

# Corpus-Derived Phraseological Lists and EAP Pedagogy: Interpretation and Comparison

David Oakey  
djoakey@iastate.edu

BALEAP PIM  
Coventry University  
21st June 2014

About Me

Handout

This is a paperless talk: come up  
and scan the QR to get a pdf at  
the end:



About Me



Handout

## Structure of today's talk

Background – the phraseology terminology problem

The lists of phraseological forms in today's talk

A suggested framework for comparison

Comparison of the phraseological features in the papers

Applicability to teaching



## The phraseology terminology problem:

- same term used by different people for different things
- same term used by one person for too wide a range of things
- different terms used for the same thing

All terms come with their own epistemological 'baggage', carrying along with them echoes of the contexts in which they have been coined, used, and/or misused by different scholars over the years

Many recent papers tend to refer the reader to Wray's (2002) list of 40 items and then use their own term or a previous term on top of this



# The phraseology terminology problem:

Terms used in the literature to describe formulaic sequences and formulaicity

---

Amalgams	Gambits	Preassembled speech
Automatic	Gestalt	Prefabricated routines and patterns
Chunks	Holistic	Ready-made expressions
Clichés	Holophrases	Ready-made utterances
Co-ordinate constructions	Idiomatic	Rote
Collocations	Idioms	Routine formulae
Composites	Irregular	Schemata
Conventionalized forms	Lexical(ised) phrases	Semi-preconstructed phrases that constitute single choices
FEIs <sup>a</sup>	Lexicalised sentence stems	Sentence builders
Fixed expressions	Multiword units	Stable and familiar expressions with specialized subsenses
Formulaic language	Non-compositional	Synthetic
Formulaic speech	Non-computational	Unanalysed chunks of speech
Formulas/formulae	Non-productive	
Fossilized forms	Petrification	
Frozen phrases	Praxons	

---

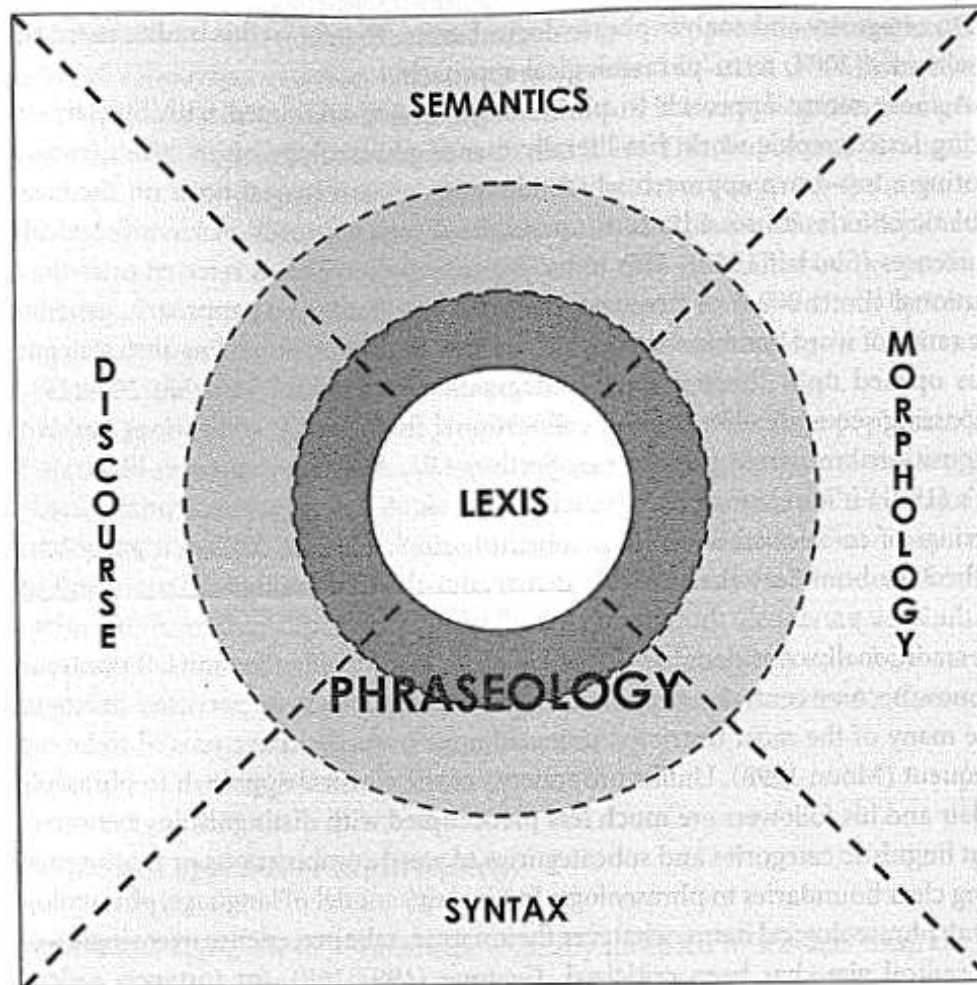
<sup>a</sup> Fixed Expressions including Idioms (Moon, 1998).

The list has since become even longer:  
institutionalized expressions, lexical bundles, formulaic bundles, vocabulary based discourse units, meaning shift units, multi-word items, multi-word expressions, ngrams, p-frames, phraseologisms, polywords, recurrent clusters, recurrent word combinations, speech formulas

# The "Phraseology Terminologizer"

Term	Author/ Article	Definition/Use/Meaning/Means of Identification	Connections to Other Uses of this Term	Connections to Other Terms, Definitions, Meanings
<b>lexical bundle</b>	Biber et al 2004	<p>a contiguous string of n words occurring more than X times per million words by more than Y speakers/writers; corpus-identified, statistical</p> <p>“To be classified as a lexical bundle, an MWU has to (a) occur frequently in a register, e.g. 10 times per one million words,<sup>3</sup> and (b) occur in multiple texts in this register” (Romer, 2009)</p> <p>“recurrent expressions, regardless of their idiomaticity, and regardless of their structural status.” (Biber et al., 1999, p. 990)</p> <p>"neither terribly functional nor pedagogically compelling" (Simpson-Vlach &amp; Ellis 2010: 491)</p>	<p>Biber et al 1999 (first use)</p> <p>Cortes 2006 - different disciplines</p> <p>Hyland 2008 - different written registers</p> <p>Nesi 2008 speech</p>	<p>similar in form to a ngram but has a different definition and doesn't cross sentence boundaries; originally used for comparing registers; Biber et al (2004) gave it a discourse function in texts which it didn't originally have</p>
<b>formulaic sequence (1)</b>	Wray & Perkins 2000; Wray 2002	<p>“A sequence, continuous or discontinuous, of words or other meaning elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar.”</p>		
<b>formulaic sequence (2)</b>	Boers et al 2006 Schmitt et al 2004	<p>An overarching term for many other items, specifically includes: multiword expressions (eg collocations), lexical phrases, multiword units, formulas, prefabricated chunks.</p>		

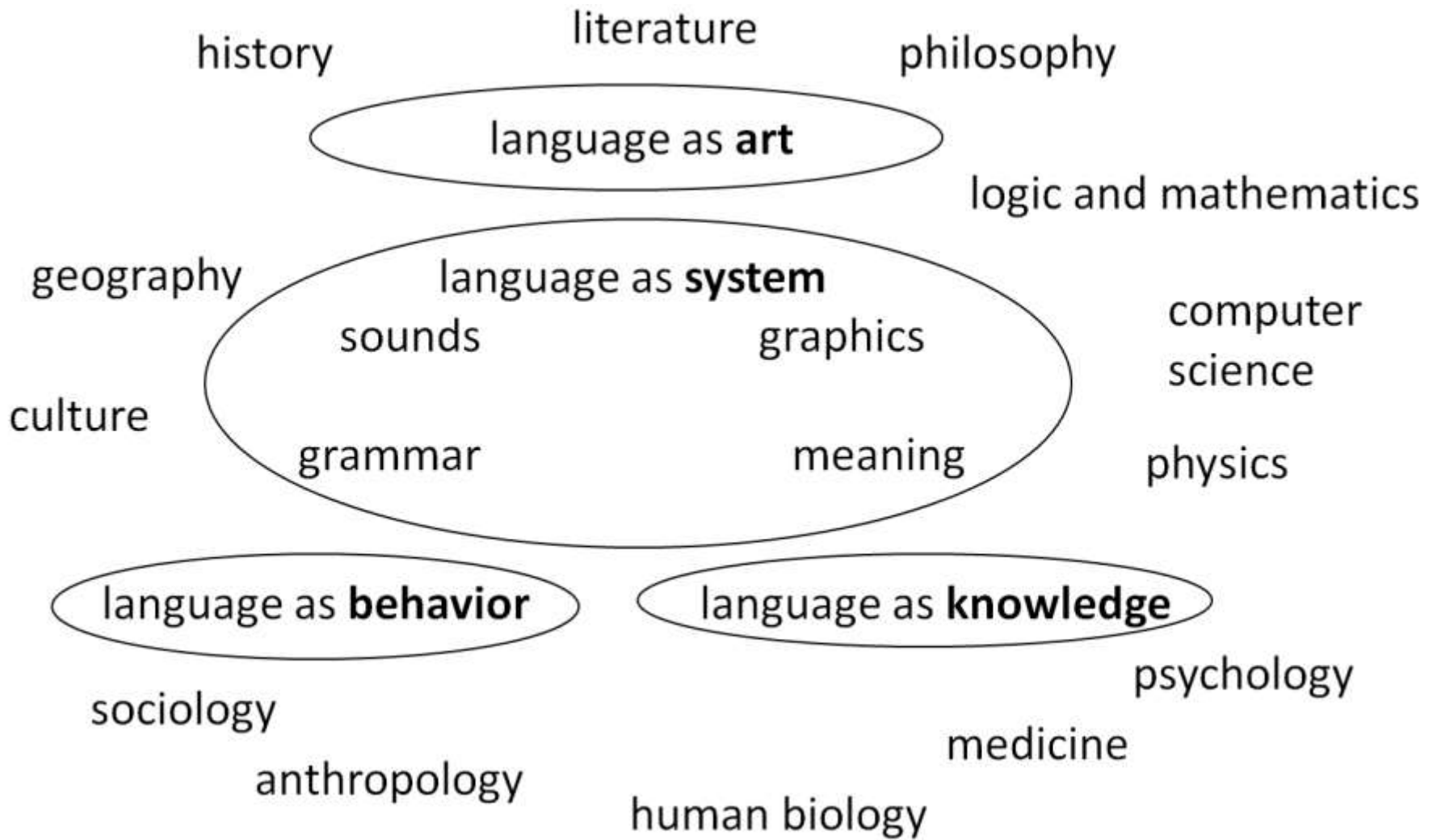
The phraseology terminology problem:  
Granger and Paquot (2008) talk about reasons for the  
'phraseological web'



About Me



Handout



Adapted from Halliday, *Language as Social Semiotic* (1978: 11)



About Me



Handout

Several recent corpus-backed lists of phraseological forms intended for use in EFL/EAP pedagogy:

'grammatical collocations' (Durrant 2009)	100 items
'academic formulas' (Simpson-Vlach & Ellis 2010)	208 items
'academic collocations' (Ackermann & Chen 2011)	2,470 items
'phrasal expressions' (Martinez & Schmitt 2012)	506 items
'multi-word constructions' (Liu 2012)	232 items

a quick look....





**Comparison Criteria:** lexical, syntactic, **semantic**, pragmatic

- Traditional "Eastern European" lexicography for identifying 'restricted collocations' (Aisenstadt 1981; Howarth 1996)
- "Empirical Firthian" lexicology for identifying 'extended lexical units' (Stubbs 2001)
- "Usage-Based" cognitive linguistics for defining 'phraseologisms' (Gries 2008)

### **Traditional "Eastern European" lexicography**

The semantic criterion for a 'restricted collocation' is that it must be partially semantically transparent, in which one element of the combination has a literal, unidiomatic meaning, e.g. *stark naked*

### **"Empirical Firthian" lexicology**

The semantic criterion for an 'extended lexical item' is that it has an observed semantic preference, such as a particular lexical set, semantically related word-form or lemma; e.g. *UNDERGO* in its right-hand environment has a discourse prosody of "unpleasantness" (Stubbs 2001:91):

### **"Usage-Based" cognitive linguistics**

The semantic criterion for a 'phraseologism' involve an item's semantic non-compositionality or non-predictability – whether the meaning of the whole phrase can be predicted from the meanings of its individual elements e.g. *kick the bucket*

## **Comparison Criteria:** lexical, syntactic, semantic, **pragmatic**

- Traditional "Eastern European" lexicography for identifying 'restricted collocations' (Aisenstadt 1981; Howarth 1996)
- "Empirical Firthian" lexicology for identifying 'extended lexical units' (Stubbs 2001)
- "Usage-Based" cognitive linguistics for defining 'phraseologisms' (Gries 2008)

### **Traditional "Eastern European" lexicography**

The pragmatic criterion for a restricted collocation is that it must be institutionalized ... somehow distinctive and memorized, e.g. *stark naked* rather than *stark nude*, *stark unclothed*, or *stark bare*

### **"Empirical Firthian" lexicology**

The pragmatic criterion for an extended lexical unit involves the item's discourse function and its distribution in text types – e.g. *UNDERGO* doesn't have a negative prosody in technical reports (Stubbs 2001)

### **"Usage-Based" cognitive linguistics**

The pragmatic criterion for identifying phraseologisms is less explicit

The pragmatic functions of phraseological forms are particularly important for EAP learners, whose linguistic choices must be appropriate to the register in which they will use English – several of these lists are organized according to discourse function

Criterion	‘Eastern European ‘ Perspective (e.g. Howarth)	Empirical Firthian Perspective (e.g. Stubbs)	Usage-Based Cognitive Linguistic Perspective (e.g. Gries)
<b>Lexical</b>	Limited commutability	Frequent collocation of individual word-forms or lemmas	-the nature of the elements involved -the degree of lexical and syntactic flexibility
<b>Syntactic</b>	Grammatically well-formed (an adverb modifying an adjective);	Observed frequent colligation of word classes	-the nature of the elements involved -the degree of lexical and syntactic flexibility -the permissible distance between the elements involved
<b>Semantic</b>	Partially semantically transparent, in which one element of the combination has a literal, unidiomatic meaning	Observed semantic preference, such as lexical sets, semantically related word-forms or lemmas	-semantic non-compositionality or non-predictability
<b>Pragmatic</b>	Institutionalization; distinctive and memorized	Discourse prosody; descriptor of speaker attitude and discourse function	not explicit
<b>Data for identification</b>	Native speaker judgments or observation of use	corpus	- the number of times an expression must be observed before it counts

Criterion	Grammatical collocations (Durrant 2009)	Academic formulas (Simpson-Vlach & Ellis 2010)	Academic collocations (Ackermann & Chen 2011)	Phrasal expressions (Martinez & Schmitt 2012)	Multi-word constructions (Liu 2012)
<b>Lexical</b>	mostly one frequent word with another frequent word	frequently-occurring multiple words	mostly one frequent word with another frequent word	frequently-occurring multiple words	frequently-occurring multiple words
<b>Syntactic</b>	well-formed; usually one lexical and one grammatical word; not necessarily fixed or contiguous	fixed, contiguous sequences of words which may or may not fit traditional boundaries	well-formed; mostly adjective + noun pairs, presumably contiguous	fixed, contiguous sequences of words which may or may not fit traditional boundaries	fixed or discontinuous sequences of words and word classes
<b>Semantic</b>	not prioritized	identified as meaningful and teachable	not specified	focuses on non-compositionality	more generalized meanings
<b>Pragmatic</b>	valid across multiple academic disciplines but no explicit discourse function	assigned spoken and written discourse functions	valid across multiple academic disciplines but no explicit discourse function	assigned with spoken and written discourse functions	assigned with spoken and written discourse functions
<b>Data for identification</b>	5m-word academic corpus compared with 85 million non-academic BNC	BNC, Hyland, MICASE spoken & written	Pearson (25m) International Corpus of Academic English	BNC	academic sections of COCA and BNC

# Applicability to teaching

1 s lacking in the present state of the design art. <move3>Computationally efficient, scalable, lumped-eleme  
2 of interpolating methods, as depicted in Fig. 1. <move3>In order to reduce the number of testing devices use  
3 unacceptable for circuit design and optimization. <move3>In Section II, we describe new approximate expressio  
4 illators has not been investigated before either. <move3>In Section II, a brief overview of conventional seco  
5 in order to make intelligent design compromises. <move3>In this paper, the construction and performance char  
6 y spoil the jitter performance of the PLL itself. <move3>In this paper, we present a digital approach for hig  
7 functions without iterative fitting optimization. <move3>In this paper, we shall present an asymmetric double  
8 a poor phase noise, for given power consumption. <move3>In this paper, we propose a frequency tuning method  
9 without resorting to numerical simulations [16]. <move3>In this paper, we develop a first-order analytical m  
10 ics of spiral inductors [8] and transformers [2]. <move3>In this paper, we study experimentally and analyze t  
11 t with measured data over a wide frequency range. <move3>In this paper, the analysis and modeling of monolith  
12 mponents were calculated using numerical methods. <move3>In this paper, we provide a broadband and a scalable  
13 operation at the lowest possible supply voltage. <move3>The feedback amplifier presented in this paper employ  
14 [16] or tunable-impedance transmission lines [17]. <move3>This paper explores the potential of the load adaptati  
15 hnology [1], allowing new circuit configurations. <move3>This paper introduces a modification of stacked induct  
16 h high power efficiency and good signal fidelity. <move3>This paper presents a broadband PA topology with a tra  
17 and magnetic coupling for series power combining. <move3>This paper presents a study of the most common passive  
18 ency band of interest for a specific application. <move3>This paper presents a new comprehensive methodology fo  
19 f discrete components found at lower frequencies. <move3>This paper reports inductors and transformers suitable

*brief overview (Ackermann & Chen)*

*this paper (Durrant)*

*a/the study of (det + N)*  
(Liu)

*in order to*  
(Simpson-Vlach & Ellis, Liu, Martinez & Schmitt)



# Closing thoughts...

## Appendix 1 Summary of the Academic Language Features in the Syllabus

Text sections	Sentences	Words and phrases	Single words
<ul style="list-style-type: none"> <li>• general-to-specific text structure</li> <li>• situation-problem-solution-evaluation text structure</li> <li>• moving from general to specific information in an introduction</li> <li>• outside information in a text</li> <li>• linking paragraphs</li> <li>• organizing long academic texts</li> <li>• common text sections</li> <li>• signaling the structure of a text</li> <li>• planning the sequence of information in a text</li> <li>• linking the introduction and conclusions in a text</li> </ul>	<ul style="list-style-type: none"> <li>• noun combinations in simple sentences</li> <li>• definitions</li> <li>• spoken and written style</li> <li>• appropriate and inappropriate academic style</li> <li>• critical sentences</li> <li>• the language of comparisons</li> <li>• the language of changes and events</li> <li>• transitivity</li> <li>• vagueness with quantities</li> <li>• relative clauses at the end and in the middle of a sentence</li> <li>• reduced relative clauses</li> <li>• long sentences containing lists</li> <li>• connecting ideas in two sentences</li> <li>• topic sentences</li> <li>• active and passive voice</li> <li>• tense and passive voice</li> <li>• beginning sentences with <i>it</i></li> </ul>	<ul style="list-style-type: none"> <li>• noun combinations</li> <li>• turning noun combinations into simple sentences</li> <li>• the sounds of words spoken together in RP and Brummie*</li> <li>• critical phrases</li> <li>• re-writing preposition and compound noun combinations</li> <li>• fixed noun combinations</li> <li>• changing clauses into noun combinations</li> <li>• words and phrases for bringing outside information into a text</li> <li>• linking words and phrases</li> </ul>	<ul style="list-style-type: none"> <li>• word classes: vocabulary words and grammar words</li> <li>• learning new words and old words used in new ways</li> <li>• guessing the meaning of new words</li> <li>• word “families”</li> <li>• common and rare words</li> <li>• head nouns</li> <li>• nouns with general meanings and nouns with specific meanings</li> <li>• academic English terminology</li> <li>• the sounds of single words in RP and Brummie</li> <li>• stressed and unstressed words</li> <li>• relative pronouns</li> <li>• verbs of movement</li> <li>• critical words</li> <li>• linking words</li> <li>• referring words</li> <li>• transitive and intransitive verbs</li> <li>• the meaning and grammar of reporting verbs</li> </ul>

So many lists, so many different names - very strong on form and function, less so on implementation for teachers and learners

Corpus findings are still seen as “enthusiastic yet fragmented efforts” which are “banked intellectual resources whose pedagogical time has yet to come” Swales (2002: 159).

The syllabus needs to situate phraseological units in the general scheme of things.

(Oakey 2011: 187)



About Me



Handout



About Me



Handout