



UNIVERSITY of CAMBRIDGE
ESOL Examinations

Cambridge English

Analysing corpora to teach financial English

Evan Frendo



PART OF THE
CAMBRIDGE ASSESSMENT
GROUP





What is a corpus?

A collection of texts, spoken or written, held in electronic format.

Cambridge English Corpus



Business English

The Cambridge Business English Corpus is 200 million word collection of written and spoken texts from, or relating to, the world of business.

The Corpus is drawn from a range of global sources and includes business meetings, minutes, reports, proposals, presentations, and articles.

What's in the Cambridge Business English Corpus?

The Cambridge Business English Corpus is made up of two parts:

- **Spoken Business English** – The Cambridge and Nottingham Spoken Business English Corpus (CANBEC) is a collection of spoken business English recorded in companies of all sizes, ranging from big multinational companies to small partnerships. Recordings include formal and informal meetings, presentations, conversations on the phone and in more informal settings.
- **Written Business English** - reports, newspaper articles, proposals, tenders, journals and books.



Hong Kong Financial Services Corpus

Welcome to the [HKFSC](#) developed by the Research Centre for Professional Communication in English of the Hong Kong Polytechnic University. The HKFSC is a large collection of texts collected from the financial services sector of Hong Kong.

There are currently **7,341,937** words in the HKFSC.

- You can search for a word, e.g. **bank**, **not**, or a phrase, e.g. **annual report**, **a lot of**, and find examples of its use in its context.
- You can also search for an additional word in combination with your search word, e.g. **equity** (search word) and **profit** (additional word), or search phrase, e.g. **financial institutions** (search phrase) and **banks** (additional word).

Enter search word or phrase

Additional word or phrase (optional)

SEARCH

Cambridge English Corpus



Cambridge Learner Corpus

The Cambridge Learner Corpus is the **world's largest Learner Corpus**.

It is made up from thousands of exam scripts written by students taking Cambridge ESOL English exams in countries around the world. It forms part of the Cambridge English Corpus. It has been built by Cambridge University Press and [University of Cambridge ESOL Examinations](#).

The Cambridge Learner Corpus currently contains over **200,000 exam scripts** from students speaking 148 different languages living in **217 different countries or territories**. The Cambridge Learner Corpus is growing all the time.

The EnronSent Corpus

Will Styler

University of Colorado - Boulder

`william.styler@colorado.edu`

May 26th, 2011

~ 500,000 messages
from 150 users
(mostly senior
management of
Enron)

Contents

1	Introduction	2
1.1	The Enron Email Corpus	2
1.2	Difficulties with the Enron Corpus for linguistic research	2
1.3	Goals of the EnronSent corpus	3
2	Creating the EnronSent Corpus	4
2.1	Retrieving and simplifying the raw emails	4
2.2	Cleaning the corpus	4
2.3	Formalizing the corpus	5
2.4	EnronSubjects	6
3	The EnronSent Corpus	6
3.1	Conclusion and Disclaimers	6
3.2	Final Corpus Statistics	6
3.3	Availability	7
4	Acknowledgements	7



BULATS
Business
Language
Testing Service



BEC
Business
English
Certificates



ICFE
International
Certificate in
Financial English



ILEC
International
Legal English
Certificate



ICFE
International
Certificate in
Financial English



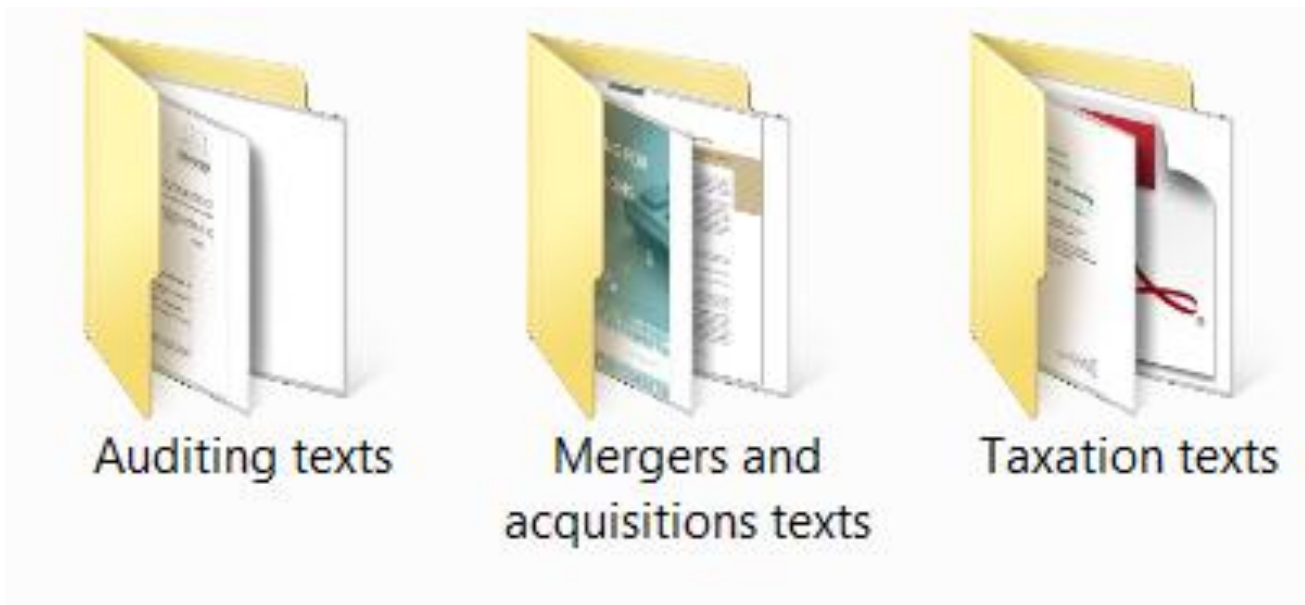
ICFE topics

Financial Reporting
Company Financial
Strategy
Risk Assessment and
Analysis
Auditing
Ethics and Professionalism
Accounting Software
Packages
Assets and Company
Valuations
Budgetary Processes
Corporate Governance

Cost and Management
Accounting
Environmental and
Sustainability Issues
Mergers and Acquisitions
Taxation
Banking
Debt Recovery and Credit
Policy
Economic Conditions and
Forecasts
Insurance
Pricing and Purchasing



Own corpora



Each corpus consists of 45-65 texts totalling approx 1.3 million words



ANALYSING CORPORA



Concordance lines

Use to find every occurrence of a word / phrase

Key Word In Context - KWIC

1 Thanks. Tomorrow ? PLEASE STOP REPLYING TO ALL USERS i sold 1,
2 1,000/d of apr/oct @ 3.02 for GD-Y-Gas Daily book. please put in that book. Thanks. Two things: 1). The
3 below trade did not go into the GD-Y-Gas Daily book, please move it into the corect book. 2). I bot a Nov16
4 Mar3 @ 3.495 to hedge the above book. can you please put it in that book. Thanks. thnk u nothing for
5 being offered by the IntercontinentalExchange. Please contact one of our following representatives to
6 : chris.dambrosio@intcx.com I can take this mission. Please advise. Andy Zipper sonjay.
7 Jeff. We'll let you know. Regards, Andy Sonjay, Please give me a call at your convienence. Thanks,
8 in other trading house are looking for a junior trader, please let me know. Deep in my heart I want to be
9 Okay, but you're still sleeping on the couch tonight... Please see revised market assignments. Additions
10 following note to the EnronOnline origination group: "Please attend an origination meeting on Tuesday,
11 It wil be Bob Shults. I will put him in contact with Joe. Please call Joe Robertson of enron Metals in NY to
12 stuff, but every long journey starts with a small step. Please let me your interest in pursuing this sort of
13 Thanks, Andy Zipper Vice President Enron Corp. Please review the spreadsheet, and look at Daily
14 need Jay involved. Kal will be finishing mine shortly. Please be prepared to do the 3 hole punch thing 20
15 -- Opening comments start promptly at 8:00 a.m., so please allow plenty of time to arrive at the Houstonian,
16 casual. If you haven't turned in your presentation yet, please bring 20 copies to the meeting, 3-hole punched
17 stuff, but every long journey starts with a small step. Please let me your interest in pursuing this sort of
18 your commencement date to December 31, 2001. (2) Please complete the first step by December 15th so



Collocates – Please (L1)

N	Word	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	YOU	836	2,045	395	400	3,687	0	10	222	790	967	3,364
2	QUESTIONS	124	219	308	52	1,867	0	0	11	85	169	261
3	ERROR	0	2	0	12	1,702	0	0	1	3	0	0
4	PM	79	154	171	428	1,238	0	0	0	0	0	1
5	RECIPIENT	4	27	0	1	997	0	0	0	0	0	0
6	ATTACHED	31	42	7	40	747	0	3	1,188	984	169	58
7	AM	75	135	125	288	662	0	0	2	5	31	24
8	SO	17	18	60	12	599	0	18	116	63	129	138
9	ALSO	14	9	58	18	481	0	139	11	12	24	14
10	INFORMATION	72	80	48	8	289	0	0	27	217	210	105
11	AND	154	167	306	383	288	0	8	1,208	755	1,732	1,673
12	20	78	79	68	122	224	0	21	32	79	45	40
13	COM	40	69	58	89	214	0	0	0	1	69	258
14	COMMENTS	18	32	138	14	201	0	1	80	207	201	211
15	BUT	20	6	29	5	200	0	1	2	1	19	11
16	KAY	12	20	49	96	192	0	0	7	18	17	15
17	YES	31	27	2	12	187	0	1	5	5	5	3
18	APPROVAL	0	4	18	30	174	0	4	21	26	61	31



Collocates – Please (R1)

N	Word	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	LET	12	12	18	0	0	0	9,040	59	318	112	204
2	CALL	18	15	10	19	67	0	3,474	38	100	1,366	255
3	CONTACT	12	12	11	12	17	0	2,868	26	66	497	370
4	REVIEW	5	23	29	5	66	0	2,024	26	27	49	70
5	ADVISE	1	3	3	1	1	0	1,930	15	68	26	17
6	SEND	15	32	14	2	4	0	1,843	46	75	54	70
7	SEE	20	19	12	12	3	0	1,829	20	26	21	78
8	FIND	9	2	9	9	0	0	1,771	2	14	1	13
9	NOTE	14	2	7	9	24	0	1,406	51	101	47	11
10	NOTIFY	1	0	0	0	0	0	1,346	365	7	6	41
11	FORWARD	5	10	21	5	16	0	1,338	25	70	64	47
12	GIVE	9	3	8	0	0	0	1,311	14	44	96	110
13	RESPOND	3	12	4	8	5	0	1,290	1	20	10	24
14	DO	35	38	35	21	34	0	1,284	21	17	50	38
15	FEEL	2	2	2	2	0	0	1,213	7	5	4	15
16	PROVIDE	5	9	4	11	7	0	994	18	98	18	31
17	TAKE	23	8	7	5	5	0	749	13	8	10	9
18	MAKE	13	13	23	20	1	0	689	15	60	41	43



Frequency lists

Use to identify most common words / clusters

Most frequent four-word clusters

	Auditing	Mergers and Acquisitions	Taxation
1	THOSE CHARGED WITH GOVERNANCE	THE FAIR VALUE OF	IN THE CASE OF
2	THE AUDITOR S REPORT	FAIR VALUE OF THE	THE END OF THE
3	OF THE FINANCIAL STATEMENTS	IN A BUSINESS COMBINATION	IS DISPLAYED AS SHOWN
4	RISKS OF MATERIAL MISSTATEMENT	OF THE ACQUISITION DATE	AT THE RATE OF
5	OF THE AUDITOR GENERAL	AS OF THE ACQUISITION	THE CORPORATE INCOME TAX
6	IN THE FINANCIAL STATEMENTS	THE VALUE OF THE	AT A RATE OF
7	IN ACCORDANCE WITH THE	FOR THE BUSINESS COMBINATION	IN THE UNITED STATES
8	OF THE AUDITOR S	ACCOUNTING FOR THE BUSINESS	DETERMINATION OF TAXABLE INCOME
9	APPLICABLE FINANCIAL REPORTING FRAMEWORK	AS A RESULT OF	FOR THE PURPOSE OF
10	ON THE FINANCIAL STATEMENTS	IN THE ACCOUNTING FOR	ARE NOT SUBJECT TO
11	INTERNATIONAL STANDARDS ON AUDITING	CROSS BORDER M AS	SALES AND USE TAX
12	THE APPLICABLE FINANCIAL REPORTING	AT THE ACQUISITION DATE	ARE SUBJECT TO A
13	SUFFICIENT APPROPRIATE AUDIT EVIDENCE	FOR A BUSINESS COMBINATION	E RELATED PARTY TRANSACTIONS
14	OFFICE OF THE AUDITOR	IN THE CASE OF	ON THE BASIS OF
15	OF THE ENTITY S	OF THE BUSINESS COMBINATION	TREATY AND NON TREATY
16	THE RISKS OF MATERIAL	ACQUISITION DATE FAIR VALUE	TREATY WITHHOLDING TAX RATES
17	THE FINANCIAL STATEMENTS AND	OF MERGERS AND ACQUISITIONS	VALUE ADDED TAX VAT
18	FINANCIAL AUDIT MANUAL PAGE	CROSS BORDER M A	AT THE END OF
19	TRUE AND FAIR VIEW	ACCOUNTING FOR A BUSINESS	IN ACCORDANCE WITH THE
20	THE NATURE OF THE	ON THE OTHER HAND	FROM THE SALE OF



Keywords

N	Key word	Freq.	Keyness
1	#	1,116,891	1,419,053.63
2	ENRON	56,066	231,764.22
3	PLEASE	57,424	171,432.20
4	PM	39,521	152,069.53
5	THANKS	45,655	151,907.45
6	COM	29,947	120,838.48
7	AM	46,364	104,368.53
8	YOU	192,087	83,287.21
9	I	211,018	68,662.66
10	EMAIL	14,586	59,782.51
11	GAS	21,684	58,662.09
12	WE	108,983	57,718.57
13	ATTACHED	18,119	55,564.65
14	ME	64,020	53,108.02
15	E	28,532	52,328.09
16	MAIL	16,322	50,417.98
17	MESSAGE	18,221	48,034.14
18	LET	26,595	44,891.36
19	HOUSTON	11,411	43,981.75
20	WILL	87,921	43,638.70

Use to find words which are significantly more frequent in a corpus (when compared to another corpus)

Tells you what the corpus is about



Keywords

Auditing	Mergers & acquisitions	Taxation
#	#	#
AUDIT	S	TAX
OR	VALUE	INCOME
FINANCIAL	BUSINESS	TAXATION
AUDITOR	ACQUISITION	TAXES
S	ASSETS	S
STATEMENTS	MARKET	TAXABLE
MAY	FINANCIAL	NON
ENTITY	COMPANY	RATE
INFORMATION	M	WITHHOLDING
INTERNAL	BANK	CORPORATE
MANAGEMENT	MERGER	FOREIGN
REPORT	MERGERS	CAPITAL
CONTROL	TARGET	RESIDENT

50 most frequent key words, taken from three ICFE topics, auditing, mergers and acquisitions, and taxation. Shaded boxes = two topics. Black boxes = three topics.



SAMPLE ACTIVITIES



Collocations

Use the words in the box to fill the gaps

transactions, engagement, evidence, statement, fraud

- 1 unauthorised, recurring, complex, routine, significant ____
- 2 specific, compilation, group, assurance, audit ____
- 3 further, documentary, additional, appropriate, audit ____
- 4 group, consolidated, summary, income, financial ____
- 5 likely, potential, alleged, possible, suspected ____



Collocations

Use the words in the box to fill the gaps

transactions, engagement, evidence, statement, fraud

- 1 unauthorised, recurring, complex, routine, significant transactions
- 2 specific, compilation, group, assurance, audit engagement
- 3 further, documentary, additional, appropriate, audit evidence
- 4 group, consolidated, summary, income, financial statement
- 5 likely, potential, alleged, possible, suspected fraud



Collocations

What word from the previous slide fits in all three gaps?

- ...when undertaking analytical procedures in a limited assurance _____ ...
- ...agreeing the terms of the compilation _____ ...
- ...during the course of an audit _____ ...



Collocations

What word from the previous slide fits in all three gaps?

...when undertaking analytical procedures in a limited assurance engagement...

...agreeing the terms of the compilation engagement...

...during the course of an audit engagement ...



Word formation

Use the words in the box to fill the gaps

acquire, acquired, acquiree, acquirer, acquisition, acquisitive

- 1 They are paying 8 million to _____ the stock.
- 2 We are talking about the net assets _____ by the buyer.
- 3 An _____ may obtain control of an _____ without transferring consideration.
- 4 The _____ date is the date the buyer obtains control of the target.
- 5 The level of _____ activities in the sector has increased.



Word formation

Use the words in the box to fill the gaps

acquire, acquired, acquiree, acquirer, acquisition, acquisitive

- 1 They are paying 8 million to acquire the stock.
- 2 We are talking about the net assets acquired by the buyer.
- 3 An acquirer may obtain control of an acquiree without transferring consideration.
- 4 The acquisition date is the date the buyer obtains control of the target.
- 5 The level of acquisitive activities in the sector has increased.



Semi-fixed expressions

Complete the phrase with as many options as you can.

At the end of the _____

audit	year
audit period	day
financial period	month
year	century
financial year	war
month	first
survey phase	season
interview	decade
annual audit cycle	week
fieldwork	world
course	line



CLOZE (GAP FILLS)

CLOZE TEST

Paste your text below:

The directors are responsible for preparing a company's annual accounts that comply with the provisions of the Companies Act 1985 (the 'Act') and give a 'true and fair view' of the company's profit or loss for the year and the state of affairs at the balance sheet date. They are also responsible for the other parts of the annual report, principally the directors' report, the report on compliance with the Combined Code on Corporate Governance, the directors' remuneration report and the operating and financial review or equivalent material.

- every words
- articles
- prepositions
- link words
- auxiliaries
- wh- words
- quantifiers
- modals
- let me choose



- Text only Interactive
- Clues No clues

Submit

clue:

Cloze test creator: <http://l.georges.online.fr/tools/cloze.html>

Text from:
 Compliance Audit Handbook (2006) Department of Environment and Conservation NSW (Downloaded from:
<http://www.environment.nsw.gov.au/resources/licensing/cahandbook0613.pdf>)



WORDLIST: The a the

___ directors are responsible for preparing ___ company's annual accounts that comply with ___ provisions of ___ Companies Act 1985 (___ "Act") and give ___ "true and fair view" of ___ company's profit or loss for ___ year and ___ state of affairs at ___ balance sheet date. They are also responsible for ___ other parts of ___ annual report, principally ___ directors' report, ___ report on compliance with ___ Combined Code on Corporate Governance, ___ directors' remuneration report and ___ operating and financial review or equivalent material.



WORDLIST: at for of on with

The directors are responsible ____ preparing a company's annual accounts that comply ____ the provisions ____ the Companies Act 1985 (the "Act") and give a "true and fair view" ____ the company's profit or loss ____ the year and the state ____ affairs ____ the balance sheet date. They are also responsible ____ the other parts ____ the annual report, principally the directors' report, the report ____ compliance ____ the Combined Code ____ Corporate Governance, the directors' remuneration report and the operating and financial review or equivalent material.



WORD CLOUDS

A word cloud featuring various terms related to corporate reporting and governance. The most prominent words are 'report', 'company's', 'responsible', 'annual', and 'directors'. Other visible words include 'state', 'accounts', 'preparing', 'give', 'year', 'true', 'balance', 'also', 'Corporate', 'principally', 'fair', 'Governance', 'Companies', 'profit', 'compliance', 'parts', 'material', 'Act', 'date', 'financial', 'provisions', 'remuneration', 'review', 'sheet', 'comply', 'operating', 'view', 'Combined', 'equivalent', 'Code', 'loss', and 'sheet'.

Created using
Wordle™
(www.wordle.net)

the directors' remuneration report and the
the Companies Act 1985
the directors are responsible for
the report on
the company's profit or loss for
other parts of the annual report
are also responsible for the
compliance with the Combined
and give a true and fair view of
Code on Corporate Governance
the report on
the company's profit or loss for
other parts of the annual report
are also responsible for the
compliance with the Combined
and give a true and fair view of
operating and financial review or equivalent material
affairs at the balance sheet date. They
the Year and the state of
the directors' remuneration report and the
the Companies Act 1985
the directors are responsible for
the report on
the company's profit or loss for
other parts of the annual report
are also responsible for the
compliance with the Combined
and give a true and fair view of
Code on Corporate Governance
the report on
the company's profit or loss for
other parts of the annual report
are also responsible for the
compliance with the Combined
and give a true and fair view of



EXPLORING LANGUAGE



Literal vs metaphorical

Look at these email extracts. Which ones use a literal meaning of “red”, and which ones are metaphorical?

1. So let me know if I still have any money coming to me or if I'm in the red. Take it easy, Kevin.
2. The revisions are in red line (red and/or blue) per your request. Let me know any issues.
3. Jeff will be leading our tasting of six red wines, each from a different winegrowing region.
4. As you can see he is raising a major red flag in regards to our non-binding offer to XXX.
5. Little wonder, therefore, that investors fear more red ink splashed over the balance sheets.
6. Can you believe this? We have more red tape than Washington on this.
7. I've already marked up w/ red pen, the HG + the Summary of terms sheet.



UNIVERSITY of CAMBRIDGE
ESOL Examinations

Cambridge English

Cambridge ESOL

E-Mail: info@cambridgeESOL.de

www.cambridgeESOL.de

www.cambridgeESOL.org



www.twitter.com/cambridgeESOLDE



www.facebook.com/cambridgeESOLDE



www.cambridgeesolde.wordpress.com



PART OF THE
CAMBRIDGE ASSESSMENT
GROUP

