



Automatic Acquisition of Knowledge About Multiword Predicates

Afsaneh Fazly and Suzanne Stevenson

Department of Computer Science

University of Toronto

We thank our collaborator Ryan North, and other members
of the CL research group at the University of Toronto.
The support of NSERC of Canada is gratefully acknowledged.



The Role of Metaphor in Language

- Allows creative expression, precise connotations:
 - *Juliet is the sun.* [Romeo & Juliet]
 - *Life's but a walking shadow.* [Macbeth]
 - Not just “flowery language”: [Lakoff and others]
 - *KMT pulls out the big guns as elections draw near.*
 - *Bush's recent fruitless trip to China.*
 - *The end of an era in football broadcasting.*
 - *Premier says he won't resign, but will step down.*
- ➔ Requires special computational attention.



“Basic” Verbs

- Express actions or states that are central to human experience:
 - *give, hear, put, see, sit, stand, take*, among others.
- Are observed crosslinguistically to be highly frequent and highly polysemous.
- Easily undergo metaphorization:
 - *The files sat on my desk all week.*
 - *I see your point of view.*
 - *The house stands at the corner of Main Street.*



Multiword Predicates (MWP)

- Basic verbs combine with many different words to form a range of **multiword predicates**:
 - *cut in line, cut (someone) a break, cut a dash*
 - *give a speech, give a groan, give ground*
 - *put (something) to rest, put one's finger on*
 - *sit in judgment, sit tight, sit on the fence*
- ➔ We focus on MWPs of the form **V_{basic} + N**:
 - frequent across a wide range of languages.
 - the basic verb takes on a range of metaphorical meaning extensions.



Computational Issues [1 of 2]

- What is the meaning contribution of the basic verb to an expression?

- **literal:** *give a present*
- **metaphorical:** *give a speech, give a groan*
- **idiomatic:** *give ground, give a wide berth*

➔ Metaphoricity affects translation, paraphrase:

- *give a present* ➔ *donner*_[give] *un*_[a] *cadeau*_[present]
- *give a groan* ➔ *gémir*_[to groan]
- *give ground* ➔ *reculer*_[to draw back]



Computational Issues [2 of 2]

- What complements can a basic verb combine with to form an MWP?
 - individual **acceptability** of potential MWPs.
 - **productivity** of combining with a semantic class of complements.
- ➔ Productivity indicates generalizability, as well as a possible meaning extension of the basic verb:
 - *give a speech, give a talk, ...* ➔ abstract transfer
 - *give a groan, give a howl, ...* ➔ emission



The Metaphoricity Continuum



literal

metaphorical

idiomatic

- Literal phrases: *give a present*
 - can be interpreted by compositional rules of grammar
- Metaphorical verb: *give a speech*, *give a groan*
 - verb contributes a metaphorical meaning
 - noun contributes a predicative meaning
- Idiomatic expression: *give ground*, *give the boot*
 - non-compositional interpretation



Distinguishing Idioms from Literal VPs

- Idiomatic MWPs conform to the grammar rules for VPs; however, they are:
 - More lexically fixed:
 - *give ground* *give earth* nor *donate ground*
 - More syntactically fixed:
 - ?? *Kiva gave the ground.*
 - ?? *Kiva gave grounds.*
 - ?? *Kiva gave tenuous ground.*
- Use statistical measures of fixedness to indicate degree of idiomaticity of a verb+noun.



Measuring Lexical Fixedness

- Use association strength (PMI) between a verb and noun as an indicator of idiomaticity.
- Compare the strength of association of the target V+N to that of V with related Ns:
 - PMI(give, ground) >> ?
PMI(give, earth), PMI(give, dirt), PMI(give, land), ...
- ➔ Novel technique for combining association strengths into single lexical fixedness measure.



Measuring Syntactic Fixedness

- We determine syntactic patterns that are resistant to variation in idiomatic MWPs.
 - Determiner use, singular/plural, modification, etc.:
 - ?? *Kiva gave the ground.*
 - ?? *Kiva gave grounds.*
 - ?? *Kiva gave tenuous ground.*
- Calculate probability distribution over patterns.
- ➔ Divergence of usage of target V+N from typical VP usage yields measure of syntactic fixedness.



Experimental Set-Up

- 28 basic verbs taken from linguistic literature.
 - *cut, find, give, kick, lose, put, smell, take, ...*
- V+N combinations extracted from BNC ($f \geq 10$).
- Idioms determined from dictionaries of idioms.
 - *cut one's losses, smell the roses, lose track*
- 100 literal and 100 idiomatic V+Ns in test set.
- Classified into two equal sets according to measures.



Literal vs. Idiomatic: Results

	Accuracy	Error Reduction
Random Baseline	.50	—
PMI [informed baseline]	.64	.28
Lexical Fixedness	.65	.30
Syntactic Fixedness	.70	.40
Lex+Syn Fixedness	.74	.48

- Syntactic and combined fixedness measures perform very well.
- Fixedness measures are less sensitive than PMI to frequency of items.



Literal vs. Idiomatic: Results

	Accuracy	Error Reduction
Random Baseline	.50	—
PMI [informed baseline]	.64	.28
Lexical Fixedness	.65	.30
Syntactic Fixedness	.70	.40
Lex+Syn Fixedness	.74	.48

- Syntactic and combined fixedness measures perform very well.
- Fixedness measures are less sensitive than PMI to frequency of items.



Literal vs. Idiomatic: Results

	Accuracy	Error Reduction
Random Baseline	.50	—
PMI [informed baseline]	.64	.28
Lexical Fixedness	.65	.30
Syntactic Fixedness	.70	.40
Lex+Syn Fixedness	.74	.48

- Syntactic and combined fixedness measures perform very well.
- Fixedness measures are less sensitive than PMI to frequency of items.



Literal vs. Idiomatic: Results

	Accuracy	Error Reduction
Random Baseline	.50	—
PMI [informed baseline]	.64	.28
Lexical Fixedness	.65	.30
Syntactic Fixedness	.70	.40
Lex+Syn Fixedness	.74	.48

- Syntactic and combined fixedness measures perform very well.
- Fixedness measures are less sensitive than PMI to frequency of items.



Literal vs. Idiomatic: Results

	Accuracy	Error Reduction
Random Baseline	.50	—
PMI [informed baseline]	.64	.28
Lexical Fixedness	.65	.30
Syntactic Fixedness	.70	.40
Lex+Syn Fixedness	.74	.48

- Syntactic and combined fixedness measures perform very well.
- Fixedness measures are less sensitive than PMI to frequency of items.



Detecting Level of Metaphoricity

- Focus on intermediate level of metaphoricity.
- MWPs vary in their degree of metaphoricity.
 - Literal *give*: physical transfer of possession.
 - In *give a speech*, *give* retains “transfer” meaning.
 - In *give a groan*, no element of “transfer”.
- These MWPs differ from literal phrases in that the **noun** is the primary source of predication:
 - *give a speech* can be paraphrased as *speak*.
 - *give a groan* can be paraphrased as *groan*.



Syntactic Fixedness Again Plays a Role

- Metaphoricity of MWP's related to fixedness:
 - Less metaphorical:
 - *Kiva gave a speech.*
 - *Kiva gave the speech.*
 - *A speech was given by Kiva.*
 - More metaphorical:
 - *Kiva gave a groan.*
 - ?? *Kiva gave the groan.*
 - ?? *A groan was given by Kiva.*
- ➔ Measure difference in strength of association between preferred and less preferred patterns.



Syntactic Fixedness Again Plays a Role

- Metaphoricity of MWP's related to fixedness:
 - Less metaphorical:
 - *Kiva gave a speech.*
 - *Kiva gave the speech.*
 - *A speech was given by Kiva.*
 - More metaphorical:
 - *Kiva gave a groan.*
 - ?? *Kiva gave the groan.*
 - ?? *A groan was given by Kiva.*
- ➔ Measure difference in strength of association between **preferred** and less preferred patterns.



Syntactic Fixedness Again Plays a Role

- Metaphoricity of MWP's related to fixedness:
 - Less metaphorical:
 - *Kiva gave a speech.*
 - *Kiva gave the speech.*
 - *A speech was given by Kiva.*
 - More metaphorical:
 - *Kiva gave a groan.*
 - ?? *Kiva gave the groan.*
 - ?? *A groan was given by Kiva.*
- ➔ Measure difference in strength of association between preferred and **less preferred** patterns.



Experimental Set-Up

- Focus on two highly frequent basic verbs in English: *give* and *take*.
- Extract "*give/take* + a/an + N" combinations (as in *give a groan*) from the BNC.
- Level of metaphoricity given by human judges.
- 147 expressions (79 for *give* and 68 for *take*):
 - metaphoricity: 54 high, 39 moderate, and 54 literal.



Level of Metaphoricity: Results

- Measure compared to human judgments using Spearman rank correlation:

	<i>give</i>	<i>take</i>
PMI [informed baseline]	.68	.63
Syntactic Fixedness	.75	.72

- All correlations are highly statistically significant.
- Improvement over PMI shows that level of metaphoricity is more than degree of collocation.



Summary: Metaphoricity Continuum



literal

metaphorical

idiomatic

- Good results in distinguishing literal from idiomatic expressions.
- Good correlations with judgments of level of metaphoricity of intermediate expressions.
- ➔ Future work: Combine measures into one score that places any V+N expression on continuum.



Summary: Metaphoricity Continuum



literal

metaphorical

idiomatic

- Good results in **distinguishing literal from idiomatic** expressions.
- Good correlations with judgments of level of metaphoricity of intermediate expressions.
- ➔ Future work: Combine measures into one score that places any V+N expression on continuum.



Summary: Metaphoricity Continuum



literal

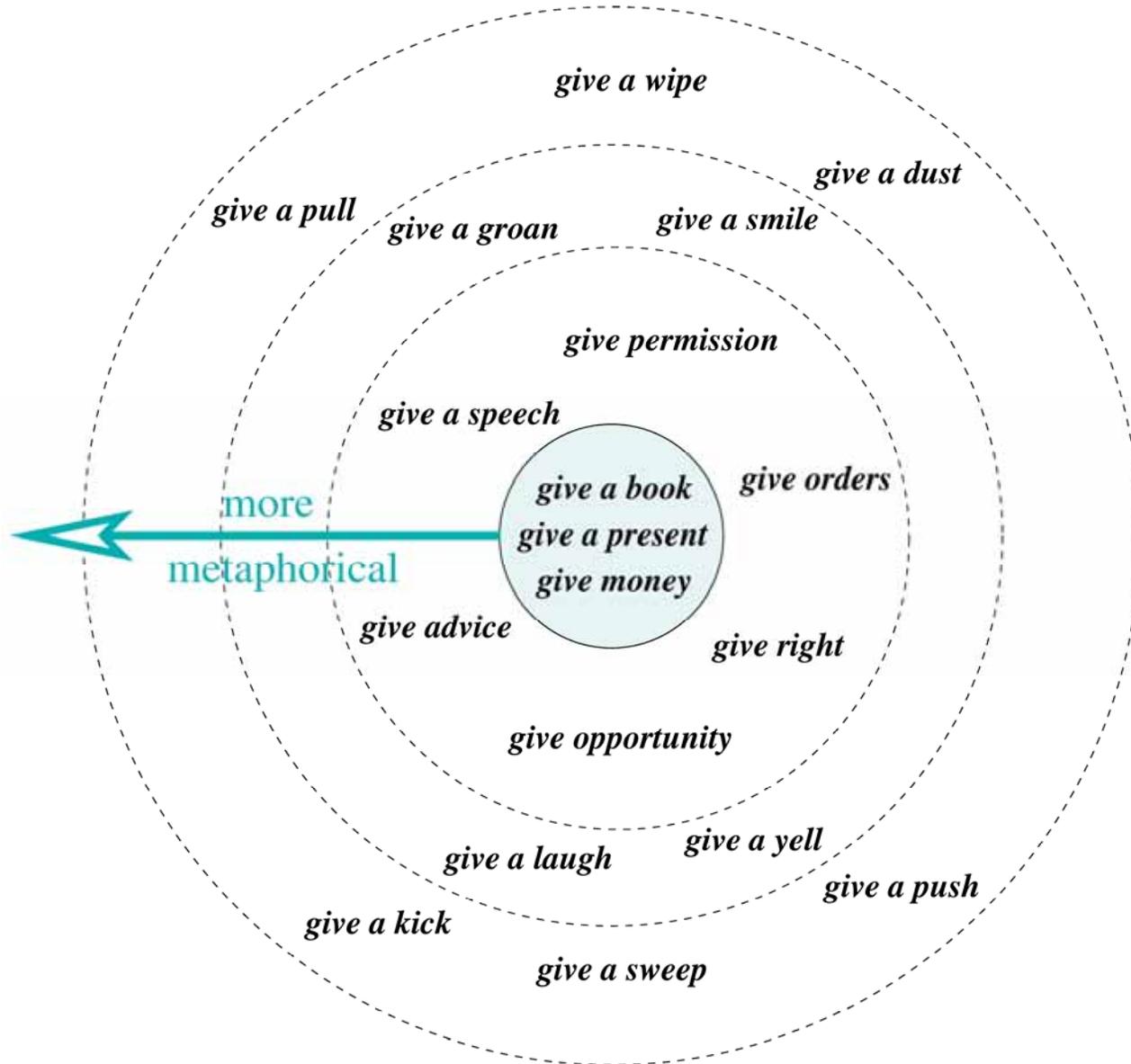
metaphorical

idiomatic

- Good results in distinguishing literal from idiomatic expressions.
- Good correlations with judgments of **level of metaphoricity** of intermediate expressions.
- ➔ Future work: Combine measures into one score that places any V+N expression on continuum.



Example Metaphority "Continuum": *give*





Semantic Patterns in MWP Formation

- Recall: What complements can a basic verb combine with to form an MWP?
 - individual **acceptability** of potential MWPs.
 - **productivity** of over a class of complements.
- Focus on intermediate metaphoricity, since these MWPs show predictability of combination.
 - *give a speech, talk, presentation, demo, ...*
 - *give a groan, howl, sigh, moan, ...*



Measures of Acceptability

- Recall: Intermediate MWP's show preferred and less preferred patterns of usage:
 - *Kiva gave a groan.*
 - ?? *Kiva gave the groan.*
 - ?? *A groan was given by Kiva.*
- **PMI:MWP** measure uses information about collocations with linguistically preferred patterns.
- **Prob:MWP** measure incorporates more linguistic information about preferred combinations.



Measures of Productivity

- Class-based behaviour can enable us to extend acceptability knowledge to new expressions:
 - observe *give a groan*, *give a howl*, *give a moan*
 - ➔ unseen *give a rasp* should be promoted.
- Extend acceptability measures to measures of productivity across a class of complements.
 - **Productivity:**
 - the proportion of class members that form acceptable MWPs with a given basic verb.



Experimental Set-Up

- Again focus on two highly frequent basic verbs in English: *give* and *take*.
- Take complements from semantically related sets of nouns in WordNet. (Four test classes.)
- Extract counts of "*give/take*+a/an+N", as well as other needed counts, from the web.
- Compare our measures to levels of acceptability and productivity given by human judges.



Acceptability and Productivity: Results

- Spearman rank correlation with human ratings :

Acceptability	<i>give</i>	<i>take</i>
PMI:MWP		
Prob:MWP		

Greyscale indicates level of correlation (.30 to over .70)

- Divergence of productivity from human judgments:

Productivity	Sum of Error ²	Mean Abs Error
PMI:MWP	.057	.11
Prob:MWP	.035	.07

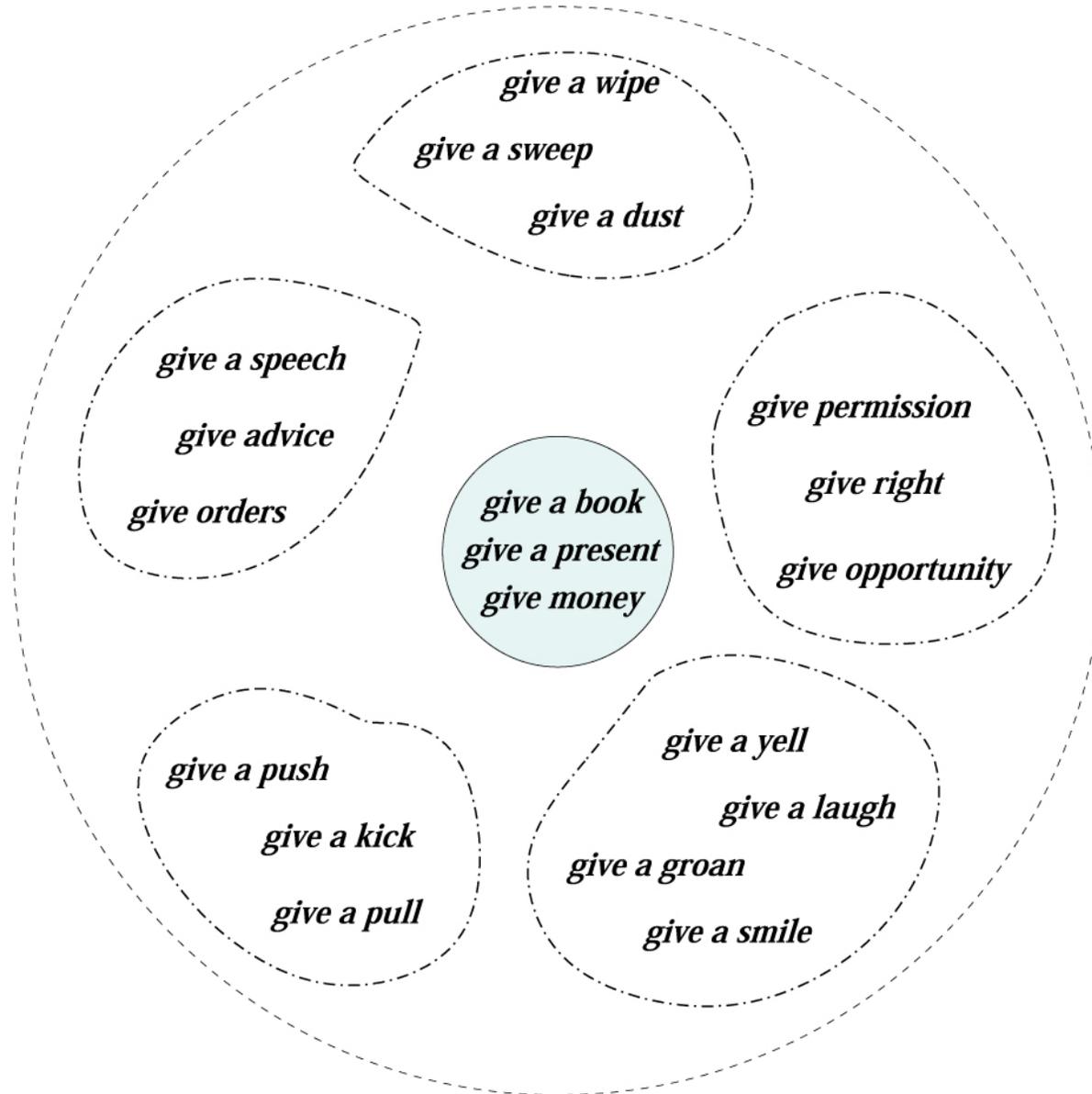


Summary: Acceptability and Productivity

- Linguistically informed probability measure has very good correlations with human acceptability.
- The same measure also shows a very good match with human judgments on productivity.
- Future work: Meaning extensions of basic verbs correspond to semantic sets of complements:
 - Our goal is to refine the semantic space of these highly polysemous verbs.



Example Semantic Refinement: *give*





Contributions: The Role of Metaphor

- We take a statistical corpus-based approach to the handling of metaphor in “everyday” language.
 - Other work lacks specific computational proposals or relies on expensive knowledge-based resources.

[Fass 91, Fellbaum et al. 05, Villavicencio et al. 04]

- We identify the central role of metaphor in the treatment of highly polysemous verbs.
 - Previous automatic acquisition techniques rely on domain distinctions that do not extend to such verbs.

[Mason 04]



Contributions: Multiword Expressions

- We focus on MWPs using basic verbs, a frequent class of expressions across diverse languages.
 - Most work on multiword expressions examines compound nouns and verb-particle constructions.
[Though see Venkatapathy & Joshi, 2005]
- We analyze linguistic properties of MWPs and relate them to their statistical behaviour.
 - Prior work is limited to surface-level collocational analysis or measurement of distributional similarity.
[Smadja 93, Baldwin et al. 03, Bannard et al. 03, McCarthy et al. 03; though see Lin 99, Wermter & Hahn 05]



Contributions: Novel Statistical Measures

- We develop measures of fixedness for placing MWPs on the metaphoricity continuum, enabling:
 - appropriate handling of their syntax and semantics.
- We devise measures for capturing behaviour over classes of potential complements, supporting:
 - generalization of lexical knowledge.
 - refinement of semantics of highly polysemous verbs.
- Our on-going work aims to extend our techniques to other languages, and other types of MWPs.