

Shorter notices

The Student Engineering English Corpus

Olga Moudraia

Lancaster University, UK

The Student Engineering English Corpus (SEEC)¹ of nearly 2,000,000 tokens was built to represent the engineering lexis encountered in English-language textbooks in basic engineering disciplines ('BED'). By 'BED', I mean those which are compulsory for all engineering students regardless of their fields of specialization. Based on the SEEC, a word list of the most frequent engineering lexis was developed. It consists of over 1,200 word families or 9,000 word-types.

The goal of the project was to provide a foundation for a reliable lexical syllabus for engineering students in order to meet the objectives of English teaching for Engineering at Walailak University in Thailand, where I worked for nearly seven years. We were in a situation quite common in Southeast Asia: lectures in most subjects were delivered in a local language (Thai, in this case) whilst textbooks were in English.

That is why, in order to build a representative corpus of Student Engineering English, I selected English-language textbooks in 'BED' (such as Engineering Mechanics, Mechanics of Materials, Engineering Materials, Mechanics of Fluids, Thermodynamics, Electrical Engineering, Manufacturing Process, Engineering Drawing and Computer Programming) that were compulsory for all engineering students at Walailak University, regardless of their fields of specialization. The main criterion for selection was that these textbooks were recommended for the engineering students, who had to read them in English.

Whole texts² were used in the SEEC, as opposed to text extracts, which was the case with most other smaller technical corpora designed for language learners (e.g. GPEC³, JDEST⁴ or HKUST⁵). In corpus construction, whole texts are preferable to text extracts, wherever possible, as this frees the researcher from concern about the validity of sampling techniques; moreover, a corpus made up

of whole documents is open to a wider range of linguistic studies than a collection of short samples (Sinclair 1991: 19).

The main stages in the project included gathering a text corpus, putting it into machine-readable form, conducting the computer analysis of the material, and building the word list. The material was analysed with the help of the *Word-Smith Tools 2.0* software – an integrated suite of programs for examining the behaviour of words in texts. The corpus comprises about two million tokens and over 18,000 types (Table 1). As suggested by the type / token ratio, the lexical density in the SEEC is quite low at 0.0092, which means that words recur rather often in the engineering textbooks, the token / type ratio being 109.14.

Table 1: Statistics on the SEEC

Tokens	1,986,595
Types	18,203
Token / Type Ratio	109.14
Type / Token Ratio	0.0092

The entries in the resulting word list were organized by word families, with a reduction of entries to about 7,700. The “word family” here is interpreted in the broadest sense – in accordance with the Bauer and Nation’s (1993) level 7 of generalization, which includes derived and inflected forms as well as compound words. The resultant entries were treated according to the cumulative frequency of occurrence of the members of the word families, and the most frequent word families (with the sum total of 100, or 0.005%) were selected. Overall, more than 1,200 of the most frequent word families comprising nearly 9,000 words were included in the Student Engineering Word List (Table 2).

Table 2: Statistics on the Student Engineering Word List

Total Entries	1,260
Types	8,850
Minimum Frequency	0.005 %

The SEEC has not yet been annotated. Its part-of-speech and semantic tagging is being carried out at the University Centre for Computer Corpus Research on

Language (UCREL) at Lancaster, using CLAWS (the Constituent Likelihood Automatic Word-Tagging System) for POS tagging and the SEMTAG program, which is part of the UCREL Semantic Analysis System (USAS), for semantic tagging⁶. The preliminary semantic profile of the SEEC (Table 3) was drawn using Wmatrix – the web front end of USAS.

Wmatrix allows us to see significant concepts in a corpus and the words related to those concepts in frequency order. It can also compare the concept frequencies against a 1.7-million-words semantically tagged subcorpus of the BNC related to Information Technology (IT). The results of the key concept comparison against the BNC IT are sorted on the log-likelihood (LL), field which shows how significant the difference is. To be statistically significant, the LL value should be over 6.63, which is the cut-off for 99 per cent confidence of significance ($p < 0.01$), or 3.84 for 95 per cent confidence of significance ($p < 0.05$).

Table 3 illustrates the most statistically significant semantic categories in the SEEC, produced in comparison between the SEEC semantic frequency list and the BNC IT. However, the current SEMTAG Lexicon will need to be enhanced in order for us to be able to assign proper semantic tags to the as yet unmatched lexis, amounting to 127,870 lexical items, or 6.4 per cent of the lexis.

Table 3: The frequencies of the semantic categories in the SEEC and the BNC IT compared

Item	Frequency	LL	Semantic category
N1	74,870	+12844.38	Numbers
O4.4	21,627	+10704.81	Shape
N3.5	7,165	+ 7287.83	Measurement: Weight
Z5	679,652	+ 7254.53	Grammatical bin
N2	11,061	+ 7152.46	Mathematics
O2	53,221	+ 6810.75	Objects generally
O1.2	6,883	+ 6703.00	Substances and materials: Liquid
O1	8,707	+ 6294.31	Substances and materials generally
N3.7	7,675	+ 4440.78	Measurement: Length & height
O4.6	3,259	+ 3805.63	Temperature
O1.3	3,667	+ 3679.58	Substances and materials: Gas

E3-	5,407	+ 3102.38	Calm/Violent/Angry
E6-	4,666	+ 2762.14	Worry, concern, confident
X6+	6,847	+ 2715.36	Deciding
B1	8,780	+ 2664.81	Anatomy and physiology
M6	23,444	+ 2541.75	Location and direction
O4.6+	2,975	+ 2318.45	Temperature
M5	3,671	+ 1880.51	Aircraft and flying
N3.2	3,734	+ 1758.18	Measurement: Size
O3	7,581	+ 1739.22	Electricity and electrical equipment
Z99	127,870	+ 1737.24	Unmatched

In conclusion, this project had three primary aims: a) to establish a representative corpus of Student Engineering lexis; b) to provide teachers and learners with a word list that could serve as the lexical syllabus foundation of English for Engineering; and c) to explore the syntactical, morphological, lexical, and discursive features of Engineering English. The first two aims have been largely accomplished; the third aim, however, is a long-term aim, requiring thorough linguistic analysis of the data. Tagged data can be beneficial for studies of semantic fields and grammatical categories. Subsequently, the material is expected to produce valuable information relevant to wide-ranging linguistic analysis.

Notes

1. More detailed information on the SEEC can be found in Moudraia (2003).
2. This required obtaining permission from the publishers for the electronic use of their texts. My acknowledgements go to *McGraw-Hill Australia* (permission dated October 12, 1998), *McGraw-Hill Companies, Inc.* (permission dated December 1, 1998), *Brooks / Cole Publishing Company* (grant No. G-09857, November 17, 1998) and *Addison Wesley Longman Limited* (ref. AP/2743, November 25, 1998) for their understanding of the nature of the project and kind permission to store their texts in an electronic format in order to create a word list.

3. GPEC is the Guangzhou Petroleum English Corpus of about 400,000 tokens (Qi-bo 1989).
4. JDEST is the Jiaotong Daxue English of Science and Technology (JDEST) Corpus of 1,000,000 tokens (Yang 1985).
5. HKUST is the Hong Kong University of Science and Technology (HKUST) Computer Science Corpus of 1,000,000 tokens (James et al. 1994).
6. For details, see Wilson and Rayson (1993) and Rayson and Wilson (1996).

References

- Bauer L. and P. Nation. 1993. Word families. *International Journal of Lexicography* 6 (3): 1–27.
- James G., R. Davidson, A. C. Heung-yeung and S. Deerwester. 1994. *English in computer science: A corpus-based lexical analysis*. The Hong Kong University of Science and Technology, Longman Asia Ltd.
- Moudraia O. 2003. The Student Engineering Corpus: Analysing word frequency. In D. Archer, P. Rayson, A. Wilson and T. McEnery (eds.), *Proceedings of the Corpus Linguistics 2003 conference*, 552–561. UCREL technical paper number 16. UCREL, Lancaster University. ISBN 1 86220 131 5.
- Moudraia O. 1999. Lexical syllabus foundation for engineering. *RELC Journal* 30 (2): 140–141.
- Moudraia O. 1999. Designing a lexical syllabus for engineering. *The 19th Annual Thailand TESOL International Conference Proceedings*, 54–55.
- Qi-bo Z. 1989. A quantitative look at the Guangzhou Petroleum English Corpus. *ICAME Journal* 13: 28–38.
- Rayson P. and A. Wilson. 1996. The ACAMRIT semantic tagging system: Progress report. In J. L. Evett and T. G. Rose (eds.), *Language Engineering for Document Analysis and Recognition*, LEDAR, AISB96 Workshop proceedings. Brighton, England, 13–20. Faculty of Engineering and Computing, Nottingham Trent University, UK. ISBN 0 905 488628.
- Sinclair J. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Wilson A. and P. Rayson. 1993. Automatic content analysis of spoken discourse: A report on work in progress. In C. Souter and E. Atwell (eds.), *Corpus based computational linguistics*, 215–226. Amsterdam: Rodopi.
- Yang H. 1985. The JDEST Computer Corpus of texts in English for science and technology. *ICAME News* 9: 24–25.

**XXIV International ICAME Conference
23–27 April, 2003, Guernsey**

**Anne Curzan and Taryn Hakala
University of Michigan**

Victor Hugo, who was apparently immediately captivated by Guernsey and lived there for over fifteen years, dedicated his book *Les Travailleurs de la Mer* to the island: “I dedicate this book to the rock of hospitality, to this corner of old Norman land where resides the noble little people of the sea, to the Island of Guernsey, severe and yet gentle...”. As the more than 85 participants in the 24th International ICAME Conference learned, it is easy to be enamored of this beautiful island and the hospitality of its inhabitants. The sheer cliffs overlooking the sea and the bustling port with its winding cobbled streets were a spectacular backdrop for a superb conference, with a rich program of innovative work which generated productive conversations at the sessions and afterwards, complemented by daily excursions which took full advantage of the island’s scenery and fascinating history. Conference organizer extraordinaire Antoinette Renouf and her marvelous team (Jay Banerjee, Ceri Davies, Andrew Kehoe, Barry Morley) from the Research and Development Unit for English Studies at the University of Liverpool hosted an impeccably organized conference at the lovely Les Cotils conference center, a beautiful Victorian hotel, with views of Guernsey’s neighboring islands and twelve acres of green belt land, complete with grazing Guernsey cows.

With a total of 33 papers (including those by invited speakers Rita Simpson and Michael Hoey), 15 presentations of work in progress, 12 posters, five software demonstrations, and one 5-person panel discussion, the conference showcased new linguistic findings based on corpus-based research, recent developments in corpora and software, important applications to language learning and teaching, and ongoing theoretical and methodological questions and debates. The conference made the island news, with an attention-grabbing (and misleadingly alarming!) headline in the *Guernsey Press* on Saturday, April 26: “Killer bug fear hits conference.” The article noted that three conference participants from New Zealand and Toronto were prohibited from attending due to the outbreak of the SARS virus – and we would add that their presence was missed. The article then went on to highlight two focal points of the 2003 conference: the use of the world-wide web as a source of linguistic data; and the study of language change, both more historical and more recent.

The article quotes Antoinette Renouf on the notable focus on diachronic linguistics at this ICAME meeting: “There is particular emphasis on the study of change in the language across time, both historically or with more recent trends, such as the suffix ‘gate’, as in Camillagate or phrases such as weapons of mass destruction.” Bas Aarts and Sean Wallis argued persuasively against any clear distinction between synchronic and diachronic language study, as language change is happening at any given moment in the language. And many papers, examining ongoing changes in English, supported their point, including studies of: the distribution of types of adjective comparison, addressing the still debated question of whether periphrastic forms are becoming more common (Tuija Nuutinen, Shunji Yamazaki); the developing frequency of phrasal verbs in newspaper language (David Minugh); the gradual decrease of modals in favor of semi-modals (Geoffrey Leech), as well as the development of modals and other forms of grammaticalization (Hans Lindquist); the ratio of singular and plural agreement with collective nouns (Magnus Levin); and semantic developments of ‘love’ and ‘hate’ (Claudia Claridge).

A host of other papers addressed older historical developments in the English language, demonstrating the ways in which historical corpus linguistic studies continue to enhance our understanding of the history of the language. Ongoing historical studies of modals in Early Modern English provided new details about the development of *shall* and *will* (Maurizio Gotti) as well as *can* (Roberta Facchinetti). Two papers addressed the traditional distinction between transitive and intransitive verbs in English, examining the historical development of mediopassive constructions in the 20th century (Marianne Hundt) and of ambitransitive verbs (Jürgen Gerner). Two genre-specific studies provided new details about the development of noun phrase structure in early scientific writing (Päivi Pahta and Irma Taavitsainen) and of advertising, drawing on the ZEN corpus (Caren Sanders). Anneli Meurman-Solin and Päivi Pahta focused on the grammaticalization and lexicalization of circumstantial adverbial clauses in the development of argumentative discourse. The application of current sociolinguistic theory in historical corpus linguistics has been an exciting development in the field, and Helena Raumolin-Brunberg demonstrated how social network theory and Labov’s work on leaders in linguistic change can be used to interpret data from the CEEC on several morphological and grammatical developments. Other papers addressed the lexical diffusion of a sound change in early Middle English based on spelling evidence (Nils-Lennart Johannesson), patterns of change in the use of *was* in plural constructions in late Middle English and Early Modern English (Terttu Nevalainen), and the development of plural quantifiers,

using the newly developed Corpus of Nineteenth-Century English (Erik Smittenberg).

This ICAME conference continued the tradition of showcasing new corpora and providing updates on corpora still under development. In the opening talk, invited speaker Rita Simpson presented a detailed overview of the Michigan Corpus of Academic Spoken English (MICASE), a relatively new, approximately 1.8 million-word corpus which is particularly valuable for the insights it provides about spoken academic English and about spoken American English more generally. Simpson outlined planned developments for the corpus, as well as ongoing work, and then provided a case study of what we can learn from the distribution of questions in MICASE, including the ways in which patterns observable in academic spoken English correspond to those in conversations and ways in which they differ (e.g., less use of tag questions). In a plenary focused on the development of gendered constructions in English, Anne Curzan presented evidence from MICASE on the multifaceted use of *guys* in spoken American English.

Other reports on new corpora highlighted the development of structured corpora, both synchronic and diachronic, of non-dominant varieties of English: ICE-Ireland (John Kirk, Anne Rooney, and Orla Lowry); the Corpus of Early Ontario English (Stefan Dollinger), which also fills gaps in history of 19th-century English, as does the historical corpus of early English in Australia, 1788–1900 (Clemens Fritz); and A Corpus of 19th Century Scottish Correspondence (Marina Dossena). Two new historical corpora promise to be of great interest for historical sociolinguistics and pragmatics: a corpus of early 19th-century pauper letters (Mikko Laitinen), and a corpus of sermons, 1350–1770 (Patricia Sift). The Corpus of EFL dictionaries, grammars and language guides, 1750–1850 (Manfred Markus) and Lexicons of Early Modern English (Ian Lancashire) will be valuable for those studying language change and the history of standardization. The ELFA Corpus promises to offer new insights into English as *lingua franca* in the academic domain (Anna Mauranen). Complementary talks addressed the ongoing development of corpus tools and methodologies, including: refining the “key word” approach to the study of language of specialized subdomains (Magnar Brekke); developments in tagging and parsing (Nelleke Oostdijk); computer recognition of varieties of World English in written texts (Clive Souter) and identification of related texts in a collection (Jay Banerjee); and the Old English Concordance Programme, which is lemma-based to better handle a tagged OE corpus (Javier Calle Martín).

The plenary talk by invited speaker Michael Hoey focused on, and captured, some of the critical implications of corpus linguistics for grammatical theory,

specifically what it means to make lexical priming central. Hoey proposed a single theory that unites grammar and lexicon. He highlighted the importance of collocation, which reflects the reproduction of patterns previously received, and of lexical priming, arguing that grammar arises from lexical priming. Michael Barlow subsequently proposed that we consider chunks as the basic unit of grammar (as opposed to lexical categories), as part of a larger argument that language in use is the best evidence for the structuring of grammar.

These papers set the stage for the panel, Grammar and Corpus Linguistics, on which Bas Aarts, Geoffrey Leech, Christian Mair, Joybrato Mukherjee, and Elena Tognini Bonelli discussed the goals and implications of a corpus-informed, corpus-based, or corpus-driven approach to grammatical description and theory. Panelists noted the importance of corpora for providing examples, for describing frequency, for distinguishing the periphery from the core, for addressing grammatical queries, and for addressing descriptive and explanatory adequacy. They also raised the difficult questions of whether more data necessarily makes a better grammar and how an electronic reference grammar might be structured.

The studies of contemporary varieties of English presented at the conference continue to refine our descriptions of lexical and grammatical features, including: high frequency nouns expressing time (Michaela Mahlberg) and noun complementation (Rhonwen Bowen); similes (Kay Wikberg); *start/begin* + perception verbs (Jean Chuquet and Caroline David); the multifunctionality of the imperative in British English as well as a comparison with Dutch (Bernard De Clerck); the distribution of English ditransitives with two NPs versus a prepositional phrase (Gabriel Ozon); and the deletion of appositive construction *that*, as related to selected syntactic and semantic factors (Naohiro Takizawa).

Returning to the *Guernsey Press*: Antoinette Renouf was also quoted on one of the other focal points of the conference: the world wide web. She explained, “The web is the only up to date source of new words and uses of words but because it’s not finite like a corpus; [sic] you can’t quantify the data. You can say what is popular but not how popular.” The plenary talk by Christian Mair effectively outlined many of the central issues and provided much fodder for discussion throughout the rest of the conference. Mair described three types of corpora – small and tidy (e.g., Brown, LOB, ICE), big and messy (e.g., COBUILD, Bank of English), and big and tidy (e.g., BNC?) – and argued that we need to and can find ways to rely on “large-but-messy” corpora to study regional diversification and language change, given our need for large amounts of data and the difficulty of compiling BNC-type reference corpora. Mair raised the question of the nature of a corpus versus a text collection – a hot-button issue

that resurfaced in many other papers and discussions about the web and web-based resources. Ylva Berglund and Martin Wynne described the adaptation of a text database like Oxford Text Archive for linguistic purposes, and Charles Meyer addressed how Lexis-Nexis can potentially be used responsibly as a historical corpus through the creation of sub-corpora. Josef Schmied proposed ways of handling the massive amounts of data on the web with WebCorps and Phrasecount, and Gunnar Bergh highlighted some of the pitfalls of Google as a linguistic search engine. Sebastian Hoffmann examined whether the *OED* quotations database could be considered a corpus. Joybrato Mukherjee's call for clarification of terminology, including *corpus*, *corpus linguistics*, and *corpus-linguistic methods* occurred in the middle of these fascinating and complicated discussions about what constitutes a corpus and how we can analyze the large amounts of "messy" data now available.

Applied corpus linguistic studies presented at the conference included a focus on language teaching and learning with the development of new learner corpora (Rita Calabrese, Vivian De Klerk) and on the improvement of EFL textbooks by closer attention to usage (Ute Römer), as well as on child language acquisition of pronouns (Sue Blackwell) and of negative verbal structures (Norbert Schläuter). Two papers addressed the use of parallel corpora for translation studies (Natalia Gvishiani) and contrastive analysis of pragmatic markers (Anna-Brita Stenström). Corpus-based analysis of issues central to pragmatics and discourse analysis also focused on: introducing characters in narratives (Anja Janoschka and Sabine Baumann); apologies in British English (Mats Deutschmann); and handling ambiguity with a term such as *recent* (Göran Kjellmer). Extending the work historically, Martti Mäkinen examined strategies of instruction in Middle English herbals.

Valuable software demonstrations featured: monolingual concordancing (MP2.2, the latest version of MonoConc Pro) and bilingual concordancing (ParaConc) by Michael Barlow; GlossNet by Cédric Fairon; the COLT CD-ROM, now available with sound files, by Knut Hofland; WebCorp, a tool for retrieving linguistic data from the web, by Barry Morley; and the ways in which IMS Corpus Workbench, existing software for large text corpora, can be extended, by Stefan Evert.

The richness of this academic program was matched by the beautiful scenery on the excursions throughout the conference and the wonderful (and rich!) meals served to us with Guernseyan style and hospitality. Among our excursions was a visit to the thirteenth-century Castle Cornet. There we strolled along the ramparts and around the courtyard awaiting the commencement of the Vin d'Honneur reception. Afterwards, we traveled by bus up the winding roads to

Jerbourg Point, where we dined at the Idlerocks Hotel. Practically perched on the side of a cliff, the view from Idlerocks includes the Islands of Herm, Sark and Jersey, and the coast of France.

Another of our excursions included a boat trip to nearby Herm Island. We were met at the dock by two charming and tuxedoed hosts and led up the hill to the elegant White House Hotel for our gourmet dinner. Our meal was followed by a buffet of desserts with everything from fresh fruit to decadent tortes and cheesecakes. The walk back down the winding and partially wooded trail was accomplished without the aid of torches, which made for a somewhat spooky trek, adding to the strange and wonderful atmosphere of Guernsey and its neighboring islands.

Guernsey's eclectic culture is a testament to its geographical location and history. Situated eighty miles from England and only thirty miles from France, Guernsey has been the territory of both nations. It is currently a Bailiwick of the United Kingdom, and businesses will accept the British pound, but Guernsey has its own currency and its own independent spirit. Juxtaposed with rustic French street names are British post boxes, and although English is now spoken widely across the island, until three generations ago, it was a minority language on Guernsey.

Guernsey French (Guernesiais or Patois) was the lingua franca of the island until the mass evacuation of Guernsey children to the British mainland during the Second World War and that generation's subsequent severing from its native tongue. Many from that generation have returned to the island, but due to the large number of people immigrating to and emigrating from Guernsey, there are far fewer families with two Patois-speaking parents to pass on the language to their children. Guernsey French is still spoken on the island, however, and we were treated with a Guernesiais poem, recited by Hazel Tomlinson (BBC Radio Guernsey newscaster for Guernsey French) during the Vin d'Honneur reception at Castle Cornet.

During her talk, Tomlinson explained that Guernesiais is a variation of Norman French and retains many of its phonological features such as /k/ in words where Modern French would have /š/ (e.g., *la vague/la vache*). Lexical items that have disappeared from Modern French also survive in Guernesiais – for example *choir* 'to fall' and *devaler* 'to go down' have not been replaced by *tomber* and *descendre*. The extensive use of diphthongs also makes Guernesiais unique compared to its Norman French relatives: for example, instead of the closed /e/ sounds of Modern French verb endings such as *-er*, *-é*, *-ez*, Guernesiais employs /ai/.

It seems fitting that an island community as unique as Guernsey's would fight to retain its unique language. Harry Tomlinson, former librarian of the Priaulx Heritage Library and Secretary of the Coumité de la Culture Guernesaise, who also spoke at the Vin d'Honneur, has been offering a Guernsey French for Beginners course for several years. The Priaulx Heritage Library holds a wide selection of Guernsey French books, including the first dictionary of Guernsey French that was produced in the nineteenth century, but only one book on Guernsey French written for children. Not surprisingly, this book, *Jimmain Va à la Banque*, was produced by Les Ravigotteurs, a group that formed in 1995 with the purpose of promoting the use and tuition of Guernsey French.

The final gala dinner brought the conference to an impressive close. The eight-inch-in-diameter crème brûlée that was served at the dinner, hosted at the Duke of Richmond Hotel, was certainly the pièce de résistance and serves as exemplary for the sumptuousness of our meals throughout our stay on the island. One particular highlight of the gala dinner was a rousing rendition of "The Corpus-User's Chorus," written by Sue Blackwell and performed by Geoffrey Leech (on the piano) and Willem Meijs (vocals), with spirited (if not always in-tune) audience participation by all gathered at the dinner. With permission of Sue Blackwell, we reprint the lyrics here for posterity:

I am the very model of a user of technology
For testing out hypotheses on grammar and morphology.
I used to do it manually, with diagrams arboreal,
But life is so much better since my research went corporeal.

A language looks quite different when processed electronically:
My lab has all the software to describe it diachronically.
I have a suite of programs which equips me with facilities
For tagging and for parsing and computing probabilities.

I surf the web each week to glean the latest innovations
In the interlanguage usages of English-speaking nations;
Yes, my Google search expressions are the height of specificity:
There's nothing that can stop me in my quest for authenticity.

I offer you my expertise in storage and retrieval,
Be your data Dutch or Danish, present-day or mediaeval.
My concordancing's unparalleled, my interface spectacular
For coding collocations in contemporary vernacular.

I'll integrate your audio and annotate your narratives;
I'll aid your exploration of Germanic ambitransitives.
And if you need comparisons – which after all is rational –
I'll help you break the ICE and turn your English international.

I follow corpus matters with a passion quite fanatical:
I'm sought as an authority on everything grammatical.
At ev'ry ICAME conference I feature as a panelist:
I am the very model of a modern corpus analyst.

And on that note, we will end with a phrase of Guernsey French: “A la perchôine!” (‘Until next time!’).

