

Automatic Interpretation of Noun Compounds via Verb Semantics

Su Nam Kim

WING, CSSE Dept.



Su Nam Kim

Content

- Introduction
- Related Work
- Motivation
- System Architecture
- Semantic Relations
- Resources
- Data Collection & Evaluation
- Discussion & Conclusion

Introduction

- **Goal:** to automatically interpret the semantic relations between head noun and modifier
- **Semantic Relations(SRs) in Noun Compounds(NCs)**
 - ★ **def** relation between head noun and modifier in NCs
 - ★ **examples**
 - GM car* : **MAKE** (car made by GM)
 - family car* : **POSSESSOR** (car belongs to family)
 - morning tea* : **TIME** (tea in the morning)
 - student protest* : **AGENT** (student protests)

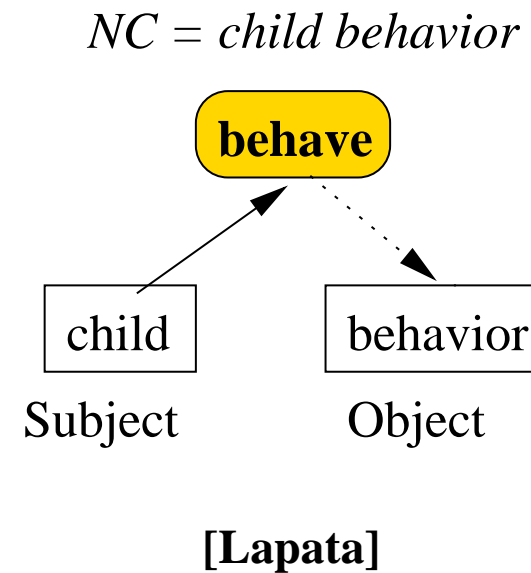
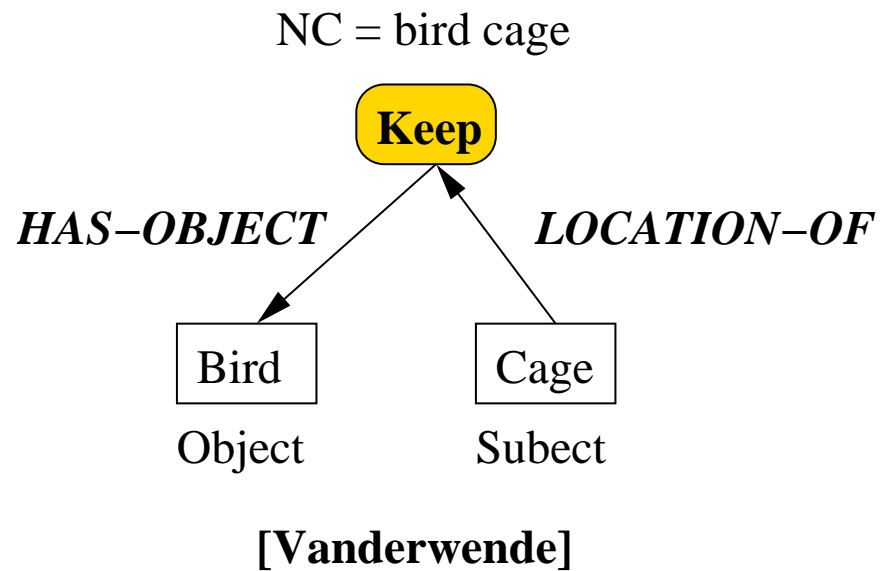
Significance of Study

- **Why do we interpret Semantic Relations in NCs?**
 - ★ Question-answering system
 - ★ Machine translation
 - ★ Information extraction
 - ★ Information understanding

Related Work(I)

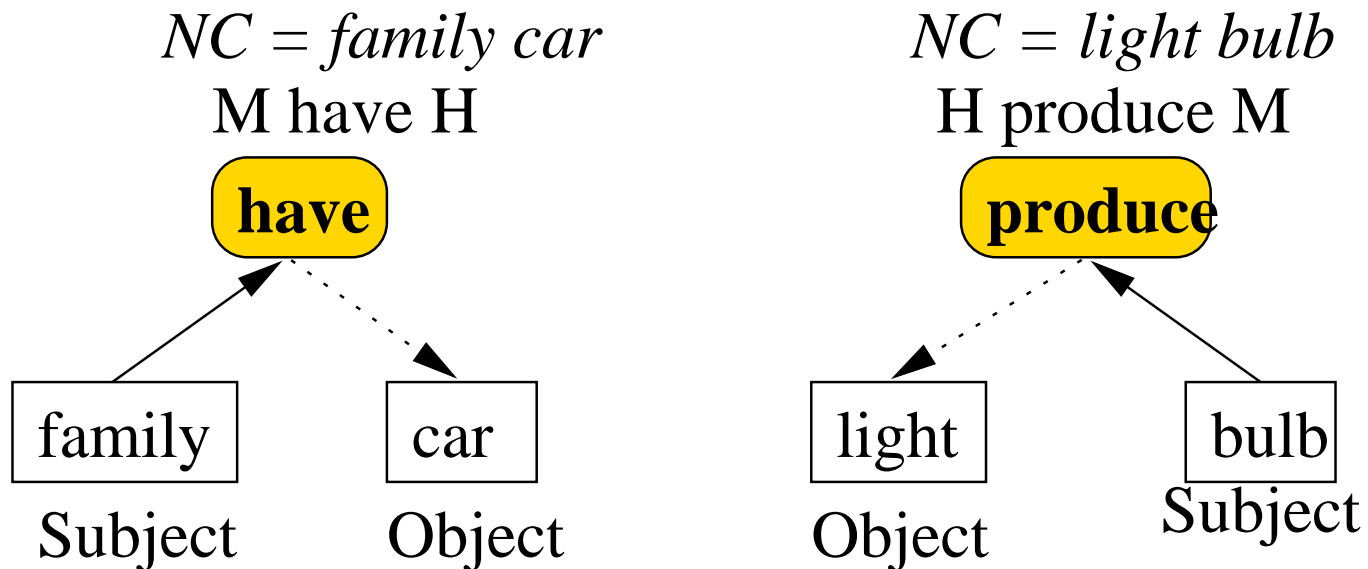
- SEMANTIC SIMILARITY BASED
(Rosario:2001, Moldovan:2004, Kim:2005, Girju:2007, Kim:2007)
- ELLIPSED PREDICATE BASED
 - ★ **Vanderwende(1994)** based on manually assigned verb semantics between head nouns and modifiers
 - ★ **Lapata(2002)** examined NCs that modifier is subject/object of verb expressed by head noun

Related Work (II)



Motivation(I)

- Using the verb semantics defined in Semantic Relations and grammatical role of head noun and modifier



Motivation(II)

(1) *family car*

case: family **owns** the car.

form: H

own

 M

relation: POSSESSOR

(2) *student protest*

case: protest is **performed** by student.

form: M is

performed

 by H

relation: AGENT

Motivation(III)

(3) *family car*

case: *Synonym=have/possess/belong to*

form: H own M

relation: POSSESSOR

(4) *student protest*

case: *Synonym=act/execute/do*

form: M is performed by H

relation: AGENT

Motivation(IV)

- **Emerging Issue** Can we have enough instances for interpretation?
- **Solution** Mapping actual verbs into verb classes in terms of SRs based on **Seed Verbs**
- **What is Seed Verbs?**
 - ★ verbs from definition of SRs and their some of synonyms
 - ★ two sets of seed verbs (57 vs 84)
 - ★ example of Seed Verbs for SR, POSSESSOR
 - (57) own/have/possess/belong to
 - (84) own/have/possess/belong to/acquire/grab/occupy

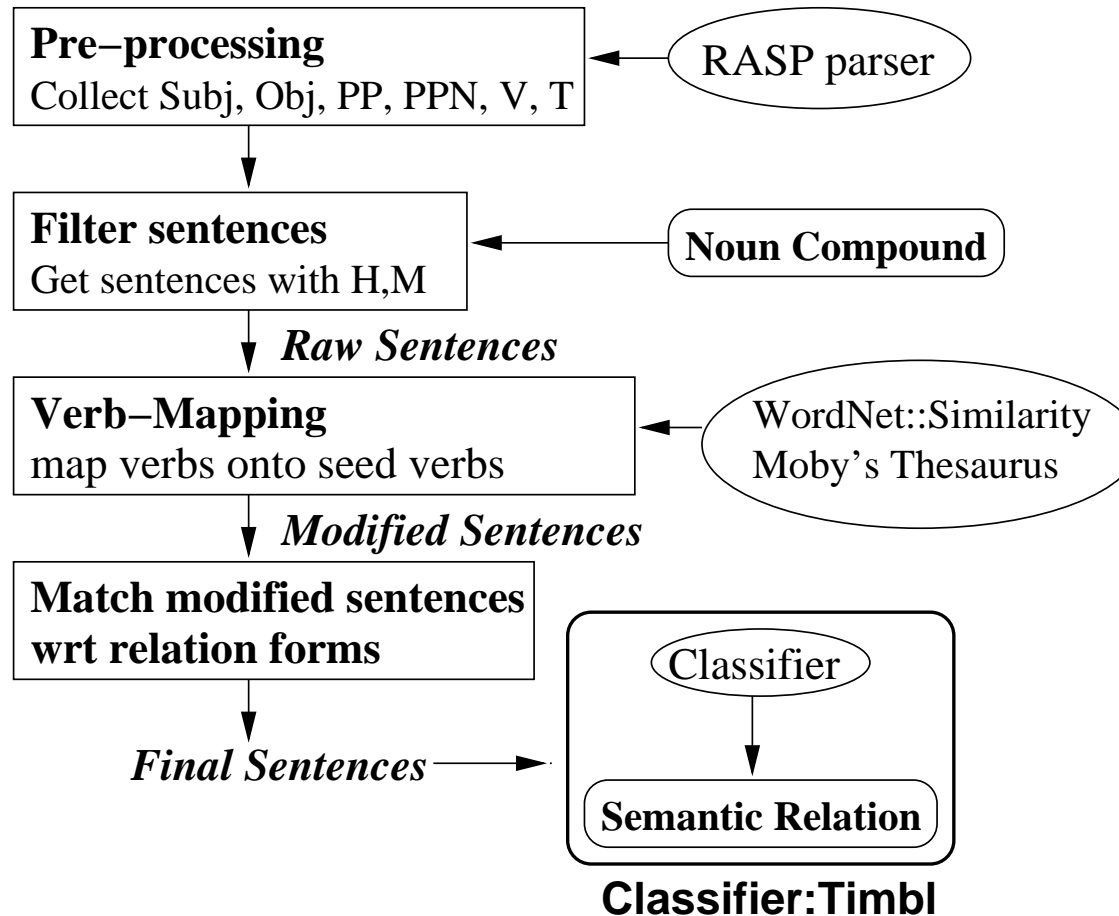
Motivation(V)

- Example of constructional templates associated with SR, POSSESSOR

$$S(\{have, own, possess\}_V, M_{SUBJ}, H_{OBJ}) \quad (5)$$

$$S(\{belong_to\}_V, H_{SUBJ}, M_{OBJ}) \quad (6)$$

System Architecture



Semantic Relations

- based on Barker, K. Szpakowicz (1998)
- two sets of semantic relations (17 vs 19)
 - ★ ignored relation : PROPERTY
 - too general (def: M is H)
 - includes some other relations such as MATERIAL (e.g. apple pie)
 - ★ separated relations : TIME, EQUATIVE
 - for TIME, use CoreLex
 - for EQUATIVE, collect sentences using conjunction (and/or)

Resources(I)

- RASP parser
 - ★ extract argument structure from the output of the dependency analysis
- WordNet::Similarity
 - ★ **RelatedNess** has-part, is-made-of, is-an-attribute-of (*lesk, vector*)
 - ★ **Similarity:path-based** is-a (*wup, lch, path*)
 - ★ **Similarity:information-based** is-a (*jcn, lin, lesk*)
 - ★ **Random** (*random*)

Resources(II)

- Moby's Thesaurus
 - ★ Based on Roget's Thesaurus, contains 30K root words and 2.5M synonyms and related words
- CoreLex
 - ★ Based on unified approach to systematic polysemy and semantic underspecification of nouns from WordNet 1.5
- TiMBL 5.1
 - ★ build a supervised classifier

Data Collection(I)

- NCs for evaluation
 - ★ POS tagged Wall Street Journal in Penntree bank
 - ★ binary NCs excluding proper nouns
 - * original : 2,166, after filter : 453
 - * test NCs : 88, train NCs : 365
- Sentences for evaluation
 - ★ sentences for 453 NCs : 7,714
 - ★ distinct main verbs from sentences : 1,165
 - ★ sentences for test and train NCs : various in terms of verb mapping methods

Data Collection(II)

- Collect Data for SR, **TIME**

★ if modifiers are tagged as 'tme(time) in CoreLex, highest priority

- Collect Data for SR, **EQUATIVE**

(7) *player coach*

case: coach **and** player

form:

$form_{OBJ} = M \boxed{\text{and}} H$

relation: EQUATIVE

Data Collection(III)

- How to compute Weight in sentential form

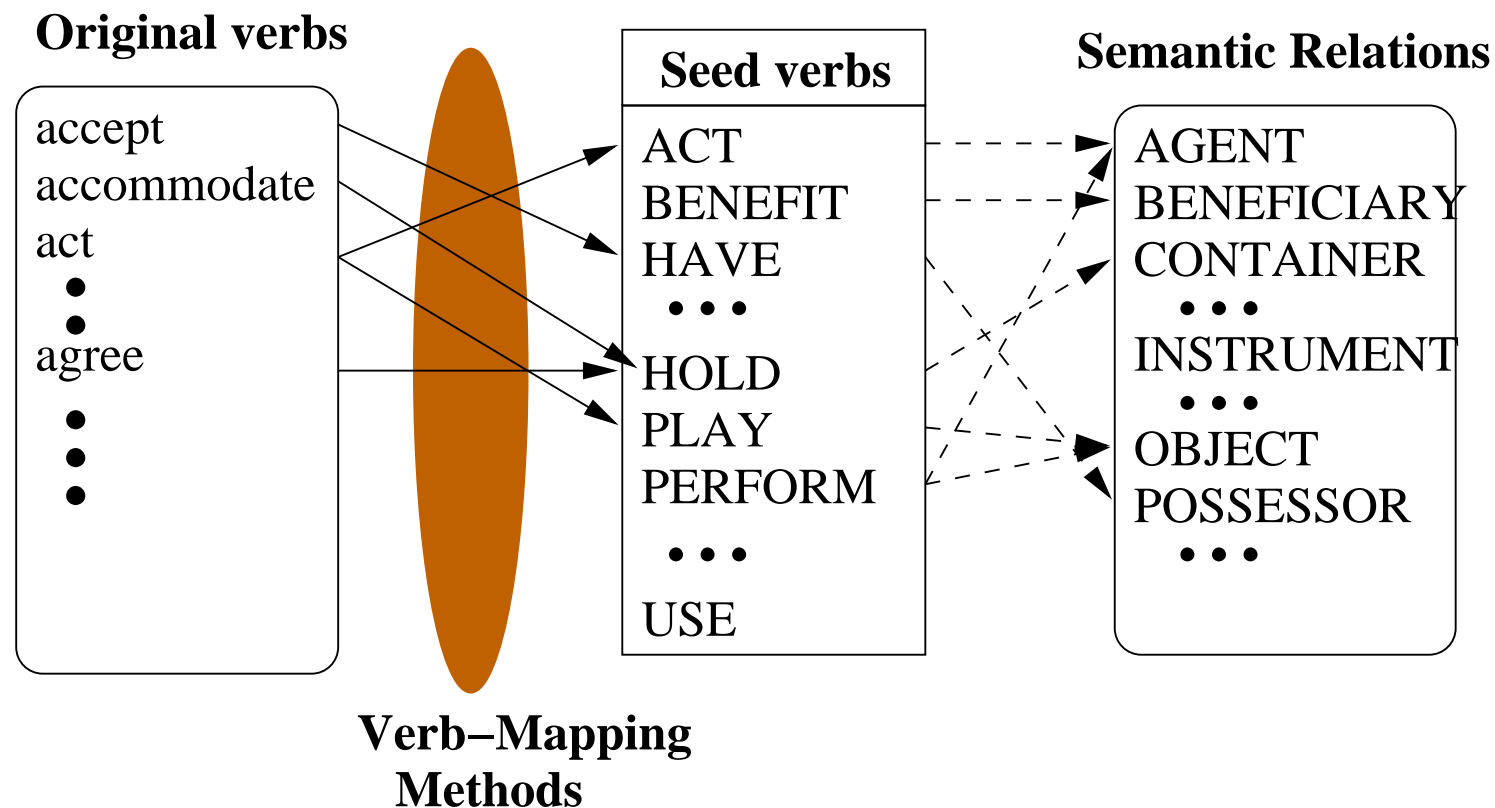
$$Weight(SeedV_j) = \frac{\sum_{i=1, n} n(H_i, SeedV_j)}{\text{total \# of pairs}} \quad (8)$$

- How to compute the weight of NCs in conjunction form

$$NC_i = -\log_2\left(\frac{\sum NC_i \text{ in Conjunction}}{\sum M \text{ in } NC_i * \sum H \text{ in } NC_i}\right) \quad (9)$$

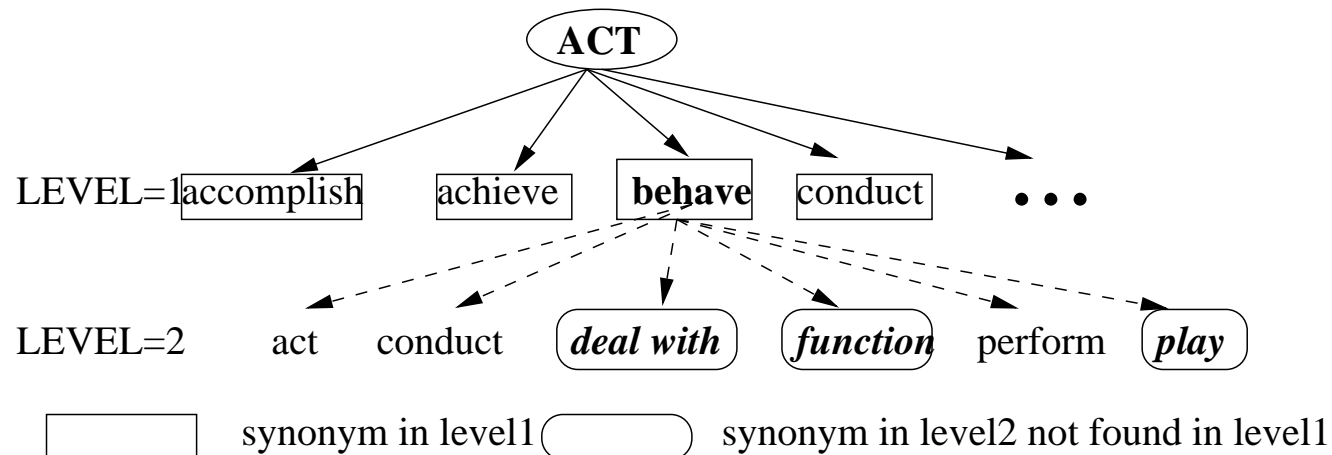
Data Collection(IV)

- Verb Mapping using WordNet::Similarity



Data Collection(V)

- Verb Mapping using Moby's Thesaurus



# of SeedVB	D-Synonym	D,I-Synonym
57	6,755(87.57%)	7,388(95.77%)
84	6,987(90.58%)	7,389(95.79%)

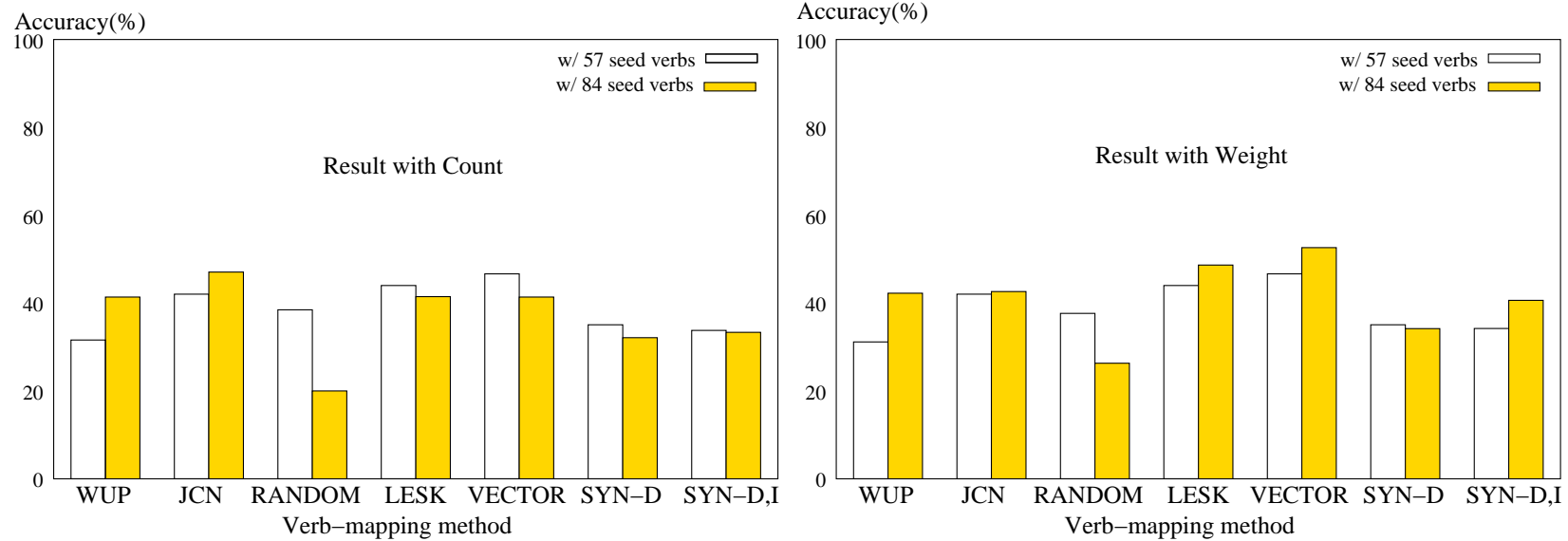
Evaluation

- Result with various Verb Mapping methods

#of SR	# SeedVB	Method	wup	jcn	random	lesk	vector	dsynonym	isynonym
17		Baseline	.4235	.4235	.4235	.4235	.4235	.4235	.4235
	57	Count	.3247	.4085	.3797	.4167	.4667	.3375	.3378
		Weight	.3205	.4085	.3718	.4167	.4667	.3375	.3421
	84	Count	.4066	.4706	.1846	.4390	.4138	.3176	.3333
		Weight	.4247	.4262	.2597	.4571	.5263	.3418	.4062
19		Baseline	.4091	.4091	.4091	.4091	.4091	.4091	.4091
	57	Count+ET	.3158	.4203	.3846	.4400	.4667	.3506	.3378
		Weight+ET	.3117	.4203	.3766	.4400	.4667	.3506	.3421
	84	Count+ET	.4138	.4706	.2000	.4146	.4138	.3214	.3333
		Weight+ET	.4394	.4464	.2800	.4865	.5263	.3562	.3934

Discussion(I)

- Comparison of the Performance with two sets of Seed Verbs using 19 SRs



Discussion(II)

- Result of Constituent Similarity method(Kim&Baldwin 2005)

#of SR	# SeedVB	WUP	LCH	JCN	LIN	RANDOM	LESK	VECTOR
17	Baseline	.4337	.4337	.4415	.4415	.4337	.4776	.4285
	57	.4499	.4217	.4156	.3377	.4096	.4697	.3448
	Baseline	.4337	.4337	.4337	.4337	.4285	.4383	.4444
	84	.4767	.4167	.4093	.3494	.2262	.4658	.3333
19	Baseline	.4186	.4186	.4303	.4303	.4186	.4776	.4138
	57	.4651	.4186	.4177	.3418	.2326	.4627	.3448
	Baseline	.4138	.4138	.4186	.4186	.4138	.4383	.4267
	84	.4713	.4138	.4070	.3488	.2184	.4658	.3200

Conclusion

- achieved 52.63% with 84 seed verbs using **VECTOR** vector mapping method from **Weight**
- investigate the effective verb mapping method to expand the instances
- test two different sets of seed verbs
- exceed Moldva(2004), Kim,Baldwin(2005)
- show performance of similarity method introduced by Kim,Baldwin(2005) over our data set