

# Vocabulary Analysis of Project Gutenberg

Zachary Booth Simpson  
May 2000

(c)2002 ZBS. <http://www.mine-control.com/zack>  
Please [sign my guestbook](#) if you find this work useful.

## Introduction

While reading Moby Dick in April 2000, I was astounded by Melville's enormous vocabulary. I wondered what was Moby Dick's total vocabulary and how it compared to other works. Thanks to the [Project Gutenberg](#), an online resource for literature, (and copious spare-time) I was able to download a considerable sample of works and perform a word analysis. The following are the results from this informal study including relative vocabulary densities and anomalous word usage.

Please [sign my guestbook](#) if you find this interesting or helpful. Thanks, Zack

## Sample Database

The works represented in this study come exclusively from the [Project Gutenberg](#) (PG). While most PG works are included, the sample is not complete; some works have been eliminated for obvious reasons (e.g. Pi to 10,000 digits) while others works were eliminated because they were malformed or unavailable. Some books in the Project Gutenberg are split into several separate volumes or alternatively several works are combined into one; this may effect the sample slightly, especially the [Anomalous Word Charts](#). In some cases, I have manually combined multiple volumes into one for logical consistency.

The sample index was derived from the [Thalason Index](#) of the Project Gutenberg because the master indices from the PG itself were inconsistent. I extend my thanks to their efforts as well as to all contributors to the Project Gutenberg.

View the sample database by [TITLE](#)  
View the sample database by [AUTHOR](#)

**NOTE:** Due to a change of server, I no longer have sufficient room to store the entire sample database on-line. My apologies.

## Total Vocabulary

'Total Vocabulary' is the measure of unique words in a book. A word is defined as a set of case-insensitive alpha characters and apostrophes (to include contractions such as can't) thus excludes numbers and punctuation. Each work is scanned in its entirety including titles, indices, and page numbers after eliminating the Gutenberg Preamble which prefixes each work.

### Largest Vocabularies (Regardless Of Book Size)

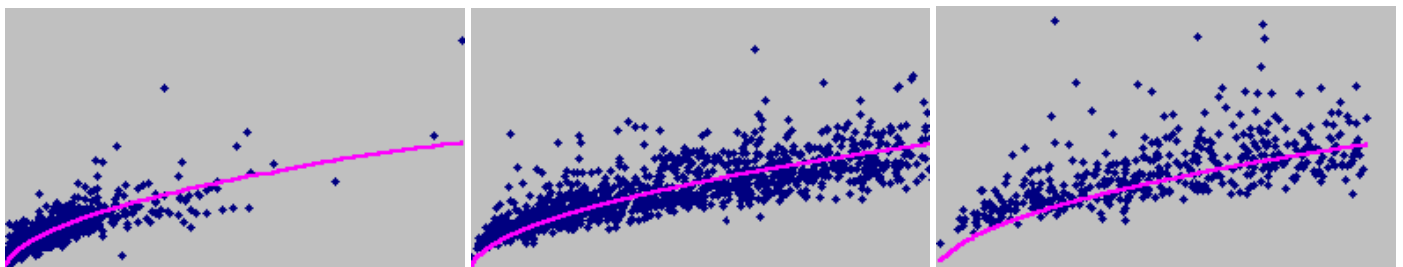
Title (click on work to view word anomalies)	Author	Vocabulary Words	
<a href="#">Decline and Fall of the Roman Empire, vol 1-6</a>	<a href="#">Gibbon, Edward</a>	43113	1543676
<a href="#">Roget's Thesaurus</a>	<a href="#">Anonymous / Various</a>	39023	203886
<a href="#">Gargantua and Pantagruel</a>	<a href="#">Rabelais, Francis</a>	25985	323013
<a href="#">1998 CIA World Factbook, The</a>	<a href="#">US CIA</a>	24220	422744
<a href="#">Les Miserables</a>	<a href="#">Hugo, Victor</a>	23334	570508
<a href="#">Anomalies and Curiosities of Medicine</a>	<a href="#">Gould/Pyle</a>	22930	393856
<a href="#">Brann The Iconoclast, vol 1,10,12</a>	<a href="#">Brann, William Cowper</a>	22507	300783
<a href="#">Plutarch's Lives, trans by A. H. Clough</a>	<a href="#">Plutarch</a>	20237	742013
<a href="#">History Of The Conquest Of Peru (2nd ver), The</a>	<a href="#">Prescott, William H.</a>	19235	300976
<a href="#">Warfare of Science/Theology</a>	<a href="#">White, Andrew Dickson</a>	19187	322799

<a href="#">Bible, Douay-Rheims Version, Challoner Revision, The Anonymous / Various</a>		18559	1029084
<a href="#">Moby Dick</a>	<a href="#">Melville, Herman</a>	17227	211763
<a href="#">Cloister and the Hearth, The</a>	<a href="#">Reade, Charles</a>	16911	282120
<a href="#">Hackers' Dictionary of Computer Jargon, The</a>	<a href="#">Anonymous / Various</a>	16757	169716
<a href="#">Sketches by Boz</a>	<a href="#">Dickens, Charles</a>	16413	262440
<a href="#">Vanity Fair</a>	<a href="#">Thackeray, William Makepeace</a>	16349	360049
<a href="#">Our Mutual Friend</a>	<a href="#">Dickens, Charles</a>	16337	338266
<a href="#">Dombey and Son</a>	<a href="#">Dickens, Charles</a>	16332	366517
<a href="#">Pickwick Papers, The</a>	<a href="#">Dickens, Charles</a>	16253	313143
<a href="#">Don Quixote (tr John Ormsby)</a>	<a href="#">Cervantes</a>	16160	425814
<a href="#">Count of Monte Cristo, The</a>	<a href="#">Dumas, père, Alexandre</a>	16110	464256
<a href="#">Terminal Compromise/NetNovel</a>	<a href="#">Schartau, Win</a>	15898	213672

## Vocabulary Density

'Vocabulary Density' is a measurement of vocabulary usage in comparison to the length of the book. This ratio is expressed as the 'Inverse Absolute Vocabulary Density' and is computed dividing the Total Words by the Unique Words (W/V). This statistic may be thought of as: 'how many words will be read on average before a new word is encountered.' For example, Moby Dick has a (W/V) score of approximately 12 -- a new word is introduced on approximately every line of the book! That is quite an accomplishment for a work that is almost a quarter of a million words long.

Ideally, the (W/V) statistic allows comparison of one book's style to another. However, this simplistic metric is complicated by the simple fact that a short work will inevitably be denser than a larger work due to the fact that practically every word in a short work is unique. To understand, consider the case of writing a multi-million word essay. Given that there are only a limited number of words in the English language (~400,000 in this sample), one would eventually run out of words and thus the vocabulary density of such a titanic treatise would drop accordingly. This effect can be seen in the flattening trend of the scatter plots seen below.



**Figure 1. 800,000 word domain**

**Figure 2. 100,000 word domain**

**Figure 3. 30,000 word domain**

Scatter-plots of inverse vocabulary density (y-axis) vs. total words (x-axis). Samples below the pink trend line have denser vocabularies than average, those above, sparser. Note that trend line fits less well for smaller works.

In order that the vocabulary densities of large and small works may be compared, a 'normalizing' curve is fit to the sample creating a 'normalized density score' useful for comparison. Unfortunately, the one-size-fits-all trend curve (found empirically by minimizing least mean square error of a square-root scale coefficient) fails to fit the smaller works well as can be seen in Figure 3. Thus, comparison of large works (> 30,000 words) to smaller ones (< 30,000) is ill-advised. Therefore, the following tables isolate these two sample groups.

**NOTE: Due to a change of server, I no longer have sufficient room to store the entire sample database on-line. My apologies.**

### Most Dense Vocabularies, Normalized For Book Size. Books Over 30,000 Words

Title (click on work to view word anomalies)	Author	Vocabulary Words	Normal Density
<a href="#">Decline and Fall of the Roman Empire, vol 1-6</a>	<a href="#">Gibbon, Edward</a>	43113	1543676 -12.92
<a href="#">Roget's Thesaurus</a>	<a href="#">Anonymous / Various</a>	39023	203886 -12.48
<a href="#">Gargantua and Pantagruel</a>	<a href="#">Rabelais, Francis</a>	25985	323013 -9.86
<a href="#">Brann The Iconoclast, vol 1,10,12</a>	<a href="#">Brann, William Cowper</a>	22507	300783 -8.14
<a href="#">1998 CIA World Factbook, The</a>	<a href="#">US CIA</a>	24220	422744 -8.04

<a href="#">Anomalies and Curiosities of Medicine</a>	<a href="#">Gould/Pyle</a>	22930	393856	-7.43
<a href="#">Hackers' Dictionary of Computer Jargon, The</a>	<a href="#">Anonymous / Various</a>	16757	169716	-6.03
<a href="#">History Of The Conquest Of Peru (2nd ver), The</a>	<a href="#">Prescott, William H.</a>	19235	300976	-5.87
<a href="#">Moby Dick</a>	<a href="#">Melville, Herman</a>	17227	211763	-5.75
<a href="#">Warfare of Science/Theology</a>	<a href="#">White, Andrew Dickson</a>	19187	322799	-5.46
<a href="#">Poems And Songs Of Robert Burns</a>	<a href="#">Burns, Robert</a>	14968	129551	-5.46
<a href="#">Les Miserables</a>	<a href="#">Hugo, Victor</a>	23334	570508	-5.17
<a href="#">Travels through France &amp; Italy</a>	<a href="#">Smollett, Tobias</a>	14625	142922	-5.05
<a href="#">Waverley</a>	<a href="#">Scott, Walter</a>	15325	185273	-4.79
<a href="#">Tracks of a Rolling Stone</a>	<a href="#">Coke, Henry J.</a>	13259	106525	-4.77
<a href="#">Terminal Compromise/NetNovel</a>	<a href="#">Schartau, Win</a>	15898	213672	-4.69
<a href="#">Main Street</a>	<a href="#">Lewis, Sinclair</a>	14580	169912	-4.51
<a href="#">Sketch-Book of Geoffrey Crayon, The</a>	<a href="#">Irving, Washington</a>	13241	129907	-4.32
<a href="#">Devil's Dictionary, The</a>	<a href="#">Bierce, Ambrose</a>	11172	60906	-4.23
<a href="#">Roads of Destiny</a>	<a href="#">O'Henry</a>	12196	98952	-4.22
<a href="#">Leaves of Grass</a>	<a href="#">Whitman, Walt</a>	12924	124036	-4.21
<a href="#">Lucasta Poems, The</a>	<a href="#">Lovelace, Richard</a>	11153	62900	-4.20

### Least Dense Vocabularies, Normalized For Book Size. Books Over 30,000 Words

Title (click on work to view word anomalies)	Author	Vocabulary Words	Normal Density
<a href="#">Book of Mormon, The</a>	<a href="#">Anonymous / Various</a>	5612	275887 28.56
<a href="#">Bible, Both Testaments, King James Version, The</a>	<a href="#">Anonymous / Various</a>	12867	790126 26.55
<a href="#">Le Morte D'Arthur, vol 2</a>	<a href="#">Malory, Thomas</a>	5717	194249 16.69
<a href="#">Bible, Douay-Rheims Version, Challoner Revision, The</a>	<a href="#">Anonymous / Various</a>	18559	1029084 15.67
<a href="#">High History of the Holy Graal, The</a>	<a href="#">Anonymous / Various</a>	5327	158488 14.14
<a href="#">Le Morte D'Arthur, vol 1</a>	<a href="#">Malory, Thomas</a>	5826	169703 12.97
<a href="#">Treaty of the European Union [Maastricht], The</a>	<a href="#">Anonymous / Various</a>	2826	59469 11.48
<a href="#">Nada the Lily</a>	<a href="#">Haggard, H. Rider</a>	5040	117857 9.92
<a href="#">White Knight: Tirant Lo Blanc (tr R.S. Rudder), The</a>	<a href="#">Martorell, Joanot</a>	6343	161871 9.74
<a href="#">Story of Burnt Njal (Njal's Saga) Icelandic, The</a>	<a href="#">Anonymous / Various</a>	5468	129135 9.52
<a href="#">Moll Flanders</a>	<a href="#">Defoe, Daniel</a>	6139	139300 8.05
<a href="#">Heimskringla [Norwegian Kings]</a>	<a href="#">Sturlson, Snorri</a>	10405	306474 7.74
<a href="#">Twilight Land</a>	<a href="#">Pyle, Ernie Howard</a>	4113	74003 7.32
<a href="#">First Book of Adam and Eve</a>	<a href="#">Platt, Rutherford</a>	2287	32820 7.25
<a href="#">On the Origin of Species</a>	<a href="#">Darwin, Charles</a>	6993	155549 6.78
<a href="#">Personal Memoirs of U.S. Grant, vol 2</a>	<a href="#">Grant, Ulysses S.</a>	6965	154177 6.74
<a href="#">Two Years in the Forbidden City</a>	<a href="#">Der Ling, Princess</a>	4962	92456 6.71
<a href="#">United States Copyright Act of 1976, The</a>	<a href="#">Anonymous / Various</a>	2271	30635 6.63
<a href="#">Princess of Cleves, The</a>	<a href="#">Lafayette, Madame de</a>	3779	61809 6.61
<a href="#">Emma</a>	<a href="#">Austen, Jane</a>	7228	161099 6.55
<a href="#">Flower Fables</a>	<a href="#">Alcott, Louisa May</a>	2501	34525 6.52
<a href="#">Parmenides</a>	<a href="#">Plato</a>	2616	36337 6.41

### Most Dense Vocabularies, Normalized For Book Size. Books Under 30,000 Words

Title (click on work to view word anomalies)	Author	Vocabulary Words	Normal Density
<a href="#">Biog Study of A. W. Kinglake</a>	<a href="#">Teikwell, Rev. W.</a>	6794	29001 -2.41
<a href="#">Waifs and Strays, etc</a>	<a href="#">O'Henry</a>	5826	29482 -1.67
<a href="#">50 Bab Ballads (vol 1)</a>	<a href="#">Gilbert, W.S.</a>	5689	28588 -1.61
<a href="#">Style</a>	<a href="#">Raleigh, Walter</a>	5385	24331 -1.60
<a href="#">Cicero's Orations [selected orations in Latin]</a>	<a href="#">Cicero</a>	4525	13219 -1.59
<a href="#">New Poems</a>	<a href="#">Thompson, Francis</a>	5392	25151 -1.55
<a href="#">Chita: A Memory of Last Island</a>	<a href="#">Hearn, Lafcadio</a>	5495	26874 -1.54

<a href="#">Poems</a>	<a href="#">Henley, William E.</a>	5301	24303	-1.53
<a href="#">Georgics [English], The</a>	<a href="#">Virgil</a>	5089	21668	-1.51
<a href="#">Letters on Literature</a>	<a href="#">Lang, Andrew</a>	5550	29479	-1.42
<a href="#">Shelley</a>	<a href="#">Waterlow, Sydney</a>	4907	21390	-1.38
<a href="#">Sword Blades and Poppy Seed</a>	<a href="#">Lowell, Amy</a>	5255	26996	-1.31
<a href="#">Bab Ballads, vol 2, The</a>	<a href="#">Gilbert, W.S.</a>	4769	20582	-1.31
<a href="#">Who Was Who: 5000 B. C. to Date</a>	<a href="#">Gordon, Irwin L.</a>	4802	21807	-1.25
<a href="#">Ginx's Baby, A Satire</a>	<a href="#">Jenkins, Edward</a>	5370	29763	-1.22
<a href="#">Bab Ballads, vol 3, The</a>	<a href="#">Gilbert, W.S.</a>	4853	23153	-1.20
<a href="#">Foolish Dictionary, The</a>	<a href="#">Wurdz, Gideon</a>	3826	11615	-1.19
<a href="#">Lays of Ancient Rome</a>	<a href="#">Macaulay, Thomas Babbington</a>	4987	25043	-1.18
<a href="#">Essay on Comedy, Comic Spirit</a>	<a href="#">Meredith, George</a>	4344	17204	-1.18
<a href="#">Reginald in Russia and Other Sketches</a>	<a href="#">Saki (H.H. Munro)</a>	4715	22184	-1.14
<a href="#">Philobiblon of Richard de Bury, The</a>	<a href="#">Bury, Richard de</a>	4921	24906	-1.13
<a href="#">Reading of Life, and Other Poems, A</a>	<a href="#">Meredith, George</a>	3878	12990	-1.12

### Least Dense Vocabularies, Normalized For Book Size. Books Under 30,000 Words

Title (click on work to view word anomalies)	Author	Vocabulary Words		Normal Density
<a href="#">New McGuffey First Reader, The</a>	<a href="#">McGuffey (compiler), W.H.</a>	630	8276	9.57
<a href="#">Ethics, part 2 (tr Elwes)</a>	<a href="#">Spinoza, Benedict de</a>	1485	18314	7.03
<a href="#">Ethics, part 3 (tr Elwes)</a>	<a href="#">Spinoza, Benedict de</a>	1866	22877	6.33
<a href="#">Somebody's Little Girl</a>	<a href="#">Young, Martha</a>	983	9795	6.08
<a href="#">Ethics, part 1 (tr Elwes)</a>	<a href="#">Spinoza, Benedict de</a>	1422	14046	5.23
<a href="#">Berne Universal Copyright Convention [1988], The</a>	<a href="#">Anonymous / Various</a>	967	8023	4.78
<a href="#">Adventures of Reddy Fox</a>	<a href="#">Burgess, Thornton W.</a>	1572	14948	4.71
<a href="#">Ethics, part 5 (tr Elwes)</a>	<a href="#">Spinoza, Benedict de</a>	1269	10805	4.44
<a href="#">Well of the Saints, The</a>	<a href="#">Synge, J.M.</a>	1695	15540	4.28
<a href="#">Lady Windermere's Fan</a>	<a href="#">Wilde, Oscar</a>	2056	19942	4.16
<a href="#">Organic Syntheses</a>	<a href="#">Conant (Editor), James Bryant</a>	2202	21695	4.08
<a href="#">Alice's Adventures in Wonderland</a>	<a href="#">Carroll (C.L. Dodgson), Lewis</a>	2649	27785	3.95
<a href="#">White People, The</a>	<a href="#">Burnett, Frances Hodgson</a>	2262	21593	3.78
<a href="#">Dreams</a>	<a href="#">Schreiner, Olive</a>	2137	19817	3.75
<a href="#">True Story of Christopher Columbus, The</a>	<a href="#">Brooks, Elbridge S.</a>	2805	29141	3.69
<a href="#">Woman of No Importance, A</a>	<a href="#">Wilde, Oscar</a>	2374	22496	3.59
<a href="#">Meno, second part</a>	<a href="#">Plato</a>	933	6200	3.56
<a href="#">Tom Sawyer Detective</a>	<a href="#">Twain (Samuel Clemens), Mark</a>	2475	23467	3.47
<a href="#">Rosmersholm</a>	<a href="#">Ibsen, Henrik</a>	2815	28053	3.40
<a href="#">Ballad of Reading Gaol</a>	<a href="#">Wilde, Oscar</a>	1196	8287	3.36
<a href="#">Deirdre of the Sorrows</a>	<a href="#">Synge, J.M.</a>	1968	16415	3.32
<a href="#">Story of Doctor Dolittle, The</a>	<a href="#">Lofting, Hugh</a>	2759	26835	3.30

## Word Anomalies

It would be interesting to know for a given book what words are used uncommonly often or, likewise, uncommonly infrequently. To compute this, the relative frequency of each words is sampled from the database at large and then compared to the frequency in each book.

Not surprisingly, these 'Anomalous Word Summaries' paint an incredibly accurate picture of the work. For example, among Moby Dick's most anomalous words are: whale, sperm, and harpooneer. Of course, proper names tend to dominate these lists; for example, ahab, stubb, and queequeg top out Moby Dick. Just as interesting is what the book is NOT about. Among Moby Dick's most infrequently used words (i.e. words which are common in other books, but not in this one) are: miss, government, happiness, smiled, and machine.

The Infrequently Used Summaries list only words which are **actually used in the work**. While it might be logical to list words that are frequently used in other books but that **never** show up in this book, it would be useless because such a list

would be dominated by anachronistic words such as 'thou' and 'thy' that are common in the database but unused in most works.

Misspellings significantly skew both the Infrequent and Unique Word Lists and are fairly common due to the use of Optical Character Recognition (OCR) software which is extremely prone such mistakes.

The following table is a sample of Word Anomalies picked by hand from the database to illustrate the technique. To view Anomaly Summaries for any work, click on the book name in either the author index or title index.

**NOTE: Due to a change of server, I no longer have sufficient room to store the entire sample database on-line. My apologies.**

View the index by [TITLE](#)

View the index by [AUTHOR](#)

(Click on any title to view the Anomaly Summary)

## Sample of Word Anomalies

### **The Bible (King James Edition); Anonymous / Various**

**Frequent:** unto, lord, isreal, shall, god, moses, jesus, david, offering, tabernacle

**Infrequent:** girl, boy, school, success, condition, listen, princess

### **Wonderful Wizard of Oz; Baum, Frank**

**Frequent:** woodman, scarecrow, witch, tin, emerald, monkeys, kansas, brains, winged

**Infrequent:** mother, money, soul, natural

### **White Fang; London, Jack**

**Frequent:** musher, beaver, sled, dogs, cherokee, snarl

**Infrequent:** letter, person, window, green, sweet, loved, party, paper

### **The Republic; Plato**

**Frequent:** guardians, unjust, true, injustice, state, gymnastic, rulers, democractical

**Infrequent:** miss, girl, boy, prince

### **Alice's Adventures In Wonderland; Carroll (C.L. Dodgson), Lewis**

**Frequent:** gryphon, turtle, caterpillar, mock, dodo, mouse, rabbit, hedgehog

**Infrequent:** death, country, happy, fair, common

### **Origin of the Species; Darwin, Charles**

**Frequent:** species, varieties, subaerial, selection, sterility, plants, modification, forms, variability

**Infrequent:** person, government, love, thinking, god, evil, fire

### **Communist Manifesto; Marx, Karl/Engels, Friedrich**

**Frequent:** bourgeois, proletariat, communists, antagonisms, revolutionising, socialism, production, class, feudal, reactionary, exploitation, conditions, crises

**Infrequent:** said, love, why, heart, mother, poor, felt

### **Paradise Lost; Milton, John**

**Frequent:** wonderous, heaven, satan, dominations

**Infrequent:** country, church, horses, sister

### **Apology; Plato**

**Frequent:** corrupter, accusers, demigods, socrates, oracle, indictment

**Infrequent:** she, work, morning, replied, body

### **Gargantua and Pantagruel; Rabelais, Francis**

**Frequent:** codpiece, catchpole, ballocks, dingdong, fart, chitterlings, gymnast, arse

**Infrequent:** smile, existence, feelings, british, professor, suffering

### **1st Inaugural Speech; Roosevelt, Franklin Delano**

**Frequent:** foreclosure, interdependence, uneconomical, leadership, outgo, unsolvable, values, redistribution, national, emergency

**Infrequent:** you, her, his

### **The Jungle; Sinclair, Upton**

**Frequent:** packingtown, packers, stockyards, fertilizer, slaughterhouses, streetcar, lituanian

**Infrequent:** influence, village, pray, gods, example

### **20,000 Leagues Under The Sea; Verne, Jules**

**Frequent:** manometer, canadian, captain, frigate, harpoon, cuttlefish, submarine

**Infrequent:** garden, justice, ladies, laughed, wife

**Time Machine; Wells, H. G.**

**Frequent:** psychologist, sphinx, traveller, machine, i, lever, dimension

**Infrequent:** mother, dear, money, friends, horse, peace

**War of the Worlds; Wells, H. G.**

**Frequent:** martians, leatherhead, artilleryman, londonward, cylinder, pit, scullery

**Infrequent:** love, king, truth, gentleman, joy, youth

**Moby Dick; Melville, Herman**

**Frequent:** whale, sperm, harpooner, pequod, leviathan, fishery

**Infrequent:** miss, fortune, happiness, smiled, angry, enemies

Previous (Prohibition on alcohol (United States)). Next (Projectile). Project Gutenberg, abbreviated as PG, is a volunteer effort to digitize, archive, and distribute cultural works. Founded in 1971, by Michael S. Hart, it is the oldest digital library. Most of the items in its collection are the full texts of public domain books. The project tries to make these as free as possible, in long-lasting, open formats that can be used on almost any computer. As of December 2007, Project Gutenberg claimed Project Gutenberg began in 1971 by Michael Hart as a community project to make plain text versions of books available freely to all. Book from Project Gutenberg: The Happy Prince and Other Tales Library of Congress Classification: PR. Project Gutenberg. 22,270 22K. Ducks at a Distance: A Waterfowl Identification Guide. - - by Hines, Robert W. texts. eye 22,270. Start of this project gutenberg ebook new word-analysis \*\*\*. Produced by Keith Edkins and the Online Distributed Proofreading Team at <http://www.pgdp.net>. NEW WORD-ANALYSIS 6. The Latin element in the English vocabulary consists of a large number of words of Latin origin, adopted directly into English at various periods. The principal periods, during which Latin words were brought directly into English are: 1. At the introduction of Christianity into England by the Latin Catholic missionaries, A.D. 596. 8 Project Gutenberg tips and tricks. 1. Preview books online with images. Before downloading any book from PG to your e-reading device, you can preview or read it in your web browser. When it comes to collections, Project Gutenberg does a fantastic job in putting together the works of popular authors. So, instead of several books (files) by Mark Twain, you can download all his titles published earlier on PG in one file. You can find here complete works of not only Mark Twain, but also William Shakespeare, Edgar Allan Poe, Jane Austen, or Winston Churchill. \* \* \* \* Welcome to Project Gutenberg Canada! The ebooks on this website are in the Canadian public domain, and are offered to you at no charge. If you live outside Canada, download an ebook only if you are certain that the book is in your country's public domain. Copyright in Canada generally lasts until 50 years after the end of the year of the author's death. Bienvenue à Projet Gutenberg Canada! Les conditions numériques du présent site vous sont offertes sans frais: elles font partie du domaine public canadien. Si vous ne vivez pas au Canada, vous devez vous assurer qu'un livre