

Corpora, phraseology and academic discourse

Ute Römer, Georgia State University, USA
uroemer@gsu.edu ■ <http://gsu.edu/ales/>

Talk outline

1. Introduction
2. Corpus analysis and phraseology
3. Phraseology and academic corpora
 - 3.1 The phraseological profile of book reviews (Römer 2010)
 - 3.2 Textual distribution of phraseological items (O'Donnell & Römer in preparation)
4. Concluding thoughts

Corpus tools for phraseological analysis

- KfNgram (William Fletcher), free, <http://www.kwicfinder.com/kfNgram/>
- Collocate (Michael Barlow), requires a license, <http://www.athel.com/>
- ConcGram (Chris Greaves), requires a license, <http://benjamins.com/#catalog/software/cls.1>
- AntConc (Laurence Anthony), free, <http://www.antlab.sci.waseda.ac.jp/software.html>
- WordSmith Tools (Mike Scott), requires a license, <http://www.lexically.net/wordsmith/>

Example n-grams and p-frames (data from the Book Reviews in Linguistics Corpus, BRILC)

Hits	4-gram
562	on the other hand
442	at the end of
428	on the basis of
411	as well as the
356	at the same time
330	the end of the
301	of the book is
288	in the case of
268	the fact that the
226	on the one hand
216	a wide range of
215	in the context of
214	the rest of the
207	in terms of the
206	to the study of

Figure 1. Extract of a BRILC 4-gram list (Collocate output)

it would be * to	101	10
it would be interesting to	44	
it would be useful to	14	
it would be nice to	11	
it would be better to	9	
it would be possible to	5	
it would be helpful to	5	
it would be fair to	4	
it would be difficult to	3	
it would be necessary to	3	
it would be good to	3	
it * be interesting to	58	3
it would be interesting to	44	
it will be interesting to	8	
it might be interesting to	6	

Figure 2. Example p-frames in BRILC, with numbers of tokens and numbers of variants (kfNgram output)

The distribution of phraseological items across BRILC texts (3.1, step 4)

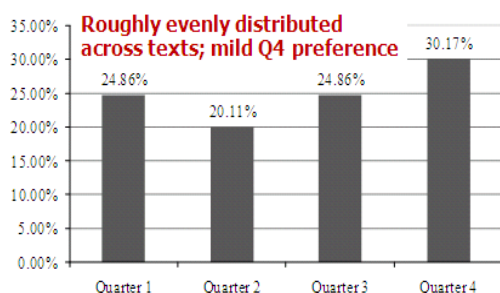


Figure 5. *At the same time* across texts

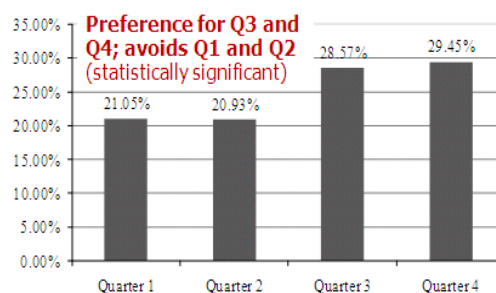


Figure 6. *On the * hand* across texts

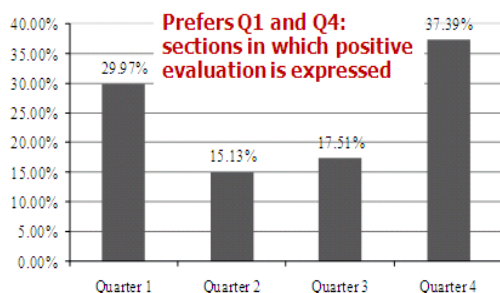


Figure 7. *A * range* of across texts

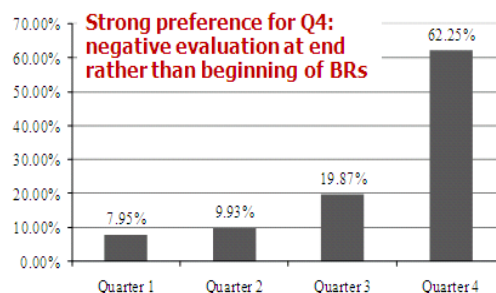


Figure 8. *It would have been* across texts

Selected references

Barlow, M. (2004). *Collocate 1.0: Locating Collocations and Terminology*. Houston, TX: Athelstan.

Fletcher, W. H. (2002-2007). *KfNgram*. Annapolis, MD: USNA.

Greaves, C. (2009). *ConcGram 1.0. A Phraseological Search Engine*. Amsterdam: John Benjamins.

Hoey, M. P. (2005). *Lexical Priming: A New Theory of Words and Language*. London: Routledge.

Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

O'Donnell, M. B. & U. Römer (In preparation). Investigating the interaction between phraseological items and textual position.

Römer, U. (2008). Identification impossible? A corpus approach to realisations of evaluative meaning in academic writing. *Functions of Language* 15(1): 115-130.

Römer, U. (2010). Establishing the phraseological profile of a text type: The construction of meaning in academic book reviews. *English Text Construction* 3(1): 95-119.

Römer, U. (2011). Observations on the phraseology of academic writing: Local patterns – local meanings? In: Herbst, T., S. Faulhaber & P. Uhrig (eds.). *The Phraseological View of Language. A Tribute to John Sinclair*. Berlin: Mouton de Gruyter. 211-227.

Sinclair, J. M. (2008). The phrase, the whole phrase, and nothing but the phrase. In: Granger, S. & F. Meunier (eds.). *Phraseology: An Interdisciplinary Perspective*. Amsterdam: John Benjamins. 407-410.

Swales, J. M. (2004). *Research Genres. Exploration and Applications*. Cambridge: Cambridge University Press.