

**Seeing the wood and the trees:
Reconciling findings from discourse and lexical analysis.**

John McKenny

Northumbria University

In this paper, I describe an exploration of a sub-corpus of learner English which views the data from widely different perspectives: at the macro or discourse level and the word or phrase level. The corpus under scrutiny is PORTICLE, the Portuguese sub-corpus of ICLE, the International Corpus of Learner English. This sub-corpus is a collection of argumentative essays written by Portuguese advanced learners of English and currently contains 126,000 running words. The data could be classed as quasi-naturalistic given that although the students did not know the purpose of the investigation, they had every right to suspect that anything they wrote might be used against them (Scholfield 1995:47).

I am fortunate to be a part of the ICLE community. This worldwide ring of investigators, with its hub in Louvain, is responsible for a growing body of research on the various sub-corpora of ICLE which are being explored individually and compared to each other and to native speaker argumentative writing. As the investigations continue, relative newcomers like myself can draw on the findings that have been published (Granger 1998) and continue to be published about established and burgeoning sub-corpora (Granger S, Hung J, Petch-Tyson S (eds) 2003) . For example, those aspects of learner English which can be ascribed to mother tongue transfer are distinguished from those features which are shared by all ICLE sub-corpora. Interestingly, there are two Portuguese sub-corpora growing up, PORTICLE in Portugal and a sub-corpus being compiled in Brazil at the Catholic University of Sao Paolo. The co-ordinator of the Brazilian sub-corpus, baptized BRICLE, is Tony Berber Sardinha. Each national team of compilers can compare their sub-corpus with the others to see which features it shares and where and how it differs.

For comparative purposes I analyse a comparable number words from another essay corpus, LOCNESS, which consists of essays written by British and American undergraduates. Incidentally, access to this corpus compiled by the Centre for English Corpus Linguistics at the Catholic University of Louvain is another perk for those belonging to the ICLE research community. The section I use is written exclusively by American undergraduate students.

I am aware that we all tend to specialize in one or several areas of linguistics, pure or applied, and can become less interested in areas more removed from our present concerns. My Ph.D. research is on the phraseology of English for Academic Purposes. When I recently engaged in discussions with colleagues about their work in Conversational Analysis and Language Ecology I initially felt less enthusiasm than I should have done. I soon discovered that openness to other areas of linguistics (and indeed to other social and pure sciences) can pay dividends in terms of new ideas and ways of seeing things. This was brought home to me at a recent seminar on the ecology of language where a colleague, Mark Garner (2003), came up with his analysis of data collected to investigate what native speakers really talk about and do with language when they meet for the first time, as opposed to what language textbooks imagine they do . I was initially surprised by the convergence between his findings about spoken language and my own on reading into my PORTICLE essays. Here is Garner's list of functions so far:

- 1) Establishing common ground
- 2) Self-disclosing (reminiscing, expressing feelings, opinions, hopes and plans)
- 3) Prompting other's self-disclosure
- 4) Keeping the conversation going

With the possible exception of 3) these functions of recently acquainted conversants are similar to those performed by writers of argumentative prose. I would suggest that 3) is never far away: the writer's inveigling the readers to admit that they know they agree with the writer . The similarity should not surprise us: the essay writer is addressing his reader as if for the first time (even Montaigne in his later essays would have to anticipate new, virgin readers). As we study the argumentation of the essays we must not lose sight of the written equivalent of number 4) above, keeping the conversation going or in writing, maintaining the readers' interest. This seems to provide a more writer-oriented model than Hoey's (1983) more reader-oriented Situation-Problem-Solution-Evaluation model of text macrostructure and I hope to apply it to the PORTICLE sub-corpus in the months ahead.

The number of essays in my learner corpus and the control corpus of native-speaker writers, approximately 240,000 words in all, means that sitting down and reading them and making notes on salient features was out of the question. Even to read a novel of comparable length, say Jane Austen's *Emma* retaining consistent linguistic and literary observations would prove a mammoth task. To get down into lexical and phraseological details while reading 'manually' would prove forbiddingly labyrinthine. Computer assistance seems to be called for.

I decided to choose a small sample of NNS and NS essays to examine at the level of discourse. The purpose of this real reading rather than electronic scanning was to seek salient features of the NS and NNS essays and to get a feel for the writers' approach to argumentation. This approach is really the key to all responsible use of learner corpora. Before setting off on bouts of computerized quantification, the researcher needs to continually re-discover the textuality and discursive flow of the data by reading it with the attention it deserves. This continual toing and froing from quantitative and qualitative analysis, with the two different approaches feeding naturally into each other, seems to be the most adequate methodology for interrogating learner corpora. This dual approach has been recommended within ICLE from the outset (Granger 1989:16).

Obviously, there are some discourse features which lend themselves to computer-assisted corpus investigation. The use of discourse markers such as 'first' or 'firstly', 'second' or 'secondly' and then 'finally' or 'in conclusion' can be measured and plotted through all the texts. A list of such text-structuring words and phrases can be built up iteratively and run through the corpus. Key argumentative words such as 'but', 'so', 'therefore', 'thus', 'then', 'since' and 'suppose' can be concordanced and tracked through the texts. Important though such anatomical takes on the texts might be, there are many discourse features which slip completely through such a net. We are dealing with a huge deficiency in most quantitative approaches. The purpose of an argumentative essay is to persuade and until such days as a Leibnizian 'universal character' has been invented, when we can translate all the essays into strings of predicate calculus symbols and say 'calcuemus' and decide exactly how cogent and persuasive each essay is, the only way such persuasiveness can be evaluated is through the engagement of an interested reader with each text. The use of the common *topoi* of logic, e.g. *modus ponens* and *modus tollens* which are a closed set of argumentative schemata, is another aspect of the essays which I intend to examine in the next stage of my investigation drawing on the work of McElholm (2002).

We can compare Wordlists and N-gram lists from the two corpora and find over- and under-used words and phrases. We can compare lexical density and suggest that the writing in one corpus tends to be more towards the spoken end of the spoken-written continuum. Let's look at the two compared wordlists of PORTICLE and LOCNESS and see what we can glean. What we find are tendencies about which we can make hypotheses which we can then go back and test on the essays. We must be wary of only testing those characteristics which lend themselves to quantitative investigation. Using a stoplist of 172 structure or grammatical words which I generated by extracting all words tagged with DT, CC, PRP, PRP\$, TO, WDT, WP\$, WRB, WP in the Brown corpus (method recommended by Diego Molla Aliod in personal communication), I calculated the number of lexical words in each essay i.e. those words which were not filtered out from the wordlists and given as types. Developing a suggestion from Halliday (1985), the lexical density of the essays in each corpus was calculated as the ratio of the lexical words in relation to the total number of words in each essay, expressed as a percentage. This Hallidayan lexical density had the same value as the Type/Token ratio in Wordsmith Tools when a stoplist of grammatical words is used. Figures of 37.9% lexical items in LOCNESS compared with 29.5% for Portuguese students show us the not too surprising fact that the native speakers have a richer lexis at their command. But it should be remembered that ICLE scholars found that NS writers can display lower lexical density than NNS writers and that in NNS essays lexical richness does not correlate highly with quality of writing. Perhaps the number of different words used is less important than what is done with the words.

When we use the Wordsmith Tools Wordlist function which compares two wordlists we are shown those words that are significantly overused or underused by the two target groups of students. Portuguese students use 'we' 1,791 times while the American NS subjects used it 273 times, i.e. well over six times more. When we add this startling fact to the 1,470 Portuguese uses of 'I' as opposed to 383 American uses of the first person singular pronoun, and look at 'our' (770 v206), 'us' (346v65), 'my' (428v129) we can see that there is a very different deployment of the pronoun system between the two groups. This overuse of 'I' is much greater than any other language group in ICLE (571 occurrences in SPICLE, the Spanish sub-corpus and 394 occurrences in the Brazilian sub-corpus (my version kindly furnished by Tony Berber Sardinha was a smaller sub-corpus of 74,617 words). Although Fanny Meunier

(Granger 1989: 32) warns against comparing corpora of different sizes this BRICLE figure for the occurrence of 'I' could be crudely converted to a percentage of the whole text (0.5%) compared to 1.4% for PORTICLE. This is startling as two groups with the same mother tongue display such a difference. If I return to the discourse level and to my role as a reader of student argumentative essays, I can point out that Portuguese students are not content to aver a proposition but seem to feel the need to put their whole being behind the asseveration. I am thinking primarily of Hunston's (2002) concept of the status of a proposition within a discourse. She describes how a writer gives a status to each clause in their essay and clauses are typically averred or attributed.

Here are a few examples from 157 examples of constructions using *I* followed by *believe*

I really believe that this is not a question of theory.
I do believe that the fantasy world is really,
I strongly believe that we must create a better world
Being so I do not truly believe that this solution could be applied.
From my point of view I believe that the years that we, nowadays,
I personally believe that it's important to help others

I often had the impression that these Portuguese writers in English were writing letters of reference for the point of view being defended and were putting themselves forward as personal advocates. "You will believe what I'm telling you because I'm a good person and you, I hope, are a good person.". There are 90 examples of *in my opinion* in PORTICLE but LOCNESS has only four exemplifications. *I'm sure* looks as if it is being used comparably often with three and four uses in LOCNESS and PORTICLE respectively but when we add the complete form *I am sure*, the combined totals for the contracted and un-contracted forms rise to five and nine. Another striking contrast is the 117 uses of *I think that* used in PORTICLE compared with a mere six instantiations found in LOCNESS. We must be careful not to get carried away with the greater frequency of this prefabricated phrase used often for averral and occasionally for hedging. We find a hundred more occurrences of this prefab in PORTICLE but it might be the case that writers in LOCNESS are carrying out a comparable number of averrals using different strategies. It must be remembered that the simplest and often the most effective way of averring something is to state it in a declarative sentence. Such a bald statement of something usually carries with it the commitment of the writer to its truth or verisimilitude. (or acceptability). Aarts and Granger (1998:137) observed 'striking differences in the way learners and native speakers begin their sentences' through their study of sentence-initial trigrams. They found that they tended to begin their sentences with something other than the subject. I think the averral systems of the two groups of students (NS and NNS) is worth further study and the Hunston model would provide a rich frame of reference for examining how the writers assemble their sentences and the value they give to each proposition they add to their prose. The work of Petch-Tyson (1998) uses different concepts (writer/reader visibility) but comes to similar conclusions about EFL written discourse. The advanced learner of English is more concerned with interpersonal involvement in their writing at the expense information content.

The differences between the two groups of arguers might be at a different level. The Portuguese writers might be thinking more of influencing their readers feelings and also maintaining a good relationship with them (including writing in an enjoyable way, sometimes they use humour, other times moral indignation or poetic imagery). The LOCNESS writers seem to pursue a more argumentative Aristotelian approach. Note the use of *argument* 194 times by LOCNESS writers and only four times by PORTICLE writers. If we lemmatize and include *arguments* and *argue* the difference in use of what could be viewed as key meta-terms is even more striking: 300 v19 uses.

One methodology for text analysis which seemed to hold out hopes for bridging the gap between discourse and the lexical level was the software and formulae evolved by Youmans. I came to Youmans's work through Michael Stubbs' (2002) analysis of the stages of James Joyce's story "Eveline" using the Vocabulary Management Profile. Although it had thus far been put only to uses within literary criticism, I thought that here might be a way to delineate more precisely the moves talked of within genre analysis (e.g. Swales 1990). I began to upload my student texts to Youmans's most accommodating website in Missouri and got back wonderful Vocabulary Management Profiles (VMPs). Below is the VMP for the first essay in PORTICLE. I include an edited version of the VMP statistics for the same essay in the appendix. As my main interest is in prefabs or formulaic sequences, I looked in each of the VMP statistics for each essay expecting to find more prefabs in the valley section of the essay profile but was surprised to

notice that peaks or their surrounding text seem to contain more prefabs than did the environs of the valleys.

Here is Youmans's explanation of why he chose VMP 2.2 and how he computes it:

VMP2.2 computes ratios wrap-around style, for the second pass through a text. Hence, the first occurrence of a word such as "the" (near the beginning of a text) occurs shortly after its last occurrence (near the end of the text); hence, its ratio is nearer to 0.0 than to 1.0. The same is true for all other repeated words; their first occurrences are assigned ratios greater than 0.0 and less than 1.0. Words that appear only once in the text are assigned ratios = 1.0. Unlike VMP2.1, VMP2.2 shows no rapid downtrend at the beginning of a text. VMP2.2s mirror our second readings of texts, when the beginnings are as familiar to us as the ends. Because we normally associate rhetorical structure with second (and subsequent) readings rather than first readings, VMP2.2 is the default program selected for this web site.

Youmans (2001:1)

I began to study the valleys and the peaks. The valley, where the type/token ratio falls to a low, is where the writer is reiterating and consolidating and recycling already used lexis. Youmans suggests that this often corresponds to the end of a paragraph or section of a story or essay. According to Youmans, less-recently-used vocabulary at the ends of moving intervals tends to correlate with new topics, whereas more-recently-used vocabulary tends to correlate with a continuation of the same topic. Hence VMP 2.2s are surprisingly sensitive indicators of the ebb and flow of new topics in discourse'. I found there to be in most essays 3 or 4 main valleys followed by a similar number of peaks. All but the final valley was also followed by a peak. (As I carried on studying the VMP, I decided to use *trough* as my metaphor for the lowest points on the graphs as peak and trough seem to collocate more frequently in writing and talk about graphs). Interestingly a small group of writers (approx. 20%) did not end with a downturn but with a rise in their introduction of new or earlier words. These were the essayists who became inspired towards the end and began to do what most teachers proscribe: introduce new ideas within the conclusion.

I noticed that Youmans's formula was no respecter of formulaic expectancies. The reader can look at words 100-102 in the appendix and see that in the case of *and so...on* the *so* is given the type-token ratio over the 35 word interval of 0.5240471, and the extremely predictable *on* gets the value 0.5466959: a rise is registered where a human assessor would suggest that *on* was completely predictable and deserved a drop or at least a plateau on the graph. In the Appendix, I only highlight some of the prefabs contained in the essay: those occurring near a peak or a trough. Although I am very impressed by the graphs and statistics, I have not managed to find a generalizable way to apply them to large collections of essays. I did find that all my essays had a peak roughly midway through at the 250-300 word point. Usually this was also the point in the essay with the highest type-token ratio and so was where the student writers marshalled their word-hoard to greatest effect. Sometimes in longer essays this climax comes later but still at the half-way stage. I haven't therefore found a good use for the VMP but will continue to look for potential applications of it to EAP. I am confident that such a good idea will not go to waste and if I fail to discover a use for it someone else will. This indeed was a challenge made by Stubbs (2002) to his readers. Here is the essay whose VMP graph appears below:

<ICLE-PT-ESE-0001.1.>

I agree with this statement because I believe that nowadays it's given a lot of value to materialism. Our modern world, which is ruled by science, technology and industrialisation is creating caps between people. We don't have the human contact that we had many years ago, nowadays we depend more of technology. It's true that this has given to us a lot of development, but and the relations between man?

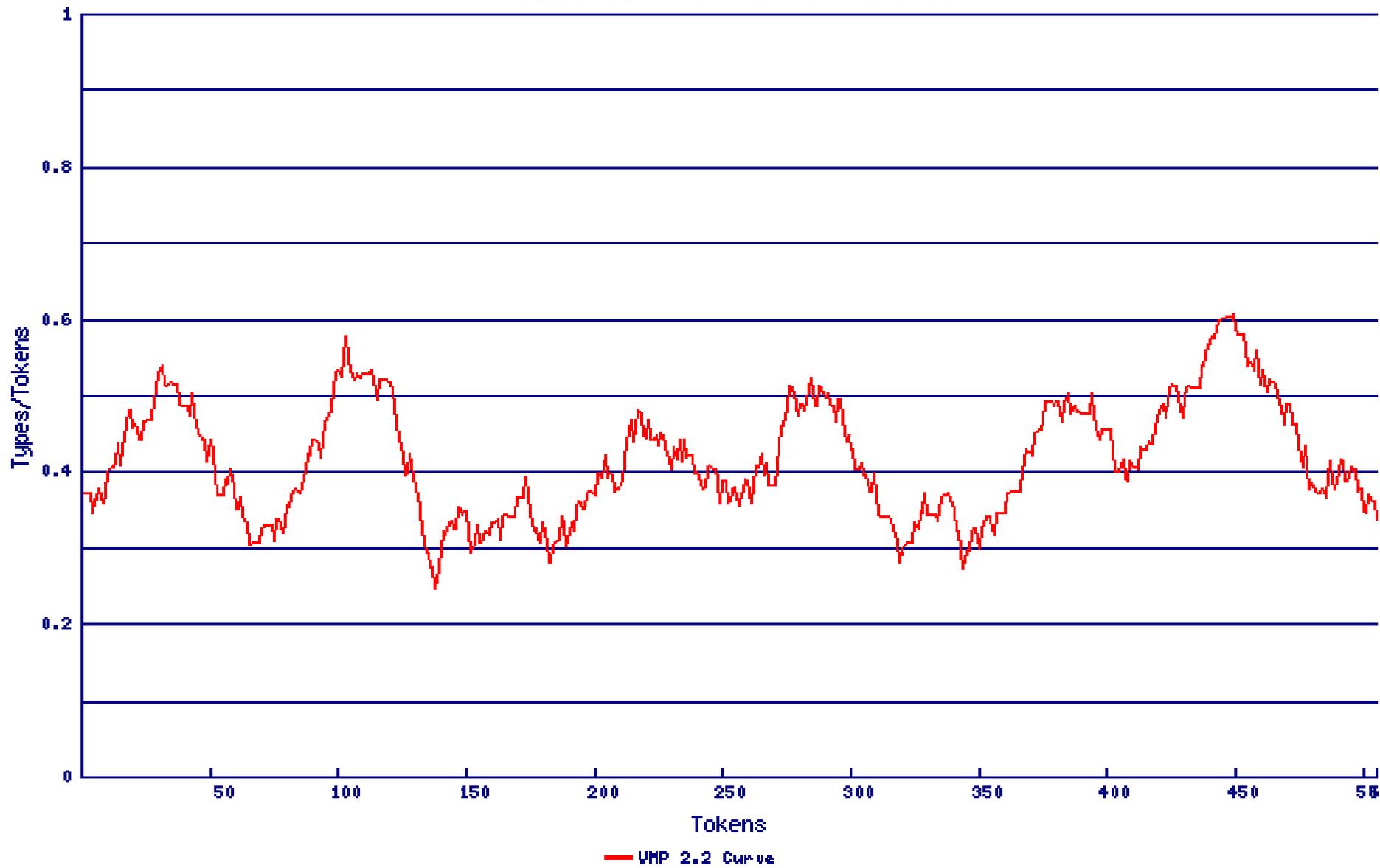
What about dreaming? What about imagination? I believe that these are the most important things that we have.

I'm going to give an example. Nowadays children don't have free time, if they aren't at school they are having classes of music, painting, ballet and so on. Parents don't have time because they are very occupied at work, so they put their children at extra-curricular classes. I believe that it's very important for children and for us to have free time just to do nothing. It's when we aren't being useful, this is, when we aren't doing nothing, that we dream, that we appeal to our imagination. Nowadays we don't have this, there is always something to do. At extra-curricular classes children do appeal to their imagination, however it's an guided imagination.

We live in a globalized world, everything is near us, everybody is next to us (due to the net), but, and what about the physical contact, the look eyes to eyes, the hugs? This is so important, even if we have the webcam, it isn't the same thing!

Nowadays we don't dream, we are more concerned with world problems, which we created, than with imagination. Many times I think that my imagination is getting sofucated with all the information that I receive through t.v, net, radio etc. Our world is starting to be extremely industrialised, work... work... work..., that's all many of us think.

PTES1001.txt - Interval: 35



The only time that I appeal to my imagination is when I'm reading a book or when I look to a picture, when I stop a bit and I "turn off" from the reality. The rest of the time I'm too concerned in my work to stop and dream. I don't like the way world is, I would have loved to live in those times when family got together and ask to the story teller to tell a story. This is an image that I have constantly in my head, nowadays we don't give value to all the old god values. We are very materialistic, we don't give value to simple things like look into the sky and try to guess what the clouds are telling us. Technology is everywhere, science is discovering everything that can be discovered, but should all the secrets be discovered? Shouldn't we leave anything in secret? I'm afraid that with this we destroy our world more than already is. I believe that sometimes we must stop and think a little in what we want, what we wish, what we are doing to nature. We must continue dreaming, we must take care of our world, because it is the world that gives everything to us!

When I examine this essay as a piece of argumentative writing I find it quite persuasive. Prefabs (marked in blue) seem to make up roughly 10% of the text. I have deliberately hedged the previous estimate as one person's prefab is another person's freshly coined phrase. The writer's phraseology is heavily influenced by Portuguese (*given a lot of value; depend more of technology; but and...; think a little in what we want*) and there are numerous "comma splices", one of the deadly sins for many EAP writers. Many of these splices would be acceptable punctuation in Portuguese.

The writer builds up a picture of how all the benefits that accrue to us from science and technology might come to naught if we lose touch with our own imagination and that of our culture. It reminds me of the gospel passage where Jesus asks the rhetorical question: what shall it profit a man if he gains the whole world and loses his soul? (Matthew 16). There are poetic images of families sitting down together to listen to a specially commissioned storyteller (as still happens in Donegal and Marrakesh) or of people looking at the clouds to see what they are telling us. This is not, by the way, a translated Portuguese idiom.

The techniques of discourse analysis consider the whole text and with the insights we glean from this kind of holistic approach we can delve down again atomistically looking for confirmation or disconfirmation of our appraisal. An important lesson that we can learn from engaging with an essay as discourse is the inter-subject variability in any corpus. There are so many ways to persuade, parody, reasoning, examples, quoting authority, irony, parable-telling, satire or creating affective bonds with the reader to name just a few. The essay I have used as a discourse and processed with Youmans's software would probably be considered fairly weak from the point of view of spelling, punctuation, collocation, the use of prepositions and phraseology. As an argumentative essay and as a piece of persuasive writing it is fairly successful (I would claim). But I would welcome some corroboration from other readers.

References

- Aarts J, Granger S 1998 Tag sequences in learner corpora: a key to interlanguage grammar and discourse. In Granger S (ed) *Learner English in corpora*. London, Addison Wesley, Longman.
- Garner M (2003) "Developing culturally appropriate language teaching". Paper delivered at Viseu Polytechnic, February 14, 2003 reporting ongoing research funded by La Trobe University, Australia.
- Granger S (ed) 1998 *Learner English in corpora*. London, Addison Wesley, Longman.
- Granger S, Hung J, Petch-Tyson S (eds) 2003 *Computer learner corpora, second language acquisition and foreign teaching*. Amsterdam, John Benjamins.
- Halliday M 1985 *Spoken and written language*. Oxford, Oxford University Press.
- Hoey M 1983 *On the surface of discourse*. London, Allen & Unwin
- Hunston S and Thompson G 2002 *Evaluation in Text* Oxford, Oxford University Press.
- McElholm D 2002 Text and Argumentation in English for Science and Technology Unpublished Ph.D. Thesis, University of Hanover.
- Meunier F 1998 Computer tools for the analysis of learner corpora. In Granger S (ed) *Learner English in corpora*. London, Addison Wesley, Longman.
- Petch-Tyson S 1989 Writer/reader visibility in EFL written discourse. In Granger S (ed) *Learner English in corpora*. London, Addison Wesley, Longman.
- Scholfield P 1995 *Quantifying language*. Clevedon, Multilingual Matters.
- Stubbs M 2002 *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford, Blackwell.
- Swales J (1990) *Genre Analysis*. Cambridge, Cambridge University Press.
- Youmans G 1991 "A New Tool for Discourse Analysis: the Vocabulary-Management Profile." *Language* 67.4:763-789
- Youmans G 2001 "The Hierarchical Structure of Discourse: A New, Improved Vocabulary-Management Profile." Manuscript:

APPENDIX

VMP2.2: "PTES1001.txt" Interval: 35		
TotalTypes=216 TotalTokens=522 Types/Tokens=0.4138		
AvgR = the ratio of Types / Tokens over the moving interval.		
Mean avgR = 0.41362		
Standard Deviation = 0.07575		
~Midpoint, AvgR, Last word in interval		
1,	0.3685221,	value , 18
2,	0.3693995,	to , 19
3,	0.3693995,	materialism , 20
4,	0.3425829,	our , 21
5,	0.3623252,	modern , 22
PEAK 1		
11,	0.4020291,	science
12,	0.4047162,	technology
13,	0.4081163,	and
14,	0.4346038,	industrialisation,
15,	0.4063614,	is , 32
16,	0.4308747,	creating , 33
17,	0.4582945,	caps , 34
18,	0.4802852,	between , 35
19,	0.4802852,	people , 36
20,	0.4546202,	we , 37
21,	0.4597751,	don't , 38
22,	0.4401974,	have , 39
23,	0.4394297,	the , 40
24,	0.4647655,	human , 41
25,	0.4647655,	contact , 42
26,	0.4655333,	that , 43
27,	0.4655882,	we , 44
28,	0.4912531,	had , 45
29,	0.5047436,	many , 46
30,	0.5260762,	years , 47
31,	0.5364409,	ago , 48
32,	0.5124212,	nowadays , 49
33,	0.5099534,	we , 50
34,	0.5126405,	depend , 51
35,	0.5165341,	more , 52
36,	0.5119276,	of , 53
37,	0.5122018,	technology , 54
38,	0.4858788,	it's , 55
39,	0.4853852,	is , 56
40,	0.4853852,	true , 57
41,	0.4846723,	that , 58
42,	0.4721689,	this , 59
43,	0.4989855,	has , 60
44,	0.4729367,	given , 61
45,	0.4466685,	to , 62
46,	0.4442007,	us , 63
47,	0.4407458,	a , 64
48,	0.4397039,	lot , 65
49,	0.4117905,	of , 66
50,	0.4400329,	development , 67
51,	0.4190842,	but , 68
52,	0.3925967,	and , 69
53,	0.3676446,	the , 70
54,	0.3676446,	relations , 71
55,	0.3667124,	between , 72
56,	0.3875514,	man , 73
57,	0.3840417,	what
58,	0.4017549,	about

59,	0.3782835,	dreaming
TROUGH 1		
60,	0.3498218,	what , 77
61,	0.3482314,	about , 78
62,	0.3633123,	imagination , 79
81,	0.3523992,	nowadays , 98
82,	0.3730738,	children , 99
83,	0.3737867,	don't , 100
84,	0.3736221,	have , 101
85,	0.3708802,	free , 102
86,	0.3783384,	time , 103
87,	0.3971483,	if , 104
88,	0.4226488,	they , 105
89,	0.4191939,	aren't , 106
90,	0.4411297,	at , 107
91,	0.4411297,	school , 108
92,	0.4358103,	they , 109
93,	0.4170003,	are , 110
94,	0.4404716,	having , 111
95,	0.4644365,	classes , 112
96,	0.4668495,	of , 113
97,	0.4800110,	music , 114
98,	0.5047436,	painting , 115
99,	0.5294763,	ballet , 116
100,	0.5307924,	and , 117
101,	0.5240471,	so , 118
102,	0.5466959,	on , 119
PEAK 2		
103,	0.5744996,	parents , 120
104,	0.5470250,	don't , 121
105,	0.5280504,	have , 122
106,	0.5188922,	time , 123
107,	0.5248697,	because , 124
108,	0.5235536,	they , 125
109,	0.5215794,	are , 126
110,	0.5261859,	very , 127
111,	0.5261859,	occupied , 128
112,	0.5256375,	at , 129
113,	0.5310667,	work , 130
114,	0.5193858,	so , 131
115,	0.4911434,	they , 132
116,	0.5170825,	put , 133
117,	0.5187826,	their , 134
118,	0.5173568,	children , 135
119,	0.5171922,	at , 136
120,	0.5168632,	extra-curricular, 137
121,	0.5031533,	classes , 138
122,	0.4854401,	i , 139
123,	0.4615300,	believe , 140
124,	0.4392103,	that , 141
125,	0.4200164,	it's , 142
126,	0.3922676,	very , 143
127,	0.3951741,	important , 144
128,	0.4222100,	for , 145
129,	0.3941870,	children , 146
130,	0.3717028,	and , 147
131,	0.3692898,	for , 148
132,	0.3453798,	us , 149
133,	0.3198245,	to , 150
134,	0.2927886,	have , 151
135,	0.2928983,	free , 152
136,	0.2726625,	time , 153

137, 0.2726625, just , 154
TROUGH 2
138, 0.2443104, to , 155
139, 0.2695915, do , 156
140, 0.2970661, any , 157
141, 0.3193858, thing , 158
142, 0.3139567, it's , 159
143, 0.3298602, when , 160
144, 0.3328763, we , 161
145, 0.3231149, aren't , 162
146, 0.3231149, being , 163
147, 0.3505347, useful , 164
148, 0.3400603, this , 165
149, 0.3453798, is , 166
150, 0.3453798, when , 167
151, 0.3171374, we , 168
152, 0.2924595, aren't , 169
153, 0.3011242, doing , 170
154, 0.3293666, nothing , 171
155, 0.3055114, that , 172
156, 0.3043597, we , 173
157, 0.3196051, dream , 174
158, 0.3165341, that , 175
159, 0.3138470, we , 176
160, 0.3300795, appeal , 177
161, 0.3304634, to , 178
162, 0.3360022, our , 179
163, 0.3097889, i , 180
164, 0.3378119, agination , 181
263, 0.4050452, sofucated , 280
264, 0.4023581, with , 281
265, 0.4211133, all , 282
266, 0.3942418, the , 283
267, 0.4088292, information , 284
268, 0.3807513, that , 285
269, 0.3812449, i , 286
270, 0.3812449, receive , 287
271, 0.4046614, through , 288
272, 0.4293392, t , 289
273, 0.4574170, v , 290
274, 0.4571429, net , 291
275, 0.4812174, radio , 292
276, 0.5096792, etc , 293
277, 0.5086921, our , 294
278, 0.4990403, world , 295
279, 0.4714012, is , 296
280, 0.4859336, starting , 297
281, 0.4865369, to , 298
282, 0.4779819, be , 299
PEAK 3
283, 0.4933918, extremely , 300
284, 0.5215794, industrialised , 301
285, 0.5023855, work , 302
286, 0.4851111, work , 303
287, 0.4847820, work , 304
288, 0.5098437, that's , 305
289, 0.4988210, all , 306
290, 0.4942693, many , 307
291, 0.4998629, of , 308
292, 0.4876885, us , 309
293, 0.4841788, think , 310
294, 0.4638333, the , 311
295, 0.4920757, only , 312
296, 0.4922402, time , 313
297, 0.4652043, that , 314

298, 0.4381683, i , 315
299, 0.4451330, appeal , 316
300, 0.4268714, to , 317
301, 0.4274198, my , 318
302, 0.4010968, imagination , 319
303, 0.4018646, is , 320
304, 0.4095969, when , 321
305, 0.3935838, i'm , 322
306, 0.3935838, reading , 323
307, 0.3712640, a , 324
308, 0.3712640, book , 325
309, 0.3962161, or , 326
310, 0.3679188, when , 327
311, 0.3400055, i , 328
312, 0.3389635, look , 329
313, 0.3379216, to , 330
314, 0.3373183, a , 331
315, 0.3373183, picture , 332
316, 0.3341925, when , 333
317, 0.3144502, i , 334
318, 0.3063340, stop , 335
319, 0.2779819, a , 336
TROUGH 3
320, 0.2971758, bi t , 337
321, 0.3032629, and , 338
322, 0.3034823, i , 339
323, 0.3034823, turn , 340
324, 0.3039759, of , 341
325, 0.3306279, from , 342
326, 0.3216891, the , 343
327, 0.3454346, reality , 344
328, 0.3435701, the , 345
329, 0.3706608, rest , 346
330, 0.3423636, of , 347
331, 0.3402797, the , 348
332, 0.3406636, time , 349
333, 0.3406087, i'm , 350
334, 0.3341376, too , 351
335, 0.3617220, concern , 352
336, 0.3673704, in , 353
337, 0.3670414, my , 354
338, 0.3685221, work , 355
339, 0.3603510, to , 356
340, 0.3489443, stop , 357
341, 0.3214149, and , 358
342, 0.3207020, dream , 359
343, 0.2932273, i , 360
TROUGH 4
344, 0.2703592, don't , 361
345, 0.2952564, like , 362
346, 0.2953660, the , 363
347, 0.3187277, way , 364
348, 0.3218536, world , 365
349, 0.3239923, is , 366
350, 0.2957499, i , 367
351, 0.3240471, would , 368
352, 0.3304634, have , 369
353, 0.3385797, loved , 370
354, 0.3391280, to , 371
355, 0.3195503, live , 372
356, 0.3145051, in , 373
357, 0.3428572, those , 374
358, 0.3428572, times , 375
359, 0.3434055, when , 376
360, 0.3434055, family , 377

361,	0.3702769,	got	, 378
362,	0.3702769,	together	, 379
363,	0.3713737,	and	, 380
364,	0.3713737,	ask	, 381
365,	0.3716479,	to	, 382
366,	0.3725802,	the	, 383
367,	0.3992322,	story	, 384
368,	0.4263230,	teller	, 385
369,	0.4253907,	to	, 386
370,	0.4253907,	tell	, 387
371,	0.4202906,	a	, 388
372,	0.4469427,	tory	, 389
373,	0.4524266,	this	, 390
374,	0.4535235,	is	, 391
375,	0.4626817,	an	, 392
376,	0.4902111,	image	, 393
377,	0.4890047,	that	, 394
378,	0.4893885,	i	, 395
379,	0.4851111,	have	, 396
380,	0.4885111,	constantly	, 397
381,	0.4890595,	in	, 398
382,	0.4629010,	my	, 399
383,	0.4876885,	head	, 400
384,	0.4932273,	nowadays	, 401
385,	0.5003016,	we	, 402
386,	0.4739786,	don't	, 403
387,	0.4841788,	give	, 404
388,	0.4767754,	value	, 405
389,	0.4770496,	to	, 406
390,	0.4735399,	all	, 407
391,	0.4738141,	the	, 408
392,	0.4738141,	old	, 409
393,	0.4738141,	god	, 410
394,	0.5000823,	values	, 411
395,	0.4720044,	we	, 412
396,	0.4447491,	a	, 413
397,	0.4399781,	e	, 414
398,	0.4536880,	very	, 415
399,	0.4536880,	materialistics	, 416
400,	0.4533589,	we	, 417
401,	0.4530847,	don't	, 418
402,	0.4252810,	give	, 419
403,	0.3974774,	value	, 420
404,	0.3980806,	to	, 421
405,	0.3980806,	simple	, 422
406,	0.4136002,	things	, 423
407,	0.3883740,	like	, 424
408,	0.3853578,	look	, 425
409,	0.4126131,	into	, 426
410,	0.4032904,	the	, 427
411,	0.4032904,	sky	, 428
412,	0.4015903,	and	, 429
413,	0.4286811,	try	, 430
414,	0.4277488,	to	, 431
415,	0.4277488,	guess	, 432
416,	0.4376748,	what	, 433
417,	0.4355909,	the	, 434
418,	0.4355909,	clouds	, 435
419,	0.4561557,	ar	, 436
420,	0.4773238,	telling	, 437
421,	0.4820949,	us	, 438
422,	0.4862627,	technology	, 439
423,	0.4677269,	is	, 440
424,	0.4952564,	everywhere	, 441
425,	0.5124212,	science	, 442

426,	0.5112147,	is	, 443
427,	0.5112147,	discovering	, 444
428,	0.4953661,	everything	, 445
429,	0.4695914,	that	, 446
430,	0.4976693,	can	, 447
431,	0.5044695,	be	, 448
432,	0.5092404,	discovered	, 449
433,	0.5066630,	but	, 450
434,	0.5066630,	should	, 451
435,	0.5088566,	all	, 452
436,	0.5090759,	the	, 453
437,	0.5368796,	secrets	, 454
438,	0.5364409,	be	, 455
439,	0.5642446,	disc	, 456
440,	0.5642446,	vered	, 457
441,	0.5744996,	shouldn't	, 458
442,	0.5734028,	we	, 459
443,	0.5967644,	leave	, 460
444,	0.5967644,	anything	, 461
445,	0.5992323,	in	, 462
446,	0.5992323,	secret	, 463
PEAK 4			
447,	0.6027968,	i'm	, 464
448,	0.6027968,	afraid	, 465
449,	0.6033452,	that	, 466
450,	0.5849191,	with	, 467
451,	0.5778996,	this	, 468
452,	0.5780641,	we	, 469
453,	0.5780641,	destroy	, 470
454,	0.5591445,	our	, 471
455,	0.5363861,	world	, 472
456,	0.5409926,	more	, 473
457,	0.5311763,	than	, 474
458,	0.5571154,	already	, 475
459,	0.5302989,	is	, 476
460,	0.5120921,	i	, 477
461,	0.5304634,	believe	, 478
462,	0.5025501,	that	, 479
463,	0.5183987,	sometimes	, 480
464,	0.5162051,	we	, 481
465,	0.5148889,	must	, 482
466,	0.5136276,	stop	, 483
467,	0.4880176,	and	, 484
468,	0.4852756,	think	, 485
469,	0.4606526,	a	, 486
470,	0.4868111,	little	, 487
471,	0.4871950,	in	, 488
472,	0.4616397,	what	, 489
473,	0.4617494,	we	, 490
474,	0.4617494,	want	, 491
475,	0.4332876,	what	, 492
476,	0.4048259,	we	, 493
477,	0.4311489,	whish	, 494
478,	0.4026871,	what	, 495

