Lexical Semantics-Syntax and Future Work Ideas

Diana McCarthy

Lexical Computing Ltd.

University of Melbourne, July 2011

Outline

1. Lexical Semantics-Syntax
   - Background: Acquistion of Predicate Argument Structure
   - SCF Acquisition
   - Selectional Preference Acquisition
   - Selectional Preferences and Diathesis Alternation Detection
   - Selectional Preferences and Compositionality Detection
2. Future Projects

Subcategorisation

She	*loaded*	*the bag*	*with chicken*
`NP`	`V`	`NP`	`PP`

Subcategorisation

She	*loaded*	*the bag*	*with chicken*
`NP`	`V`	`NP`	`PP_with`

Subcategorisation

She	*loaded*	*the bag*	*with chicken*
`NP`	`V`	`NP`	`PP_with`
He	*loaded*	*chicken*	*into the bag*
`NP`	`V`	`NP`	`PP_into`

Selectional Preferences

She	*loaded*	*the bag*	*with chicken*
NP	V	NP	PP

Selectional Preferences

She	*loaded*	*the bag*	*with chicken*
NP	V	NP	PP
	load		with ?

explosive ammunition scrap fish supplies brick fat food water . . .

Selectional Preferences

She	*loaded*	*the bag*	*with chicken*
NP	V	NP	PP
	load		with ?
	load	NP	with ?

explosive ammunition scrap fish supplies brick fat food water . . .

Semantic Role Labelling

She	*loaded*	*the bag*	*with chicken*
`NP`	`V`	`NP`	`PP`

FrameNet style labels [Ruppenhofer et al., 2010]

agent	predicate	object / goal	theme

Propbank style labels [Palmer et al., 2005]

Arg0	predicate	Arg2	Arg1

SRL identify the arguments of a given verb and assign them semantic labels describing the roles they fulfil

Diathesis Alternations

She	*loaded*	*the bag*	*with chicken*
She	*loaded*	*chicken*	*into the bag*

Lexical Information: Verb Class

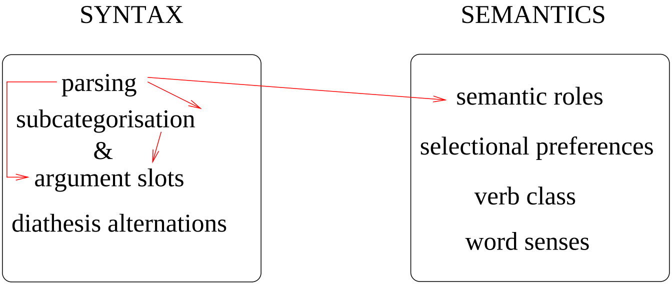
Pour Verbs: *dribble, drop, pour, slop, slosh, spew, spill, spurt*

Causative Alternation:
I pour water into the pot $\leftrightarrow$ *Water poured into the pot*
*Locative Alternation:
I pour water into the pot $\leftrightarrow$ **I poured the pot with water*
*Conative Alternation:
I pour water into the pot $\leftrightarrow$ **I poured at water into the pot*

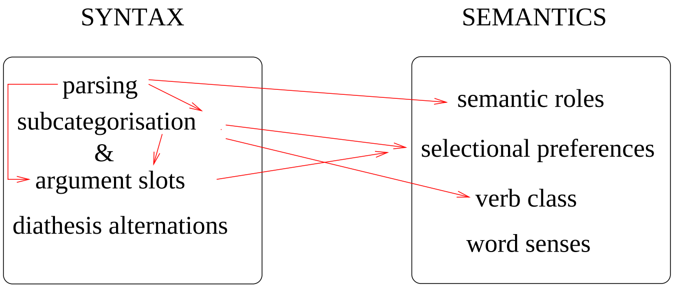
Lexical Acquisition Dependencies

SYNTAX	SEMANTICS
parsing subcategorisation & argument slots diathesis alternations	semantic roles selectional preferences verb class word senses

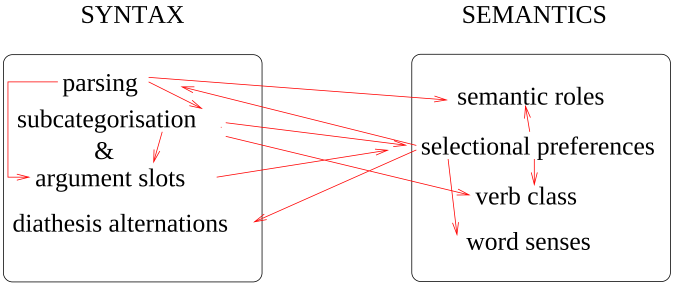
Lexical Acquisition Dependencies



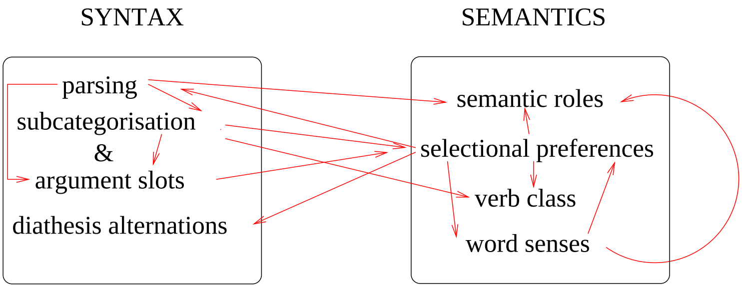
Lexical Acquisition Dependencies



Lexical Acquisition Dependencies



Lexical Acquisition Dependencies



Subcategorisation Acquisition

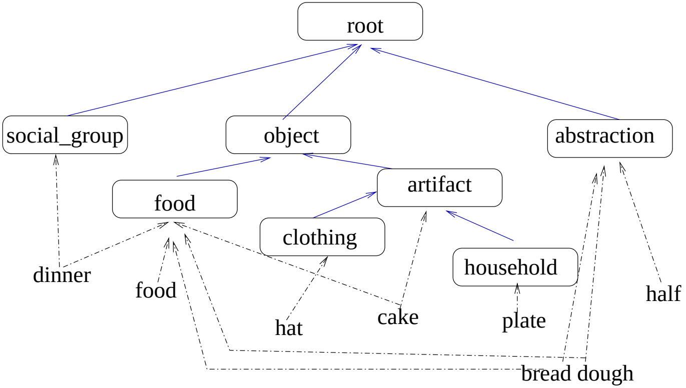
- unambiguous instances [Brent, 1991]
- parsing [Briscoe and Carroll, 1997]
- statistical filtering [Briscoe and Carroll, 1997, Korhonen et al., 2000]
- use of semantic classes for generalising [Korhonen, 2002]
- use of WSD for SCF acquisition [Preiss and Korhonen, 2002]

Selectional Preference Acquisition

- use:
  - slots e.g. direct object [Resnik, 1993] or
  - slots in SCF [McCarthy, 2001]
- generalise argument heads with
  - WordNet [Resnik, 1993, Li and Abe, 1998]
  - distributional similarity [Erk, 2007, McCarthy et al., 2007]

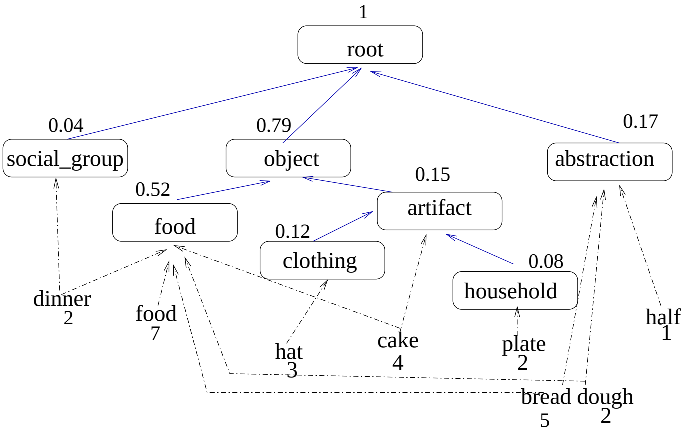
WordNet Based Models: example *eat*

food 7, bread 5, cake 4, hat 3, dinner 2, dough 2, plate 2, half 1



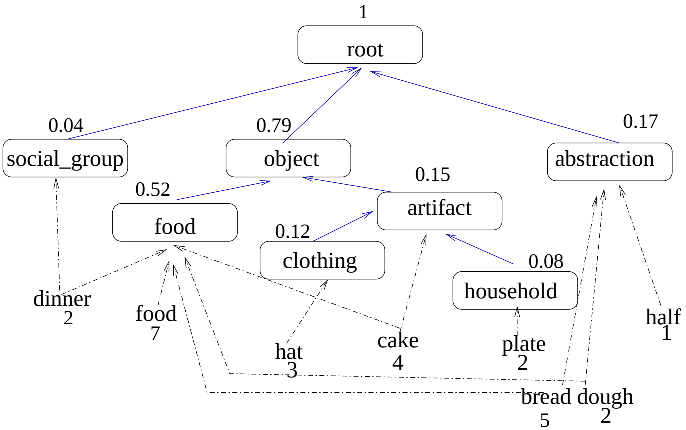
WordNet Based Models: example *eat*

food 7, bread 5, cake 4, hat 3, dinner 2, dough 2, plate 2, half 1



WordNet Based Models: example *eat*

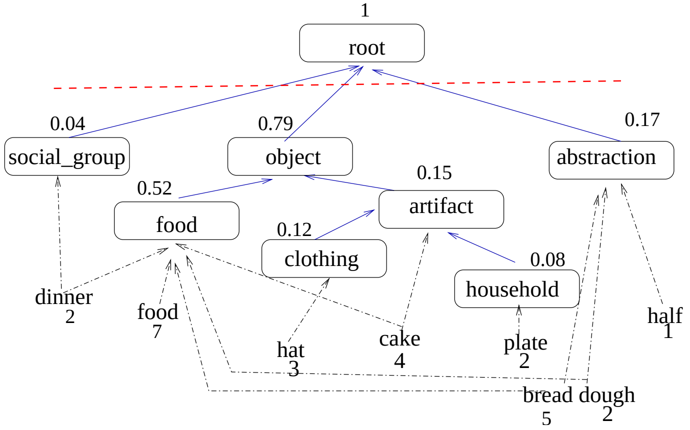
food 7, bread 5, cake 4, hat 3, dinner 2, dough 2, plate 2, half 1



noise from polysemous words, multiwords and other sources

WordNet Based Models: example *eat*

food 7, bread 5, cake 4, hat 3, dinner 2, dough 2, plate 2, half 1



Use frequency to find classes for representing preference and calculate probability distribution over these classes

Distributional Models

bread: loaf 0.195, cheese 0.179, cake 0.169, potato 0.158, butter 0.155, meat 0.153, toast 0.148, flour 0.143, bean 0.139, vegetable 0.138
van: truck 0.230, lorry 0.229, car 0.222, vehicle 0.196, bus 0.191, taxi 0.172, train 0.160, tractor 0.150, boat 0.148, cab 0.147

Distributional Models

bread: loaf 0.195, cheese 0.179, cake 0.169, potato 0.158, butter 0.155, meat 0.153, toast 0.148, flour 0.143, bean 0.139, vegetable 0.138

van: truck 0.230, lorry 0.229, car 0.222, vehicle 0.196, bus 0.191, taxi 0.172, train 0.160, tractor 0.150, boat 0.148, cab 0.147

use these directly [Erk, 2007]

Distributional Models

bread: loaf 0.195, cheese 0.179, cake 0.169, potato 0.158, butter 0.155, meat 0.153, toast 0.148, flour 0.143, bean 0.139, vegetable 0.138

van: truck 0.230, lorry 0.229, car 0.222, vehicle 0.196, bus 0.191, taxi 0.172, train 0.160, tractor 0.150, boat 0.148, cab 0.147

use these directly [Erk, 2007]

or build prototypical classes [McCarthy et al., 2007]

example: object slot of *park*

class ($p(c)$)	disambiguated objects (freq)
van (0.86)	car (174) van (11) vehicle (8) ...
backside (0.02)	backside (2) bum (1) butt (1) ...

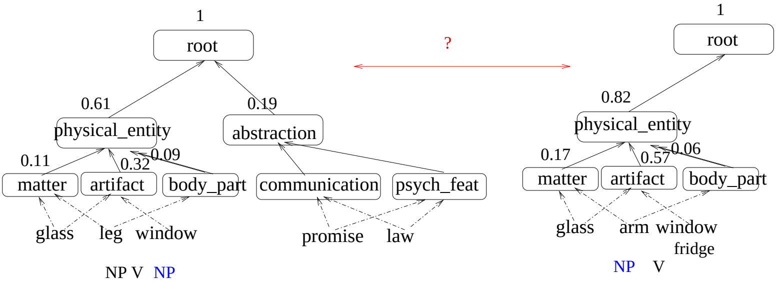
Verb Class Acquistion

- decision trees using syntactic and semantic features [Merlo and Stevenson, 2001]
- clustering SCF [Schulte im Walde, 2006]
- clustering SCF and selectional preferences [Sun and Korhonen, 2009]

Diathesis Alternations

- alternate ways in which arguments are expressed
  e.g. the causative alternation
  the boy broke the window $\leftrightarrow$ *the window broke*
- link between syntax and lexical semantics
- uses in NLP:
  - classification, prediction, recovery of predicate-argument structure,
  - subcategorization and selectional preference acquisition,
  - generation

Diathesis Alternation Detection: example *break*



[McCarthy, 2000, McCarthy, 2001]: scope

- Role Switching Alternations (RSAs): Where a particular argument type switches to a different grammatical slot in the alternating variants.
- We focus on RSAs involving NPs and PPs

e.g. <u>causative alternation</u>
the boy broke the window ↔ *the window broke*
obj of transitive ↔ subj of intransitive
e.g. <u>conative</u>
the boy pulled at the rope ↔ *the boy pulled the rope*
NP in PP ↔ obj of transitive

[McCarthy, 2000, McCarthy, 2001]: scope

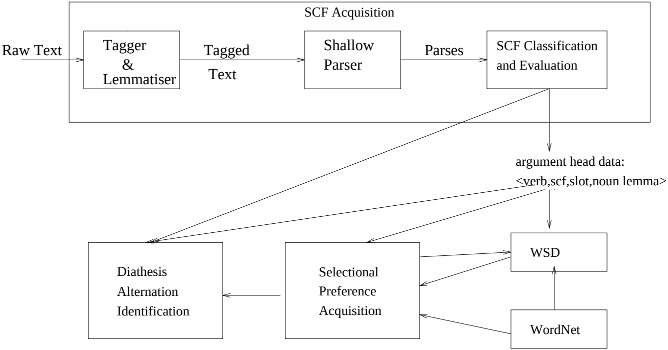
... and not:

- unexpressed object:
  the girl ate the pizza $\leftrightarrow$ *the girl ate*
- those without detail at phrase level e.g.
  I confused Maria with Anna $\leftrightarrow$ *I confused Maria and Anna*

Approach

- acquire subcategorization frame (SCF) information
- find candidates for a given alternation
- acquire selectional preferences at target slots
- use similarity of data at target slots
- e.g. causative: direct object of transitive and subject of intransitive

System Overview



Preprocessing and Parser

Preprocessor

- tokeniser e.g. numerical expressions, sentence boundaries, punctuation and abbreviations
- HMM POS tagger (sign_VV0)
- lemmatiser (e.g. doctor+s)

Parser

- unification based shallow grammar
- returns partial parses
- disambiguation - context sensitive LR Parser

SCF Acquisition

Briscoe and Carroll (1997)

1. patternset extractor - extracts SCF patterns including head lemmas of constituents
2. pattern classifier - assigns patterns to SCF classes
3. 161 classes (superset of those in ANLT and COMLEX)
4. patternset evaluator - binomial filter

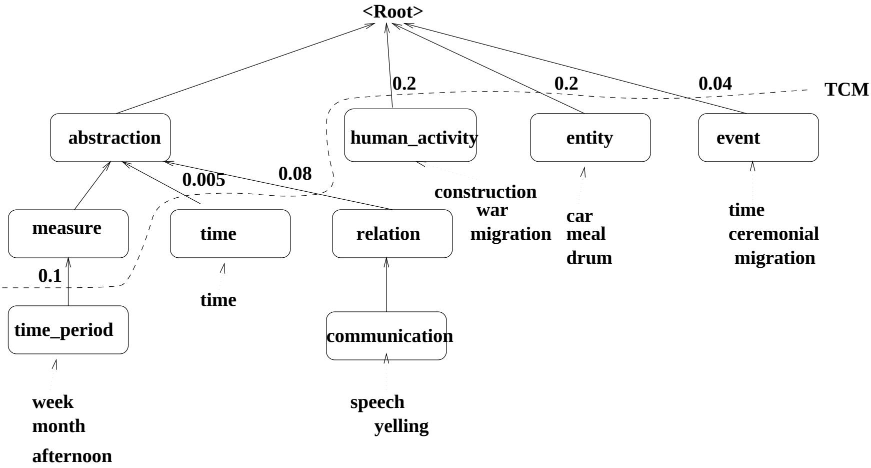
SCF lexicon entry for *bake* transitive class

```
#S(EPATTERN :TARGET |bake| :SUBCAT (VSUBCAT NP)
            :CLASSES ((24 51 161) 5293) :RELIABILITY 1.0
            :FREQSCORE 0.0 :FREQCNT 30
            :TLTL
            (VVG VV0 VV0 VV0 VVD VV0 VV0 VV0 VV0 VV0 VV0 VVG VVG VVG
             VVD VV0 VV0 VV0 VV0 VV0 VV0 VV0 VV0 VV0 VV0 VV0 VV0 VV0
             VV0 VV0)
            :SLTL
            (((|she| PPHS1)) ((|woman| NN1)) ((|i| PN1))
             ((|they| PPHS2)) ((|you| PPY)) ((|you| PPY))
             ((|society| NN)) ((|teaspoon| NN2)) ((|it| PPH1))
             ((|mother-in-law| NN1)) ((|you| PPY)) ((|you| PPY))
            ...
            :OLT1L
            (((|scone| NN2)) ((|cake| NN2)) ((|cake| NN1))
             ((|them| PPHO2)) ((|bread| NN1)) ((|cake| NN1))
             ((|anything| PN1)) ((|cake| NN2)) ((|potato| NN2))
            ...
            :OLT2L NIL :OLT3L NIL :LRL 0)
```

Selectional Preference Acquisition

- input from SCF lexicon
- preferences as Tree Cut Models (TCM)s Li and Abe (1995,1998)
- cuts across the WordNet noun hyponym hierarchy with associated probability distribution ($p(c|v)$)
- all word senses fall at or beneath a class on the cut

TCM for the object slot of the transitive of *start*



Selectional Preference Acquisition

- input: tuples <verb, noun, GR-SCF>
- output: set of WordNet classes across noun hyponym hierarchy with conditional probability distribution for each 'context'
- where context is given by verb and GR-SCF
- frequency data is used to populate WordNet noun classes
- frequency of superordinate classes includes that of hyponyms

- The probability distribution associated with a TCM:

$$\sum_{c \in \mathrm{TCM}} p(c|verb\,\mathrm{GR\text{-}SCF}) = 1$$

- Minimum description length (MDL) principle is used to obtain appropriate level of generalisation [Li and Abe, 1998, Li and Abe, 1995]
- MDL Principle [Rissanen, 1978]:
- best model is that which minimises the description length (DL), in bits, of the model and the data when encoded in the model.

$$\begin{aligned} \mathrm{DL} = & \quad \frac{k}{2} \times \log|S| & -\sum_{n \in S} \log p(n) \\ = & \quad model\,\mathrm{DL} & data\,\mathrm{DL} \end{aligned}$$

2 Methods + baseline

We expect similarity of argument heads at target slots e.g. causative: `[np1 v np2]` $\leftrightarrow$ `[np2 v ]`

- **MDL** compare encoding costs
- **Similarity** compare similarity of TCMs at target slots
- **Lemma-based** (baseline) compare overlap of argument heads

MDL Method

To determine how homogenous the data is we use the **cost** of encoding the data in the preference models :

- is it cheaper to combine the data in one TCM or not?
- this assumes implicit threshold at cost of the two separate models

MDL Causative detection for the verb *begin*

SCF : slot	object of transitive <->	subject of intransitive	combined = object of transitive + subject of intransitive
sample of data at slot	project celebration ... <->	holiday meeting ...	project celebration holiday meeting ...
Freq in WN	452	785	1237
Cost of TCM	7250.08	11729.05	18978.43
	(18979.13)		
part of the TCM	0.026 event 0.29 human action 0.02 location	0.04 event 0.22 human action 0.01 location	0.035 event 0.24 human action 0.01 location

Similarity-based Method

- We tried a number of scores defined for zero values
- results shown here use α-skew divergence [Lee, 1999] :

$$\alpha \mathrm{sd}(p1(x), p2(x)) = D(p2(x)||((\alpha \times p1(x)) + ((1-\alpha) \times p2(x))))$$

where

$$D(p2(x)||p1(x)) = \sum_x p2(x) \times \log \frac{p2(x)}{p1(x)}$$

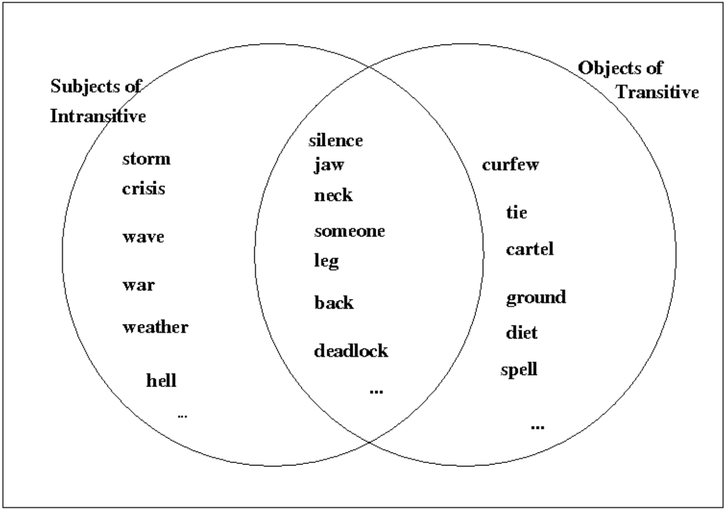
- we obtained similar results using
  - Euclidian Distance
  - L_1 norm
  - cosine

Lemma Overlap (LO) Baseline

$$\textsc{lo}(\mathrm{A},\ \mathrm{B}) = \frac{|multiset\ intersection(A\ B)|}{|smallest\ set(A, B)|} \quad (1)$$

$0 <= \textsc{lo}(\mathrm{A},\ \mathrm{B}) <= 1$
e.g. $\mathrm{A} = \{person,\ person,\ person,\ child,\ man,\ speaker\}$
$\mathrm{B} = \{person,\ person,\ child,\ chair,\ collection\}$
$\mathrm{intersection(A\ B)} = \{person,\ person,\ child\}$ $\textsc{lo} = \frac{3}{5}$

Lemma causative detection for the verb *break*



Experiments

- lexicon 1, 10.8 M words of BNC, parsed with PCP
- lexicon 2 19.3 M words of BNC, parsed with probabilistic LR parser

Parser	Zero crossings (% sents.)	Mean crossings per sent.	Bracket recall (%)	Bracket precision (%)
LR	57.2	1.11	82.54	83.00
PCP	54.2	1.13	82.50	82.68

Alternations Used:

- those RSAs with relevant frames identified by SCF acquisition system
- those with significant agreement among human judges (Gerald Gazdar, John Carroll, Stephen Clark, Bill Keller)

For each alternation we required:

- roughly even split positives and negatives
- at least 3 verbs in each category,
- each verb with 10 or more classifiable argument heads
- only verbs with $>$ 75% agreement from human judges

Sparse data issues:

- problems of sparse data: low frequency verbs, with low frequency frames e.g. substance/source alternation: *heat radiates from the sun* ↔ *the sun radiates heat* *belch* (12), *bleed* (82), *bubble* (64), *dribble* (12), *drip* (73), *drool* (9), *emanate* (64), *exude* (30), *gush* (30), *leak* (84), *ooze* (41), *pour* (449), *puff* (30), *radiate* (55), *seep* (65), *shed* (125), *spew* (8), *spout* (5), *sprout* (43), *spurt* (14), *squirt*(4), *steam* (64), *stream* (64), *sweat* (67)
- some alternations have only a few verbs e.g. blame alternation *Ann blamed the mess on Jo* ↔ *Ann blamed Jo for the mess*

- Syntactic information was sufficient for some alternations:
  dative *award*, *give*, *hand*, *lend*, *offer*, *owe*
  benefactive *award*, *earn*, *give*
- we used causative and conative for the following experiments:

Lemma Overlap 1

lexicon 1 (10.8 M words)

causative: 54 positive 56 negative
Mann Whitney U test for significance - not significant z score 1.007 p = 0.16 by chance

conative: 4 positive 4 negative for both *on* and *at*

- *on* p = 0.17, not significant
- *at* p = 0.1 not significant
- *on* and *at* p = 0.03, significant at 5% level

MDL Experiment: Causative

Causative	accuracy	sample coverage	sample size
No Filtering	63%	100%	110
With filtering	77%	35%	39

filtering option to remove those with similar preferences at subject and object slots in transitive frame e.g. *help*

MDL Experiment: Conative

	accuracy	sample coverage	sample size
conative (on)	62%	100%	8
conative (at)	50%	100%	8

conative (at)... all positive

Relative Frame Frequencies

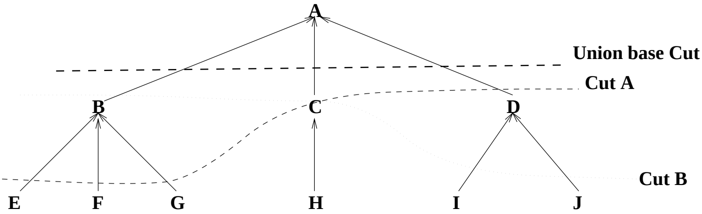
$$\text{average frequency ratio}_X = \frac{\sum_{v \in verbs} \frac{\text{freq}(v,\mathrm{SCF}1_X)}{\text{freq}(v,\mathrm{SCF}2_X)}}{|verbs|} \qquad (2)$$

Alternation	Average Frequency Ratio
Causative	1.16
Conative 'on'	28.99
Conative 'at'	32.72

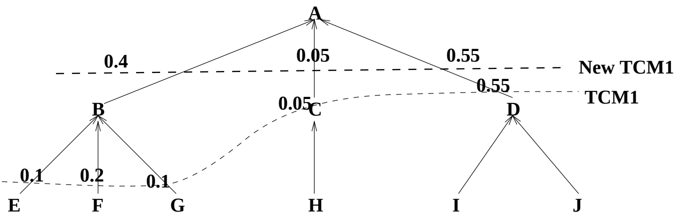
Similarity-based Experiments

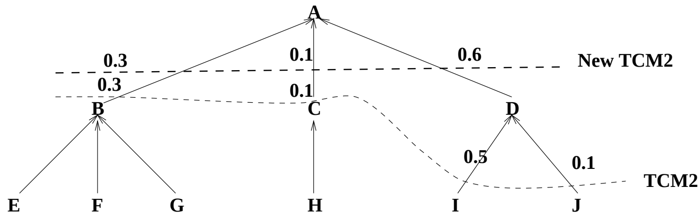
- similarity of probability distributions on 2 TCMs
- TCMs unified to common set of classes
- two common sets used:
  1. "root base cut" (RBC) - at 11 WordNet roots
  2. "union base cut" (UBC) -
- estimate probabilities on unified cut by summing estimates for hyponyms on original cut.

A Union Base Cut



New TCMs at the UBC





Results

- Mann Whitney U test to see if significant relationship between $\alpha\mathrm{SD}$ and participation
- mean and median thresholds for accuracy
- causative samples: 46 positive 53 negative
- conative samples: 6 positive 6 negative

Causative Identification with αSD

	Mann Whitney z	sign (p)	mean	median
root base cut				
αSD	-4.03	0.0003	71	63
union base cut				
	Mann Whitney z	sign (p)	mean	median
αSD	-4.3	0.00003	73	70

Lemma Overlap gave significant difference a 5% level and lower accuracy

Conative Identification

root base cut				
	M.W. sum	significance	mean	median
no WSD	26	0.02	67	83
WSD	22	0.002	83	83
union base cut				
no WSD	34	0.2	58	67
WSD	22	0.0022	83	83

Lemma Overlap was not significantly correlated with participation

Related Work

- Verb Classification
  - [Schulte im Walde, 1998, Schulte im Walde, 2006]
  - [Rooth et al., 1999]
  - [Sun and Korhonen, 2009]
  - [Stevenson and Merlo, 1999, Merlo and Stevenson, 2001]
- Identifying Participation
  - [Resnik, 1993]
  - [Lapata, 1999]
- Unsupervised Semantic Role Induction [Lang and Lapata, 2010]

Diathesis Acquistion Conclusions

- relationship between participation and similarity of preferences at alternating slots
- problems of sparse data
- WSD helps a little in cases of sparse data, but rather inconclusive
- no relation between polysemy (WordNet) and misclassification
- would help to combine alternations, target verb class and exploit correlations
- use unexpressed object [Resnik, 1993]
- use syntactic information at phrase level

Outline

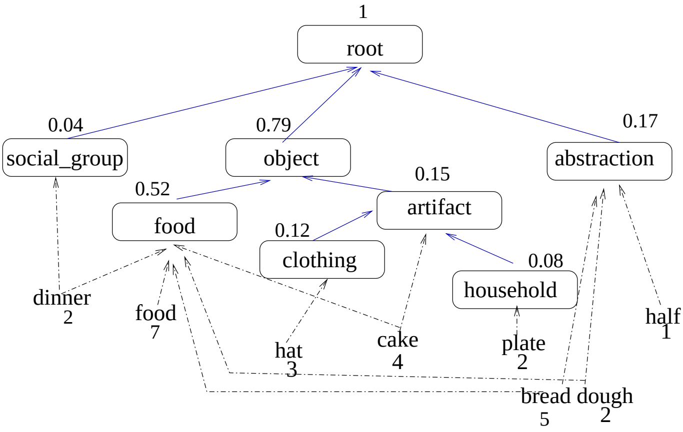
Selectional Preferences for Compositionality: verb-object

[McCarthy et al., 2007] e.g. *shoot the breeze* vs *shoot the gun*

- measure likelihood of verb object combinations
- does the verb have a preference for this sort of object?
- compare WordNet and distributional similarity preference models
- follows earlier pioneering work [Bannard, 2002] on selectional preferences and compositionality (hampered by overly general models so other approaches gave better results)

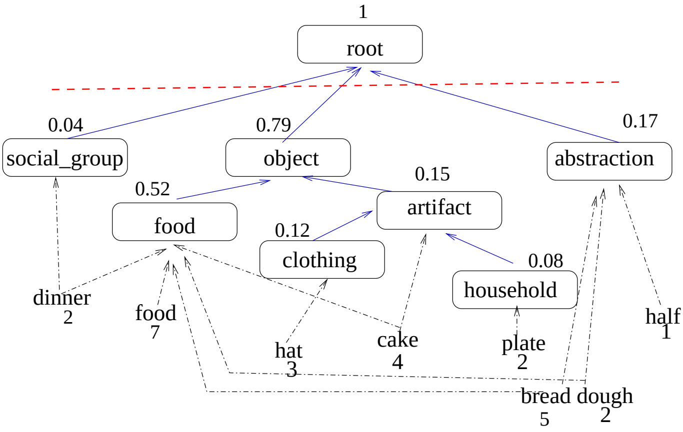
WordNet based models: example *eat*

food 7, bread 5, cake 4, hat 3, dinner 2, dough 2, plate 2, half 1 [Resnik, 1993]

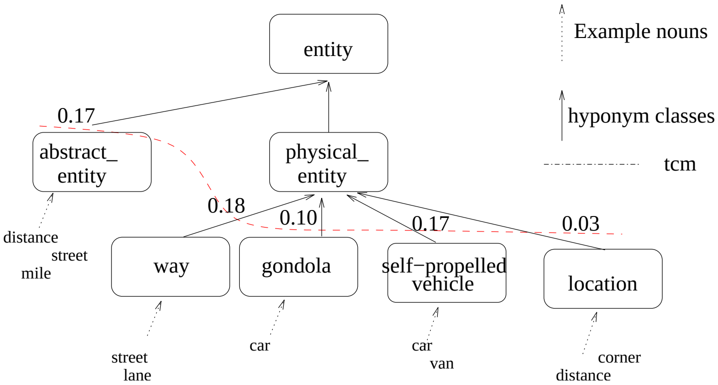


WordNet based TCMs: example *eat*

food 7, bread 5, cake 4, hat 3, dinner 2, dough 2, plate 2, half 1
[Li and Abe, 1998]



Portion of TCM for object of *park*



- Noise from *car* which occurs 174 times (out of 345).
- Contrast tokens (TCM) and type (WNproto) to obtain classes for representation, (tokens to estimate probability).

WNprotos

- prototypical classes, not coverage of all tokens
- disambiguate using "type ratio" of class C :

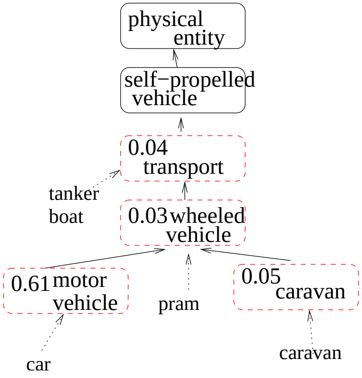
$$\frac{|noun\ types\ at\ or\ under\ C|}{|number\ of\ types\ in\ WordNet\ at\ or\ under\ C|}$$

- use types to determine classes for representing preference
  eat: food bread cake hat dinner dough plate half
- then use token frequency for associating probability distribution with these classes e.g. *eat*:
  food 7 bread 5 cake 4 hat 3 dinner 2 dough 2 plate 2 half 1

WNproto algorithm

- classes with at least 2 types
- a noun is disambiguated by whichever class it is at or under that has the largest type ratio
- only use nouns which can be disambiguated
- classes which have at least 2 disambiguated nouns are used in the model
- the disambiguated nouns are used to calculate probability over the classes in the model

WNproto for object slot of *park*



DSprotos

[McCarthy et al., 2007]

- nouns are listed in thesaurus built from parses of the BNC
  van: truck 0.230, lorry 0.229, car 0.222, vehicle 0.196, ...
  bread: loaf 0.195, cheese 0.179, cake 0.169, potato 0.158, ...
- each listing is considered a grouping or "class"
- classes with at least 2 types
- argument head nouns are disambiguated by whichever class has largest type ratio
- the noun frequency is used to calculate probability over the classes in the model

DSproto for object slot of *park*

class ($p(c)$)	disambiguated objects (freq)
van (0.86)	car (174) van (11) vehicle (8) ...
mile (0.05)	street (5) distance (4) mile (1) ...
yard (0.03)	corner (4) lane (3) door (1)
backside (0.02)	backside (2) bum (1) butt (1) ...

Evaluating Verb-Object Compositionality [Venkatapathy and Joshi, 2005]

- following [McCarthy et al., 2003] collecting judgements on a scale for 111 phrasal verbs (1-10, 3 subjects)
- Venkatapathy and Joshi collected graded judgments (1-6) on
- 2 fluent English speakers
- 765 verb objects
- agreement ρ=0.71

Evaluating DSprotos

[Venkatapathy and Joshi, 2005] data

method	ρ	$p <$ (one tailed)
selectional preferences		
TCM	0.090	0.0119
WNproto	0.223	0.00003
DSproto	0.398	0.00003
features from V&J		
frequency (f1)	0.141	0.00023
MI (f2)	0.274	0.00003
Lin [Lin, 1999] (f3)	0.139	0.00023
LSA2 (f7)	0.209	0.00003
combination		
f2,3,7	0.413	0.00003
f1,2,3,7	0.419	0.00003
DSproto f1,2,3,7	0.454	0.00003

Thanks to Advisors and Collaborators

- Gerald Gazdar
- John Carroll
- Ted Briscoe
- Anna Korhonen
- Genevieve Gorelle
- Sriram Venkatapathy
- Aravind Joshi

Outline

1. Lexical Semantics-Syntax
   - Background: Acquistion of Predicate Argument Structure
   - SCF Acquisition
   - Selectional Preference Acquisition
   - Selectional Preferences and Diathesis Alternation Detection
   - Selectional Preferences and Compositionality Detection
2. Future Projects

Future work ideas?

including

- WSD automatic detection of sense entropy integrated with probability from the context
- entropy detection alongside WSI
- unsupervised contextual clues (for predefined and induced inventories)
- contextual evidence with distributional similarity for LEXSUB
- filter antonyms from synonyms (alternatives to patterns) . . .

Future work ideas?

including

- analysis of CLLS and LEXSUB systems, what approach works <u>when</u>
- contrast CLLS and CLWSD data (clustering)
- evaluation of inventories and automatic clustering with Usim
- diathesis/ SRL induction / verb classification using distributional models

Bannard, C. (2002).
Statistical techniques for automatically inferring the semantics of verb-particle constructions.
Technical Report WP-2002-06, University of Edinburgh, School of Informatics.
http://lingo.stanford.edu/pubs/WP-2002-06.pdf.

Brent, M. R. (1991).
Automatic acquisition of subcategorization frames from untagged text.
In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, pages 209–214.

Briscoe, T. and Carroll, J. (1997).
Automatic extraction of subcategorization from corpora.
In *Proceedings of the Fifth Applied Natural Language Processing Conference*, pages 356–363.

Erk, K. (2007).
A simple, similarity-based model for selectional preferences.
In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 216–223, Prague, Czech Republic. Association for Computational Linguistics.

Korhonen, A. (2002).
Subcategorization Acquisition.
PhD thesis, University of Cambridge.

Korhonen, A., Gorrell, G., and McCarthy, D. (2000).
Statistical filtering and subcategorization frame acquisition.
In *Proceedings of the 2000 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora.*, pages 199–206, Hong Kong. ACL.

Lang, J. and Lapata, M. (2010).

Unsupervised induction of semantic roles.
In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 939–947, Los Angeles, California. Association for Computational Linguistics.

Lapata, M. (1999).
Acquiring lexical generalizations from corpora: a case study for diathesis alternations.
In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pages 397–404.

Lee, L. (1999).
Measures of distributional similarity.
In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pages 25–32.

Li, H. and Abe, N. (1995).

Generalizing case frames using a thesaurus and the MDL principle.
In *Proceedings of the International Conference on Recent Advances in Natural Language Processing*, pages 239–248, Bulgaria.

Li, H. and Abe, N. (1998).
Generalizing case frames using a thesaurus and the MDL principle.
Computational Linguistics, 24(2):217–244.

Lin, D. (1999).
Automatic identification of non-compositional phrases.
In *Proceedings of ACL-99*, pages 317–324, Univeristy of Maryland, College Park, Maryland.

McCarthy, D. (2000).

Using semantic preferences to identify verbal participation in role switching alternations.
In *Proceedings of the First Conference of the North American Chapter of the Association for Computational Linguistics. (NAACL)*, pages 256–263, Seattle,WA.

McCarthy, D. (2001).
Lexical Acquisition at the Syntax-Semantics Interface: diathesis alternations, subcategorization frames and selectional preferences.
PhD thesis, University of Sussex.

McCarthy, D., Keller, B., and Carroll, J. (2003).
Detecting a continuum of compositionality in phrasal verbs.
In *Proceedings of the ACL 03 Workshop: Multiword expressions: analysis, acquisition and treatment*, pages 73–80.

McCarthy, D., Venkatapathy, S., and Joshi, A. (2007).

Detecting compositionality of verb-object combinations using selectional preferences.
In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 369–379.

Merlo, P. and Stevenson, S. (2001).
Automatic verb classification based on statistical distribution of argument structure.
Computational Linguistics, 27(3):373–408.

Palmer, M., Gildea, D., and Kingsbury, P. (2005).
The proposition bank: A corpus annotated with semantic roles.
Computational Linguistics, 31(1):71–106.

Preiss, J. and Korhonen, A. (2002).
Improving subcategorization acquisition with WSD.

In *Proceedings of the ACL Workshop on Word Sense Disambiguation: Recent Successes and Future Directions*, Philadelphia, USA.

Resnik, P. (1993).
Selection and Information: A Class-Based Approach to Lexical Relationships.
PhD thesis, University of Pennsylvania.

Rissanen, J. (1978).
Modelling by shortest data description.
Automatica, 14:465–471.

Rooth, M., Riezler, S., Prescher, D., Carroll, G., and Beil, F. (1999).
Inducing a semantically annotated lexicon via EM-based clustering.

In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pages 104–111.

Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. R., and Scheffczyk, J. (2010).
FrameNet II: Extended theory and practice.
Technical report, International Computer Science Institute, Berkeley.
http://framenet.icsi.berkeley.edu/.

Schulte im Walde, S. (1998).
Automatic semantic classification of verbs according to their alternations.
In *Inducing Lexicons with the* EM *algorithm*, volume Aims Report 4(3), pages 55–74. IMS, University of Stuttgart.

Schulte im Walde, S. (2006).

Experiments on the automatic induction of german semantic verb classes.
Computational Linguistics, 32(2):159–194.

Stevenson, S. and Merlo, P. (1999).
Automatic verb classification using distributions of grammatical features.
In *Proceedings of the Nineth Conference of the European Chapter of the Association for Computational Linguistics*, pages 45–52.

Sun, L. and Korhonen, A. (2009).
Improving verb clustering with automatically acquired selectional preferences.
In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 638–647, Singapore. Association for Computational Linguistics.

Venkatapathy, S. and Joshi, A. K. (2005).
Measuring the relative compositionality of verb-noun (v-n) collocations by integrating features.
In *Proceedings of the joint conference on Human Language Technology and Empirical methods in Natural Language Processing*, pages 899–906, Vancouver, B.C., Canada.