Statistical Modeling of Multiword Expressions

Su Nam Kim

LT-group, CSSE Dept.

Adviser: Dr. Timothy Baldwin

THE UNIVERSITY OF MELBOURNE

Ph.D Completion Seminar

Research Outline

Linguistics in MWEs

Statistical Approaches

Resources

Summary of Modeling Tasks

Interpreting NCs via Semantic Similarity

Interpreting NCs via Interpretationality

Word Sense Disambiguation in NCs via Substitutability

Identifying VPCs via Linguistic properties

Conclusion

Research Outline

- **Aim in MWEs** Modeling syntax and semantics of Multiword Expressions (MWEs) using statistical approaches

- **Significance**

  - ⋆ resolving the syntax and semantics of words as processing units
  - ⋆ number of MWEs is equivalent to simplex words (Jackendoff 1997)
  - ⋆ reusability (e.g. *take away/off/up..*), economics (e.g. *winter school*), new vocabularies (e.g. *shock and awe, cell phone*), reliability and better expression (e.g. *piss me off*)
  - ⋆ fluency, robustness & better language understanding for NLP

Examples of English MWEs

1. (**NC**) The subject is about an *language learning system design*.
2. (**VPC**) Kim *took* her pen *out*.
3. (**LVC**) She *took a* long *bath* for relaxation after taking a long exam.
4. (**Idiom**) He will inherit when his grandfather *kicks the bucket*.
5. (**D-PP**) The survey shows that *by and large* people skip breakfast.

Open Issues, Related Work & Limits

- **Identification**
  - ⋆ determine whether multiple simplex words form a MWE in the context at token level (*put the sweater on* vs. *put the sweater on the table*)
  - ⋆ confusing with simplex words (e.g. *VPCs*, *LVCs*, *idioms*)

- **Extraction**
  - ⋆ recognize MWEs as word units at type level
  - ⋆ feed word repositories such as dictionary

- **Detecting/Measuring Compositionality**

- ⋆ denote the degree of relations among the components of MWEs
- ⋆ close relationship with **semantic contribution** of parts
- ⋆ (assumption) meanings of MWEs and their parts are specified
- ⋆ hard to measure the degree of compositionality & to utilize it

- **Semantic Classification**

  - ⋆ predict the semantics of MWEs involving understanding the degree of compositionality in MWEs
  - ⋆ (assumption) meanings of MWEs are unspecified
  - ⋆ e.g. particle semantics such as spatial and temporal information (Bannard 2003, Cook 2006)

- **Semantic Interpretation**

- ⋆ interpret the semantic association among components in MWEs
- ⋆ e.g. interpret the semantic relations in NCs, semantic classes of D-PPs such as media and manner
- ⋆ in case of NC interpretation, no standard set of SRs, conducted under their own assumptions

- **Cross-over/Cross-lingual Study**
  - ⋆ Utilize study outcomes of a type/language of MWEs to another types/language of MWEs
  - ⋆ few cases shown cross-lingual study, hard to find the same features among various MWE types (Venkatapathy 2006, Kim&Baldwin 2007)

Difficulties on Modeling MWEs

- syntactic, semantic, and pragmatic idiomaticity
  - ⋆ *family cars, He took of the coat, He kicked the bucket*
- syntactic and semantic flexibility
  - ⋆ *Eat quickly up dinner, make a big mistake*
- high productivity in language processing
  - ⋆ *orange/apple/lemon/chocolate... juice*
- different linguistic features w.r.t. various types of MWEs

Scope & Approaches of Thesis

- **Scope of Research**
  - ⋆ English Multiword Expressions only
    - ∗ due to resource availability
    - ∗ due to syntactically & semantically high productivity
  - ⋆ Noun Compounds & Verb-Particle Constructions due to the size

- **Our Approaches**
  - ⋆ using **Statistical** methods + symbolic methods
  - ⋆ minimize human labor & maximize benefits of existing resources (e.g. WordNet, CoreLex)

Our Aim & Contribution

- to shed light on underlying linguistic processes giving rise to MWEs across constructions and across languages
- to generalize techniques, abstract away from individual MWE types to develop general purpose interpretation methods
- to cross-compare alignment of pre-existing MWE classifications
- exemplify the utility of MWE interpretation within general NLP tasks
- w/ **NCs** : NC interpretation, Bracketing, WSD in NCs
- w/ **VPCs** : identification, detecting compositionality

Research Outline

English MWEs: properties & types

- ***MWEs*** : lexical items that can be decomposed into multiple simplex words and display lexical, syntactic, semantic, pragmatical and/or statistical idiosyncrasies

- **collocation** and **anti-collocation**
  - ⋆ collocation : any statistically significant word co-occurrence (Sag et al. 2002) (e.g. ~~*red tape*~~)
  - ⋆ anti-collocation : a word which must *not* be used with the target words (Pearce 2001) (e.g. *frying fan* vs. ~~*frying pot*~~)

- **Properties of English MWEs** (Fillmore 1988,Liberman 1992, Nunberg et al. 1994, Sag et al 2002)

- **Idiomaticity**: the syntactic, semantic, pragmatic, and statistical irregularity (e.g. *apple pie* vs ~~*by and large*~~)
- **Institutionalization or Cnoventionalisation**:syntactically and semantically predictable but used with a high frequency in a particular context (e.g. *black and white* vs ~~*white and black*~~)
- **Non-identifiability**: the meaning cannot be easily predicted from the surface form (components) (e.g. *kick the bucket*→*die??*)
- **Situatedness**: expressions which are associated with a fixed pragmatic content (e.g. *good morning, all aboard*)
- **Figuration**: an attribute found in encoded expressions such as metaphors, metonymies and hyperboles (e.g. *red tape*=*bureaucratic*)
- **Single-word paraphrasability**: paraphrasable MWEs enables

substitution with a single word (e.g. *leave out*=*omit*)

- ⋆ **Proverbiality**: describe and implicitly to explain a recurrent situation of particular social interest (e.g. *informality, affect*)
- ⋆ **Prosody**: have distinctive stress patterns that diverge from the norm (e.g. *soft spot* vs *first aid* vs ~~*dental operation*~~)

- **Types of English MWEs** (Sag et al 2002)
  - ⋆ Lexicalized Phrase
    - ∗ fixed expression. no morphosyntactic variation nor internal modification
    - ∗ semi-fixed expression. lexically variable. non-decomposable idiom, CNs, proper name
      - ○ various inflection (e.g. *make a speech vs.* ~~*a speech is made*~~)

  - ○ various reflexive form (e.g. *in her/his/their shoes*)
- ∗ syntactically-flexible expression. VPCs, LVCs
  - ○ variety w.r.t. verb tense (*a demo was given*), extraction (*how many demos did he give?*), internal modification (*give a clear demo*)

⋆ Institutionalized Phrase

- ∗ syntactically and semantically compositional but used with a unexpectedly high frequency in a particular context
- ∗ e.g. *salt and pepper, many thanks, telephone booth*
- ∗ *traffic light* vs. *traffic director, intersection regulator* due to statistical perspective

Compound Nouns (CNs)

- CN is a noun made up of two ore more lexemes (cf. **NCs**=lexemes are all nouns)
- Type of English CNs

combination	example	combination	example
noun+noun	*morning tea*	verb+noun	*swimming pool*
adjective+noun	*monthly ticket*	preposition+noun	*over·coat*
noun+verb	*hair·cut*	adjective+verb	*dry·cleaning*
preposition+verb	*out·put*	noun+preposition	*hanger on*

- Syntactic Variation
  - ⋆ split(*full moon*) vs joined(*bed·room*) vs. both(*post·man, post man*) vs joined w/ hyphen(*check-in*)
- Modification such as plurality & genitive (*family cars* vs ~~*families car*~~)

Verb-Particle Constructions (VPCs)

- VPC is a verb with its obligatory particle(s)
  - **intransitive:** *Kim calmed down.*
  - **transitive:** *Kim handed in the paper./Kim handed the paper in./Kim gets Sandy down.*

- Linguistic Properties of VPCs
  - Transitive VPCs undergo the particle alternation (*hand in the paper.* vs. *hand the paper in.*)
  - With transitive VPCs, pronominal objects must be expressed in the split configuration (*hand it in.* vs. ~~*hand in it*~~)
  - Manner adverbs cannot occur between the verb and particle (~~*hand it promptly in*~~)

Light-Verb Constructions (LVCs)

- LVC is a verb whose meaning is bleached to some degree & appear a complement of light verbs
- occur in many languages such as English, Dutch, Japanese
- In English, often occur with *do, get, give, have, make, put, take*
- Examples of English LVCs
  - ⋆ *do a memo* → *memo*
  - ⋆ *give a bath* → *bath(passive)*
  - ⋆ *take a bath* → *bath(active)*
  - ⋆ *make a decision* → *decide*

Idioms

- a MWE whose meaning is not predictable from the usual meanings of its parts
- categorized into *compositional* vs. *non-compositional*
  - ⋆ compositional : *take advantage of*, *spill the beans*
  - ⋆ non-compositional : *in one's shoes*, *kick the bucket*
- detected by non-compositionality, non-substitutability (~~*spill the nuts*~~), non-modifiability (~~*several thanks*~~)

Determinerless-Prepositional Phrase (D-PPs)

- a MWE constructed with a preposition & a singular noun w/o a determiner
- Syntactically Markedness : (non-)productive & (non-)modifiable (e.g. *by car/bus/plane/...* vs. ~~*on very top*~~)
- Nominal Modifiability
  - ⋆ fully fixed expressions(~~*on chilly ice*~~) vs. obligatory modification(*on summer vacation*)
- Semantically Markedness
  - ⋆ institutional(*at school, in church*), media(*on TV, off screen*), metaphor(*on ice, at large*), temporal(*on holiday, by day*), means/manner(*by car, via radio*)

Co-occurrence Properties

- uses the co-occurrence of parts in the target lexical themselves for computational tasks

- implicitly use collocation and/or anti-collocation

- related to susbstitutability, often measured by statistical test

- verb, *propose* & co-occurring words (Lin 1999)
  - ⋆ million(458), billion(438), accord(296), increase(260), call(239), year(201), change(198), support(178), proposal(154), percent(154), money(143), plan(142), cut(139), aid(130), program(124), people(122)

Substitutability

- the ability to replace parts of lexical items with alternatives
- alternatives can be similar or opposite words w.r.t tasks & approaches
- can be used when parts in lexical items occur with unusually high frequency
- subset of co-occurrence, explitly use the collocation and/or anti-collocation
- MWEs and Non-MWEs using substitutability (Pearch 2001)

MWEs	→	***Non-MWEs***
frying fan	→	*frying* ***pot***
salt and pepper	→	*salt and* ***sugar***
many thanks	→	***some*** *thanks*

Distributional Similarity

- a method to extract the semantic similarity using the context
- when two words are similar, then their context words are also similar
- Examples of Distributional Similarity

Example	Type	Context Words
kick the bucket	**MWE**	mourn, sad, blue
	Non-MWE	run, ball, accident
put on	**MWE**	clothes, garment
	Non-MWE	objects

Semantic Similarity

- based on the semantics of parts to deal with whole
- Underlying assumption of semantic similarity : the similarity of the parts represents the semantics of whole
- Examples with NCs
  - ⋆ modifier = *fruit*, head noun = *liquid* (SR:MAKE)
    e.g. *apple juice, orange juice, grapes nectar*
  - ⋆ modifier = *location*, head noun = *liquid* (SR:LOCATION)
    e.g. *Fuji apple, California orange, Australian wine*

Interpretationality

- a way to use the semantics of parts while building constructs which put parts together
- when simplex words are put together in a MWE, their relation or connection could be useful to identify MWEs
- correlated with compositionality
- w/ NC, *virus infection* → SR, CAUSE (Levi 1979)
  1. infection (virus causes infection)
  2. infection (infection is caused by virus) → Passive
  3. infection (infection is virus-caused) → Compound adjective Formation
  4. infection (which is virus-caused) → Relative Clause Formation

Linguistic Properties

- linguistic features can be the strong clues for lexical acquisition
- Syntactic & semantic features are used as linguistic properties
- local information vs. global information (distributional similarity)
- Examples with VPCs

possibility	marked	example
particle position	(O)	*pick* a broken lead pencil *up*
	(X)	*pick* a disease *up*
particle modifiability	(O)	*pick* a pencil *straight/right/back up*
	(X)	*pick* a disease *straight/right/back up*
nominalization	(O)	*feedback*, *backup*
	(X)	*boilup*

Resources (1)

- Corpus : preparsed by RASP parser
  - ⋆ British National Corpus
  - ⋆ Brown Corpus
  - ⋆ Wall Street Journal at Penntree Bank

Resources (2)

- Lexical Resources
  - ⋆ WordNet
    - ∗ lexical reference system whose design is inspired by current psycholinguistic theories of human lexical memory (Fellbaum:1998)
  - ⋆ Moby's Thesaurus
    - ∗ Based on Roget's Thesaurus, contains 30K root words and 2.5M synonyms and related words
  - ⋆ CoreLex
    - ∗ systematic polysemy and semantic underspecification of nouns from WordNet 1.5 (Buitelaar:1998)

Resources (3)

- Tools
  - ⋆ 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)
  - ⋆ TiMBL
    - ∗ statistical learner to build a classifier (Daelemans:2004)
  - ⋆ RASP parser, Minipar, Chaniak parser
    - ∗ extract argument structure from the output of the dependency analysis

Summary of Modeling Tasks

- constituent similarity method using **Semantic Similarity** †
  - ⋆ Similar NCs could have same SR
  - ⋆ e.g. *apple juice*, *banana milk* → SR=MAKE
- verb semantics method using **Interpretionality** †
  - ⋆ Using the verb semantics defined in Semantic Relations and grammatical role of head noun and modifies
  - ⋆ e.g. *GM car*=MAKE → *car made by GM*
- constituent substitution method using **Substitutability, Semantic Similarity, Co-occurrence**

  - ⋆ expand the interpreted NCs by a substitution based on the sense collocation and bootstrapping
  - ⋆ e.g. *apple juice* = MAKE → *fruit/cranapple/orange juice* = MAKE

- benchmarking & hybridizing NC interpretation methods using **Semantic Similarity, Substitutability, Co-occurrence**
  - ⋆ with sense collocation, constituent similarity and constituent substitution methods, hybrid and benchmark these using SemEval-2007 data

- WSD in NCs using **Substitutability, Semantic Similarity** †
  - ⋆ using sense collocation, roles of parts and heuristics (**one sense**

per collocation)

- ⋆ e.g. $(TOPIC|WS_{art},\ WS_{museum}) \rightarrow (WS_{art}|WS_{museum},\ TOPIC/grammatical\_role_{art})$
- ⋆ e.g. *art museum* → *artifact/creation/skill/visual museum*

- Identifying VPCs using **Linguistic Properties** †
  - ⋆ using linguistic properties of associated nouns of VPCs and Verb-PPs associated with distinct **selectional preferences**
  - ⋆ e.g. *put the coat on* vs. *put the coat on the chair*

- Detecting Compositionality of VPCs using **Semantic Similarity**
  - ⋆ using Semantic Similarity of combination of Verb and Particle
  - ⋆ e.g. *call up*:compositional → *ring up*:compositional

Using Constituent Similarity

- Intuition: Similar NCs could have same SR

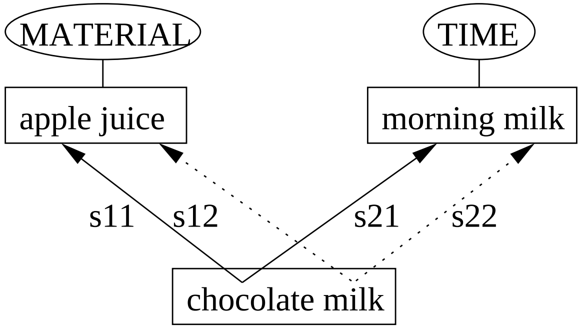
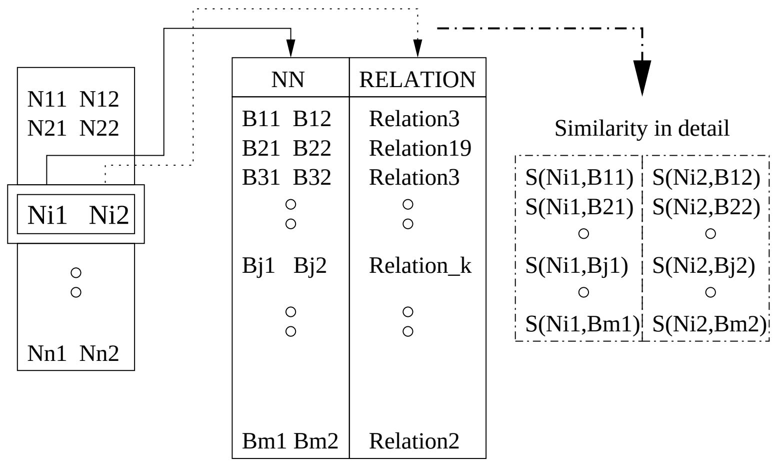


Figure 1: w/ *chocolate milk*

	Training noun	Test noun	S_{ij}
n_1	apple	chocolate	0.71
n_2	juice	milk	0.83
n_1	morning	chocolate	0.27
n_2	milk	milk	1.00

Method

- Compute the Similarity
  - $\star$ $S((N_{i,1}, N_{i,2}), (B_{j,1}, B_{j,2})) = \frac{((\alpha S1+S1)\times((1-\alpha)S2+S2))}{2}$
- Find the SR for test NC
  - $\star$ $rel(N_{i,1}, N_{i,2}) = rel(B_{m,1}, B_{m,2})$ where $m = \text{argmax}_j S((N_{i,1}, N_{i,2}), (B_{j,1}, B_{j,2}))$
- Similarity between i_{th} NC in test NC and j_{th} NC in training NC



Data

- Noun Compounds from Wall Street Journal at Penntree Bank
  - ⋆ POS tagged Wall Street Journal at Penntree Bank
  - ⋆ 2 term NCs only (noun-noun pairs)
  - ⋆ exclude proper nouns
- final number : training NCs (1,088), test NCs (1,081)

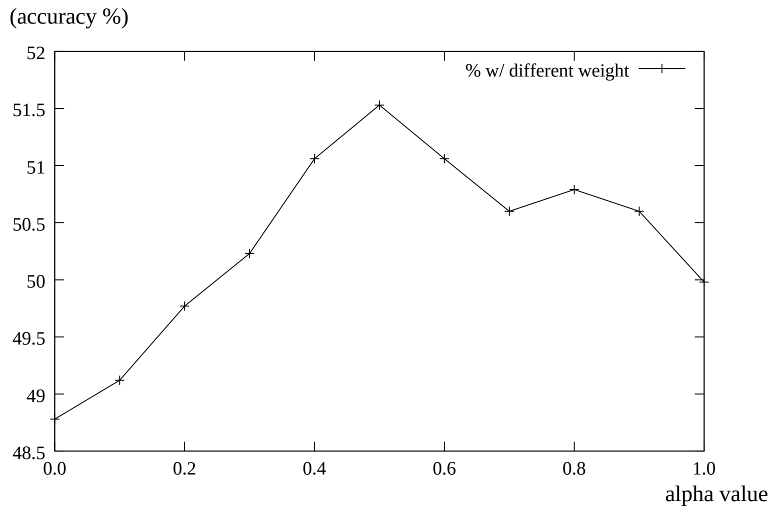
Experiment on 2-term NCs

- Accuracy of NC interpretation the different WordNet-based similarity measures

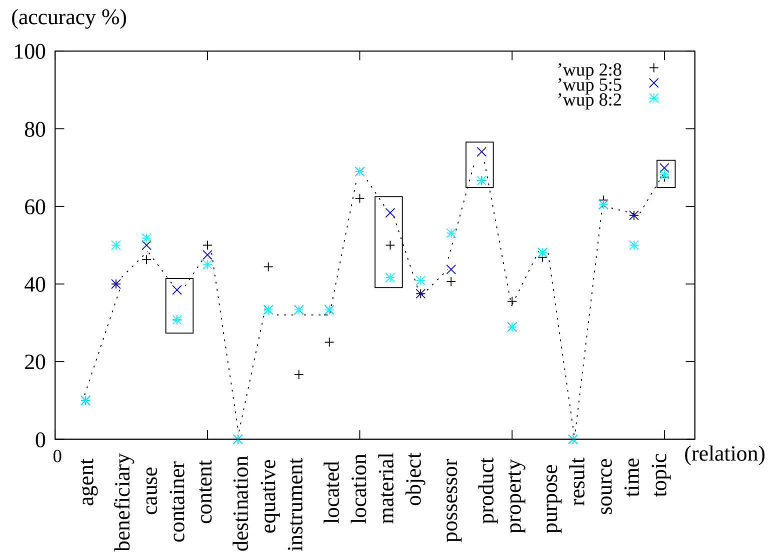
Basis	Method	Accuracy
human annotation	Inter-annotator agreement	52.30%
Zero-R	Baseline	43.00%
path-based	WUP	**53.30%**
	LCH	52.90%
information content-based	JCN	46.70%
	LIN	47.40%
relatedness	LESK	42.44%
	VECTOR	39.22%
random	RANDOM	21.83%

Experiment on Relative Contribution

- Accuracy at different α values



- Accuracy for each semantic relation at different α values



Summary of Constituent Similarity Method

- Achieved higher performance than previous results (2005)
- Confirm the relative contribution of parts w.r.t. SRs
- test the method over 3-term NCs‡
- Successfully adopt other techniques (bootstrapping & K-nearest algorithm)‡
- Show the utilization of SRs for bracketing task‡

NC interpretation via Verb Semantics (1)

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

(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

(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

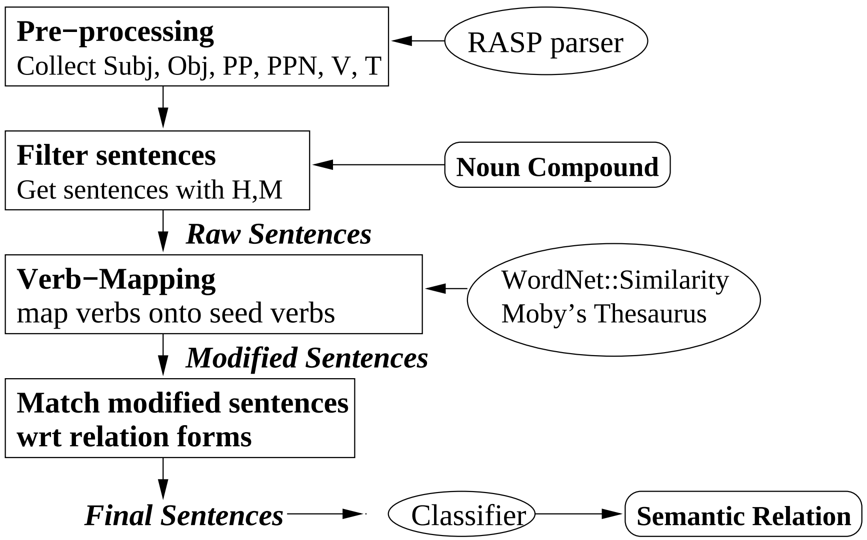
relation: AGENT

NC interpretation via Verb Semantics (2)

- **Emerging Issue** Can we have enough instances for interpretation?
- **Solution** Mapping actual verbs onto 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

Method & Architecture

- Example of constructional templates associated with SR, POSSESSOR
  - ⋆ $S(have, own, possess_{verb}, M_{subj}, H_{obj})$, $S(belong\_to_{verb}, H_{subj}, M_{obj})$



Data Collection (1)

- 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 (2)

- Collect Data for SR, TIME
  - ⋆ if modifiers are tagged as tme(time) in CoreLex, highest priority
- Collect Data for SR, EQUATIVE

  (5) *player coach*

  case: *coach* ***and*** *player*

  form: H [and] M

  relation: EQUATIVE

- PROPERTY is ignored due to higher class concept

Data Collection (3)

- How to compute Weight in sentential form

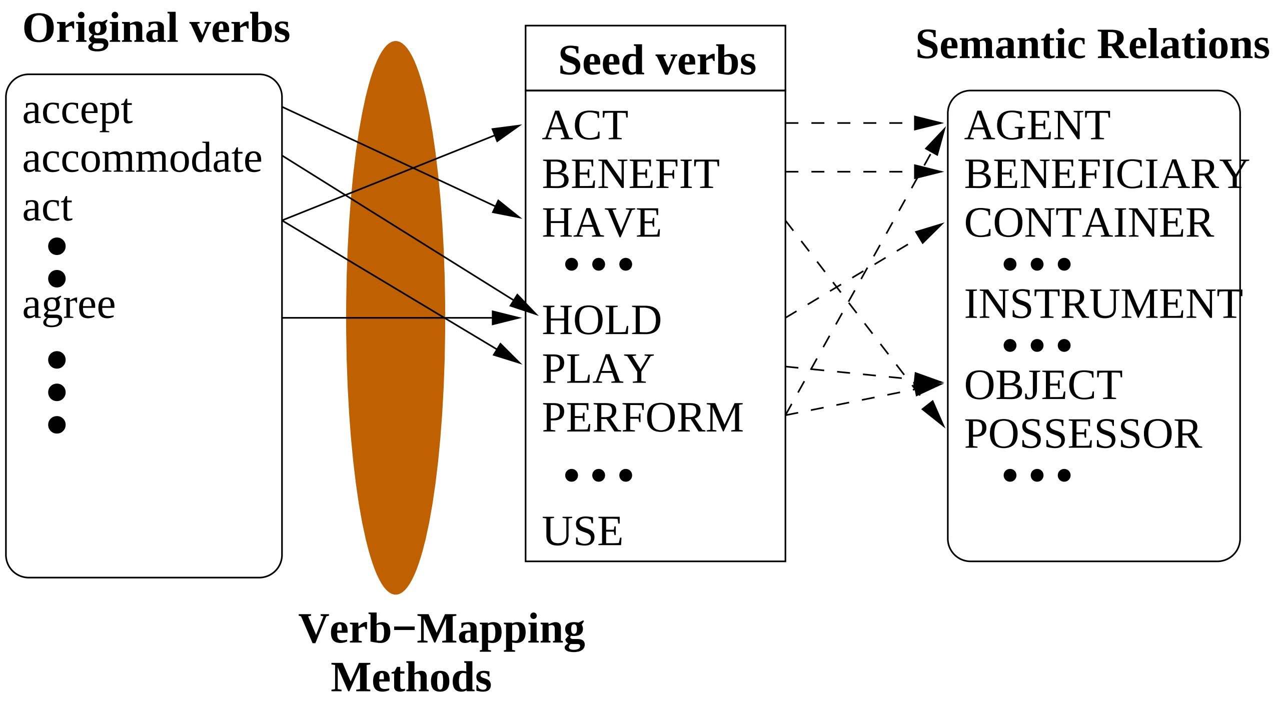
$$Weight(SeedV_j) = \frac{\sum i = 1, n(H_i, SeedV_j)}{total\ \#\ of\ pairs}$$

- How to compute the weight of NCs in conjunction form

$$NC_i = -log_2(\frac{\sum NC_i\ in\ Conjunction}{\sum M\ in\ NC_i * \sum H\ in\ NC_i})$$

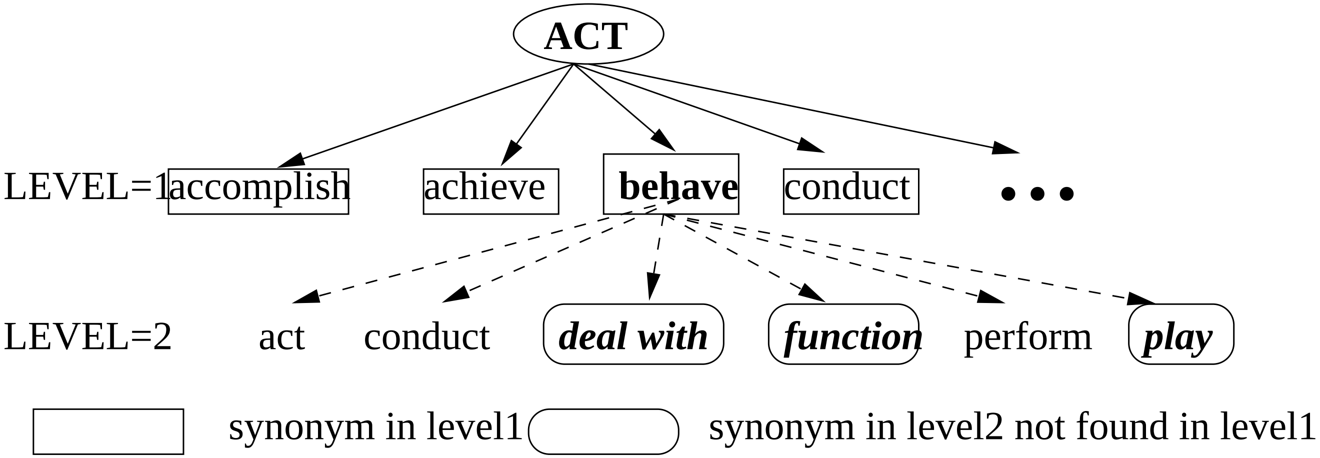
Data Collection (4)

- Verb Mapping using WordNet::Similarity



Data Collection(5)

- 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 (1)

- 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

Evaluation (2)

- Result of Constituent Similarity method as benchmarking

#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

Summary of Verb Semantics Method

- 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
- Outperformed previous methods, (Moldovan 2004) & (Kim&Baldwin 2005) (2006)
- Show performance of similarity method introduced by (Kim&Baldwin 2005) over our data set

Word Sense Disambiguation for NCs

- **Aim**: to investigate the interaction between word sense and interpretation in English NCs
  - ⋆ to automatically disambiguate polysemous nouns in NCs
  - ⋆ to improve NC interpretation performance through word sense

Observation (1)

- The sense distribution of nouns in NCs differs from simplex usages
- The sense distribution of modifier and head nouns also differs, e.g. *art* and *day* (based on SemCor and WordNet2.1):

WordNet sense	*art* mod	*art* head	*art* SemCor
ws_1	**.85**	**.62**	**.67**
ws_2	.11	.04	.22
ws_3	.00	.03	.08
ws_4	.04	.31	.03

WordNet sense	*day* mod	*day* head	*day* SemCor
ws_1	.13	.04	**.41**
ws_2	.02	.04	.20
ws_3	**.80**	.00	.12
ws_4	.00	**.91**	.20
ws_5	.04	.01	.05
ws_6	.00	.00	.03

Observation (2) One Sense per Collocation

- One Sense per Collocation heuristic of Yarowsky (1995)
  - ⋆ words almost always occur with the same sense across all token instances of a given word collocation
  - ⋆ accuracy of 90-99% over a range of binary disambiguation bootstrapping tasks

- One Sense per Collocation for NC
  - ⋆ apply the heuristic to the full WordNet sense inventory rather than coarse-grained binary distinctions
  - ⋆ apply to NCs at the type level (i.e. no linguistic claims made for different senses based on context)

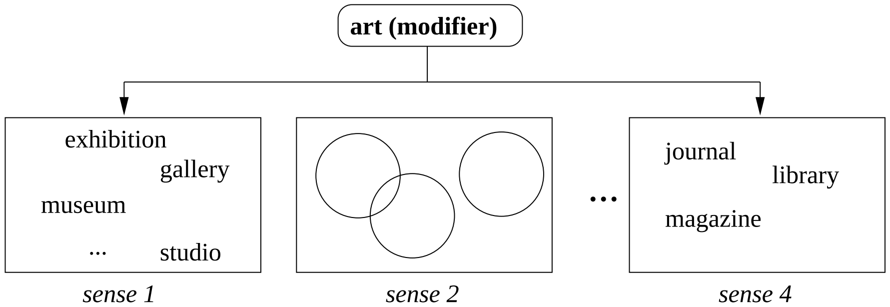
Approach I : Supervised (1)

- Use sense combinatoric method of Moldovan et al. (2004):

$$sr^{*} = \text{argmax}_{sr_i} P(sr_i | ws(n_1), ws(n_2)) (1)$$

$$ws^{*}(n_i) = \text{argmax}_{ws(n_i)} P(ws(n_i) | ws(n_j), sr) (2)$$

- Replace sr in (2) with the **grammatical role** of the polysemous noun (gr)



Approach I : Supervised (2)

- Experiment with two sense inventories:
  - ⋆ CoreLex, e.g. *apple* = FOOD (61.6% coverage)
  - ⋆ first sense and its three hypernyms in WordNet2.1

art **lesson**

Sense 1 lesson
=> teaching, instruction,...
=> education
=> profession
=> ...

Sense 2 example, deterrent example..

Approach II : Unsupervised

- Replace a polysemous noun with its synonyms and calculate the probability of each underlying word sense by web frequency
$$ws^*(n_1) = \text{argmax}_{s_i \in ws(n_1)} \frac{\sum_{n_j \in ss(s_i) \setminus \{s_i\}} freq(n_j, n_2)}{|ss(s_i) \setminus \{s_i\}|}$$

- Similar to (Mihalcea & Moldovan 1999) and (Agirre & Martinez 2000)

- Example of substitution method with *art museum*

sense	substituted NCs
1	**craft/artifact** museum
2	**artistic production/creative activity** museum
3	**artistry/superior skill** museum
4	**artwork/graphics/visual communication** museum

Data Collection

- Target nouns
  - ⋆ 9 polysemous nouns which occur in at least 50 NC token instances in both the head noun & modifier roles in the British National Corpus
- Sentences containing test & training NCs
  - ⋆ 50 sentences containing the 9 nouns for each role
  - ⋆ (9 target nouns) $\times$ (head noun vs. modifier) $\times$ (50 sentences) $=$ 900 sentences

Experiment (1) :Word Sense Disambiguation

- WSD accuracy over each target noun in the modifier and head noun positions (the best-performing method in each row is indicated in **boldface**), R is baseline by random, F is baseline by first sense, M is baseline by Zero-R, C is using CoreLex for monosemous nouns, W is using WordNet for monosemous nouns

Target noun	Role in NC	Baseline			Supervised		Unsupervised	SENSELEARNER
		Random	First	Majority	CoreLex	WordNet		
art	modifier	.25	.68	.68	.64	**.70**	.44	.54
	head noun	.25	**.54**	**.54**	.48	.51	.30	.50
	both	.25	**.61**	**.61**	.56	**.61**	.37	.52
authority	modifier	.14	.06	**.78**	.70	.77	.18	.06
	head noun	.14	.08	**.60**	.52	.54	.36	.08
	both	.14	.07	**.69**	.61	.65	.27	.07
bar	modifier	.07	.46	.46	**.54**	.47	.20	.46
	head noun	.07	.30	.24	**.46**	.40	.24	.28
	both	.07	.38	.35	**.50**	.43	.22	.37

Target noun	Role in NC	Baseline Random	Baseline First	Baseline Majority	Supervised CoreLex	Supervised WordNet	Unsupervised	SenseLearner
channel	modifier	.13	**.24**	**.24**	**.24**	.18	.26	.22
	head noun	.13	.16	.26	**.28**	.24	.30	.12
	both	.13	.20	.25	**.26**	.21	.28	.17
child	modifier	.25	**.72**	**.72**	.50	.69	.24	.60
	head noun	.25	**.78**	**.78**	.76	.76	.38	.76
	both	.25	**.75**	**.75**	.63	.73	.31	.68
circuit	modifier	.17	**.68**	**.68**	.62	.61	.62	.66
	head noun	.17	.54	.54	.48	**.57**	.42	.52
	both	.17	**.61**	**.61**	.55	.59	.52	.59
day	modifier	.10	.18	**.68**	.64	.62	.24	.14
	head noun	.10	.06	**.90**	.88	.89	.16	.06
	both	.10	.12	**.79**	.76	.75	.20	.10
nature	modifier	.20	.04	**.70**	**.70**	**.70**	.30	.04
	head noun	.20	.34	.14	**.44**	.38	.20	.32
	both	.20	.19	.42	**.57**	.54	.25	.18
stress	modifier	.20	.02	.48	**.50**	.46	.30	.02
	head noun	.20	.08	.08	.24	**.27**	.28	.08
	both	.20	.05	.28	**.37**	.36	.29	.05
Total	modifier	.16	.34	**.60**	.59	.58	.31	.30
	head noun	.16	.32	.45	**.50**	**.50**	.29	.30
	both	.16	.33	.53	**.55**	.54	.30	.30

Experiment (2): NC Interpretation

- SR annotation initial agreement : 52.31%, baseline = Zero-R
- Use (Kim&Baldwin 2005) as benchmark system
- tested three WSD outputs (*system-tagged* vs. *first-sense* vs. *hand-tagged*)

Method	CoreLex	WordNet
baseline	.273	.273
similarity	.346	.346
system-tagged	.402	**.426**
first-sense	**.403**	.425
hand-tagged	.447	.540

Summary of WSD in NCs

- The proposed (supervised) WSD method works well over NCs
  - ⋆ best performance = 55% accuracy
  - ⋆ tested semantics of non-polysemous nouns → first sense and its hypernyms is more practical choice
- Off-the-shelf WSD methods do not apply well to MWEs
  - ⋆ SenseLearnerperformed poorly over NCs (accuracy = 30%)
- WSD improves NC interpretation performance
  - ⋆ indication there is room for more improvement

Identifying VPCs

- **Aim**: to automatically distinguish between **verb-particle construction (VPC)** and **verb-prepositional phrase (V-PP) token** instances in corpus text

  He ***put*** *his coat* ***on*** vs. *He* ***put*** *his coat* ***on*** *the table*

- **Basic hypothesis**
  - ⋆ For a given verb–preposition combination ambiguous between a VPC and a V-PP analysis (e.g. *put on*), the two analysis will be associated with distinct **selectional preferences**

Capturing Selectional Preferences

(6) *put* = place

EX: *Put* *the* *book* *on* *the table.*

ARGS: $book_{\mathrm{OBJ}}$ = book, publication, object

ANALYSIS: verb-PP

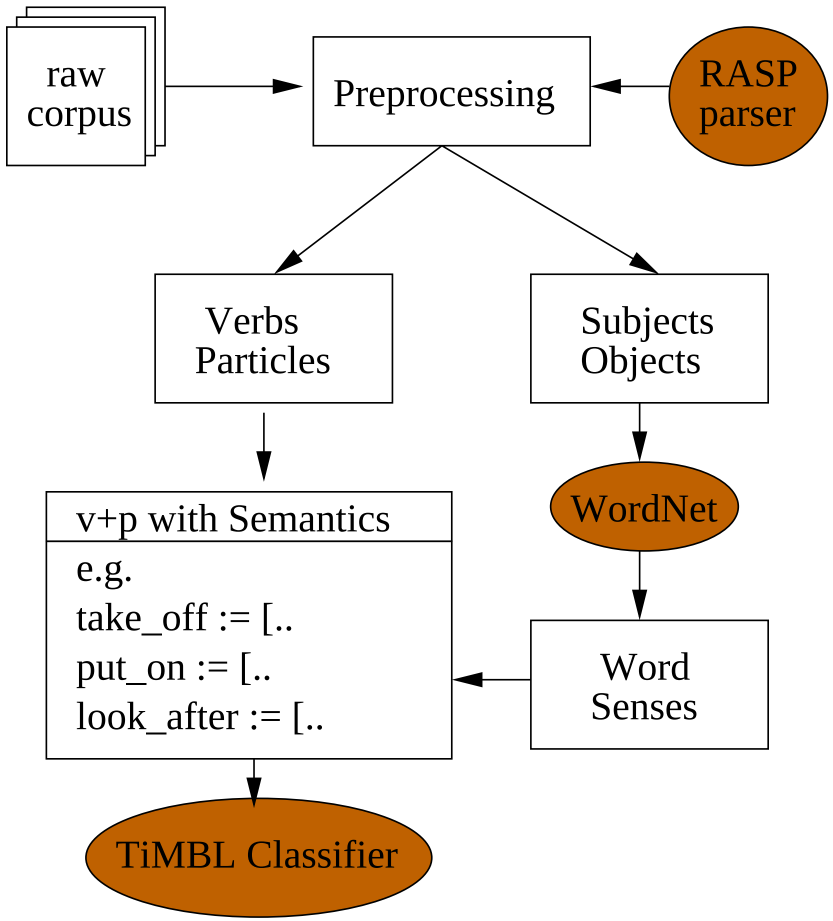
(7) *put on* = wear

EX: *Put* *on* *the* *sweater*.

ARGS: $sweater_{\mathrm{OBJ}}$ = garment, clothing

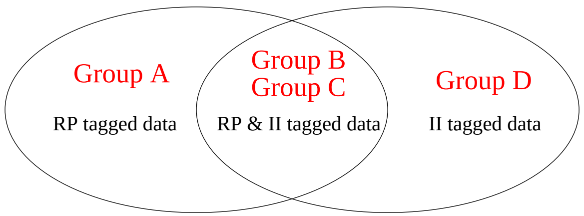
ANALYSIS: verb particle construction

System Architecture



Data (1)

- Classify each V–P token instance according to:
  - ⋆ Group A = P tagged as a particle (RP) only
  - ⋆ Group B = P tagged as a particle (RP), but co-occurs with V elsewhere as a transitive preposition (II)
  - ⋆ Group C = P tagged as a transitive preposition (II), but co-occurs with V elsewhere as a particle (RP)
  - ⋆ Group D = P tagged as a transitive preposition (II) only



Data (2)

- data frequency (f)

	$f \geq 1$		$f \geq 5$	
	VPC	V-PP	VPC	V-PP
Group A	5,223	0	3,787	0
Group B	1,312	0	1,108	0
Group C	0	995	0	217
Total	6,535	995	4,895	217

- False positive rate (FPR), false negative rate (FNR) and inter-annotator agreement

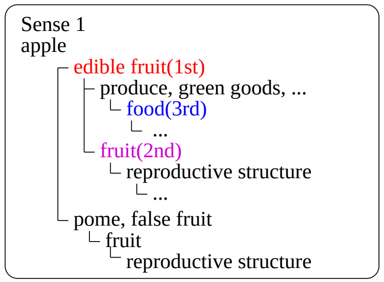
	FPR	FNR	Agreement
Group A	4.08%	—	95.24%
Group B	3.96%	—	99.61%
Group C	—	10.15%	93.27%
Group D	—	3.4%	99.20%

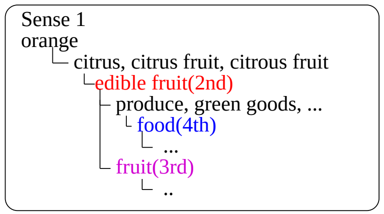
Analysis of Arguments of V–P

- Types of noun arguments (subject + object):

Type	A&B	C	Type	A&B	C	Type	A&B	C
common nn	7,116	1,239	proper nn	156	18	*who*	94	6
personal prn	629	79	demonstrative prn	127	1	*which*	32	0
						No sense (*what*)	11	0

- Word senses of nouns (subject + object) : 1_{st} & 3 first-sense hypernyms
- for Proper nouns, get hypernyms (e.g. *GM:company, Canada:country*)





Evaluation (1)

- Data selection for the evaluation: groups B, BA, BC and BAC

- Data set sizes at different frequency cutoffs:

Group	Frequency of VPCs	Size	
B	$(f_{\geq 1})$	test:	272
	$(f_{\geq 5})$	train:	1,040
BA	$(f_{\geq 1}$ & $f_{\geq 1})$	test:	1,327
	$(f_{\geq 5}$ & $f_{\geq 5})$	train:	4,163
BC	$(f_{\geq 1}$ & $f_{\geq 1})$	test:	498
	$(f_{\geq 5}$ & $f_{\geq 1})$	train:	1,809
BAC	$(f_{\geq 1}$ & $f_{\geq 1}$ & $f_{\geq 1})$	test:	1,598
	$(f_{\geq 5}$ & $f_{\geq 5}$ & $f_{\geq 1})$	train:	5,932

Evaluation (2)

- Results for VPC identification only:

Data	Frequency	Precision	Recall	F-Score
RASP	$f_{\geq 1}$	95.90%	95.50%	95.70%
BC	$f_{\geq 1} f_{\geq 1}$	80.99%	84.56%	82.73%
	$f_{\geq 5} f_{\geq 1}$	83.66%	92.28%	87.76%
BAC	$f_{\geq 1} f_{\geq 1} f_{\geq 1}$	96.21%	96.21%	96.21%
	$f_{\geq 5} f_{\geq 5} f_{\geq 1}$	96.50%	98.40%	**97.44%**

- Results for VPC (=VPC) and Verb-PP (=VPP) identification:

Data	Frequency	Type	Precision	Recall	F-Score
RASP	$f_{\geq 1}$	PV	93.30%	–	–
BC	$f_{\geq 1} f_{\geq 1}$	PV	80.68%	80.33%	80.51%
	$f_{\geq 5} f_{\geq 1}$	PV	86.53%	85.29%	85.91%
BAC	$f_{\geq 1} f_{\geq 1} f_{\geq 1}$	PV	86.60%	86.60%	86.60%
	$f_{\geq 5} f_{\geq 5} f_{\geq 1}$	PV	92.72%	88.36%	**90.54%**

Evaluation (3)

- Results with hypernym expansion (4WS) and only the first sense (1WS)

Freq	Type	#	Precision	Recall	F-score
$f_{\geq 1}$	VPCs	4WS	96.2%	96.2%	96.2%
		1WS	95.8%	96.9%	96.3%
$f_{\geq 1}$	Verb-PPs	4WS	76.9%	76.9%	76.9%
		1WS	80.0%	*74.3%*	77.0%
$f_{\geq 5}$	VPCs	4WS	96.4%	98.3%	**97.4%**
		1WS	95.0%	97.3%	*96.2%*
$f_{\geq 5}$	Verb-PPs	4WS	88.9%	78.3%	**83.2%**
		1WS	81.3%	61.4%	*74.9%*

Results Analysis & Effects of Compositionality

- Expectation: selectional preferences are marked different for VPCs of low compositionality
- Error rate reduction for VPCs of varying compositionality
- 117 VPCs scored w.r.t. compositionality (McCarthy et al. 2003)

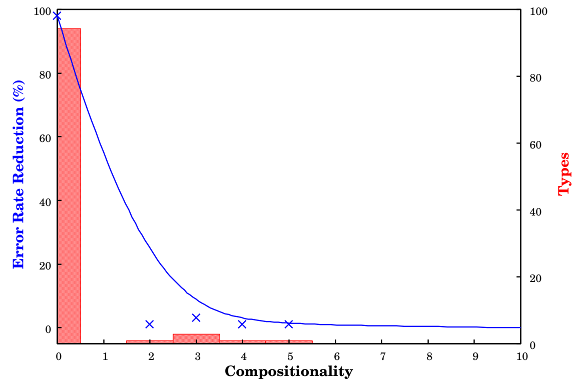


Figure 2: Compositionality & VPC identification

Summary of Identifying VPCs

- Proposed method for VPC identification based on selectional preferences
- Exceed baseline RASP performance & exceed previously-published results for VPC identification (F-score=97.4%)
- Be boosted with hypernym expansion (4WS vs. 1WS)
- Correlate (somewhat) with the relative compositionality of the VPC

Conclusion

- Interpreting Noun Compounds using
  1. constituent similarity (Kim&Baldwin 2005-IJCNLP)
  2. verb semantics (Kim&Baldwin 2006-ACL/Coling)
  3. sense collocation & bootstrapping (Kim&Baldwin 2007-PACLING, Kim&Baldwin 2007-SemEval)
  4. sense collocation & similar words (Kim&Mistica&Baldwin 2007-ALTW)
  5. benchmarking (Kim&Baldwin 2008-IJCNLP)

- Word sense disambiguation(Kim&Baldwin 2007-AAAI)

- Identifying VPCs (Kim&Baldwin 2006-EACL)

- Detecting Compositionality of VPCs (Kim&Baldwin 2007-PACLING)

Applications

- MWEs as semantic units for summarization, Question-Answer (QA) & Information Retrieval(IR)

- SRs for QA & IR
  - ⋆ provide the clues (e.g. *What state is he from?* → location)
  - ⋆ filter the candidates (e.g. *virus infection*:CAUSE=combined with sentence classification)
  - ⋆ enrich queries (e.g. *GM car* → *GM vehicle* is added as a query)

- Compositionality of MWEs for Machine Translation, QA & IR
  - ⋆ provide the clues for word to word alignment (e.g. Venkatapathy&Joshi 2006)
  - ⋆ enrich the queries for IR & QA (e.g. *a piece of cake* as a query)
  - ⋆ fluency for text generation (e.g. *eat* vs *eat up*)

Direction of Further Study

- expand the investigated methods for better performance
- integrate outcomes into NLP applications & crossover/crosslingual study
- related to NCs
  - ⋆ investigate unsupervised methods
  - ⋆ determine & propose a reliable set of SRs along with comparison methods
  - ⋆ deal with SR pragmatism
  - ⋆ utilize the research outcome into a real-world NLP applications
- related to VPCs
  - ⋆ investigate unsupervised methods to extract/identify VPCs
  - ⋆ deal with the measure of degree of VPC compositionality
  - ⋆ utilize the research outcome into a real-world NLP applications

Reading List

- related to **NC interpretation**

1. **Su Nam Kim**, Timothy Baldwin, *Automatic Interpretation of Semantic Relations in Compound Nouns using WordNet Similarity*, 2nd International Joint Conference on Natural Language Processing (IJCNLP), 2005, Jeju island, Republic of Korea, pp.945–956
2. **Su Nam Kim**, Timothy Baldwin, *Interpreting Semantic Relations in Noun Compounds via Verb Semantics*, The Joint Conference of the International Committee on Computational Linguistics and the Association for Computational Linguistics (Coling/ACL), 2006, Sydney, Australia, pp.491–498
3. **Su Nam Kim**, Timothy Baldwin, *MELB-KB: Nominal Classification as Noun Compound Interpretation*, 4th International Workshop on Semantic Evaluations (SemEval), 2007, Prague, Czech Republic, pp.231–236
4. **Su Nam Kim**, Timothy Baldwin, *Interpreting Noun Compound using Bootstrapping and Sense Collocation*, Conference of the Pacific Association for Computational Linguistics (PACLING), 2007, Melbourne, Australia, 129-136
5. **Su Nam Kim**, Timothy Baldwin, *Benchmarking Noun Compound Interpretation*, 3rd International Joint Conference on Natural Language Processing (IJCNLP), 2008, Hyderabad, India (to appear)
6. **Su Nam Kim**, Meladel Mistica, Timothy Baldwin, *Australian Language Technology Workshop*, Melbourne, Australia (to appear)
7. **Su Nam Kim**, Timothy Baldwin, *Noun Compound Interpretation : Feasibility Study of Syntax and Semantics in Noun Compounds*, Journal of Natural Language Engineering (NLE), Cambridge (in preparation)

- related to **VPC**

1. **Su Nam Kim**, Timothy Baldwin, *Automatic Extraction of Verb-Particles Using Linguistic Features*, 11th Conference of European Chapter of the Association for Computational Linguistics : 3rd ACL-SIGSEM Workshop on Preposition, 2006, Trento, Italy, pp.65–72
2. **Su Nam Kim**, Timothy Baldwin, *Detecting Compositionality of English Verb-Particle Constructions using Semantic Similarity*, Conference of the Pacific Association for Computational Linguistics (PACLING), 2007, Melbourne, Australia, pp.40-48
3. **Su Nam Kim**, Timothy Baldwin, *Identifying English Verb-Particle Constructions via Linguistic Features*, Special issue of the International Journal of Language Resources and Evaluation (LRE) (in preparation)

- related to **WSD**

1. **Su Nam Kim**, Timothy Baldwin, *Disambiguating Noun Compound*, 22nd AAAI Conference on Artificial Intelligence (AAAI), 2007, British Columbia, Canada, pp.901-906
2. David Martinez, **Su Nam Kim**, Timothy Baldwin, *MELB-MKB:Lexical Substitution system based on Relatives in Context*, 4th International Workshop on Semantic Evaluations (SemEval), 2007, Prague, Czech Republic, pp.237–240
3. Timothy Baldwin, **Su Nam Kim**, Francis Bond, Sanae Fujita, David Martinez and Takaaki Tanaka, *MRD-based Word Sense Disambiguation: Further Extending Lesk*, 3rd International Joint Conference on Natural Language Processing (IJCNLP), 2008, Hyderabad, India (to appear)
4. Timothy Baldwin, **Su Nam Kim**, Francis Bond, Sanae Fujita, David Martinez and Takaaki Tanaka, *A Reexamination of MRD-based Word Sense Disambiguation*, ACM Transactions on Asian Language Information Processing (in preparation)

Acknowledgment

- full credit to Dr.Timothy Baldwin for all the work we've been done together
- partial credit to Dr.David Martinez for word sense disambiguation task
- partial credit to Meladel Mistica for noun compound interpretation task
- many thanks to Dr.Timothy Baldwin, Dr. Steven Bird, Dr.David Martinez and people in LT group