

Developing a phrasebank: A critical examination of phrasal lists

INTRODUCTION

This poster reports on an examination of the 10 lists of formulaic phrases below. The examination aimed to determine the characteristics of the lists, their strengths and weaknesses, and whether any particular type of list is lacking, with a view to assisting teachers in understanding their usefulness. To achieve these aims, a set of nine criteria for interrogating each of the lists was produced, drawing on factors discussed in the literature on word- and phrasal-list development. The first four factors examined relate to the process of producing the list; the latter five factors relate to the product itself. The results for each are shown in the following tables.

Author(s)	Year of publication	Formulaic phrase list
Ackermann and Chen	2013	Academic Collocations List
Durrant	2009	Top 1000 key academic collocations
Garnier and Schmitt	2015	Phrasal Verb Pedagogical List
Hammond	2018	Individual Development Narrative Phrasebank
Hsu	2014	The most frequent opaque formulaic sequences in English-medium college textbooks
Liu	2012	The most frequently-used constructions in academic writing by frequency
Martinez and Schmitt	2012	A Phrasal Expressions List
Morley	2020	The Academic Phrasebank
Shin and Nation	2008	The most frequent collocations in spoken English
Simpson-Vlach and Ellis	2010	An Academic Formulas List

What was the purpose of creating the list?

	Ackermann and Chen (2013)	Durrant (2009)	Garnier and Schmitt (2015)	Hammond (2018)	Hsu (2014)	Liu (2012)	Martinez and Schmitt (2012)	Morley (2020)	Shin and Nation (2008)	Simpson-Vlach and Ellis (2010)
For learners	✓		✓	✓	✓	✓	✓	✓	✓	
For teachers/material writers	✓		✓				✓		✓	✓
For testing							✓			
The phrasal selection process	✓	✓								

What type of items are included in the list?

	Ackermann and Chen (2013)	Durrant (2009)	Garnier and Schmitt (2015)	Hammond (2018)	Hsu (2014)	Liu (2012)	Martinez and Schmitt (2012)	Morley (2020)	Shin and Nation (2008)	Simpson-Vlach and Ellis (2010)
Formulaic sequences					✓					✓
Collocations	✓	✓							✓	
Formulaic frames				✓						
Phrasal verbs			✓							
Multi-word constructions						✓				
Multi-word expressions							✓			
Phrases								✓		

What metrics were used to identify items?

	Ackermann and Chen (2013)	Durrant (2009)	Garnier and Schmitt (2015)	Hammond (2018)	Hsu (2014)	Liu (2012)	Martinez and Schmitt (2012)	Morley (2020)	Shin and Nation (2008)	Simpson-Vlach and Ellis (2010)
Raw corpus frequency	✓	✓	✓		✓	✓	✓	✓	✓	✓
Statistical measures of corpus frequency		✓			✓					✓
Human judgement	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Occurrence in dictionaries and previous studies	✓		✓		✓	✓	✓		✓	

Was a validation of the list carried out?

	Ackermann and Chen (2013)	Durrant (2009)	Garnier and Schmitt (2015)	Hammond (2018)	Hsu (2014)	Liu (2012)	Martinez and Schmitt (2012)	Morley (2020)	Shin and Nation (2008)	Simpson-Vlach and Ellis (2010)
Yes	✓		✓		✓				✓	✓
No		✓		✓		✓	✓	✓		

How many items are included in the list?

	Ackermann and Chen (2013)	Durrant (2009)	Garnier and Schmitt (2015)	Hammond (2018)	Hsu (2014)	Liu (2012)	Martinez and Schmitt (2012)	Morley (2020)	Shin and Nation (2008)	Simpson-Vlach and Ellis (2010)
Number	2468	1000	150	102	475	228	505	2030 ^a	4698	438
Justification given	✓	✓	✓	✓	✓		✓			✓

^aThe approximate number of phrases at the time of counting

Who or for what context is the list targeted towards?

	Ackermann and Chen (2013)	Durrant (2009)	Garnier and Schmitt (2015)	Hammond (2018)	Hsu (2014)	Liu (2012)	Martinez and Schmitt (2012)	Morley (2020)	Shin and Nation (2008)	Simpson-Vlach and Ellis (2010)
Targeted group				✓						
Specified field, e.g. EAP	✓	✓			✓	✓		✓		✓
Proficiency level	✓							✓	✓	
General			✓		✓					
Unspecified							✓			

Are the items in the list intended for receptive and/or productive knowledge?

	Ackermann and Chen (2013)	Durrant (2009)	Garnier and Schmitt (2015)	Hammond (2018)	Hsu (2014)	Liu (2012)	Martinez and Schmitt (2012)	Morley (2020)	Shin and Nation (2008)	Simpson-Vlach and Ellis (2010)
Receptive	✓			✓	✓		✓		✓	
Productive	✓	✓		✓		✓		✓	✓	

How are items presented in the list?

	Ackermann and Chen (2013)	Durrant (2009)	Garnier and Schmitt (2015)	Hammond (2018)	Hsu (2014)	Liu (2012)	Martinez and Schmitt (2012)	Morley (2020)	Shin and Nation (2008)	Simpson-Vlach and Ellis (2010)
Alphabetical	✓		✓		✓					
Frequency			✓			✓	✓		✓	
Function						✓				✓
Spoken/Written										✓
Moves/steps of paper				✓				✓		
Length (of phrase)					✓					
Unspecified		✓								

What, if any, additional information is provided about in the list?

	Ackermann and Chen (2013)	Durrant (2009)	Garnier and Schmitt (2015)	Hammond (2018)	Hsu (2014)	Liu (2012)	Martinez and Schmitt (2012)	Morley (2020)	Shin and Nation (2008)	Simpson-Vlach and Ellis (2010)
Meaning/function	✓		✓	✓		✓		✓		
Examples			✓	✓			✓	✓		
Genre/register info	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Form variability			✓	✓	✓	✓	✓	✓	✓	

FINDINGS

The examination revealed lists fell into two distinct categories:

- Resource-oriented:
- rigorous identification of a narrowly defined item type
 - intended principally as a resource for teachers, materials writers and curriculum designers
- Learner-oriented:
- less rigorous identification of loosely-defined items
 - emphasis on the creation of a user-friendly product

Of the lists examined, only Hammond (2018) and Morley (2020) fell into the learner-oriented category; the remainder were found to be resource-oriented.

There was very little crossover between the two types. Therefore, there is a clear need for lists that are both rigorous in the processes used to identify phrases and committed to organizing and presenting the resultant list of phrases in a manner that is maximally informative and accessible to users.