Compositionality and Multiword Expressions: Six of One, Half a Dozen of the Other?

Timothy Baldwin

THE UNIVERSITY OF MELBOURNE

POSTERA CRESCAM LAUDE

INTRODUCTION

What are Multiword Expressions (MWEs)?

- *Definition:* A **multiword expression** (MWE) is:

  1. decomposable into multiple simplex words
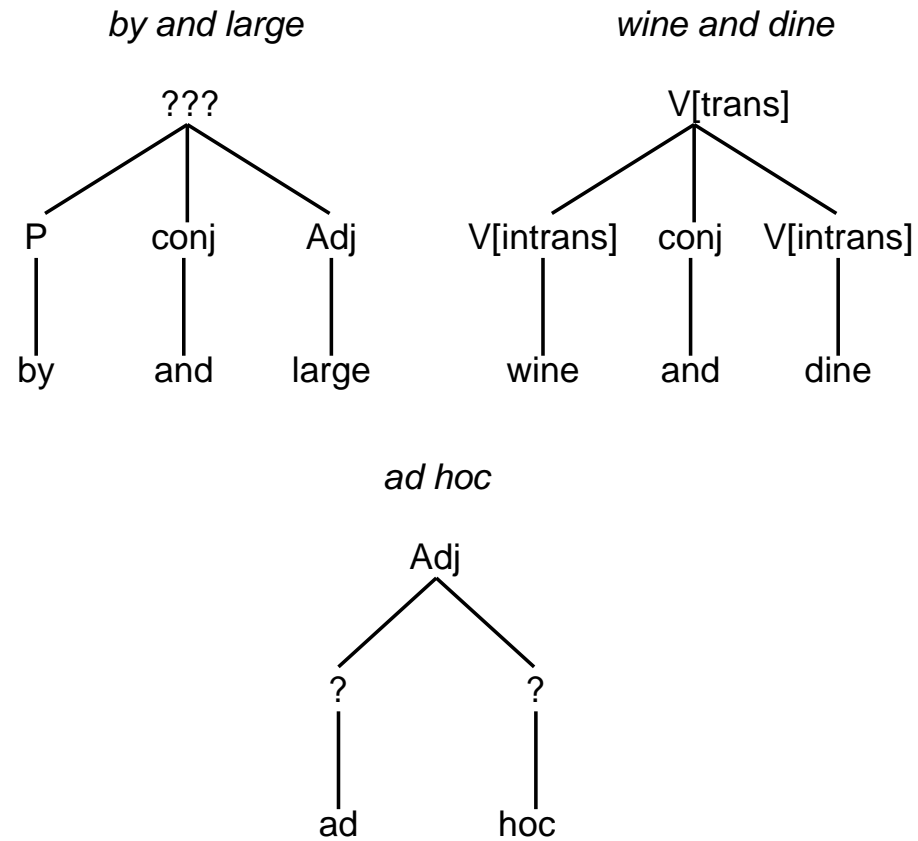 2. lexically, syntactically, semantically, pragmatically and/or statistically idiosyncratic

Some Examples

- *San Francisco, ad hoc, by and large, Where Eagles Dare, kick the bucket, part of speech, in step, the Oakland Raiders, trip the light fantastic, telephone box, call (someone) up, take a walk, do a number on (someone), take (unfair) advantage (of), pull strings, kindle excitement, fresh air, ...*
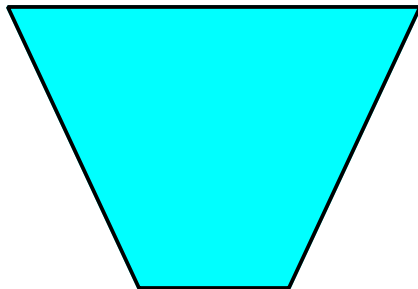
MWE or not MWE?

... there is no unified phenomenon to describe but rather a complex of features that interact in various, often untidy, ways and represent a broad continuum between non-compositional (or idiomatic) and compositional groups of words. (Moon 1998)
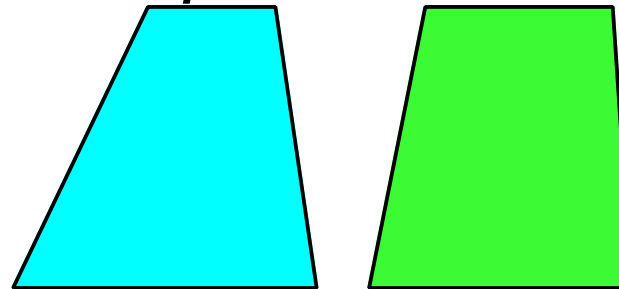
Lexicosyntactic Idiomaticity

by and large

```
        ???
       / |  \
      P  conj  Adj
      |    |    |
      by  and  large
```

wine and dine

```
        V[trans]
       /   |    \
 V[intrans] conj V[intrans]
      |     |     |
     wine  and   dine
```

ad hoc

```
       Adj
      /   \
     ?     ?
     |     |
     ad   hoc
```

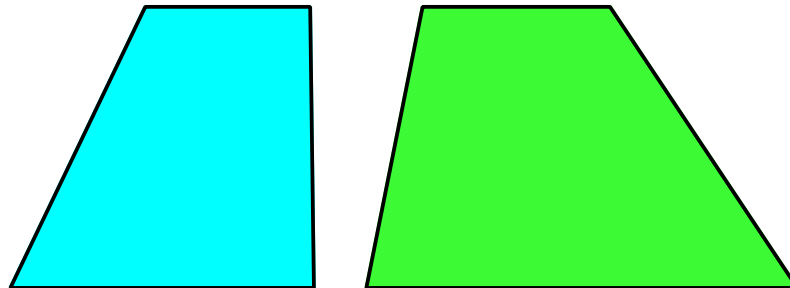# Semantic Idiomaticity

kick the bucket

`die'`

spill the beans

`reveal'(secret')`

kindle excitement

`kindle'(excitement')`

Pragmatic idiomaticity

- Situatedness: the expression is associated with a fixed pragmatic point

  ⋆ situated MWEs: *good morning, all aboard*
  ⋆ non-situated MWEs: *first off, to and fro*

- The "Wheel of Fortune" factor — how to represent the jumble of phrases stored in the mental lexicon?

- The "Monty Python" factor — mish-mash of evocative language fragments

Sag *et al.* (2002); Jackendoff (1997)

Statistical Idiomaticity

	unblemished	spotless	flawless	immaculate	impeccable
eye	−	−	−	−	+
gentleman	−	−	?	−	+
home	?	+	−	+	?
lawn	−	−	?	+	−
memory	−	−	+	−	?
quality	−	−	−	−	+
record	+	+	+	+	+
reputation	+	−	−	+	+
taste	−	−	−	−	+

Adapted from Cruse (1986)

MWE Markedness

MWE	Markedness				
	Lex	**Syn**	**Sem**	**Prag**	**Stat**
ad hominem	☑	?	?	?	☑
at first	☒	☑	☒	☒	☒
first aid	☒	☒	☑	☒	?
salt and pepper	☒	☒	☒	☒	☑
good morning	☒	☒	☒	☑	☑
cat's cradle	☑	☑	☑	☒	?

Other Indicators of MWE-hood

- Institutionalisation/conventionalisation

- Non-identifiability: meaning cannot be predicted from surface form

  ⋆ idiom of decoding (non-identifiable): *kick the bucket, fly off the handle*
  ⋆ idiom of encoding (identifiable): *wide awake, plain truth*

- **Figuration:** the expression encodes some metaphor, metonymy, hyperbole, etc

    * figurative expressions: *bull market, beat around the bush*
    * non-figurative expressions: *first off, to and fro*

- **Single-word paraphrasability:** the expression has a single word paraphrase

  - ⋆ paraphrasable MWEs: $leave\ out = omit$
  - ⋆ non-paraphrasable MWEs: $look\ up$
  - ⋆ paraphrasable non-MWEs: $take\ off\ clothes = undress$

- Proverbiality: the expression is used "to describe—and implicitly, to explain—a recurrent situation of particular social interest ... in virtue of its resemblance or relation to a scenario involving homely, concrete things and relations" (Nunberg *et al.* 1994)

  - ⋆ informality: the expression is associated with more informal or colloquial registers
  - ⋆ affect: the expression encodes a certain evaluation of affective stance toward the thing it denotes

- Prosody: the expression has a distinctive stress pattern which diverges from the norm

  - ⋆ prosodically-marked MWE: *soft spot*
  - ⋆ prosodically-unmarked MWE: *first aid, red herring*
  - ⋆ prosodically-marked non-MWE: *dental operation*

Fillmore *et al.* (1988); Liberman and Sproat (1992); Nunberg *et al.* (1994)
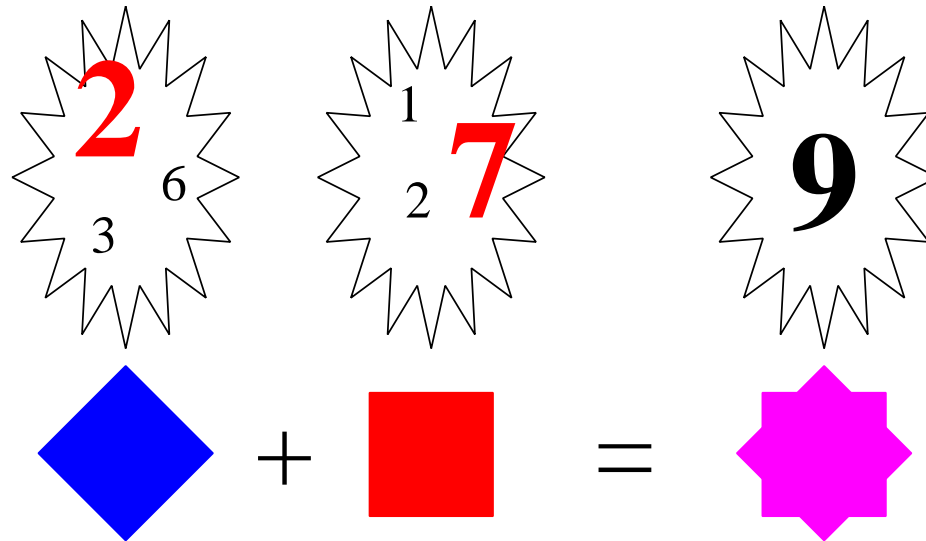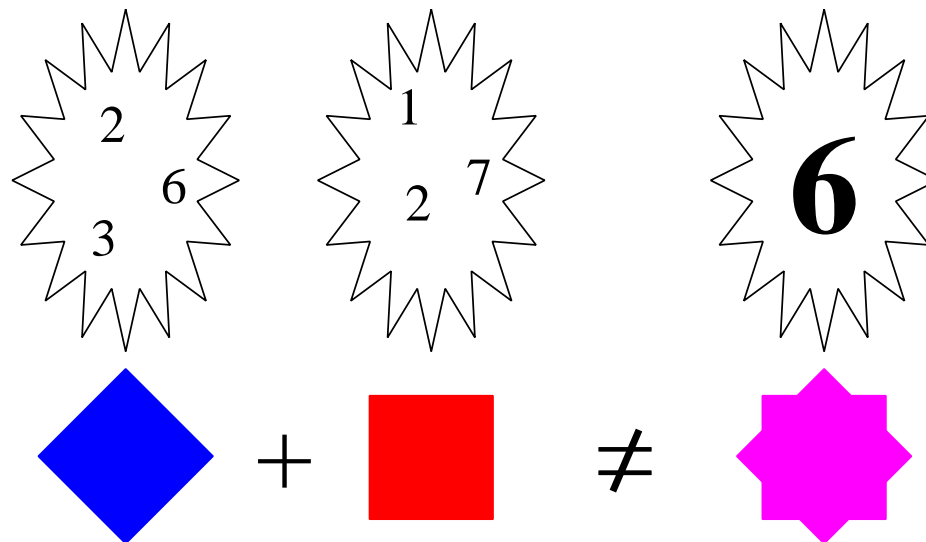
COMPOSITIONALITY

Compositionality

- **Definition:** *degree to which the features of the parts of an MWE combine to predict the features of the whole*
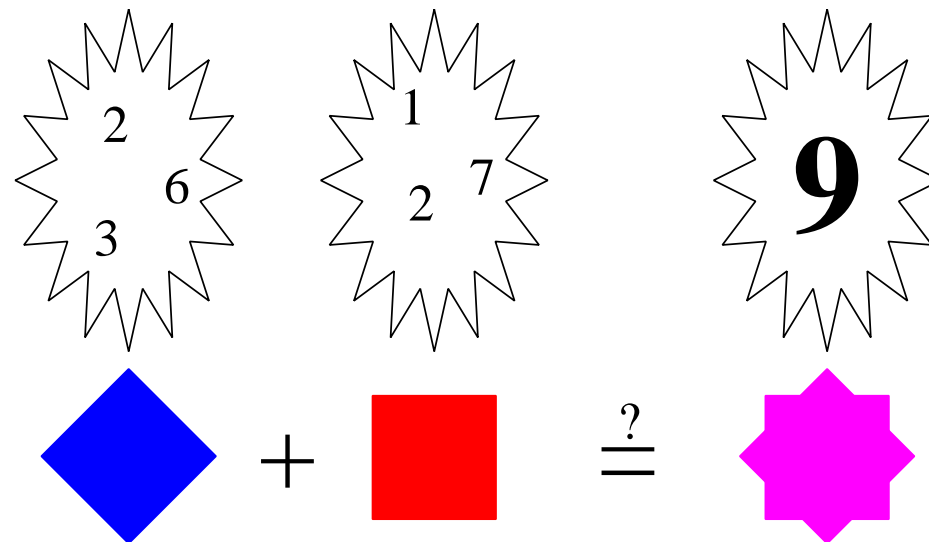
Compositionality

- **Definition:** *degree to which the features of the parts of an MWE combine to predict the features of the whole*

Compositionality

- **Definition:** *degree to which the features of the parts of an MWE combine to predict the features of the whole*

Compositionality

- Generally considered in the context of semantic compositionality, but we can equally talk about:

  ⋆ lexical compositionality
  ⋆ syntactic compositionality
  ⋆ pragmatic compositionality

Example: Syntactic Compositionality

- **Definition:** *Degree to which the* <u>*syntactic*</u> *features of the parts of an MWE combine to predict the syntax of the whole*

  - ⋆ Fixed expressions: *by and large, San Francisco*
  - ⋆ Verb particles: *eat up* vs. *chicken out*

- Syntactic compositionality binary effect; non-compositional MWEs lexicalised

Question

- Given that compositionality extends over all aspects of markedness that affect MWEs, it is the be all and end of all of MWEs?

Question

- Given that compositionality extends over all aspects of markedness that affect MWEs, it is the be all and end of all of MWEs?

  Almost, but there are subtleties due to:
  - ⋆ statistical markedness
  - ⋆ decomposability

Statistical Markedness (Revisited)

- Statistical markedness is (often) a reflection of a **lack** of statistical **non**-compositionality, rather than a lack of compositionality:

$$p(\text{impeccable N}) \times p(\text{Adj eye}) \approx p(\text{impeccable eye})$$

BUT

$$p(\text{unblemished N}) \times p(\text{Adj eye}) \gg p(\text{unblemished eye})$$
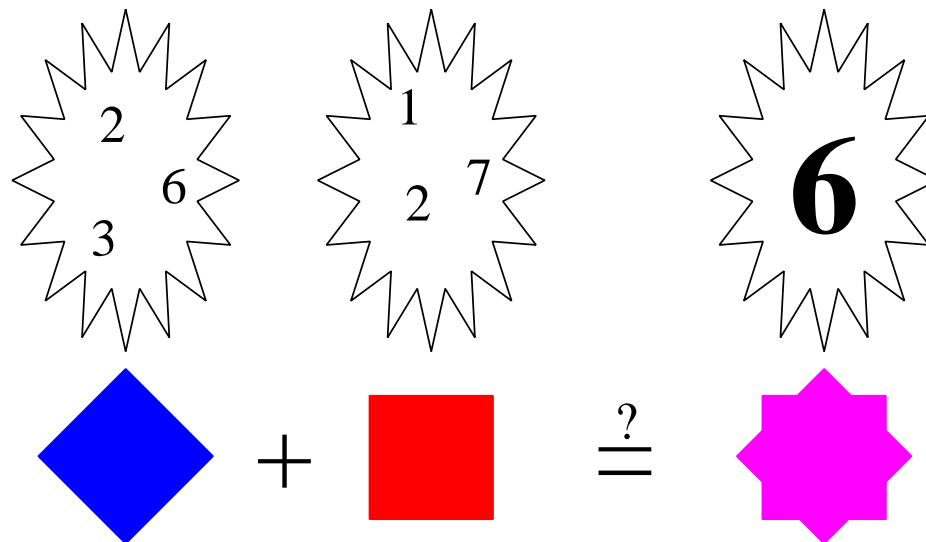$$p(\text{spotless N}) \times p(\text{Adj eye}) \gg p(\text{spotless eye})$$
$$p(\text{flawless N}) \times p(\text{Adj eye}) \gg p(\text{flawless eye})$$
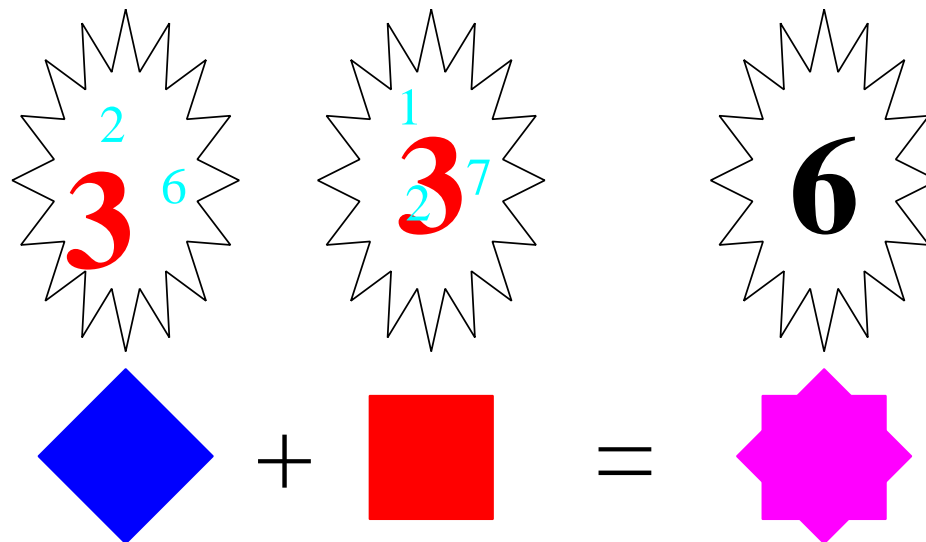$$\vdots$$

Decomposability

- Decomposability = *degree to which the features of an MWE can be ascribed to those of its parts*
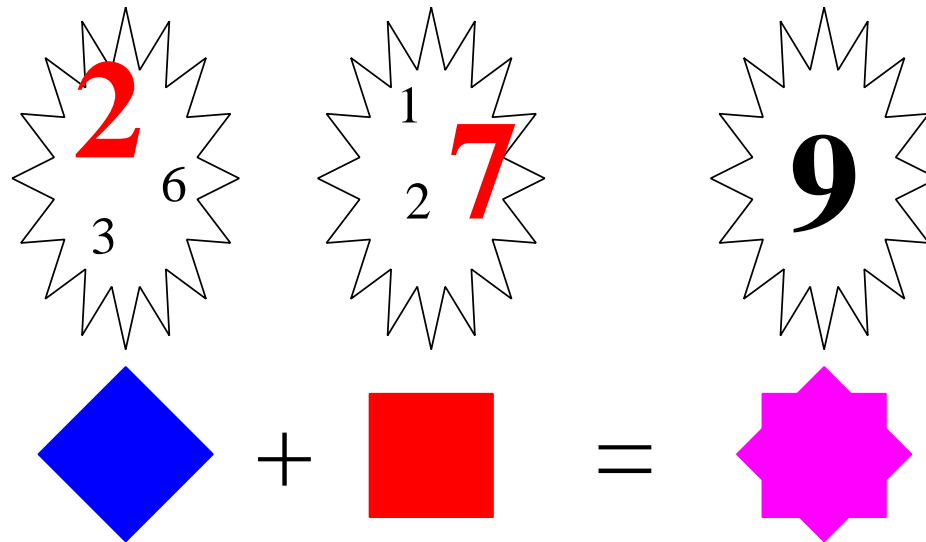
Decomposability

- Decomposability = *degree to which the features of an MWE can be ascribed to those of its parts*

Decomposability

- Decomposability $=$ *degree to which the features of an MWE can be ascribed to those of its parts*

Decomposability

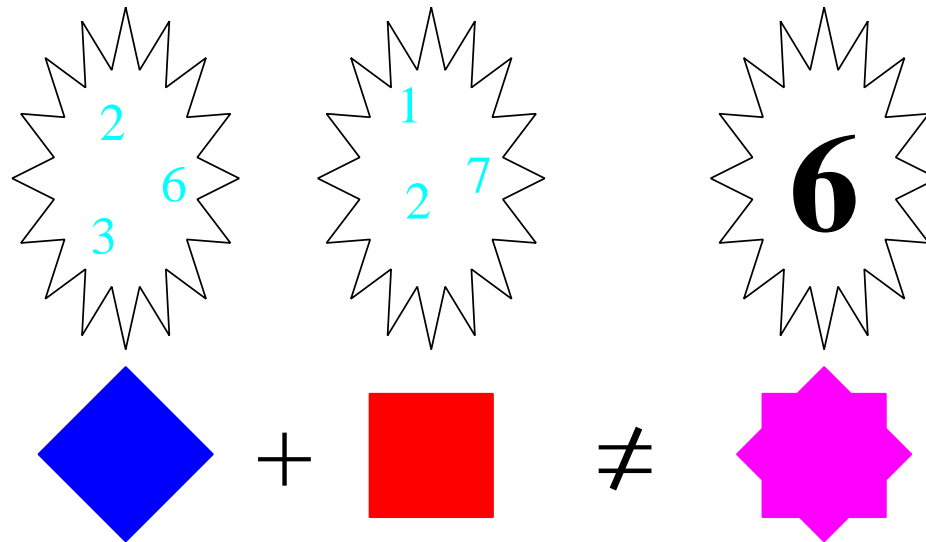- Decomposability = *degree to which the features of an MWE can be ascribed to those of its parts*

Decomposability and Semantic Idiomaticity

kick the bucket

`die'`
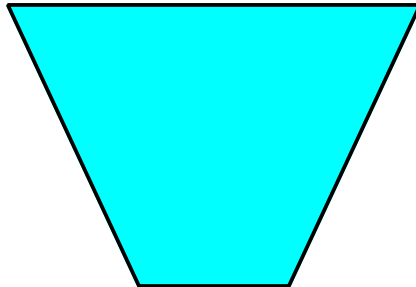
spill the beans
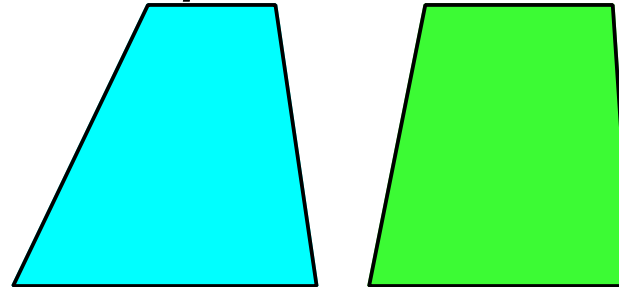
`reveal'(secret')`
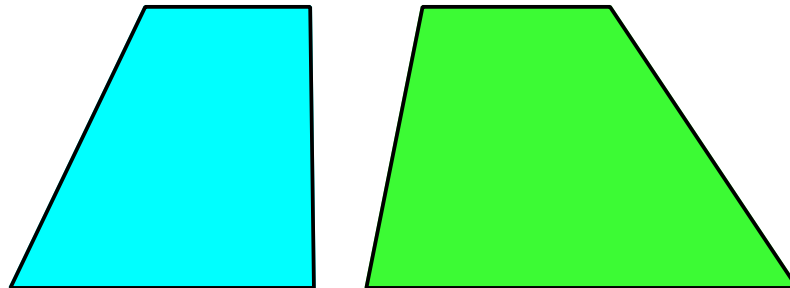
kindle excitement

`kindle'(excitement')`

Decomposability: Three Classes of MWE

- Classification of MWEs into 3 classes:

1. **non-decomposable MWEs** (e.g. *kick the bucket, shoot the breeze, hot dog*)
2. **idiosyncratically decomposable MWEs** (e.g. *spill the beans, let the cat out of the bag, radar footprint*)
3. **simple decomposable MWEs** (e.g. *kindle excitement, traffic light*)

- There is a cline of "markedness" for idiosyncratically decomposable MWEs (e.g. *chicken out* vs. *home office* vs. *radar footprint*)

Decomposability and Syntactic Flexibility

- Consider:

  the bucket was *kicked* by Kim
  Strings were *pulled* to get Sandy the job.
  The FBI *kept* closer *tabs* *on* Kim than they *kept* *on* Sandy.
  ... the considerable *advantage* that was *taken* *of* the situation

- The syntactic flexibility of an idiom can generally be explained in terms of its decomposability

So What was the Answer to our Question?

- Yes and no:

  ⋆ simple compositionality is adequate for describing many instances of lexical, syntactic, semantic and pragmatic markedness

  ⋆ BUT our notion of compositionality is significantly different for statistically-marked MWEs

  ⋆ AND decomposability diffuses the markedness boundary

Nunberg *et al.* (1994)

And Why was it we Care about Compositionality?

- For all the reasons we care about MWEs:

  ⋆ Lexicography/dictionary making
  ⋆ Idiomaticity (coherent semantics)
  ⋆ Overgeneration
  ⋆ Undergeneration
  ⋆ Relevance in applications, including MT, IR, QA, ...

REPRESENTING AND MODELLING COMPOSITIONALITY

Methods for Representing Compositionality

- **Dictionary based**: binary evaluation, based on prediction that non-compositional MWEs will be lexically listed

- **Ontology based**: relative similarity of the parts to the whole (e.g. relative to WordNet)

$$\text{sim}(pig\ metal\ ,metal) \gg \text{sim}(pig\ metal\ ,pig)$$

- **Entailment based**: binary evaluation, based on whether the whole "entails" the parts or not

  Susan finished up her paper
        $\models$ *Susan finished her paper*
        $\not\models$ *Susan's paper was up*

- **Ranking based**: describe MWE compositionality by way of continuous/discrete scale of compositionality

  $$\text{comp}(\textit{put up}) \geq \text{comp}(\textit{eat up}) \geq \text{comp}(\textit{gun down})$$

(Lin 1999; Evert and Krenn 2001; Krenn and Evert 2001; Bannard *et al.* 2003; McCarthy *et al.* 2003; Copestake 2003)

- **Class based**: interpret MWEs relative to a discrete set of semantic classes, each of which is (implicitly) associated with varying levels of compositionality

  home brewing vs. *home town* vs. *home stretch* vs. *home secretary*

(Lin 1999; Evert and Krenn 2001; Krenn and Evert 2001; Bannard *et al.* 2003; McCarthy *et al.* 2003; Copestake 2003)

Ability of the Different Methods to Represent Compositionality

Method	Markedness				
	Lex	Syn	Sem	Prag	Stat
Dictionary	☑	☑	☑	☑	☑
Ontology	?	?	☑	☑	?
Entailment	?	?	☑	☒	☒
Ranking	☑	☑	☑	☑	☑
Class	☑	☑	☑	☑	☑

- Question of the context-sensitivity of these methods

Methods for Modelling Compositionality

- Substitutability

- Distributional similarity

- Semantic similarity

- Interpretational

- Statistical tests

- Linguistic properties

- Co-occurrence properties

Substitutability

- **Assumption:** MWEs stand in opposition to **anti-collocations**, i.e. expressions derived through synonym/word order substitution which occur with markedly lower frequency than the base MWE (or not at all):

Substitutability

Lexicalisation   **Concept**
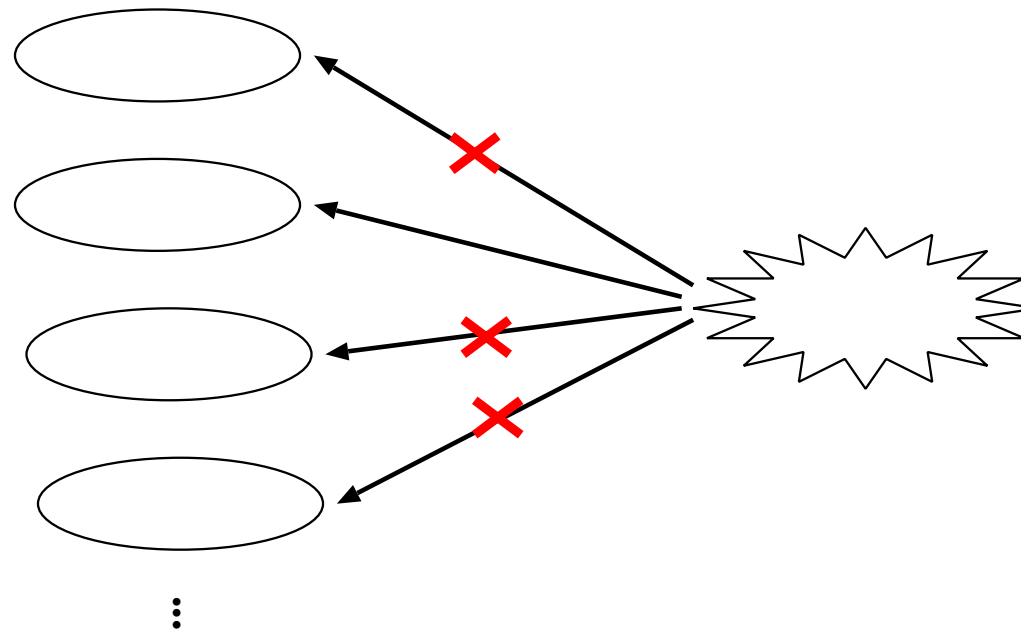
Substitutability

- Is substitution really a good test for (non-) compositionality?

  ⋆ institutionalised phrases: *frying pan*, *salt and pepper*, *many thanks*
  ⋆ productive MWEs: *call/phone/ring up*

Pearce (2001b); Pearce (2001a); Bannard *et al.* (2003); McCarthy *et al.* (2003)

Distributional Similarity

- **Assumption:** if an MWE is compositional, it will occur in the same lexical context as its parts

- Simple extension of the distributional hypothesis (as standardly applied to simplex words)

Operationalisation of Distributional Similarity: Examples

- **overlap:** relative overlap between the top N neighbours of the VPC and its simplex verb

- **sameparticle:** the number of VPCs which select for the same particle as the given VPC amongst the top N neighbours of that VPC

- **sameparticle** $-$ **simplex:** the value for **sameparticle** minus the number of top N neighbours of the simplex

verb which select for that same particle

- **simplexasneighbour:** does the simplex verb occur in the top 50 neighbours of the VPC?

- **rankofsimplex:** what is the rank of the simplex verb in the neighbours of the VPC?

- **overlapS:** the overlap of neighbours in the top N neighbours of the VPC and simplex verb, where VPC neighbours are converted to simplex verbs in the VPC case

Semantic Similarity

- **Assumption:** similarity of the parts to the whole (e.g. relative to WordNet)

$$\text{sim}(pig\ metal\ ,metal) \gg \text{sim}(pig\ metal\ ,pig)$$

- Problems due to the limited coverage of MWEs in ontologies such as WordNet

Interpretational

- **Assumption:** in interpreting MWEs relative to a discrete set of semantic classes, each of which is (implicitly) associated with a fixed degree of compositionality, we will model their relative compositionality

- Difficulties in identifying the relative compositionality of the different semantic classes

- Difficulties in interpretation (e.g. compound nominals)

Statistical Tests

- **Assumption:** pick up on word combinations which occur with "significantly" high relative frequency when compared to the frequencies of the individual words (i.e. $f(x, y)$ as compared to $f(x)$ and $f(y)$)

Statistical Tests Commonly Used

- Simple frequency: $f(x, y)$

- Pointwise/specific mutual information: $\log \frac{P(x,y)}{P(x)P(y)}$

- Dice's coefficient: $\frac{2\, f(x,y)}{f(x)f(y)}$

- (Student's) t score

- (Pearson's) chi-square (χ^2)

- Log likelihood

- Selectional association

  ⋮

  Finding of Evert and Krenn (2001) that simple frequency is as good as a wide range of collocation extraction measures over German Adj-N and P-N-V triple extraction tasks

Why so many Statistical Tests?

- Complications in evaluation

    ⋆ hard to say which is the "best" test
    ⋆ conflicting results from different researchers

- Different corpora have different distributional idiosyncracies

- Different tests have different statistical idiosyncracies

Bigram Results from the WSJ

Rank	Frequency	Mutual information	χ^2	t test
1	of the	Quadi Doum	Posse Comitatus	of the
2	in the	Wrongful Discharge	LORIMAR TELEPICTURES	in the
3	to NUMB	Seh Jik	Petits Riens	to NUMB
4	for the	Noo Yawk	Wrongful Discharge	on the
5	to the	WESTDEUTSCHE LANDESBANK	Tupac Amaru	the company
6	of NUMB	Naamloze Vennootschap	Sary Shagan	about NUMB
7	on the	Caisses Regionales	Outlaw Biker	said it
8	NUMB to	Centenaire Blanzy	GEMINI SOGETI	for the
9	that the	Guillen Landrau	Centenaire Blanzy	to be
10	the company	Ea Matsekha	Smith-Corona Typewriters	a share
⋮				

Linguistic Properties

- **Assumption:** there exist overtly-expressed linguistic properties that correlate (+vely or -vely) with compositionality

- Possibilities for verb particle constructions:

  ⋆ particle position (e.g. *pick a broken lead pencil up* vs. *?pick a disease up*)
  ⋆ particle modifiability (e.g. *pick the pencil straight/right/back/ up* vs. *pick a disease ?right/?back/*way up*)
  ⋆ nominalisation (e.g. *feedback, backup* vs. **boilup*)

Dehé (2002); Dehé *et al.* (2002)   47

Co-occurrence Properties

- **Assumption:** for combinatorial MWEs (e.g. VPCs, NNs), signature patterns of interpretations or simple co-occurrence are good predictors of compositionality

		$V2_1$	$V2_2$	$V2_3$	$V2_4$	$V2_5$	$V2_6$	$V2_7$	$\ldots$	$V2_j$
	$V1_1$					A				
	$V1_2$					S				
	$V1_3$					–				
V1	$V1_4$	D		$\ldots$	?	**?**	S	$\ldots$		–
	$V1_5$					S				
	$\vdots$					$\vdots$				
	$V1_i$					DS				

(column header: **V2**)

Ability of the Different Methods to Capture Compositionality

Method	Markedness				
	Lex	Syn	Sem	Prag	Stat
Substitutability	?	☑	☑	☑	☑
Distributional sim	☒	☑	☑	?	☑
Semantic sim	?	☒	?	☑	?
Interpretational	☑	☑	☑	☑	☑
Statistical test	☑	?	☑	☑	☑
Linguistic	?	☑	☑	?	?
Co-occurrence	☑	☑	☑	☑	☑

CONCLUSION

Conclusion

- Compositionality is a predominant factor in discussing MWEs, but we need to be aware of the subtleties (notably statistical markedness and decomposability)

- Compositionality is not the exclusive domain of semantics

- Various methods have been proposed for representing and modelling compositionality, although not all are applicable to all forms of compositionality

- There is still lots to be done, with lots of room for all!

References

BALDWIN, TIMOTHY, COLIN BANNARD, TAKAAKI TANAKA, and DOMINIC WIDDOWS. 2003. An empirical model of multiword expression decomposability. In *Proc. of the ACL-2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, 89–96, Sapporo, Japan.

BANNARD, COLIN, TIMOTHY BALDWIN, and ALEX LASCARIDES. 2003. A statistical approach to the semantics of verb-particles. In *Proc. of the ACL-2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, 65–72, Sapporo, Japan.

COPESTAKE, ANN. 2003. Compounds revisited. In *Proc. of the 2nd International Workshop on Generative Approaches to the Lexicon*, Geneva, Switzerland.

CRUSE, D. ALAN. 1986. *Lexical Semantics*. Cambridge, UK: Cambridge University Press.

DEHÉ, NICOLE. 2002. *Particle Verbs in English: Syntax, Information, Structure and Intonation*. Amsterdam, Netherlands/Philadelphia USA: John Benjamins.

——, RAY JACKENDOFF, ANDREW MCINTYRE, and SILKE URBAN (eds.) 2002. *Verb-particle explorations*. Berlin/New York: Mouton de Gruyter.

EVERT, STEFAN, and BRIGITTE KRENN. 2001. Methods for the qualitative evaluation of lexical association measures. In *Proc. of the 39th Annual Meeting of the ACL and 10th Conference of the EACL (ACL-EACL 2001)*, 188–95, Toulouse, France.

FILLMORE, CHARLES J., PAUL KAY, and MARY C. O'CONNOR. 1988. Regularity and idiomaticity in grammatical constructions. *Language* 64.501–38.

JACKENDOFF, RAY. 1997. *The Architecture of the Language Faculty*. Cambridge, USA: MIT Press.

KRENN, BRIGITTE, and STEFAN EVERT. 2001. Can we do better than frequency? A case study on extracting PP-verb collocations. In *Proc. of the ACL/EACL 2001 Workshop on the Computational Extraction, Analysis and Exploitation of Collocations*, 39–46, Toulouse, France.

LIBERMAN, MARK, and RICHARD SPROAT. 1992. The stress and structure of modified noun phrases in English. In *Lexical Matters – CSLI Lecture Notes No. 24*, ed. by Ivan A. Sag and A. Szabolcsi. Stanford, USA: CSLI Publications.

LIN, DEKANG. 1999. Automatic identification of non-compositional phrases. In *Proc. of the 37th Annual Meeting of the ACL*, 317–24, College Park, USA.

MCCARTHY, DIANA, BILL KELLER, and JOHN CARROLL. 2003. Detecting a continuum of compositionality in phrasal verbs. In *Proc. of the ACL-2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, Sapporo, Japan.

MOON, ROSAMUND E. 1998. *Fixed Expressions and Idioms in English: A Corpus-based Approach*. Oxford, UK: Oxford University Press.

NUNBERG, GEOFFREY, IVAN A. SAG, and TOM WASOW. 1994. Idioms. *Language* 70.491–538.

PEARCE, DARREN. 2001a. Synonymy in collocation extraction. In *Proc. of the NAACL 2001 Workshop on WordNet and Other Lexical Resources: Applications, Extensions and Customizations*, Pittsburgh, USA.

——. 2001b. Using conceptual similarity for collocation extraction. In *Proc. of the 4th UK Special Interest Group for Computational Linguistics (CLUK4)*.

SAG, IVAN A., TIMOTHY BALDWIN, FRANCIS BOND, ANN COPESTAKE, and DAN FLICKINGER. 2002. Multiword expressions: A pain in the neck for NLP. In *Proc. of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2002)*, 1–15, Mexico City, Mexico.

SCHONE, PATRICK, and DAN JURAFSKY. 2001. Is knowledge-free induction of multiword unit dictionary headwords a solved problem? In *Proc. of the 6th Conference on Empirical Methods in Natural Language Processing (EMNLP 2001)*, 100–108, Pittsburgh, USA.

UCHIYAMA, KIYOKO, TIMOTHY BALDWIN, and SHUN ISHIZAKI. 2005. Disambiguating Japanese compound verbs. *Computer Speech and Language, Special Issue on Multiword Expressions* 19.497–512.