

Table 1: Description of participants

Participants categories	Number	Percentage %
Gender		
Males	90	53
Females	81	47
Age		
15–20 years	129	75
21 years and above	42	25
Courses		
BA Education	3	2
BSc Health Sciences	12	7
BSc (grouped courses)	76	44
BA Arts	80	47
Home language		
Northern Ghanaian languages	13	8
English	16	9
Ewe	23	13
Ga-Adangme	24	14
Akan	95	56

Table 2: Essay content and marks awarded

Marks	Essay content
23–25	Good introduction; 4 developed body paragraphs and a concluding paragraph; devoid of errors
20–22	Good introduction; 4 developed paragraphs, including a conclusion; errors do not generally affect meaning
17–19	Introduction; 3 fully developed paragraphs and a conclusion; errors may affect meaning
14–16	3 developed paragraphs, including a conclusion; may contain several errors that affect meaning
11–13	3 paragraphs, including a conclusion; may contain several errors
0–10	Attempts to write an essay but does not finish

Table 3: Hyland's (2005a, p. 49) classification of MD

Category	Function	Example
<i>Interactive resources</i> (5 sub-categories)	Help to guide the reader through the text	
i. Transitions	Express semantic relation between clauses	In addition/but/thus/and
ii. Frame markers	Refer to source discourse acts, sequence or text stages	In conclusion/finally
iii. Endophoric markers	Refer to information in other parts of the text	Noted above/see fig./in section 2...
iv. Evidentials	Refer to source information from other texts	According to X/Z states that
v. Code glosses	Help readers grasp functions of ideational material	Namely/e.g./such as/in other words
<i>Interactional resources</i> (5 sub-categories)	Involve the reader in the argument	

Category	Function	Example
i. Attitude markers	Express writer's attitude to proposition	Unfortunately/I agree/surprisingly
ii. Engagement markers	Explicitly refer to or build relationship with reader	Consider/note that/you can see that
iii. Self-mentions	Explicit reference to authors	I/we/your/our
iv. Hedges	Withhold writer's full commitment to proposition	Might/perhaps/possible
v. Boosters	Express certainty	We proved/definitely/beyond doubt/clearly

Table 4: List of MD markers manually identified in the learner corpus

Category	Example
Interactive resources (5 sub-categories)	
i. Frame markers	in a nutshell, this essay seeks, the third, next, in conclusion, discussed below, thirdly, to begin, first, finally, to end, last but not least, to conclude, to commence, in short, to continue, above all, a second, a fourth, my purpose here is to, third, in summary, second, fourth, firstly, overall, secondly, last but not least, to start, in the following, to sum up, lastly, on the whole
ii. Code glosses	in other words, for example, such as, for instance, namely, an example is

Category	Example
iii. Endophoric markers	from the above discussion, in section..., as mentioned above, as mentioned before, see fig..., as mentioned earlier, noted above
iv. Evidentials	according to, argue that, reveal, reveals, explains, defines, states that, suggest that, suggests that, claim that
v. Transitions	among other, among others, also, and, again, additionally, although, but, therefore, in addition, though, thus, in a similar, similar, similarly, on the other hand, nonetheless, as a result, contrary, even though, likewise, another, moreover, in contrast, hence, consequently, however, unlike, and furthermore
Interactional resources (5 sub-categories)	
vi. Hedges	broadly, generally, may, around, fairly, relatively, tend, largely, mostly, sometimes, somewhat, nearly, many, about, apparently, appear, appears, perhaps, almost, seem, seems, might, approximately, likely, possible, possibly, presumably, probably, quite, some, typically, tends, usually, most, often
vii. Attitude markers	hopefully, remarkable, there is no doubt, fortunately, surprisingly, undoubtedly, no wonder that, unfortunately, understandably, ideally, I agree

Category	Example
viii. Engagement markers	consider, note that, imagine, you can see that
ix. Boosters	of course, definitely, in fact, it is proved that, indeed, it is clear, it is confirmed that, as a matter of fact
x. Self-mentions	my, our, I, we

Table 5: A random set of concordance lines for ‘also’ in the UGCD

Filename	Left	Node	Right
UGCD 2.txt	health of these youth, but	also	the economy of the country
UGCD 158.txt	not easily recycled. This is	also	one major global challenge when
UGCD 16.txt	dealing with their plastic bags.	Also	recycling of these plastic bags
UGCD 130.txt	throwing them away, we can	also	use jute, cloth or paper
UGCD 107.txt	the production of paper bags	also	led to an alarming increase
UGCD 59.txt	in this world. We are	also	aware that when man returns
UGCD 78.txt	Reduce, Re-use and Re-cycle; and	also	remember that, change begins from
UGCD 136.txt	by the toxicity in plastics.	Also,	plastic bags tend to disrupt
UGCD 95.txt	not only acutely damaging but	also	seriously harmful agriculture. It may
UGCD 175.txt	food source for marine creatures.	Also,	creatures could be entangled-restricting

Table 6: Description of the native speakers' corpus

Name	Language	Texts	Tokens (running words)
Native speakers' corpus	English	207	167 404

Table 7: LC and NSC

Name	Language	Texts	Tokens/running words
Learner corpus	English	197	98 069
Native Speakers' corpus	English	207	167 404

Table 8: Summary of participants' information

Categories	Variables	Number of students
Gender	Male	16
	Female	4
Age	17–20	16
	21–25	4
Home language	Ga/Dangme	6
	Akan	10
	Ewe	2
	Gurene (Northern Ghanaian language)	2
Course of study	Accounting finance	14
	Business administration	4
	Law	2

Table 9: Interactive and interactional metadiscourse markers in the pilot cloze test

	Number of correct selections of metadiscourse markers in categories	Percentage of correct selections of metadiscourse markers in categories
Interactive markers		52%
Evidentials	8	40%
Transitions	10	50%
Code glosses	12	60%
Endophoric markers	12	60%
Frame markers	10	50%
Interactional markers		58%
Hedges	8	40%
Boosters	12	60%
Engagement markers	8	40%
Attitude markers	12	60%
Self-mentions	18	90%

Table 10: Research questions, data collected, and the corresponding analysis

Research question	Data	Analysis	Objectives
RQ1: To what extent were students able to select appropriate and correct metadiscourse markers to complete a cloze exercise?	Cloze test	Quantitative (SPSS)	To determine students' correct selection of metadiscourse markers in a passage
RQ2: How were metadiscourse markers utilised in the learner corpus?	Learner corpus	Quantitative (#LancsBox analysis)	Frequencies of metadiscourse markers used in essays
		Qualitative (context analysis)	Types of metadiscourse markers used in essays
RQ3: How did students' use of metadiscourse markers compare to that of native speakers?	Learner corpus and native speakers' corpus	Quantitative (#LancsBox analysis)	Types and frequencies of MD markers in learner corpus and native speakers' corpus
		Qualitative (manual analysis)	Appropriate use of MD markers in learner corpus and native speakers' corpus

Table 11: Results of the cloze test

Gap number	Total number of students	Number of students who responded correctly	Percentage	Word (category)
1	171	42	25%	X explains (evidential)
2	171	121	71%	Such as (code gloss)
3	171	68	40%	In addition (transition)
4	171	99	58%	May (hedge)
5	171	95	55%	It is clear that (booster)
6	171	105	62%	Note that (engagement marker)
7	171	112	66%	Unfortunately (attitude marker)
8	171	55	33%	Noted above (endophoric marker)
9	171	124	73%	Finally (frame marker)
10	171	113	67%	I (self-mention)

Table 12: Presentation of correct answers for gender

Gender	N	Mean	Std. deviation
Female	83	50.48	25.514
Male	88	57.84	22.866
Total	171	54.27	24.397

Table 13: Presentation of correct answers for age

Age	Number	Mean	Std. deviation
17–20	129	54.11	25.667
21+	42	54.76	20.271
Total	171	54.27	24.397

Table 14: Presentation of correct answers for home language

Home language	Number	Mean	Std. deviation
Northern languages	13	53.03	27.804
English	16	54.38	27.072
Ewe	23	55.22	25.382
Ga-Adangme	24	52.92	26.943
Akan	95	54.53	23.053
Total	171	54.27	24.397

Table 15: Presentation of correct answers for course of study

Grouped courses	Number	Mean	Std. deviation
BA Arts	80	52.7	23.06
BA Education	3	66.67	5.77
BSc Applied Sciences	7	54.87	26.00
BSc Health Sciences	12	57.50	26.33
Total	171	54.27	24.40

Table 16: Frequencies and percentages of interactive resources in the LC

Interactive resources	Frequency	Percentage
Frame markers	447	7.35%
Code glosses	334	5.49%
Endophoric markers	3	0.04%
Evidentials	70	1.15%
Transitions	5 221	86.94%
Total	6 075	100%

Table 17: Frequencies and percentages of interactional resources in the LC

Interactional resources	Frequency	Percentage
Hedges	1 470	59.1%
Attitude markers	34	1.5%
Engagement markers	8	0.32%
Boosters	33	1.33%
Self-mentions	921	37.34%
Total	2 466	100%

Table 18: Results of descriptive analysis of metadiscourse use by home language

Home language	MD markers	N	Mean	Std. deviation	Total MD markers
Akan	Interactive	121	29.34	6.275	3 545
	Interactional	121	12.10	6.160	1 490
English	Interactive	16	32.8	9.831	524
	Interactional	16	12.56	7.202	227
Ewe	Interactive	23	31.87	8.165	740
	Interactional	23	11.22	4.451	275
Ga-Adangme	Interactive	24	33.04	5.599	800
	Interactional	24	9.17	6.391	237
Northern Langs.	Interactive	13	35.46	7.125	467
	Interactional	13	17.00	9.055	236

Table 19: Hypothesis test summary for home language

Null hypothesis	Test	p-value	Decision
1. The distribution of interactive resources is the same across categories of home language.	Independent-samples Kruskal-Wallis test	.008	Reject the null hypothesis
2. The distribution of interactional markers is the same across categories of home language.	Independent-samples Kruskal-Wallis test	.017	Reject the null hypothesis

Table 20: Results of descriptive analysis of metadiscourse use by course of study

Course of study	MD markers	N	Means	Std. deviation	Total
BSc Applied Sciences	Interactive	104	30.94	6.366	3 218
	Interactional	104	12.04	7.075	1 253
BA Education	Interactive	3	26.00	3.000	78
	Interactional	3	11.00	8.185	33
BSc Health Sciences	Interactive	12	32.67	8.026	392
	Interactional	12	9.67	5.193	116
BA Arts	Interactive	78	30.33	8.352	2 387
	Interactional	78	12.35	6.479	1 064

Table 21: Hypothesis test summary for course of study

Null hypothesis	Test	p-value	Decision
1. The distribution of interactional resources was the same across courses of study.	Independent Kruskal-Wallis test	.523	Reject the null hypothesis
2. The distribution of interactive resources was the same across courses of study.	Independent Kruskal-Wallis test	.299	Reject the null hypothesis

Table 22: Performance levels and use of MD markers in the LC

MD markers	Performance levels	N	Mean	Std. deviation	Total MD markers
Interactive resources	High	70	31.97	6.646	2 250
	Average	70	31.57	7.496	2 219
	Low	57	28.19	7.259	1 606
	Total N	197	30.74	7.284	6 075
Interactional resources	High	70	11.46	5.999	836
	Average	70	11.79	6.799	859
	Low	57	12.93	6.668	771
	Total N	197	12.00	6.479	2 465
Interactive and interactional					8 541

Table 23: Hypothesis test summary for performance levels

Null hypothesis	Test	p-value	Decision
1. The distribution of interactive resources was the same across the three performance levels.	Independent Kruskal-Wallis test	.003	Reject the null analysis
2. The distribution of interactional resources was the same across the three performance levels.	Independent Kruskal-Wallis test	.387	Accept the null analysis

Table 24: Pairwise comparisons of performance levels

Sample 1–Sample 2	p-value
Low–average	.006
Low–high	.001
Average–high	.563

Table 25: Summary of quantitative details of LC and NSC

Corpus	No. of words	RF of MD	NF of MD	MD density
LC	98 069	8 541	87 091	8.70
NSC	167 404	11 531	68 881	6.88

Table 26: Distribution of interactive resources in the LC and the NSC

Interactive resources	Learner corpus				Native speakers' corpus			
	RF	NF	% (NF)	Ranking	RF	NF	% (NF)	Ranking
Endophoric markers	3	30	0.03	5th	1	5	0.00	5 th
Evidentials	70	713	0.81	4th	180	1 075	1.56	4 th
Code glosses	334	3 405	3.90	3rd	206	1 230	1.78	3 rd
Frame markers	447	4 558	5.23	2nd	410	2 449	3.55	2 nd
Transitions	5 221	53 238.	61.12	1st	6 078	36 307	52.70	1 st
Total	6 055	61 743	71.09		6 875	41 068	59.59	

Table 27: LL values for interactive resources across the LC and the NSC

Interactive resources	% NF (LC)	% NF (NSC)	LL value	Significance status
Endophoric markers	0.05	0.01	2.40	Not significant
Evidentials	1.15	1.56	8.94	Significant
Code glosses	5.51	1.79	137.23	Significant
Frame markers	7.38	3.56	81.93	Significant
Transitions	61.23	52.71	405.07	Significant

* Significance level: 3.84

Table 28: Statistical details of interactional resources in the LC and the NSC

Interactional resources	Learner corpus				Native speakers' corpus			
Category	RF	NF	%	Ranking	RF	NF	%	Ranking
Engagement markers	8	81	0.09	5th	45	266	0.39	5 th
Boosters	33	336	0.39	4th	128	764	1.11	3 rd
Attitude markers	34	346	0.40	3rd	62	336	0.54	4 th
Self-mentions	921	9 389	10.78	2nd	1 828	10 791	15.85	2 nd
Hedges	1 470	14 989	17.21	1st	2 593	15 307	22.49	1 st
Total	2 466	25 140	28.87		4 656	27 464	40.38	

Table 29: LL values for interactional resources across the LC and the NSC

Interactional resources	% NF (LC)	% NF (RC)	LL value	Significance status
Engagement markers	0.09	0.39	12.45	Significant
Boosters	0.39	1.11	20.44	Significant
Attitude markers	0.40	0.54	0.10	Not significant
Self-mentions	10.78	15.85	14.44	Significant
Hedges	17.21	22.49	1.01	Not significant

* Significance level: 3.84

Table 30: Frequency and LL results for identified frame markers

Frame markers	RF LC	RF NSC	NF LC	NF NSC	LL value	Significance status
The third	1	12	10	71	6.01	Significant
Above all	1	1	10	5	0.14	Not significant
In the following	1	0	10	0	1.99	Not significant
On the whole	1	1	10	5	0.14	Not significant
A fourth	1	2	10	11	0.02	Not significant
Next	2	43	20	256	27.27	Significant
To commence	2	0	20	0	3.98	Significant
Fourth	2	8	20	47	1.35	Not significant
This essay seeks	3	0	30	0	5.98	Significant
Discussed below	3	0	30	0	5.98	Significant
To end	3	12	30	71	2.03	Not significant

Frame markers	RF LC	RF NSC	NF LC	NF NSC	LL value	Significance status
In a nutshell	3	0	30	0	5.98	Significant
Over all	4	14	40	101	1.81	Not significant
In short	5	4	50	23	1.28	Not significant
To continue	6	11	61	65	0.02	Not significant
A second	7	5	71	29	2.25	Not significant
In summary	8	0	81	0	15.93	Significant
Last but not least	8	0	81	0	15.93	Significant
Third	10	30	101	179	2.60	Not significant
Second	11	44	112	262	7.44	Significant
To start	11	6	112	35	5.37	Significant
To sum up	12	0	122	0	23.90	Significant
Thirdly	13	2	132	11	15.96	Significant
Lastly	26	3	265	17	35.26	Significant
Firstly	28	5	285	29	32.31	Significant
To conclude	31	6	316	35	34.48	Significant
Secondly	35	4	365	23	47.60	Significant
To begin	37	7	377	41	41.59	Significant
First	39	154	397	919	25.44	Significant
Finally	47	26	479	155	22.51	Significant
In conclusion	86	10	876	59	116.35	Significant
Total	447	410	4 558	2 449	81.93	Significant

RF: Raw frequency; NF: Normalised frequency; NSC: Native Speakers' corpus; LC: Learner corpus

Table 31: Frequency and log-likelihood results for identified code glosses

Code glosses	RF LC	RF NSC	NF LC	NF NSC	LL value	Significance status
An example	3	1	30	5	2.40	Not significant
Namely	4	6	40	35	0.04	Not significant
In other words	9	4	91	23	5.57	Significant
For example	19	66	193	394	8.38	Significant
For instance	26	14	265	83	12.90	Significant
Such as	273	115	2 783	686	178.15	Significant
Total	334	206	3 405	1 230	137.23	Significant

Table 32: Frequency and log-likelihood results for identified endophoric markers

Endophoric markers	RF LC	RF NSC	NF LC	NF NSC	LL value	Significance status
Noted above	0	1	0	5	0.92	Not significant
As mentioned before	1	0	10	0	1.99	Not significant
As mentioned earlier	2	0	20	0	3.98	Significant
Total	3	1	30	5	2.40	Not significant

Table 33: Frequency and Log-likelihood results for identified evidentials

Evidentials	RF LC	RF NSC	NF LC	NF NSC	LL value	Significance status
Define(s)	1	6	10	35	1.78	Not significant
Argue(s)	2	35	20	209	20.70	Significant
Claim(s)	2	25	20	149	12.78	Significant
Reveal(s)	3	3	30	17	0.42	Not significant
Explain(s)	4	5	40	29	0.21	Not significant
State(s)	4	30	40	179	11.00	Significant
Suggest(s)	14	11	142	65	3.730	Not significant
According to	40	65	407	388	0.06	Not significant
Total	70	180	713	1 075	8.94	Significant

Table 34: Frequency and log-likelihood results for identified transitions

Transitions	RF LC	RF NSC	NF LC	NF NSC	LL value	Significance status
In a similar	1	0	10	0	1.99	Not significant
Contrary	2	10	20	59	2.39	Not significant
Similar	3	20	30	119	6.61	Significant
Nonetheless	3	0	30	0	5.98	Significant
Likewise	3	7	30	41	0.21	Not significant
Similarly	4	6	40	35	0.04	Not significant
Consequently	4	8	40	47	0.07	Not significant
In contrast	5	4	50	23	1.28	Not significant
Among other(s)	8	4	81	23	4.35	Significant
Additionally	13	0	132	0	25.89	Significant
On the other hand	14	28	142	167	0.24	Not significant
Unlike	17	8	173	47	9.89	Significant
Even though	19	34	193	203	0.03	Not significant
Again	33	59	336	316	0.05	Not significant
Though	51	80	520	477	0.2	Not significant
In addition	54	16	550	95	47.05	Significant
Moreover	57	6	581	35	79.43	Significant
Although	60	67	611	400	5.62	Significant
Furthermore	64	9	652	53	81.25	Significant
Another	66	180	672	1 075	11.32	Significant
Thus	85	55	866	328	32.41	Significant
However	93	218	948	1 302	6.82	Significant

Transitions	RF LC	RF NSC	NF LC	NF NSC	LL value	Significance status
As a result	98	24	999	143	96.34	Significant
Hence	98	9	999	53	141.70	Significant
Therefore	147	105	1 498	627	47.30	Significant
But	198	662	2 018	3 954	76.82	Significant
Also	554	421	5 649	2 514	158.20	Significant
And	3 467	4 038	35 352	24 121	268.38	Significant
Total	5 221	6 078	53 238	36 307	405.07	Significant

Table 35: Frequency and log-likelihood results for identified hedges

Hedges	RF LC	RF NSC	NF LC	NF NSC	LL value	Significance status
Somewhat	0	11	0	65	10.14	Significant
Apparently	0	6	0	35	5.53	Significant
Fairly	1	9	10	53	3.79	Not significant
Presumably	1	1	10	5	0.14	Not significant
Perhaps	2	58	20	346	39.93	Significant
Possibly	2	21	20	125	9.76	Significant
Probably	2	53	20	316	35.68	Significant
Relatively	3	9	30	53	0.78	Not significant
Generally	5	14	50	101	0.97	Not significant
Appear(s)	5	26	50	155	6.54	Significant
Typically	5	4	50	23	1.28	Not significant

Hedges	RF LC	RF NSC	NF LC	NF NSC	LL value	Significance status
Nearly	8	20	81	119	0.87	Not significant
Might	8	59	81	352	21.34	Significant
Largely	9	2	91	11	9.34	Significant
Approximately	12	5	122	29	7.91	Significant
Seem(s)	13	135	132	8 064	62.33	Significant
Likely	13	37	132	221	2.71	Not significant
Possible	21	63	214	376	5.45	Significant
Quite	28	29	281	173	3.51	Not significant
Almost	32	32	326	191	4.52	Significant
Mostly	35	9	356	53	33.42	Significant
Usually	43	45	438	268	5.19	Significant
Sometimes	44	31	448	185	14.52	Significant
Often	80	81	815	483	10.85	Significant
Tend(s)	81	24	825	143	70.57	Significant
About	85	339	866	2 025	57.03	Significant
Around	100	70	1 019	418	33.37	Significant
May	156	237	1 590	1 415	1.27	Not significant
Some	218	344	2 222	2 054	0.82	Not significant
Most	218	302	2 222	1 804	5.45	Significant
Many	240	517	2 447	3 088	9.11	Significant
Total	1 470	2 593	14 989	15 489	1.01	Not significant

Table 36: Frequency and log-likelihood results for identified attitude markers

Attitude markers	RF LC	RF NSC	NF LC	NF NSC	LL value	Significance status
Remarkable	0	2	0	11	1.84	Not significant
Surprisingly	0	2	0	11	1.84	Not significant
I agree	0	8	0	47	7.38	Significant
Fortunately	1	2	10	11	0.02	Not significant
Hopefully	1	12	10	71	6.01	Significant
Ideally	1	1	10	5	0.14	Not significant
Understandably	2	0	20	0	3.98	Significant
No wonder	3	1	30	5	2.40	Not significant
Undoubtedly	8	6	81	35	2.35	Not significant
There is no doubt	9	2	91	11	9.34	Significant
Unfortunately	9	26	91	155	2.00	Not significant
Total	34	62	346	370	0.10	Not significant

Table 37: Frequency and log-likelihood results for identified engagement markers

Engagement markers	RF LC	RF NSC	NF LC	NF NSC	LL value	Significance status
Note that	2	3	20	17	0.02	Not significant
Imagine	3	6	30	35	0.05	Not significant
Consider	3	36	30	212	18.02	Significant
Total	8	45	81	268	12.45	Significant

Table 38: Frequency and log-likelihood results for identified boosters

Boosters	RF LC	RF NSC	NF LC	NF NSC	LL value	Significance status
Definitely	3	17	30	101	4.74	Significant
Indeed	3	24	30	143	9.27	Significant
It is clear that	4	6	40	35	0.04	Not significant
Of course	5	40	50	238	15.45	Significant
In fact	6	41	61	244	13.86	Significant
It is proved that	6	0	61	0	11.95	Significant
As a matter of fact	6	0	61	0	11.95	Significant
Total	33	128	336	764	20.44	Significant

Table 39: Frequency and log-likelihood results for identified self-mentions

Self-mentions	RF LC	RF NSC	NF LC	NF NSC	LL value	Significance status
My	4	219	40	1 308	169.84	Significant
I	38	679	226	4 056	404.66	Significant
We	279	527	2 844	3 148	1.89	Not significant
Our	600	403	6 118	2 407	215.15	Significant
Total	921	1 828	9 391	10 919	14.14	Significant