New York, October 9, 2001

Relating Lexical Semantics to Statistical Generation

Maria Lapata

Department of Computational Linguistics
Saarland University

mlap@coli.uni-sb.de

UNIVERSITÄT
DES
SAARLANDES

Overview

- Lexical semantics:
  - symbolic generation systems
  - stochastic generation systems
- Levin's classification:
  - class ambiguity
  - frame ambiguity
  - class-word mapping
- Probabilistic modeling:
  - assumptions
  - parameter estimation
- Results
- Relevance for NLG
- Conclusions

Lexical Knowledge in Generation Systems

- important for abstract semantic representations;
- variety of interdependent information sources:
  - syntactic restrictions;
  - argument structure restrictions;
  - collocational restrictions;
  - mapping between words and concepts;
  - word usage and combination;
- requires linguistic sophistication;
- relatively flexible.

Application	Theory
RealPro (Lavoie and Rambow 1997)	Meaning Text Theory (Mel'čuk 1988)
MT (Palmer and Wu 1995)	Verb classes (Levin 1993)
MT (Dorr and Olsen 1996)	Verb classes (Levin 1993)
Generation (Stede 1998)	Verb classes (Levin 1993)
Generation (McDonald and Busa 1994)	Generative Lexicon (Pustejovsky 1995)
Generation (Viegas and Bouillon 1994)	Generative Lexicon (Pustejovsky 1995)

Lexical Knowledge Cont'd

For 99% of commercial systems text generation is done via templates (Reiter 1995).

- labour intensive;
- lexical semantic theories are not generation specific;
- different theories handle different lexical phenomena;
- underlying representations differ (Pustejovsky vs Levin).

Application	Theory
MT (Berger et al. 1996)	corpus
Generation (Bangalore and Rambow 2000)	LTAG, corpus
Generation (Ratnaparkhi 2000)	attribute-value pairs, corpus
Generation (Langkilde and Knight 1998)	semantic representation, corpus

Levin's Classification

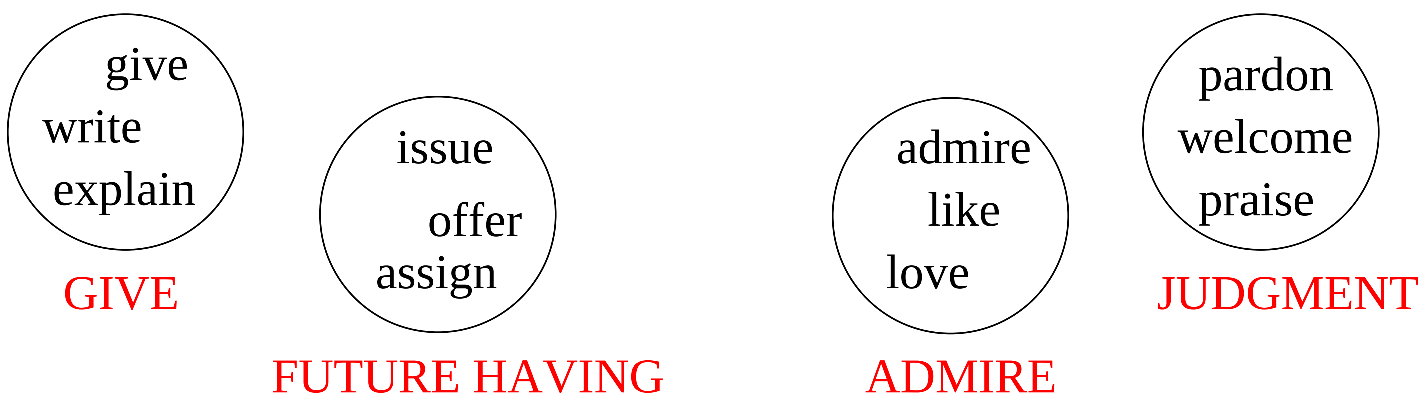
Diathesis alternations: changes in the realization of the argument structure of a verb (Levin 1993).

(1) a. John offered shares to Beth.
    b. John offered Beth shares.

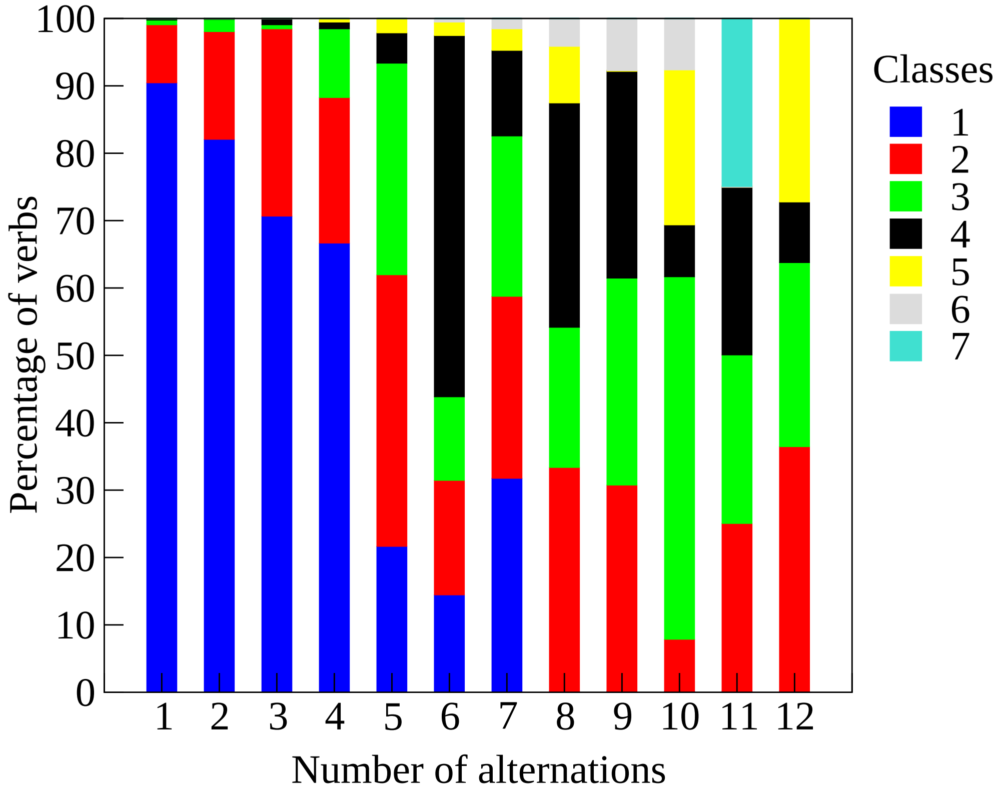
(2) a. I admire John for his courage.
    b. I admire John's courage.

Assumptions:

- argument structure correlates with meaning;
- verbs with same alternations form a semantically coherent class;
- verbs in same class share meaning components.



Class Ambiguity

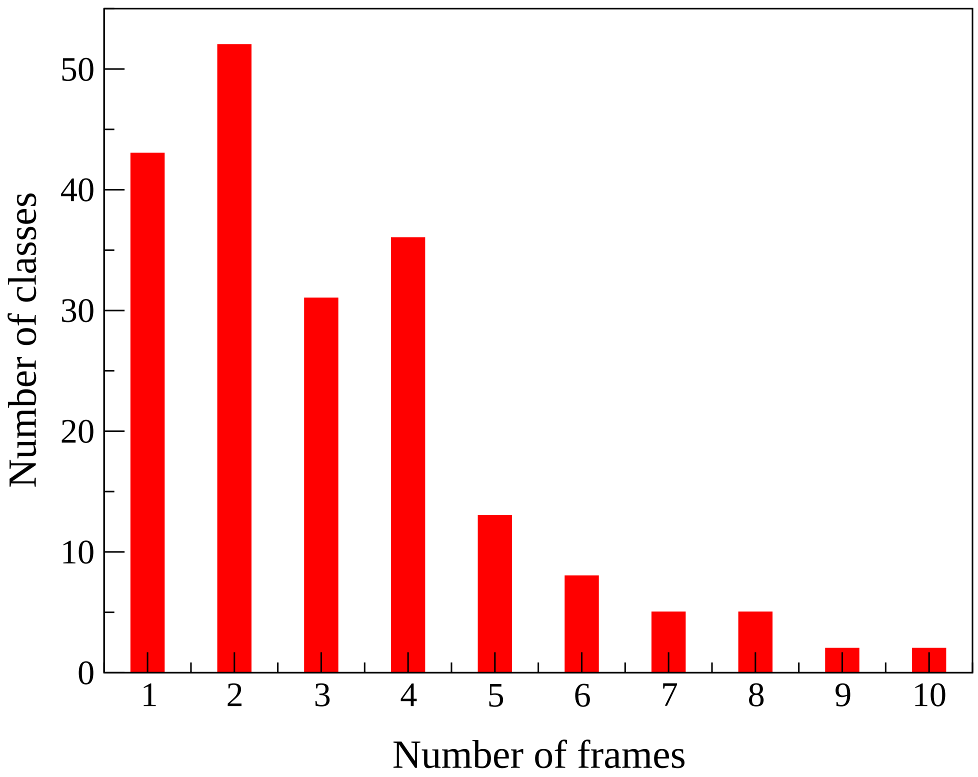


(3) a. A solicitor wrote him a letter at the airport. [PERFORMANCE, MSG. TRANSFER]
b. Un abogado escribió una letra para él en el aeropuerto. PERFORMANCE
c. Un abogado le escribió una letra en el aeropuerto. MSG. TRANSFER

Disambiguate class in a probabilistic framework; combine Levin with corpus data

(Lapata and Brew 1999).

Frame Ambiguity

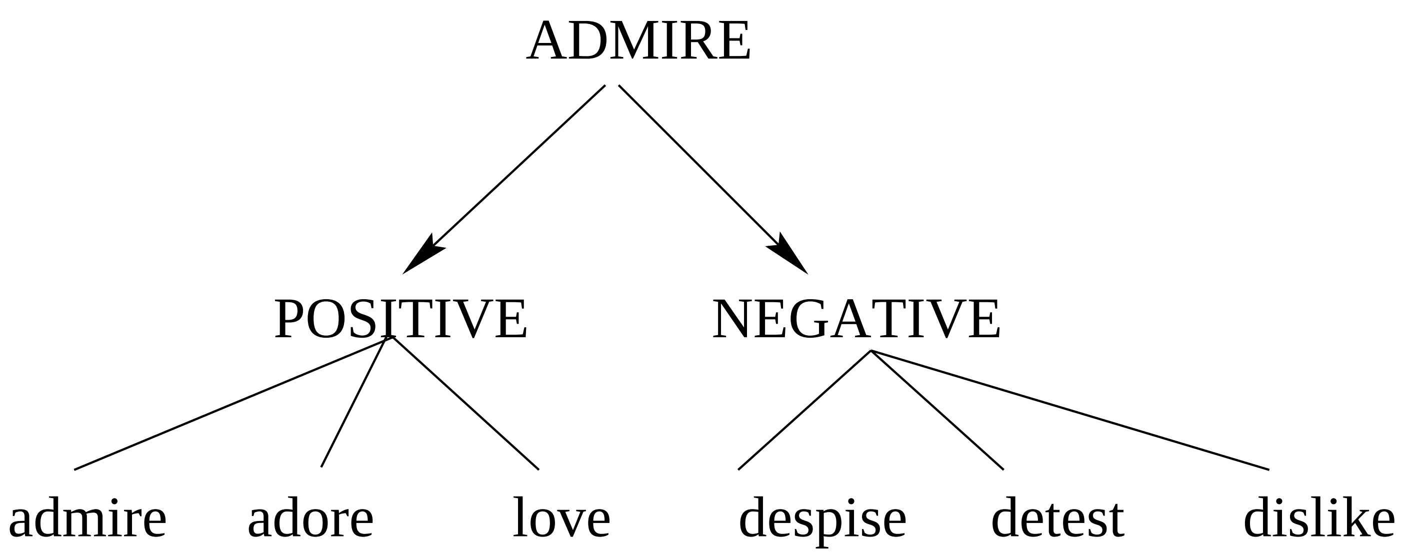


(4) a. I appreciated the mayor's honesty. [ADMIRE, NP]
b. I appreciated the mayor for his honesty. [ADMIRE, NPforNP]
c. I appreciated the honesty in the mayor. [ADMIRE, NPinNP]

Select frame in a probabilistic framework; combine Levin with corpus data (Lapata 2001).

Surface Realization

- Mapping between classes and verbs;
- Classes may be coarse-grained for this mapping;
- Intersective Levin classes (Dang et al. 1998).



Select verbs in a probabilistic framework (Langkilde and Knight 1998).

Probabilistic framework

$$P(c,f,v) = P(v) \cdot P(f|v) \cdot P(c|f,v)$$

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

Independence Assumption: verbs of the same class uniformly subcategorize for a given frame; gross simplification of Levin's hypothesis.

$$P(c|v,f) \approx P(c|f)$$

Bayes Law: $P(c|f) = \frac{P(f|c) \cdot P(c)}{P(f)}$

$$P(c,f,v) \approx \frac{P(v) \cdot P(f|v) \cdot P(f|c) \cdot P(c)}{P(f)}$$

Parameter Estimation

Method:

- Corpus: POS-tagged, lemmatized version of the BNC (100M words).
- Parser: uses a PCFG for English (Carroll and Rooth 1998):
  - information about the lexical heads of constituents;
  - extract verb and its arguments from most probable parse.

Simplifications:

$$f(c) = \sum_i f(v_i, c) \quad f(v,c) \approx \frac{f(v)}{|classes(v)|}$$

$$P(f|c) = \frac{f(f,c)}{f(c)} \qquad f(f,c) \approx \sum_{v \in c} \frac{f(f,v)}{|classes(v)|}$$

Examples

$$*c = \underset{c_j \in C}{\text{argmax}}\, P(c, f, v)$$

$P(c, \text{NPNP}, \textit{leave})$	F.HAVE	FULFILL	GET
$P(c, \text{NPtoNP}, \textit{convey})$	SEND	SAY	

$$*f = \underset{f_j \in F}{\text{argmax}}\, P(c, f, v)$$

$P(\text{GIVE}, f, \textit{give})$	NPNP	NPtoNP	NP
$P(\text{ADMIRE}, f, \textit{criticize})$	NP	NPforNP	NPinNP

$$*v = \underset{v_j \in V}{\text{argmax}}\, P(c, f, v)$$

$P(\text{ADMIRE}, \text{NPforNP}, v)$	support	love	admire	like
$P(\text{ADMIRE}, \text{NP}, v)$	like	support	love	enjoy

$$P(\text{SAY}, \text{INTR}, \textit{say})$$

Results

Verb	**Class**				K
leave √	GET 6	FULFILL 14	F. HAVE 56	OTHER 23	.80
write √	MSG. TRANS. 54	PERFORM 19	OTHER 18		.85
find √	DECLARE 36	GET 47	OTHER 17		.70

Frame	Baseline	Model
NP1 V NP2 NP3	50.0%	78.6%
NP1 V NP to NP3	43.8%	68.8%
NP1 V NP for NP3	00.0%	100.0%
NP1 V at NP2	100.0%	100.0%
NP1 V NP2	47.1%	73.5%
Combined	46.2%	74.6%

Baseline: most likely class

Results

Alternation typicality: measures how likely a verb is to alternate.

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

- a verb is typical if both frames are equally frequent: AT $\approx$ 0.5;
- AT $\approx$ 0 or 1 for untypical verbs;
- Generalises over classes: $\sum_{v \in c} P(f|v)$.

GIVE	NPtoNP	NPNP	AT
give	3735	13430	.23
feed	40	98	.29
serve	58	85	.40
pay	567	632	.47
repay	17	12	.58
lend	648	343	.65
sell	982	294	.76

Results

- production study using WebExp (Keller et al. 1998);
- participants (140) are given verbs taken from Levin classes and produce sentences;
- comparison between corpus-based AT and production data.

Dative Alternation

Class	AT	
	Corpus	Subjects
BRING-TAKE	.78	.69
GIVE	.56	.60
MSG. TRANS.	.52	.62
CARRY	.96	.93
THROWING	.71	.68
SEND	.68	.72
INSTR. COM.	.76	.79
SLIDE	.90	.83

Possessor Object Alternation

Class	AT	
	Corpus	Subjects
JUDGMENT	.44	.44
ASSESSMENT	.12	.05
WANT	.14	.04

Relevance for Generation

- Rerank Nitrogen's Nbest lists; supply missing information.
- Empirical derivation of defaults; domain sensitive.

```
(A / "give"
:AGENT (J / "judge")
:DESTINATION (d / |dog,canid|)
:PATIENT (B / |os,bone|))
```

the judge gave the bones in the dogs.
the judge gave the bone in the dogs.
the judge gave the bone in dogs.
the judge gave bones in the dogs.
the judge gave bone in the dogs.
the judge gave bone in dogs.
the judge gave the bone to the dogs.
the judge gave a bone in the dogs.

-8.7	P(GIVE,NPNP,give)
-10.6	P(GIVE,NPtoNP,give)
-13.7	P(GIVE,NPinNP,give)

```
(A / "knock"
 :AGENT (V / "visitor")
 :PATIENT (D / "door"))
```

visitors knocked to do.
visitors are knocking to do.
visitors knocked doors.
visitors were knocking to do.
visitors knocked the doors.
visitors knock the doors.
the visitors knocked to do.
the visitor knocked to do.

-17.1	P(HIT,NPon,knock)
-18.1	P(HIT,NP,knock)
-23.0	P(HIT,Sto,knock)

Conclusions

Approach

- complementary to linguistic theory;
- well-defined framework of statistical inference.
- cheap: assumes no semantic annotation.
- combines data-intensive techniques with human evaluation;

Data-intensive methods

- test linguistic predictions;
- quantify generalizations;
- discover novel facts about lexical units;
- enrich linguistic classifications.

Future Work

- better estimation of model parameters;
- include selectional restrictions;
- interface with a generation system.

References

Bangalore, Srinivas, and Owen Rambow. 2000. Exploiting a Probabilistic Hierarchical Model for Generation. In COLING (2000), ??–??

Berger, Adam, Stephen A. Della Pietra, and Vincent J. Della Pietra. 1996. A Maximum Entropy Approach to Natural Language Processing. *Computational Linguistics* 22(1): 39–71.

Carroll, Glenn, and Mats Rooth. 1998. Valence Induction with a Head-lexicalized PCFG. In Nancy Ide and Atro Voutilainen, eds., *Proceedings of the 3rd Conference on Empirical Methods in Natural Language Processing*, 36–45. Granada, Spain.

COLING. 2000. *Proceedings of the 18th International Conference on Computational Linguistics*, Saarbrücken, Germany.

COLING/ACL. 1998. *Proceedings of the 17th International Conference on Computational Linguistics and 36th Annual Meeting of the Association for Computational Linguistics*, Montréal, Canada.

Dang, Hoa Trang, Karin Kipper, Martha Palmer, and Joseph Rosenzweig. 1998. Investigating Regular Sense Extensions Based on Intersective Levin Classes. In COLING/ACL (1998), 293–299.

Dorr, Bonnie, and Mari Broman Olsen. 1996. Multilingual Generation: The Role of Telicity in Lexical Choice and Syntactic Realization. *Machine Translation* 11: 37–74.

IWNLG. 1994. *In Proceedings of the 7th International Workshop on Natural Language Generation*, Kennebunkport, Maine.

Keller, Frank, Martin Corley, Steffan Corley, Lars Konieczny, and Amalia Todirascu. 1998. WebExp: A Java Toolbox for Web-Based Psychological Experiments. Technical Report HCRC/TR-99, Human Communication Research Centre, University of Edinburgh.

Langkilde, Irene, and Kevin Knight. 1998. Generation that Exploits Corpus-Based Statistical Knowledge. In COLING/ACL (1998), 704–710.

Lapata, Maria. 2001. The Acquisition and Modeling of Lexical Knowledge: A Corpus-based Investigation of Systematic Polysemy. Ph.D. thesis, University of Edinburgh.

Lapata, Maria, and Chris Brew. 1999. Using Subcategorization to Resolve Verb Class Ambiguity. In *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*, 266–274. College Park, MD.

Lavoie, Benoit, and Owen Rambow. 1997. A Fast and Portable Realizer for Text Generation Systems. In *Proceedings of the 5th Conference on Applied Natural Language Processing*, 265–268. Washington, DC.

Levin, Beth. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.

McDonald, David, and Frederica Busa. 1994. On the Creative Use of Language: The form of Lexical Resources. In IWNLG (1994), 81–90.

Mel'čuk, Igor. 1988. *Dependency Syntax: Theory and Practice*. New York: State University of New York Press.

Palmer, Martha, and Zhibiao Wu. 1995. Verb Semantics for English-Chinese Translation. *Machine Translation* 10: 59–92.

Pustejovsky, James. 1995. *The Generative Lexicon*. Cambridge, MA: The MIT Press.

Ratnaparkhi, Adwait. 2000. Trainable Methods for Surface Natural Language Generation. In *Proceedings of the 1st North American Annual Meeting of the Association for Computational Linguistics*, 194–201. Seattle, WA.

Reiter, Ehud. 1995. NLG vs. Templates. In *In Proceedings of the 5th European Workshop on Natural Language Generation*, 95–105. Leiden, The Netherlands.

Stede, Manfred. 1998. A Generative Perspective on Verb Alternations. *Computational Linguistics* 24(3): 401–430.

Viegas, Evelyne, and Pierrette Bouillon. 1994. Semantic Lexicons: The Cornerstone for Lexical choice in Natural Language Generation. In IWNLG (1994), 91–98.