

TEACHING THROUGH CORPORA

Pierfranca Forchini
pierfranca.forchini@unicatt.it

What is meant by CORPORA?

a **CORPUS** is a collection of texts
in an **electronic** data-base
designed for **linguistic analysis**
(for a particular representative function)

The METHODOLOGY which studies
LANGUAGE through CORPORA
is called

CORPUS LINGUISTICS

In principle, **ANY COLLECTION** of more than one **text** can be called a **corpus** (corpus being Latin for "body", hence a corpus is any body of text),

BUT the term CORPUS when used in the context of modern linguistics tends to have more specific connotations than this simple definition.

Four main characteristics of the modern corpus:

- 1) Sampling and Representativeness
- 2) Finite Size
- 3) Machine-readable form
- 4) A standard reference



1) Sampling and Representativeness

A corpus **has to be maximally representative of the variety under examination**: it has to give as accurate a **picture as possible of the tendencies of that variety**, as well as their proportions

> **a broad range of authors and genres**, for example, which, when taken together, may be considered to provide a reasonably accurate picture of the entire language population the corpus represents

2) Finite Size

Usually a corpus has a **finite number of words** which is determined at the beginning of a corpus-building project

3) Machine-readable form

Corpus data are usually **stored on a computer**, hence they:

- can be searched and manipulated at speed
- can be enriched with extra information

4) A standard reference

If corpora are **built according to standard criteria** then they can serve as **standard references** and can become available to other researchers

SOME EXAMPLES of ONLINE CORPORA

THE CORPUS OF CONTEMPORARY AMERICAN ENGLISH

400+ MILLION WORDS, 1990-2009



MARK DAVIES

BRIGHAM YOUNG UNIVERSITY

ENTER

<http://corpus.byu.edu/coca/>

P. Forchini *Teaching through Corpora*

CORPUS OF CONTEMPORARY AMERICAN ENGLISH

Mark Davies / Brigham Young University

EMAIL
PASSWORD
LOG IN (HELP)

DISPLAY ?

☒ LIST ☐ CHART ☐ KWIC ☐ COMPARE

SEARCH STRING ?

WORD(S) ?

CONTEXT ?

POS LIST ?

SECTIONS ?

1

IGNORE

SPOKEN
FICTION
MAGAZINE
NEWSPAPER
ACADEMIC

2

IGNORE

SPOKEN
FICTION
MAGAZINE
NEWSPAPER
ACADEMIC

SORTING AND LIMITS ?

SORTING

MINIMUM ☐ 10 ?

CLICK TO SEE OPTIONS ?

NOW AVAILABLE: FREQUENCY LISTS AND DICTIONARY You can now get the printed [book \(sample\)](#) or the expanded eBook ([sample](#)) -- both of which are based on frequency data from the Corpus of Contemporary American English. On this page you can see sample entries from the [20,000 entry eBook](#), which also contains 20-30 collocates and synonyms for each word.

16758 slog v
noun ,mud, ,way, ,snow, day, foot, ,road, ,swamp, mile, hour, ,rain, soldier,, ,sand, ,jungle, month,, crew, *misc* through, on, across, through, along, away, toward, long, uphill, around, ahead, hard, up, spend, forward
• trudge, plod, tramp, trek || work, labor, toil, struggle
584 | 0.92

[SEE ANOTHER WORD](#) [SEARCH](#) [FREE LIST](#) [EBOOK](#) [MORE ...](#)

FREQUENCY LISTS [More information...](#)

[WHERE SHOULD I START?]

[COMPARE TO OTHER CORPORA / ARCHITECTURES]

The **Corpus of Contemporary American English** (COCA) is the largest freely-available corpus of English, and the only **large and balanced** corpus of American English. COCA was released in 2008 and it is now used by **tens of thousands** of users every month (linguists, teachers, translators, and other researchers). COCA is also related to **other large corpora** that we have created or modified, including the **British National Corpus** (our architecture and interface) and the 100 million word **TIME Corpus** (1920s-2000s).

The corpus contains more than **400 million words** of text and is equally divided among spoken, fiction, popular magazines, newspapers, and academic texts. It includes **20 million words each year from 1990-2009** and the corpus is also **updated** once or twice a year (the most recent texts are from Summer 2009). Because of its design, it is perhaps the only corpus of English that is suitable for looking at **current, ongoing changes** in the language.

The interface allows you to search for **exact words** or **phrases, wildcards, lemmas, part of speech**, or any combinations of these. You can **search for surrounding words** (collocates) within a ten-word window (e.g. all nouns somewhere near *faint*, all adjectives near *woman*, or all verbs near *feelings*), which often gives you good insight into the meaning and use of a word.

The corpus also allows you to easily **limit searches by frequency** and **compare the frequency** of words, phrases, and grammatical constructions, in at least two main ways:

```

    <w>NN1</w> lemma="the"
    <w>lemma="cost">cost
    <w>NN1</w> lemma="issue">is
    <w>legal</w> <w>w type="P"
    <w>lemma="should">should
    it of">out of</w> <w type="
    </w><w type="NN1"> lemma="
    nt to">pursuant to</w> <w>
    <CJC"> lemma="and">and</
    </w> <w type="AJ0"> le
    <w><w type="PUL"></<
    <w>ulations</w>
    <w>AJ0">

```

© 2011-2012 Nuova Secondaria – La Scuola SPA – Tutti i diritti riservati

American National Corpus

[first release](#) [second release](#) [oanc](#) [resources](#) [about](#)

What's New

ANC.org

The American National Corpus now owns the **anc.org** domain name! Our web address is now www.anc.org.

We would like to thank the **Animal News Center** for transferring the domain to us. In gratitude, the American National Corpus project has made a donation to the [Humane Society of the United States](#) in the name of the "other" ANC.

ANC Tool Update

October 22, 2008: Version 1.2.5 of the [ANC Tool](#) is now available. The new version fixes a problem that prevented it from starting on Mac OS X.

July 24, 2008: Version 1.2.3 of the [ANC Tool](#) is now available. The new version includes better support for selecting the Unicode character encoding, a few bug fixes, and (experimental) NLTK output.

The Open ANC

The open portion of the ANC (approximately 15 million words of text, with annotations) is now [available for download](#).

2nd Release Frequency Counts

Frequency counts for the second release are now available and can be downloaded [here](#).

The American National Corpus (ANC) project is creating a massive electronic collection of American English, including texts of all genres and transcripts of spoken data produced from 1990 onward. The ANC will provide the most comprehensive picture of American English ever created, and will serve as a resource for education, linguistic and lexicographic research, and technology development.

When completed, the ANC will contain a core corpus of at least 100 million words, comparable across genres to the [British National Corpus](#) (BNC). The corpus will also include an "opportunistic" component of potentially several hundreds of millions of words, chosen to provide both the broadest and largest selection of texts (and, where available, annotations) possible.

ANC Status

The ANC has so far released 22 million words of American English, which is available from the [Linguistic Data Consortium](#)--please consult the [LDC Catalog entry](#). The ANC has also released an "Open" portion of the full ANC consisting of approximately 15 million words, which is [freely available for download](#). All ANC and OANC data include annotations for word and sentence boundaries, part of speech (4 tagsets), and noun and verb chunks. Parts of the corpus are annotated for additional linguistic features.

Contribute Data and Annotations to the ANC

➡ CONTRIBUTE TEXTS

The ANC is actively soliciting contributions of written texts and spoken transcripts in American English that were produced in or after 1990, to be included in the ANC and OANC.

Those who have for any purpose developed corpora of post-1989 American English are encouraged to contribute their unrestricted data to be included in the ANC. Authors can consult the [frequently asked questions page](#) to learn more about how the data will be used, and why you should consider contributing your work to the ANC.

<http://www.americannationalcorpus.org/#>

WHAT CAN WE DO WITH CORPORA?

DIFFERENT PURPOSES

answer questions
at various linguistic levels

= an empirical basis, evidence for
linguistic description and argumentation

≠ text archive/database

(a text repository, normally not structured)

AUTHENTIC DATA

. Empirical Data vs. INTUITION:

Data allow for the description
of words in context and, especially, reflect
the actual combinatorial possibilities of
language
which **introspection cannot discern**

. Quantitative & Qualitative Analyses:

computerized corpus analyses +
functional (qualitative) interpretations of
quantitative data

. **Electronic Processing:**

- time-saving procedures
- wide-ranging storing
- fast and complex analyses
- replication of experimental results
- reusability and sharing of resources

CORPUS-BASED APPROACH

“[...] the theory pre-exists the corpus examination” (p71)

“[...] the term *corpus-based* is used to refer to a methodology that avails itself of the corpus mainly to expound, test or exemplify theories and descriptions that were formulated before large corpora became available to inform language study” (p65)

CORPUS-DRIVEN APPROACH

“The *corpus-driven* approach builds up the theory step by step *in the presence* of the evidence. The observation of certain patterns leads to a hypothesis, which in turn leads to the generalisation in terms of rules of usage and finally finds unification in a theoretical statement” (p17)

“The theory has no independent existence from the evidence [...]” (pp84-85)

WHAT DO WE NEED?

STEP 1

WE NEED A CORPUS

CORPUS TYPES

- General / Reference
- Specialized
- Comparable
- Parallel
- Monolingual
- Bilingual
- Multilingual



➔ FOR EXAMPLE:

BNC - General / Referential

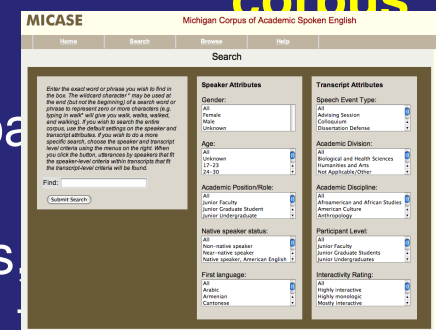
= a corpus of texts of many types likely to include as wide a spread of texts as possible from one or many countries, of spoken or written language



<http://www.natcorp.ox.ac.uk/>

MICASE - Specialized corpus

= a corpus of texts of a particular type such as newspaper editorials, textbooks, academic articles, which is representative of a given type of language and which is used to investigate a particular type of language



<http://micase.elicorpora.info/>

TITLE [USA/Australia, 2001 Directed by Soderberg]	OCEAN'S ELEVEN	OCEAN'S ELEVEN
Corridor, North Jersey State Prison.		
VOICE	One con under escort, open gate one, man walking. Open gate two. Let's go, face the wall. Got one under escort	Un detenuto sotto scorta, aprire cancello uno. Uomo sotto scorta, aprire cancello due. Andiamo, voltatevi verso il muro. Un detenuto sotto scorta!
Parole board hearing room, North Jersey State Prison.		
BOARD MEMBER 1 (WOMAN) to DANNY	Good morning.	Buon giorno.
DANNY	Good morning.	Buon giorno.
BOARD MEMBER 1 (WOMAN) to DANNY	Please state your name for the record.	Il suo nome per il verbale prego.
DANNY	Daniel Ocean.	Daniel Ocean.
BOARD MEMBER 1 (WOMAN) to DANNY	Thank you. Mr. Ocean, the purpose of this hearing, is to determine whether if released you are likely to break the law again. While this was your first conviction, you have been implicated, though, never charged, in over a dozen other confidence schemes and frauds. What can you tell us about this?	Grazie. Signor Ocean, lo scopo di questa udienza, è stabilire se una volta rilasciato, lei sarà incline a violare nuovamente la legge. Anche se questa è stata la sua prima condanna, lei è stato implicato sebbene, mai incriminato in più di dodici truffe. Cosa può dirci a questo proposito?
DANNY	As you say ma'am I was never charged.	Come ha detto lei, non sono mai stato incriminato.

Comparability	The Financial Times	Il Sole 24 Ore
Genre	Quality Daily Newspaper	Quality Daily Newspaper
Content	Finance and General	Finance and General
Year	2003	2003
Size	35.146.523 Words	31.245.954 Words
Language	English	Italian

Bilingual Comparable

Multilingual Parallel

= 2.

```
<text>
<body lang=EN id=Bible>
<div id="b.GEN" type=book>
<div id="b.GEN.1" type=chapter>
<seg id="b.GEN.1.1" type=verse>In the beginning God created the heaven and the earth.
</seg>
<seg id="b.GEN.1.2" type=verse>And the earth was without form, and void; and darkness was upon the face of the deep. And the Spirit of God moved upon the face of
the waters.
</seg>
<seg id="b.GEN.1.3" type=verse>And God said, Let there be light: and there was light.
</seg>
</div>
</div>
</body>

<body lang=LA id=Bible>
<div id="b.GEN" type=book>
<div id="b.GEN.1" type=chapter>
<seg id="b.GEN.1.1" type=verse>in principio creavit Deus caelum et terram
</seg>
<seg id="b.GEN.1.2" type=verse>terra autem erat inanis et vacua et tenebrae super faciem abyssi et spiritus Dei ferebatur super aquas
</seg>
<seg id="b.GEN.1.3" type=verse>dixitque Deus fiat lux et facta est lux
</seg>
</div>
</div>
</body>

<body lang=SV id=Bible>
<b id="GEN ">
<div id="b.GEN.1" type=chapter>
<seg id="b.GEN.1.1" type=verse>I begynnelsen skapade Gud himmel och jord.
</seg>
<seg id="b.GEN.1.2" type=verse>Och jorden var öde och tom, och mörker var över djupet, och Guds Ande svävade över vattnet.
</seg>
<seg id="b.GEN.1.3" type=verse>Och Gud sade: »Varde ljus»; och det vart ljus.
</seg>
</div>
</div>
</body>

<body lang=DN id=Bible>
<div id="b.GEN" type=book>
<div id="b.GEN.1" type=chapter>
<seg id="b.GEN.1.1" type=verse>I Begyndelsen skabte Gud Himmelen og Jorden.
</seg>
<seg id="b.GEN.1.2" type=verse>Og Jorden var øde og tom, og der var Mørke over Verdensdybet. Men Guds Ånd svævede over Vandene.
</seg>
<seg id="b.GEN.1.3" type=verse>Og Gud sagde: "Der blive Lys!" Og der blev Lys.
</seg>
</div>
</div>
</body>
```

SOME ONLINE CORPORA

CORPUS OF CONTEMPORARY AMERICAN ENGLISH

<http://corpus.byu.edu/coca/>

TIME MAGAZINE CORPUS

<http://corpus.byu.edu/time/>

MICASE

<http://micase.elicorpora.info/>

CORPUS OF CONTEMPORARY AMERICAN ENGLISH (COCA)

<http://corpus.byu.edu/coca/>

CORPUS OF CONTEMPORARY AMERICAN ENGLISH

425 MILLION WORDS, 1990-2011

ACCESS: 0/5

history | lists | **SEE PROFILE** | login

DISPLAY

☒ LIST ☐ CHART ☐ KWIC ☐ COMPARE

SEARCH STRING

WORD(S)

COLLOCATES

POS LIST

SECTIONS ☐ SHOW

1 IGNORE

SPOKEN

FICTION

MAGAZINE

NEWSPAPER

ACADEMIC

2 IGNORE

SPOKEN

FICTION

MAGAZINE

NEWSPAPER

ACADEMIC

SORTING AND LIMITS

SORTING

MINIMUM

CLICK TO SEE OPTIONS

Download COCA data for powerful searches offline (e.g. noun+noun, phrasal verbs, and much more)

N-grams (frequency of strings of words) NEW	Word frequency (frequency of individual words)
Free: 1,000,000 most frequent 2, 3, 4, and 5-grams	Free: Top 5,000 lemmas, 500,000 word forms
Purchase: All 2, 3, 4, 5-grams, along with their frequency (hundreds of millions of rows of data).	Purchase: Top 60,000 lemmas (including frequency by genre and sub-genre), collocates (4.8 million node/collocate pairs)
You can also download data from the 400 million word Corpus of Historical American English (COHA). See the frequency of every word, 2, 3, 4, and 5-gram that occurs at least three times in the corpus, along with its frequency in each of the 20 decades (1810s-2000s).	

OVERVIEW: PIERFRANCA FORCHINI

Help / information / contact

Personal information

Institution	UCSC Milan Update
Country	ITALY Update
Short profile (edit)	I have an MA in Foreign Languages and Literatures, an MA in Theoretical and Applied Linguistics, and a PhD in Linguistic and Literary Sciences. My main interests are corpus linguistics, language varieties, applied linguistics, and contrastive linguistics. I currently lectures in English Linguistics at Catholic University of Milan (Italy).
Activity	
Last login (before this time)	12/9/2011 9:52:40 AM (Utah time; same as Denver)
Queries (24 hours / 30 days)	12 / 12
Corpora used (30 days)	TIME (12)
KWIC lists (24 hours saved)	0/500 [0%] 0/15000 [0%]
Customized word lists	0
Status	

<http://corpus.byu.edu/time/>

P. Forchini *Teaching through Corpora*

MICASE

<http://micase.elicorpora.info/>

MICASE Michigan Corpus of Academic Spoken English

HomeSearchBrowseHelp

Search

Enter the exact word or phrase you wish to find in the box. The wildcard character * may be used at the end (but not the beginning) of a search word or phrase to represent zero or more characters (e.g. typing in walk* will give you walk, walks, walked, and walking). If you wish to search the entire corpus, use the default settings on the speaker and transcript attributes. If you wish to do a more specific search, choose the speaker and transcript level criteria using the menus on the right. When you click the button, utterances by speakers that fit the speaker-level criteria within transcripts that fit the transcript-level criteria will be found.

Find:

Speaker Attributes

Gender:

All
Female
Male
Unknown

Age:

All
Unknown
17-23
24-30

Academic Position/Role:

All
Junior Faculty
Junior Graduate Student
Junior Undergraduate

Native speaker status:

All
Non-native speaker
Near-native speaker
Native speaker, American English

First language:

All
Arabic
Armenian
Cantonese

Transcript Attributes

Speech Event Type:

All
Advising Session
Colloquium
Dissertation Defense

Academic Division:

All
Biological and Health Sciences
Humanities and Arts
Not Applicable/Other

Academic Discipline:

All
Afroamerican and African Studies
American Culture
Anthropology

Participant Level:

All
Junior Faculty
Junior Graduate Students
Junior Undergraduates

Interactivity Rating:

All
Highly interactive
Highly monologic
Mostly interactive

STEP 2

WE NEED A SOFTWARE PROGRAM
(A CONCORDANCER)

SOFTWARE

AntConc
Concorder
Wordsmith Tools

...

Multiconcord
Multilingual Tool

...

ANTCONC

<http://www.antlab.sci.waseda.ac.jp/software.html>

Laurence Anthony's Website

Home
Resume
Publications >>
Software >>
Classes
Photo Albums >>
Links
Contact

Software

AntConc

A freeware concordance program for Windows, Macintosh OS X, and Linux.

The AntConc Homepage (including previous versions, tutorials, and help)

Platform	Download
	AntConc3.2.4w AntConc3.2.4 - Readme Screenshots <i>Tested on Windows 98, 2000, ME, XP, Vista, Win 7</i>
	AntConc 3.2.4m AntConc3.2.4 - Readme Essential guide to installing AntConc on Macintosh OS X Download X11 (for OS X 10.4) Screenshots <i>Tested on Macintosh OS X 10.6.8</i>
	AntConc3.2.4u AntConc3.2.4 - Readme AntConc3.2 - Linux Quick Install Guide Written by Laurence Anthony Brief guide to installing AntConc on Linux Written by Graham Ranger Screenshots <i>Tested on Ubuntu 10 (Linux Mint 10)</i>

STEP 3

WE NEED
QUANTITATIVE + QUALITATIVE
ANALYSES

SOME SAMPLE ACTIVITIES

1. LEARNING FREQUENT WORD TYPES USING WORD LISTS

**CHECK
THE MOST FREQUENT WORDS
IN THE GENERAL CORPUS:**

Longman Spoken American Corpus		
N	Word	Occurrence Per 1,000 W
1	I	34.50
2	THE	29.76
3	YOU	27.59
4	AND	26.59
5	TO	23.50
6	IT	18.99
7	THAT	18.20
8	A	18.14
9	OF	12.52
10	YEAH	10.90
11	IN	10.69
12	IS	10.33
13	KNOW	09.68
14	WAS	09.16
15	HAVE	08.77
16	LIKE	08.58
17	SO	08.47
18	WE	07.88
19	OH	07.69
20	THEY	07.59
21	THIS	07.57
22	IT'S	07.54
23	WHAT	07.27
24	JUST	07.20
25	DO	07.00
26	BUT	06.94
27	WELL	06.52
28	UH	06.44
29	ON	06.35
30	HE	06.29

1. LEARNING FREQUENT WORD TYPES USING WORD LISTS

**CHECK
THE MOST FREQUENT WORDS
IN THE GENERAL CORPUS:**

ARE THEY MOSTLY
FUNCTION WORDS
OR CONTENT/LEXICAL WORDS?

WHAT GENERALIZATIONS
CAN YOU MAKE ABOUT
LANGUAGE AS A WHOLE?

Longman Spoken American Corpus		
N	Word	Occurrence Per 1,000 W
1	I	34.50
2	THE	29.76
3	YOU	27.59
4	AND	26.59
5	TO	23.50
6	IT	18.99
7	THAT	18.20
8	A	18.14
9	OF	12.52
10	YEAH	10.90
11	IN	10.69
12	IS	10.33
13	KNOW	09.68
14	WAS	09.16
15	HAVE	08.77
16	LIKE	08.58
17	SO	08.47
18	WE	07.88
19	OH	07.69
20	THEY	07.59
21	THIS	07.57
22	IT'S	07.54
23	WHAT	07.27
24	JUST	07.20
25	DO	07.00
26	BUT	06.94
27	WELL	06.52
28	UH	06.44
29	ON	06.35
30	HE	06.29

2. VOCABULARY LEARNING

USING WORD LISTS

**CHECK
THE MOST FREQUENT
WORDS
IN THE SPECIALIZED
CEOL08 LETTERS CORP**

N	Word	Freq.	%
1	THE	25007	4.382518291
2	AND	21487	3.765632391
3	TO	17650	3.093191862
4	#	17441	3.056564331
5	IN	15873	2.781769514
6	OF	15834	2.774934769
7	OUR	15618	2.737080336
8	WE	12309	2.15717268
9	A	9250	1.621077895
10	FOR	6254	1.096023917
11	THAT	5032	0.881866336
12	AS	4159	0.728871644
13	WITH	3995	0.700130403
14	IS	3990	0.699254155
15	ARE	3763	0.659471989
16	ON	3718	0.651585698
17	YEAR	3060	0.536270082
18	THIS	3050	0.534517586
19	HAVE	3043	0.533290803
20	WILL	2828	0.495611697
21	BY	2638	0.46231389
22	BUSINESS	2471	0.433046848
23	MORE	2274	0.398522288
24	NEW	2065	0.361894667
25	FROM	2033	0.356286615
26	AT	1901	0.333153397
27	GROWTH	1795	0.314576745
28	BE	1780	0.311947972
29	AN	1763	0.308968693
30	COMPANY	1709	0.299505085

2. VOCABULARY LEARNING

USING WORD LISTS

**CHECK
THE MOST FREQUENT
WORDS
IN THE SPECIALIZED
CEOL08 LETTERS CORP**

ARE THEY SPECIALIZED
OR NON-SPECIALIZED
WORDS?

WHAT CAN YOU
CONCLUDE THEN?
IS THE TEXT SPECIALIZED
OR NON-SPECIALIZED?

N	Word	Freq.	%
1	THE	25007	4.382518291
2	AND	21487	3.765632391
3	TO	17650	3.093191862
4	#	17441	3.056564331
5	IN	15873	2.781769514
6	OF	15834	2.774934769
7	OUR	15618	2.737080336
8	WE	12309	2.15717268
9	A	9250	1.621077895
10	FOR	6254	1.096023917
11	THAT	5032	0.881866336
12	AS	4159	0.728871644
13	WITH	3995	0.700130403
14	IS	3990	0.699254155
15	ARE	3763	0.659471989
16	ON	3718	0.651585698
17	YEAR	3060	0.536270082
18	THIS	3050	0.534517586
19	HAVE	3043	0.533290803
20	WILL	2828	0.495611697
21	BY	2638	0.46231389
22	BUSINESS	2471	0.433046848
23	MORE	2274	0.398522288
24	NEW	2065	0.361894667
25	FROM	2033	0.356286615
26	AT	1901	0.333153397
27	GROWTH	1795	0.314576745
28	BE	1780	0.311947972
29	AN	1763	0.308968693
30	COMPANY	1709	0.299505085

3. VOCABULARY LEARNING THROUGH COLLOCATION

Learning information about the context of use for particular words / phrases through the company a word keeps

Black & White or White & Black?
MAKE a mistake or DO a mistake?

Check them in the Time Magazine Corpus
<http://corpus.byu.edu/time/>



3. VOCABULARY LEARNING THROUGH COLLOCATION

TIME MAGAZINE CORPUS

100 MILLION WORDS, 1923-2006

COMPARE: COCA COHA TIME BNC **GOOGLE**

ACCESS: 0/5

[history](#) | [lists](#) | [SEE PROFILE](#) | [logout](#)

DISPLAY

☒ LIST
☐ CHART
☐ KWIC
☐ COMPARE
☐ SEARCH STRING
☐ WORDS
☐ COLLO
☐ POS LIST
☐ RANDO
☐ SECTION
☒ SHOW

SEE CONTEXT: CLICK ON WORD (ALL SECTIONS), NUMBER (ONE SECTION), OR [CONTEXT] (SELECT)

☐ **CONTEXT**

TOT ☐

1920s ☐

1930s ☐

1940s ☐

1950s ☐

1960s ☐

1970s ☐

1980s ☐

1990s ☐

2000s ☐

1	☐	BLACK AND WHITE	778	47	34	22	102	163	162	104	99	45
---	--------------------------	-----------------	-----	----	----	----	-----	-----	-----	-----	----	----

0.609 seconds

SEE CONTEXT: CLICK ON WORD (ALL SECTIONS), NUMBER (ONE SECTION), OR [CONTEXT] (SELECT)

☐ **CONTEXT**

TOT ☐

1920s ☐

1930s ☐

1940s ☐

1950s ☐

1960s ☐

1970s ☐

1980s ☐

1990s ☐

2000s ☐

1	☐	WHITE AND BLACK	156	7	7	15	20	31	40	16	16	4
---	--------------------------	-----------------	-----	---	---	----	----	----	----	----	----	---

1

--IG
2000
1990
1980
1970
1960
1950

Help / information / contact

3. VOCABULARY LEARNING THROUGH COLLOCATION

TIME MAGAZINE CORPUS

ACCESS: 0/5

100 MILLION WORDS, 1923-2006

COMPARE: COCA COHA TIME BNC **GOOGLE**

[history](#) | [lists](#) | [SEE PROFILE](#) | [logout](#)

DISPLAY SEE CONTEXT: CLICK ON WORD (ALL SECTIONS), NUMBER (ONE SECTION), OR [CONTEXT] (SELECT)

[\[HELP...\]](#)

	☐	CONTEXT	TOT ☐	1920s ☐	1930s ☐	1940s ☐	1950s ☐	1960s ☐	1970s ☐	1980s ☐	1990s ☐	2000s ☐
1	☐	MAKE A MISTAKE	62	3	7	4	12	10	7	5	7	7

0.703 seconds

DISPLAY

☒ LIST ☐ CHART ☒ KWIC ☐ COMPARE

SEARCH STRING

WORD(S)

COLLOCATES

POS LIST

RANDOM

SEARCH

RESET

SECTIONS ☒ SHOW

SECTION

☒ SHOW

1 --IG
2000
1990
1980
1970
1960
1950

Sorry, there are no matching records. You may wish to do one of the following:

1. Make sure the frequency limits are set correctly (MIN FREQ and checkbox)
2. Select a less restrictive set of sections (genres or time periods)
3. [Check the query syntax](#)

4. VOCABULARY LEARNING THROUGH SEMANTIC PROSODY

Semantic prosody is the “consistent aura of meaning with which a form is imbued by its collocates” (Louw 1993).

What is the SEMANTIC PROSODY of the verb
undergo?



which truly valued women would have had to present a composition portfolio and calves. [p] [p] Cows kept for milk production Other studies have shown that families who Erdington estate. [p] Mr Forbes was forced to on contact with air. Even a full bottle will moderately advanced glaucoma. They would then bright and cheerful" in hospital, where he'll [p] Once a year each operative had to arriving in the other world still has to Most alternative practitioners have to concrete forms of representation themselves with Joni when she saw what Joni had to [p] Before joining the Host company, managers against the emotional roller-coaster he would Not every company car driver needs to also tighten tests that all AC wines have to with parents so that they can ask a student to of the 15,000-ton Severodvinsk class will unchanged for 14 years. [p] Yet the LPC is to September may be an odd month during which you in the hit film Mermaids, was challenged to distressing disease. [p] Many of them British-born star has been told she must now latter, the current storylines will have to know that, in time, my daughters will have to with three days in hospital where the women sport safer. One is that every fighter must then a stock with a beta of, say, 0.8 should experience a high of euphoric energy and can to go, he would walk through the living room, doctoral degree in Spanish universities had to turns out that the woman really didn't want to other STL s, anyone with such growths should family adjustment and adaptation. Families to Milledgeville to live with her mother,

undergo a deep and important change of heart. [p] undergo a viva-voce examination. The portfolio must undergo a continuous cycle of pregnancies, producing undergo a period of constant day-to-day hassles are undergo an emergency operation to remove a blood undergo chemical changes if left in direct light. undergo further tests to confirm the diagnosis and undergo a brain scan before leaving on Monday. [h] undergo the most rigorous test of all -- to pit his undergo many tribulations before being able to undergo lengthy training, and it would be just as undergo a sort of intrinsic mystical transformation undergo, day after day, week after week. Jackie, undergo training of up to three weeks `to update and undergo if he launches an appeal. [p] Leeson, who undergo such training, however. We estimate that undergo and remove the interested parties from the undergo analysis if there are well-founded undergo trials within the next two years. [p] The undergo a review by the Law Society, with relevant undergo a crisis of confidence or find yourself undergo the change for the Trading Places charity. undergo disfiguring surgery and every year 15,000 undergo a course of exhausting chemotherapy. [p] undergo a lobotomy if they're to sustain us for a undergo tests. But you can only hope if they have undergo a one-and-a-half hour breast enlargement undergo a mandatory yearly brain scan - and if undergo a change of around 8 (10 [o] 0.8 8). Now you undergo a transcendent, positive change in attitude undergo the usual shouts and insults, and pretend to undergo, and teases them quite without mercy. This undergo sterilization, but did so at the insistence undergo thorough medical evaluation to check for the undergo repeated cycles of adjustment/crisis/ undergo treatment, raise peacocks, and write. The

which truly valued women would have had to undergo a deep and important change of heart. [p]
 present a composition portfolio and undergo a viva-voce examination. The portfolio must
 calves. [p] [p] Cows kept for milk production undergo a continuous cycle of pregnancies, producing
 Other studies have shown that families who undergo a period of constant day-to-day hassles are
 Erdington estate. [p] Mr Forbes was forced to undergo an emergency operation to remove a blood
 on contact with air. Even a full bottle will undergo chemical changes if left in direct light.
 moderately advanced glaucoma. They would then undergo further tests to confirm the diagnosis and
 bright and cheerful" in hospital, where he'll undergo a brain scan before leaving on Monday. [h]
 [p] Once a year each operative had to undergo the most rigorous test of all -- to pit his

Semantic prosody: *NEGATIVE*

Reason: *UNDERGO* usually co-occurs with words
 expressing *unpleasant situations*

(cf. examination, test, training, treatment, evaluation, ...)

latter, the current storylines will have to undergo a lobotomy if they're to sustain us for a
 know that, in time, my daughters will have to undergo tests. But you can only hope if they have
 with three days in hospital where the women undergo a one-and-a-half hour breast enlargement
 sport safer. One is that every fighter must undergo a mandatory yearly brain scan - and if
 then a stock with a beta of, say, 0.8 should undergo a change of around 8 (10 [o] 0.8 8). Now you
 experience a high of euphoric energy and can undergo a transcendent, positive change in attitude
 to go, he would walk through the living room, undergo the usual shouts and insults, and pretend to
 doctoral degree in Spanish universities had to undergo, and teases them quite without mercy. This
 turns out that the woman really didn't want to undergo sterilization, but did so at the insistence
 other STL s, anyone with such growths should undergo thorough medical evaluation to check for the
 family adjustment and adaptation. Families undergo repeated cycles of adjustment/crisis/
 to Milledgeville to live with her mother, undergo treatment, raise peacocks, and write. The

FURTHER EXERCISES

- **collocation**

strong tea

The tea is strong

- **colligation**

in <quantifier> cases

*some
several
many
most
...*

- **semantic preference**

seek <relief>

*help
advice
support
asylum
.....*

- **semantic prosody**

cause <unpleasant events>

*problem
death
damage
concern
...*