to be better than those with low probs.
- Probability bands yield desired differences!
  - Post-hoc Tukey tests show that the differences between all pairs of conditions were significant; $\alpha = 0.1$
  - Comparison between our model and human judgements yields a Pearson correlation coeffecient of 0.64 ($p < 0.01$, $N = 174^2$).

So model correlates reliably with human judgements.

What's the Upper Bound? Inter-Subject Agreement

Correlations computed via 'leave-one-out cross-validation':

- Start with *m* subjects
- Divide subjects into 2 groups of size $m - 1$ and 1
- Correlate mean ratings of the responses of first subject group with that of the latter subject.
- Repeat *m* times.

Result

- Gives average inter-subject agreement of .74
- So model doing OK (scoring .64), given this upper bound.

The Lower Bound

- Naive baseline model is to just take verb-noun co-occurrence data into account.
  - Don't use $f(e, v)$ to estimate $P(e, v, n)$
- Like assuming that the metonymic verb is semantically empty.
  - Cf. *begin the house* vs. *enjoy the house*

Results:

- Naive model has Pearson correlation coefficient of 0.42
- Difference with our model is statistically significant ($p < 0.05$).
- And correlation between naive model and our model is relatively low; 0.46

Semi-Productivity
Examples from Verspoor 1997, Pustejovsky 1995, Lascarides and Copestake 1998

John began a chair	$\longrightarrow$	sitting in/on
John began the tunnel	$\longrightarrow$	driving through
John began a keyboard	$\longrightarrow$	typing on
John began the trees	$\longrightarrow$	growing/planting/watering
John began the highway	$\longrightarrow$	driving on
John began the film	$\longrightarrow$	watching
John began the nails	$\longrightarrow$	hammering in
John began the door	$\longrightarrow$	opening/walking through
John began the dictionary	$\longrightarrow$	reading
John enjoyed the path	$\longrightarrow$	hiking
Mary began the rock	$\longrightarrow$	???

What Grice (1975) Predicts

Maxim of Manner

- When metonymy is acceptable, you are relatively more likely to use metonymic construction than its paraphrase.
- The opposite is true when metonymy is unacceptable.

So does our model comply with this prediction?

Model-derived Paraphrases for ‘Bad’ Examples

$P(e, n, v)$	e_1	e_2	e_3	e_4	e_5
$P(e, \mathit{chair}, \mathit{begin})$	fidget on	sink into	rise from	take	move
$P(e, \mathit{tunnel}, \mathit{begin})$	waddle through	dig	walk towards	emerge from	build
$P(e, \mathit{keyboard}, \mathit{begin})$	use	play	operate	assemble	tune
$P(e, \mathit{tree}, \mathit{begin})$	climb	climb towards	sing in	run towards	grow
$P(e, \mathit{highway}, \mathit{begin})$	obstruct	regain	build	use	detach
$P(e, \mathit{film}, \mathit{begin})$	make	appear in	show	develop	work on
$P(e, \mathit{nail}, \mathit{begin})$	bite	dig	chew	dig in	drive
$P(e, \mathit{door}, \mathit{begin})$	open	walk towards	knock on	close	move towards
$P(e, \mathit{dictionary}, \mathit{begin})$	compile	flick through	use	publish	advance
$P(e, \mathit{path}, \mathit{enjoy})$	walk	follow	ride	take	travel on
$P(e, \mathit{rock}, \mathit{begin})$	crunch across	climb	run towards	percolate through	dissolve

BNC Frequencies

Odd	$f(v,n)$	$f(v,e,n)$
begin chair	0	9
begin tunnel	0	4
begin keyboard	0	0
begin tree	1	13
begin highway	0	2
begin film	0	7
begin nail	0	4
begin door	0	18
begin dictionary	0	3
begin rock	0	17
enjoy path	1	2

Well-formed	$f(v,n)$	$f(v,e,n)$
begin book	35	17
begin sandwich	4	0
begin beer	2	1
begin cigarette	0	0
begin coffee	0	4
begin speech	21	4
begin solo	1	1
begin song	19	8
begin story	31	15
enjoy symphony	34	30
enjoy film	16	5
enjoy movie	5	1
enjoy coffee	8	1
enjoy book	23	9
like movie	18	3
want cigarette	18	3
want beer	15	8
want job	116	60

- Suggestive, but not conclusive: *begin keyboard* vs. *begin cigarette*
- This is a model of interpretation, not of grammaticality

Other Case Studies

Metonymic Adjectives *fast scientist*, *good soup*...

- Similar model, with additional SUBJ vs. OBJ parameter:
  OBJ: *good soup* is soup that tastes good
  SUBJ: *good programmer* is programmer who programs well
- Performance of model very similar to metonymic verbs

Compound Nouns *hospital admission*, *patient admission*...

- Experiments performed on Medline. Exploit
  - Occurrences of grammatical relations of modifier to corresponding verb in the corpus;
  - WordNet and UMLS (to smooth over sparse data)
- Achieve 72.5% accuracy

Conclusions

- Regular polysemy is pervasive; it needs to be modelled.
- Manual modelling is undoable
- Machine learning can help, because
- you can exploit meaning paraphrases and surface cues, and this makes unsupervised training feasible.
- Probabilistic models *rank* interpretations
- and the rankings correlate reliably with human judgements about meaning.