

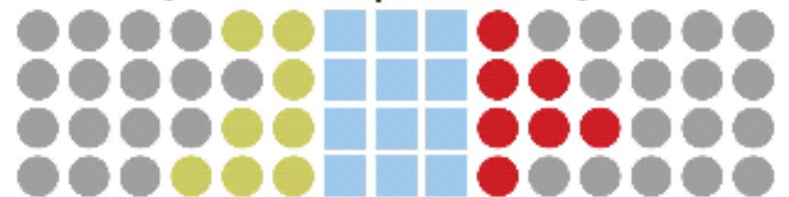
# Phraseological items in apprentice academic writing: Does nativeness matter?

Ute Römer

AACL Edmonton- 9 October 2009

[uroemer@umich.edu](mailto:uroemer@umich.edu)  
[www.ELIcorpora.info](http://www.ELIcorpora.info)

Michigan Corpus Linguistics



# 1. Introduction: E. as an academic lg.

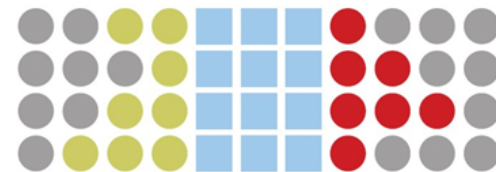
- Nowadays: a large and growing number of academic texts produced by NNS of English
- Research world is becoming more Anglicized
- Large numbers of "non-Anglophones" (Swales 2004: 46) produce academic English
- Still unclear: **status of nativeness** in this context
- Does nativeness matter when we are dealing with English in academia?
- What challenges do NNS novice writers face in academic settings? (similar to NS novice writers?)
- **How different are NNS and NS academic writing?**



# 1. Introduction: E. as an academic lg.

## This paper will...

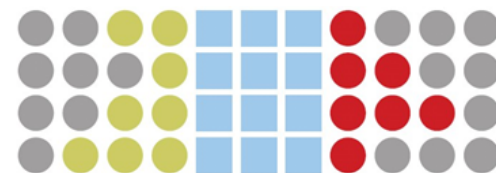
- ... examine successful NNS and NS apprentice academic writing
- ... discuss whether nativeness has an effect on academic writing proficiency if other factors (genre, discipline, years of university instruction) are controlled
- ... focus on **frequent phraseological items**  
"Language as phraseology" (Hunston 2002: 137)  
"the normal carrier of meaning is the phrase" (Sinclair 2005)
- ... compare findings from apprentice and expert performance data (Bazerman 1994; Scott & Tribble 2006)
- What are the most frequent phraseological items in NNS/NS apprentice (AAW) and expert academic writing (EAW)?



## 2.1 Data and method

Data retrieved from 3 corpora:

- **CHALC**: Cologne–Hanover Advanced Learner Corpus (see Römer 2007), 45 essays and term papers by upper–level university students (L1=German) in English Linguistics and Literary Studies; ~200,000 w
- **MICUSP\_EL**: subset of 91 English and Linguistics papers (term papers, essays, lit reviews) of the Michigan Corpus of Upper–level Student Papers, only papers by seniors, L1=English; ~200,000 words
- **Hyland\_Ling**: the Linguistics subsection of the Hyland Corpus, 30 published Linguistics research articles (by NSs or NS–checked); ~210,000 words

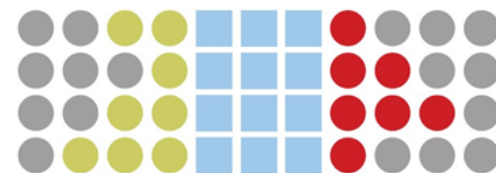


## 2.1 Data and method

### Extraction of phraseological items from corpora

- Tool used: *kfNgram* (Fletcher 2002–2007)
- *KfNgram* creates lists of n-grams and phrase-frames (with variants) of different lengths
- Extraction of **n-grams and p-frames**, spans 3–5
- Report only on spans 3 and 4 (short 5-gram lists)
- Also: **key n-gram** identification
- Tool used: *WordSmith Tools*, 'compare 2 wordlists' function (4.0, Scott 2004)
- First part of the analysis: examination and comparison of lists of **frequent 3-/4-grams in CHALC and MICUSP\_EL**

Ute Römer (uroemer@umich.edu)



## 2.2 N-grams in NS and NNS AAW

Top-20 3-grams in CHALC and MICUSP\_EL

CHALC	hits (pmw)	MICUSP_EL	hits (pmw)
<b>IN ORDER TO</b>	<b>761.1</b>	<b>IN ORDER TO</b>	<b>371.4</b>
<b>THE FACT THAT</b>	<b>615.7</b>	<b>ONE OF THE</b>	<b>321.2</b>
<i>the present perfect</i>	<b>426.6</b>	<b>THE FACT THAT</b>	<b>311.1</b>
<b>AS WELL AS</b>	<b>383</b>	<b>AS WELL AS</b>	<b>281</b>
<b>on the other</b>	<b>373.3</b>	<i>the end of</i>	<b>261</b>
<i>of the clause</i>	<b>353.9</b>	<i>end of the</i>	<b>230.8</b>
<b>the other hand</b>	<b>339.4</b>	<b>THERE IS A</b>	<b>225.8</b>
<b>THE USE OF</b>	<b>334.5</b>	<i>at the end</i>	<b>215.8</b>
<b>PART OF THE</b>	<b>329.7</b>	<b>THE USE OF</b>	<b>190.7</b>
<b>has to be</b>	<b>305.4</b>	<i>the united states</i>	<b>180.7</b>
<b>in the following</b>	<b>300.6</b>	<i>in the novel</i>	<b>160.6</b>
<i>of functional grammar</i>	<b>276.3</b>	<b>in which the</b>	<b>160.6</b>
<b>ONE OF THE</b>	<b>271.5</b>	<b>it is not</b>	<b>160.6</b>
<i>of the verb</i>	<b>261.8</b>	<b>there is no</b>	<b>160.6</b>
<b>in terms of</b>	<b>252.1</b>	<i>in the first</i>	<b>155.6</b>
<b>in contrast to</b>	<b>247.2</b>	<b>to be a</b>	<b>150.6</b>
<i>position of the</i>	<b>242.4</b>	<i>nicaraguan sign language</i>	<b>145.5</b>
<b>THERE IS A</b>	<b>242.4</b>	<b>PART OF THE</b>	<b>145.5</b>
<b>due to the</b>	<b>237.6</b>	<b>that it is</b>	<b>145.5</b>
<i>in the clause</i>	<b>237.6</b>	<i>that she is</i>	<b>145.5</b>

Ute Römer (uroemer@umich.edu)

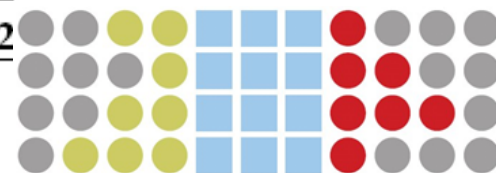


## 2.2 N-grams in NS and NNS AAW

Top-20 4-grams in CHALC and MICUSP\_EL

CHALC	hits (pmw)	MICUSP_EL	hits (pmw)
<b>ON THE OTHER HAND</b>	339.4	<b>THE END OF THE</b>	185.7
<b>AT THE END OF</b>	174.5	<b>AT THE END OF</b>	170.6
<b>in the case of</b>	164.8	<i>in r j baumgardner</i>	100.4
<b>with the help of</b>	164.8	<i>r j baumgardner ed</i>	100.4
<i>of the present perfect</i>	150.3	<b>ON THE OTHER HAND</b>	90.3
<b>THE END OF THE</b>	145.4	<i>collected poetry and prose</i>	80.3
<i>the specifier position of</i>	135.7	<b>it is clear that</b>	80.3
<i>quirk et al 1985</i>	130.9	<i>of nicaraguan sign language</i>	80.3
<b>to the fact that</b>	130.9	<b>one of the most</b>	75.3
<i>specifier position of the</i>	126	<b>the beginning of the</b>	75.3
<b>the fact that the</b>	116.4	<b>AT THE SAME TIME</b>	70.3
<i>the seven deadly sins</i>	116.4	<i>the gospel of mark</i>	70.3
<b>due to the fact</b>	111.5	<i>the universal access plan</i>	70.3
<i>of the english language</i>	111.5	<i>the university of michigan</i>	70.3
<i>the meaning of the</i>	111.5	<i>the vicar of wakefield</i>	70.3
<b>AT THE SAME TIME</b>	106.7	<b>as a result of</b>	65.2
<b>on the one hand</b>	106.7	<i>in the united states</i>	65.2
<i>the structure of the</i>	97	<b>it is important to</b>	65.2
<b>on the basis of</b>	92.1	<i>new york w.w norton</i>	65.2
<i>the subject of the</i>	87.3	<i>the english language in</i>	65.2

Ute Römer (uroemer@umich.edu)



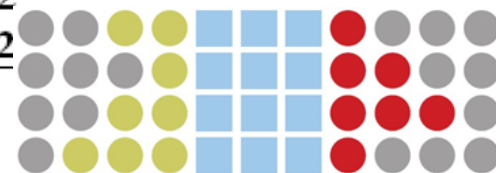
## 2.2 N-grams in NS and NNS AAW

Top-20 key n-grams (span: 3-6) in CHALC and MICUSP\_EL

KEY n-grams!

key n-grams in CHALC (reference corpus: MICUSP_EL)	keyness value	key n-grams in MICUSP_EL (reference corpus: CHALC)	keyness value
<i>halliday # #</i>	810.4	<i>the united states</i>	183
<i>the present perfect of the clause</i>	466.4	<i>nicaraguan sign language</i>	176.5
<i>Giegerich # #</i>	407.6	<i>indo pakistani english</i>	147
<i>harris # #</i>	317.4	<i>university of michigan</i>	145.6
<i>of functional grammar</i>	310.9	<i>in the water</i>	123.5
<i>in the clause</i>	303.1	<i>gen # #</i>	122
<i>quirk et al</i>	267.8	<i>j baumgardner ed</i>	117.6
<i>of the verb</i>	261.2	<i>in genesis #</i>	117.6
<i>state of affairs</i>	231.5	<i>poetry and prose</i>	116.2
<i>the structure of</i>	214.2	<i>to the reader</i>	111.8
<i>the analysis of</i>	213.8	<i>throughout the novel</i>	110.3
<i>halliday # # #</i>	208.9	<i>the book of</i>	108.8
<i>the speaker s</i>	207.7	<i>of the narrative</i>	104.4
<i>et al #</i>	201.1	<i>in the novel</i>	96.5
<i>position of the</i>	198.5	<i>of adult input</i>	95.6
<i>dik # #</i>	186.8	<i>Universal access plan</i>	92.4
<i>martin luther king</i>	186.8	<b>IT SEEMS THAT</b>	91.4
<b>I E THE</b>	182.8	<i>the pardoner is</i>	91.2
<i>givón # #</i>	182.3	<i>of nicaraguan sign</i>	88.2
	177.6	<i>vicar of wakefield</i>	88.2

Ute Römer (uroemer@umich.edu)





## 2.2 N-grams in NS and NNS AAW

Top-20 key n-grams (span: 3-6) in CHALC and MICUSP\_EL, compared with Hyland\_Ling

KEY n-grams!

key n-grams in CHALC (reference corpus: Hyland_Ling)	keyness value	key n-grams in MICUSP_EL (reference corpus: Hyland_Ling)	keyness value
<i>halliday # #</i>	791.9	<i>in the novel</i>	195.5
<i>the present perfect of the clause</i>	426.7	<i>nicaraguan sign language</i>	177.7
<i>giegerich # #</i>	398.5	<i>that she is</i>	173.3
<i>harris # #</i>	339.5	<i>indo pakistani english</i>	148.1
<i>of functional grammar</i>	332.5	<i>of the novel</i>	130.3
<i>quirk et al</i>	324.1	<i>in the water</i>	124.4
<i>in the clause</i>	279.4	<i>gen # #</i>	122.9
<i>halliday # # #</i>	252.3	<i>in genesis #</i>	118.5
<b>HAS TO BE</b>	222.1	<i>j baumgardner ed</i>	118.5
<i>dik # #</i>	207	<i>poetry and prose</i>	117
<i>martin luther king</i>	199.7	<i>throughout the novel</i>	111.1
<i>position of the</i>	195.6	<i>the book of</i>	109.6
<i>chapter # #</i>	193.3	<b>IS ABLE TO</b>	105.4
<i>givón # #</i>	192.8	<i>of the narrative</i>	105.2
<i>spencer # #</i>	190	<i>the gospel of</i>	100.7
<i>in chapter #</i>	188.6	<i>the story of</i>	100.7
<i>the specifier position</i>	180.2	<i>genesis # #</i>	100.7
<i>the mood element</i>	178.8	<i>of women in</i>	100.7
<i>selkirk # #</i>	171.8	<i>of adult input</i>	96.3
	163.4	<i>Universal access plan</i>	93.3

Ute Römer (uroemer@umich.edu)

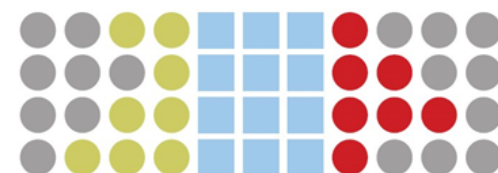


## 2.2 N-grams in NS and NNS AAW

Non-topic related key n-grams (span: 3-6) in Hyland\_Ling, compared with CHALC and MICUSP\_EL

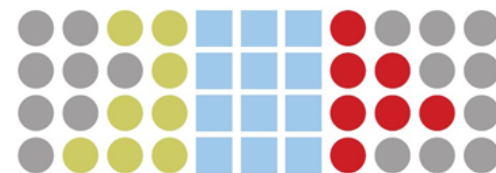
KEY n-grams!

key n-grams in Hyland_Ling (reference corpus: CHALC)	keyness value	key n-grams in Hyland_Ling (reference corpus: MICUSP_EL)	keyness value
<b>OF THE RESEARCH</b>	112.9	<b>IN THIS STUDY</b>	108.7
<b>OF THIS STUDY</b>	110	<b>OF THIS STUDY</b>	105.2
<b>IN THIS STUDY</b>	84.5	<b>WHAT WAS SAID</b>	84.3
an attempt to	71.6	<b>OF THE RESEARCH</b>	79.9
were able to	68.3	the analysis of	76.5
is most likely	66.1	on the other	69.2
<b>WHAT WAS SAID</b>	64.5	the other hand	66.7
has been a	60.6	in terms of	65.5
I think that	60.6	the results of	62.3
<b>THERE HAS BEEN</b>	60.6	in this respect	58.3
in light of	60.6	<b>THERE HAS BEEN</b>	57
<b>OF THIS RESEARCH</b>	56.4	<b>OF THIS RESEARCH</b>	53.1



## 2.3 P-frames in native and non-native speaker AAW

- Look at pattern variability
- What are the **most frequent p-frames** in our NS and NNS AAW corpora?
- Following tables list the top-20 3- and 4-p-frames in CHALC and MICUSP\_EL together with their **numbers of variants**
- E.g. 1,791 instances of *the \* of* in CHALC and 209 different realizations (*the use of, the structure of...*)
- High number of variants → productive p-frame

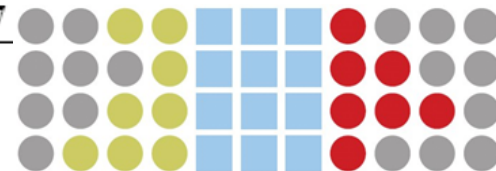


## 2.3 P-frames in NS and NNS AAW

Top-20 3-p-frames in CHALC and MICUSP\_EL

CHALC	hits (tokens)	variants (types)	MICUSP_EL	hits (tokens)	variants (types)
THE * OF	1791	209	THE * OF	1448	220
* OF THE	1191	157	* OF THE	940	137
OF THE *	1173	149	OF THE *	835	151
IN THE *	760	94	IN THE *	634	95
* IN THE	547	98	* IN THE	413	85
* TO THE	488	68	IT IS *	286	42
TO THE *	346	69	A * OF	228	39
* TO BE	321	30	TO * THE	201	42
IT IS *	297	38	* TO THE	176	37
A * OF	289	43	* IT IS	165	25
IN * TO	283	8	the * that	160	20
TO * THE	256	52	* TO BE	156	31
can be *	246	38	TO THE *	154	35
TO BE *	243	37	is * to	138	22
on the *	237	34	that * is	132	9
of a *	221	36	at the *	130	13
* IT IS	213	27	TO BE *	128	18
the * is	210	36	IN * TO	128	7
* that the	207	29	the * and	126	33
is * to	206	29	to * a	118	17

Ute Römer (uroemer@umich.edu)



## 2.3 P-frames in NS and NNS AAW

Top-20 4-p-frames in CHALC and MICUSP\_EL

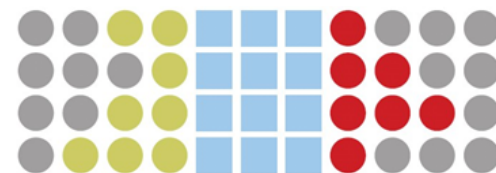
CHALC	hits (tokens)	variants (types)	MICUSP_EL	hits (tokens)	variants (types)
THE * OF THE	381	56	THE * OF THE	227	45
IN THE * OF	128	17	IN THE * OF	125	22
* THE FACT THAT	93	12	AT THE * OF	67	8
on the * hand	92	2	IN ORDER TO *	45	10
AT THE * OF	72	8	* <i>the end of</i>	41	3
IN ORDER TO *	65	13	* <i>end of the</i>	40	2
to the * of	63	16	THE END OF *	40	2
THE FACT THAT *	57	8	of the * of	32	9
* <i>the present perfect</i>	53	5	it is * that	31	5
the * of a	51	12	* THE FACT THAT	27	5
With the * of	49	6	IT IS * TO	25	3
IT IS * TO	47	6	TO THE * THAT	24	5
* <i>of the clause</i>	44	8	as a * of	23	4
<i>of the clause</i> *	42	7	THE FACT THAT *	23	3
TO THE * THAT	40	3	The * in which	22	2
in * to the	39	5	as well as *	22	3
* <i>position of the</i>	38	3	* <i>the united states</i>	22	3
THE END OF *	38	3	that there is *	21	3
<i>the present perfect</i> *	37	8	<i>j baumgardner ed</i> *	20	2
<i>of the present</i> *	36	2	a * of the	19	5

Ute Römer (uroemer@umich.edu)



### 3. Conclusion: Nativeness and academic writing expertise

- Paper explored the **phraseological profile** of NS and NNS apprentice academic writing
- It aimed to examine how **nativeness and expertise** affect language patterning
- Central results of (key) n-gram/p-frame analyses:
  - Considerable **overlap** between CHALC and MICUSP\_EL lists
  - Very few items refer to differences in acad. writing styles
  - Most differences explainable on the basis of topic-related differences between papers included in corpora
  - **NS and NNS AAWriters differ in similar ways from expert writers**, esp. in terms of p-frame use
  - CHALC writers closer to Hyland\_Ling than MICUSP\_EL writers



### 3. Conclusion: Nativeness and academic writing expertise

- When dealing with advanced-level acad. writing, we move **beyond the native/non-native distinction**
- Expertise a more important aspect than nativeness
- A need to disentangle "communicative performance in research settings from mother-tongue status per se" (Swales 2004: 57)
- NS and NNS AAWs appear to develop academic discourse competence in similar ways
- The **native academic writer** does not seem to exist



### 3. Conclusion: Nativeness and academic writing expertise

- Since NS and NNS AAWs lack similar sets of expert phraseological items, both groups may need similar training
- EAP teachers "will be better served by using the notion of **expertise** [...] rather than the notion of the native-speaker." (Tribble 2008: 307)
- In teaching academic writing:
  - put more emphasis on expertise than nativeness
  - base writing instruction on **samples of successful (or expert) writing by native or non-native speakers**



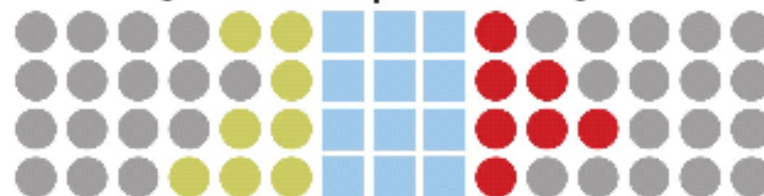


# More 4-grams...

Freq	Collocation
927	<b>thank</b> you very much
218	<b>thank</b> you <b>thank</b> you
55	<b>thank</b> you for that
53	<b>thank</b> you mr chairman
52	well <b>thank</b> you very
51	very much <b>thank</b> you
42	<b>thank</b> you for your
41	you very much <b>thank</b>

[uroemer@umich.edu](mailto:uroemer@umich.edu)  
[www.ELIcorpora.info](http://www.ELIcorpora.info)

Michigan Corpus Linguistics



# References

- Bazerman, C. (1994). Systems of genres and the enactment of social intentions. In: A. Freedman & P. Medway (eds.). *Genre and the New Rhetoric*. London: Taylor and Francis. 79–101.
- Hyland, K. (1998) *Hedging in Scientific Research Articles*. Amsterdam: John Benjamins.
- Römer, U. (2007). Learner language and the norms in native corpora and EFL teaching materials: A case study of English conditionals. In: S. Volk–Birke & J. Lippert (eds.). *Anglistentag 2006 Halle. Proceedings*. Trier: Wissenschaftlicher Verlag Trier. 355–363.
- Römer, U. (2009). The inseparability of lexis and grammar: Corpus linguistic perspectives”. *Annual Review of Cognitive Linguistics* 7: 141–163.
- Scott, M. & C. Tribble. (2006). *Textual patterns. Key words and corpus analysis in language education*. Amsterdam: John Benjamins.
- Swales, J. M. (2004). *Research Genres. Exploration and Applications*. Cambridge: Cambridge University Press.

[uroemer@umich.edu](mailto:uroemer@umich.edu)  
[www.ELIcorpora.info](http://www.ELIcorpora.info)

