UNIVERSITÀ DEGLI STUDI DI NAPOLI
"FEDERICO II"

FACOLTÀ DI SCIENZE POLITICHE
DIPARTIMENTO DI SCIENZE STATISTICHE
SEZIONE LINGUISTICA

DOTTORATO DI RICERCA IN

LINGUA INGLESE PER SCOPI SPECIALI

XX CICLO

TESI DI DOTTORATO

"COGITAMUS ERGO SUMUS"
Web 2.0 Encyclopaedi@s: the case of Wikipedia
A Corpus Based Study

CANDIDATA
dott.ssa Antonella Elia

RELATORE
dott.ssa Cristina Pennarola

COORDINATORE
prof.ssa Gabriella di Martino

NAPOLI 2007

ACKNOWLEDGEMENTS

A doctoral degree is not simply writing a thesis. It is an enriching, although sometimes terrifying journey in which, luckily, you are not alone since along with you, there are many other people. The supervision, expertise, encouragement and patience of my travel companions have been indeed essential for me to conclude my journey and reach my final destination.

For this reason I wish to thank the doctoral board of *English For Special Purposes*, and in particular, Prof. Gabriella Di Martino who placed her trust in my research abilities, Dr. Cristina Pennarola for guiding me through the intellectual and empirical process of my research work and supporting me throughout while sharing my enthusiasm for my topic and findings. My thanks are also due to Prof. Vanda Polese, who gave me precious suggestions during these years and Mrs Mena Vilardi for dealing with all the administrative duties and deadlines of my doctorate. I also wish to express my gratitude to Prof. Jocelyne Vincent, my long life mentor and to all my new and numerous Wikipedian friends to whom I apologize in advance for not including them individually in this list.

On a more personal note, I would like to thank my family and my friends, for believing in me even when others did not (myself included), and for having constantly supported me throughout this stage of my life. I would also like to thank my partner, Giuseppe who grinned and bore me through what was at times a very hard life not only for me but also for the people next to me.

Last but not least, I thank the authors, editors and readers of both *Encyclopædia Britannica* and *Wikipedia*. Collecting knowledge is a noble, challenging and endless task, which they achieve with the best of intentions and infinite enthusiasm.

ABSTRACT

This doctoral research is a corpus based study focused on a new genre: Web 2.0 online encyclopaedias. In particular, the attention is focused on the English edition of Wikipedia, a multilingual, web-based, co-authored encyclopaedic project. In the introduction, the encyclopaedic genre and its milestones are presented from a diachronic point of view and the genre evolution and its migration from the paper format to the web is explored. Then, a general overview of Wikipedia and Encyclopaedia Britannica Online is carried out, followed by a presentation of wiki, as a new textual genre, with its new collaborative writing model. In the first part of this research, the linguistic analysis focuses on an intra-genre investigation which compares Wikipedia vs. Britannica encyclopaedic articles dealing on with the same topic. Index of Readability and Web Usability are then explored. In the second part, an inter-genre analysis contrastively analyzes Wikipedia talk pages and encyclopaedic entries. The WikiSpeak, the spoken-written language used by Wikipedians in their backstage community, is also taken into account. Findings of this research show to what extent Wikipedia's co-authored articles prove to be formal and standardized in a way not very dissimilar from Encyclopaedia Britannica Online. By contrast, talk pages and WikiSpeak can be considered a new writing space where a novel variety of the NetSpeak Jargon is conveyed. Encyclopaedic articles and WikiLanguage, talk pages and WikiSpeak, can be considered, in McLuhan's terms, the "medium and the message" of the new Web 2.0 collaborative environments. Thanks to them a new Computer Mediated Discourse Community with its specific linguistic peculiarities is coming to life.

TABLE OF CONTENTS

1. INTRODUCTION

This doctoral research is a corpus based study focused on Wikipedia, a free content multilingual web encyclopaedia written collaboratively by contributors around the world.

My specific interest in Wikipedia and online encyclopaedias grew after reading an article written by Emigh and Herring (2005) dedicated to this specific subject. It prompted me to develop and research this area in more depth. Other fundamental episodes were my meetings with some of the most eminent representatives of the Wiki world. First of all with Tommaso Tozzi, professor of Multimediality and Visual Communication in the Faculty of Education at the University of Florence. As supervisor of my second degree in Multimedia Education in 2004 and as main representative of *WikiartPedia*[1]*,* he introduced me to the Web 2.0 culture ideals and to the virtual activist philosophy. Secondly, it was influential my acquaintance with Jimbo Wales, the founder of Wikipedia, at the *Second Wikimania International Conference* at Harvard University in Boston (Cambridge, Massachusetts 1-3, August 2006). All the above mentioned factors have contributed in some way to better define my specific interest and to identify research objectives.

Cybergenre studies investigate new media as well as emergent forms. Nevertheless, scholarship on encyclopaedia-making in the contexts of the Internet and Computer Mediated Communication has not adequately addressed Wikipedia as a transitional genre, being the result of the hybridization process of Wikis web 2.0 technologies and traditional encyclopaedias.

Encyclopaedic entries in traditional encyclopaedias, such as Britannica, are written by individual scholars, professionals, and experts whereas articles in Wikipedia are written collaboratively by volunteers and sometimes by anonymous contributors. The differences in the authorial and writing process have stimulated this study and contributed to identify the following specific research questions:

- To what extent does the different nature of authorship and the dissimilar mechanisms of individual or collaborative writing influence the formal expository style conveyed in the Encyclopaedic genre and particularly in Wikipedia vs. Britannica Online?
- Can the *encyclopaedic expository style* be quantified? And if so, how and to what extent does it differ in the two above mentioned encyclopaedias?
- Are *Index of Readability* and the *Web Usability* similar or different in the two online encyclopaedias?

[1] *WikiartPedia* is an Italian wiki project dedicated to the research and documentation of Art and Network cultures http://www.wikiartpedia.org

- Do Wikipedia contributors use the same linguistic register in the encyclopaedia and inside the community? In other words, can a variation be quantitatively recorded between the *WikiLanguage*, the formal expository style of encyclopaedic entries and the *WikiSpeak*, the language spoken-written by contributors inside the community?

This doctoral research is mainly organized in two different areas. The first one offers an exploratory profile and a descriptive and quantitative analysis of Wikipedia as an online collaborative encyclopaedia. Specifically the intragenre analysis (Wikipedia vs. Britannica Online) explores the formal expository style of the encyclopaedic production and shows to what extent the *WikiLanguage* which is expository, formal, neutral and objective, is used by Wikipedians when they write official encyclopaedic articles in document mode pages. The study, which provides a systematic comparison to Britannica and an analysis of *Index of Readability* and *Web Usability*, identifies Wikipedia as an emergent encyclopaedic genre that joins traditional stylistic principles of reference works with web-only communication technologies.

The second part of this research analyses *Wikipedia* as a web 2.0 online community. In particular, the intergenre analysis (talk pages *vs.* Wikipedia entries) has focused the attention on the linguistic features of what has been here defined as *WikiSpeak,* the spoken-written language through which contributors express themselves during discussions in their backstage community (specifically in the talkpages associated to encyclopaedic entries). Compared to the *WikiLanguage*, *WikiSpeak* shows to use a more informal, involved and high context interactive style, through which contributors freely convey their personal writing style.

1. Encyclopaedias: A General Overview

The term encyclopaedia comes from the Greek words ἐγκύκλιος παιδεία (enkyklios paideia) which means *comprehensive education.* Owing to different orthographic conventions, both the spellings *encyclopaedia* and *encyclopaedia* are used in British and American English. The *æ* ligature (*encyclopædia),* frequently used in the 19th century, is rare today; nevertheless it is retained in product titles such as *Encyclopædia Britannica* and others. *Merriam-Webster Online Dictionary* defines Encyclopaedia as:

> a comprehensive written compendium that contains information on all branches of knowledge or treats comprehensively a particular branch of knowledge usually in articles arranged alphabetically often by subject.

Encyclopaedias are conceived as single works, in which the contents and relations of the various arts and sciences are systematically explained. They can cover many different areas of interest, or can focus only on a particular field of study. It is a genre notoriously difficult to produce, mainly because it is necessarily a hybrid genre consisting of numerous small entries that give the reader basic information on a particular subject. The sum of these parts is supposed to equal a more universal body of knowledge. Although attempts to produce books of this kind were made more than 2.000 years ago, nevertheless, the name *encyclopaedia* was not given to such works until the 16th century.

2. Previous Studies on Encyclopaedias

A brief overview of previous studies on encyclopaedias is provided in this section.

McArthur (1986) in *Worlds of reference: lexicography, learning and language from the clay tablet to the computer* analyses the conventions and forms of reference texts, dictionaries and encyclopaedias in a diachronic perspective over thousands of years starting from the ancient clay tablets and concludes with two chapters on technological effects on reference books. Though he writes before the World Wide Web, McArthur predicts that new technologies will change the producer/consumer relationship stating that the entire relationship will undergo a "profound sea change" (McArthur 1986:171). Smith (1989) in *Wholly new forms of encyclopaedias* argues that by the end of the 19th century the encyclopaedia's genre was already well defined with almost universally accepted principles of its form: text written in the national language, alphabetical ordered contents, articles written by employed specialists, inclusion of living people's biographies and illustrations, maps, plans, bibliographies and analytical indexes. Up-to-date articles and textual cross references supplemented the main work. Before the existence of the Web, when discussing the possible impacts of hypertext technologies on the encyclopaedia genre, Smith predicted that in electronic hypertext-based encyclopaedias article sequence would not be linear and multiple paths would not be provided, author and reader roles would be blurred while author contributions will be augmented by reader annotations, and article bibliographies would be partially replaced by direct hyperlinks to the source documents.

A comprehensive work is Collison's (1966) *Encyclopaedias: Their History throughout the Ages.* Centring his work on Bacon, Diderot and Encyclopaedia Britannica, Collison includes a complete chronology up to the 1960's and references not only to western encyclopaedias but also to Asian and Arabic works.

Kister's (1994) *Best Encyclopaedias: A Guide to General and Specialized Encyclopaedias* is essentially an annotated list of encyclopaedias, with their bibliographic facts, an evaluation and ways to purchase over seventy reference books. What is interesting in Kister's work is the introductory section where he includes a short description of an encyclopaedia and why anyone would need one. He then asks the question: *won't computers replace encyclopaedias?* (p.10). He concludes that computers, though becoming more pervasive, can never replace encyclopaedias. Computerized encyclopaedias are just electronic versions of printed material that are very expensive and require advanced technology like modems and *cd platters* (p. 11). It is evident how Kister did not imagine the forthcoming changes.

It is worth quoting Soojung-Kim Pang (2000)'s *The work of the encyclopaedia in the age of electronic reproduction*. He examines how the digitization of literature affects the craft of editing, and the everyday work of content producers. It particularly focuses its attention on Encyclopaedia Britannica, which like all encyclopaedias has been profoundly affected by the emergence of cds and Internet. In concluding, this short overview, Crawford in *Encyclopaedia* (2001)*,* proposes a very useful checklist base on a selected criteria to assess the quality of reference works: *scope* (purpose, subject coverage, audience, arrangement and style), *format*, *uniqueness*, *authority*, *accuracy* (accuracy and reliability, objectivity), *currency*, and *accessibility* (indexing). Two additional features also to be kept in mind are *relevance* to user needs and *cost*.

3. History of Encyclopaedias

3.1 Printed Encyclopaedias

The 'impulse' of collecting the world's knowledge into a single work has always been rooted in mankind since encyclopaedias emerged in the ancient world and later in the Middle Ages.

Early encyclopaedias were intended for continuous reading and study and represented the accumulated learning of their individual authors. They were designed to be all-inclusive textbooks and thus very different from modern encyclopaedias, which serve chiefly as references and are generally the product of many scholars' cooperative work (Kister, 1994).

The Greek philosopher Aristotle is often considered the father of encyclopaedias since he attempted to summarize the then existing knowledge in a single work. Nevertheless, the first encyclopaedia is said to have been compiled in the 4th century BC by the Greek philosopher Speusippus, a disciple of the Greek philosopher Plato. However, nothing remains of his work. The oldest complete encyclopaedia still in existence is the *Historia Naturalis* (about AD 77), written by the Roman writer Pliny the Elder. It is a natural science [2] encyclopaedia extremely popular in western Europe in the Middle Ages (Bolter, 2001). The most important of all the early encyclopaedias which sums up the learning of the time is the *Speculum majus* [3] (1220-1244), compiled by the Dominican friar Vincent of Beauvais. All these early compilations of knowledge and many of their successors were unsystematic or disorganized in both form and substance. Coordination and systematization of all branches of science remained an unsolved problem until modern times.

[2] *Historia Naturalis* remained popular for almost 1,500 years. The topics, treated in 37 books, include mathematical and physical descriptions of the world, anthropology, human physiology, botany, zoology and mineralogy.

[3] *Speculum majus* consists of 80 books and is made up of four parts. It represents the writings of 450 Greek, Hebrew, and Roman scholars. In the Renaissance, the English William Caxton translated it and printed it as *The Myrrour of the Worlde*.

In 1620, the English philosopher Francis Bacon devised a structure for his *Instauratio Magna*, which was intended to represent a reference work of all the available knowledge. His work can be considered one of the first efforts made to build a comprehensive work with a philosophic organization and an adequate method. Unluckily, he never completed the project.

The modern concept of encyclopaedia was largely the result of the age of the Enlightenment. The 18th-century was a period of intellectual curiosity and experimentation and one of the dominant trends of this century was the creation of reference works useful to a wide audience. Although the dominant arrangement was by subjects, most of the reference works started to be organized in alphabetical order and their structure became similar to that of a dictionary (Britannica, 2007). The definition *dictionary* started to be used in the title of many encyclopaedic works. During the Enlightenment, encyclopaedias became a work of reference in the strictest sense of the word: a work for occasional use, in which readers could find alphabetically ordered information on a particular topic (Kister, 1994).

To varying degrees, modern works have been based on this methodology[4]. The alternative approach has given rise to encyclopaedias based on a collection of monographs. Most modern encyclopaedias employ both principles to varying degrees, but they tend more toward the dictionary format since it can serve both specialized and general audience. In England the dictionary format was followed in Ephraim Chambers' *Cyclopaedia*[5] (1728), commonly considered the father of English encyclopaedias (Kister, 1994). A French translation of Chambers's *Cyclopaedia* was the foundation of the famous *Encyclopédie ou dictionnaire raisonné des sciences, des arts et des métiers* commonly called the *Encyclopédie* (fig. 1).

The task of revising the translation of Chambers's *Cyclopaedia* was given to French encyclopaeist Denis Diderot. He worked with a group made up of the most distinguished scholars of that age such as d'Alembert, Rousseau, Daubenton and others.

The purpose of the *Encyclopédie,* which was essentially an encyclopaedic dictionary was to:

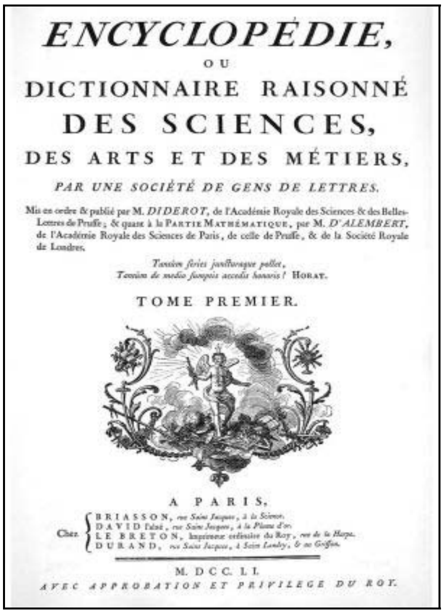
> […] exhibit as far as possible the order and system of human knowledge, and as a *dictionnaire raisonné* [descriptive dictionary] of the sciences, the arts, and trades, to contain the fundamental principles and the most essential details of every science and every art, whether liberal or mechanical.

This description, which is part of the *Encyclopédie's Preliminary Discourse* (1751), was written by d'Alembert to describe the structure of the articles included in the *Encyclopédie* and their

[4] The *Grand dictionnaire universel du XIX^{e} siècle* of Pierre Athanase Larousse is an extreme example of the 18th century modern encyclopaedic dictionary.

[5] *Cyclopaedia* was the product of many contributors and was the first example of modern encyclopaedia with the systematic collaboration of many scholars. *Cyclopaedia* was published in two volumes and had different editions during Chambers's lifetime.

philosophy, as well as to provide the reader with a background in the history of the cultural works which contributed to the knowledge of the time.



ENCYCLOPÉDIE,
OU
DICTIONNAIRE RAISONNÉ
DES SCIENCES,
DES ARTS ET DES MÉTIERS,
PAR UNE SOCIÉTÉ DE GENS DE LETTRES.

Mis en ordre & publié par M. *DIDEROT*, de l'Académie Royale des Sciences & des Belles-Lettres de Pruſſe ; & quant à la PARTIE MATHÉMATIQUE, par M. *D'ALEMBERT*, de l'Académie Royale des Sciences de Paris, de celle de Pruſſe, & de la Société Royale de Londres.

Tantùm ſeries juncturaque pollet,
Tantùm de medio ſumptis accedit honoris! HORAT.

TOME PREMIER.

A PARIS,
Chez BRIASSON, *rue Saint Jacques, à la Science.*
DAVID l'aîné, *rue Saint Jacques, à la Plume d'or.*
LE BRETON, Imprimeur ordinaire du Roy, *rue de la Harpe.*
DURAND, *rue Saint Jacques, à Saint Landry, & au Griffon.*

M. DCC. LI.
AVEC APPROBATION ET PRIVILEGE DU ROY.

Fig. 1 The title page of the *Encyclopédie*

The *Encyclopédie* presented explicit philosophical ideals and for this reason was considered revolutionary by the conservatives, who condemned it and its editors to persecution. This aspect of the *Encyclopédie* has given it an important place in the history of modern thought. Those who accepted its views became identified as Encyclopaedists, a term that denotes a social philosophy and a defined movement. The *Encyclopédie* was first published in 28 volumes between 1751 and 1772 and was followed by many editions (Britannica, 2007).

3.2 Encyclopaedia Britannica

As the dictionary-style encyclopaedia grew in importance, so did the monographic encyclopaedia. A major example is the *Encyclopædia Britannica* (Kister, 1994) which contained distinct treatises and long articles but also included definitions in alphabetical order. These general characteristics have been retained in each of the following editions since the 18th century.

The *Britannica* is the oldest English-language encyclopaedia that is still in print and it has been a trusted reference work for scholars for more than two centuries. It was born in Scotland in the 18th century in the middle of the great intellectual ferment of the Scottish Enlightenment.

It was in this setting that Colin Macfarquhar, a printer, and Andrew Bell, an engraver, decided to create an encyclopaedia arranged alphabetically, compiled upon a new plan in which the different Sciences and Arts were organized into distinct treatises (Britannica, 2007).

The first edition of the *Britannica* was published one section at a time, in fascicles, over a three-year period, beginning in 1768.

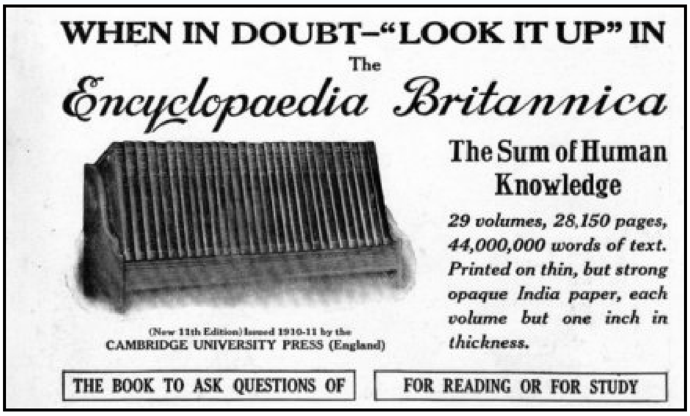

WHEN IN DOUBT–"LOOK IT UP" IN
The
Encyclopaedia Britannica
The Sum of Human Knowledge
29 volumes, 28,150 pages, 44,000,000 words of text. Printed on thin, but strong opaque India paper, each volume but one inch in thickness.
(New 11th Edition) Issued 1910-11 by the CAMBRIDGE UNIVERSITY PRESS (England)
THE BOOK TO ASK QUESTIONS OF
FOR READING OR FOR STUDY

Fig. 2 U.S. Advertisement of *Encyclopaedia Britannica*'s 11th edition (1913)

The three-volume set, completed in 1771, was quickly sold out. Encouraged by this success, the publishers issued the second edition in ten volumes (1777-84).

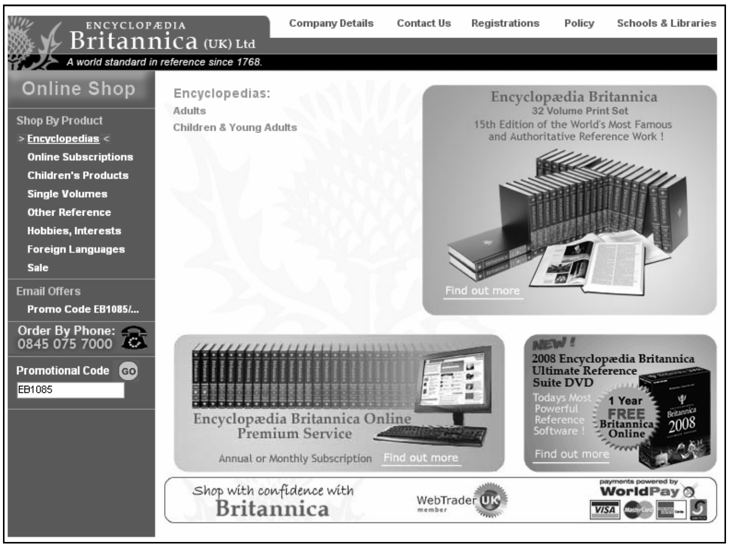

ENCYCLOPÆDIA Britannica (UK) Ltd
A world standard in reference since 1768.
Company Details
Contact Us
Registrations
Policy
Schools & Libraries
Online Shop
Shop By Product
> Encyclopedias <
Online Subscriptions
Children's Products
Single Volumes
Other Reference
Hobbies, Interests
Foreign Languages
Sale
Email Offers
Promo Code EB1085/...
Order By Phone: 0845 075 7000
Promotional Code GO
EB1085
Encyclopedias:
Adults
Children & Young Adults
Encyclopædia Britannica
32 Volume Print Set
15th Edition of the World's Most Famous and Authoritative Reference Work !
Find out more
Encyclopædia Britannica Online Premium Service
Annual or Monthly Subscription
Find out more
NEW !
2008 Encyclopædia Britannica Ultimate Reference Suite DVD
Todays Most Powerful Reference Software !
1 Year FREE Britannica Online
Find out more
Shop with confidence with Britannica
WebTrader UK member
payments powered by WorldPay

Fig. 3 Online advertisement of Encyclopaedia Britannica's products

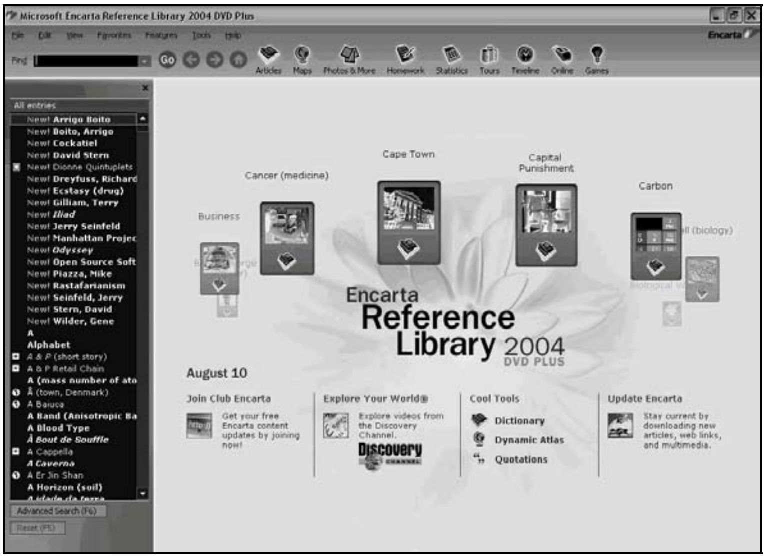
Its rising stature helped in recruiting eminent contributors and, both the 9th edition (1875-1889) and the 11th edition (1911) are regarded as landmark encyclopaedias for scholarship and literary style. In 1901, the encyclopaedia was purchased by the American publishers Horace Hooper and Walter Jackson and in 1920 Britannica was bought by Sears, Roebuck and Co., retaining Horace Hooper as its publisher (Wikipedia, 2006).

Beginning with the 11th edition, the *Britannica* gradually shortened and simplified its articles to make them more accessible and to expand its North American market. In 1933, it became the first encyclopaedia to adopt a continuous revision policy, in which the complete work is continually reprinted and every article is updated on a regular schedule.

In today's multimedia and wired up world, Britannica, for two centuries the undisputed repository of all human knowledge, being heavily engaged by other competitors, has had to come to terms with the advent of home computers in the 1990s making a tremendous effort to become one of the prime resources of the new Information Age. As a result, Encyclopaedia Britannica nowadays is published in paper form (32 volumes containing 65,000 articles), on cd-rom or dvd-rom and online (about 100,000 articles) (fig. 3). Brief article summaries can be read for free on the net, while the full text is available only for monthly or yearly paying individual subscribers.

3.3 Electronic Encyclopaedias

The advent of home computers has hopelessly undermined shelf-load encyclopaedias and door-to-door encyclopaedias' salespeople have become extinct as working class.



Get your free Encarta content updates by joining now!

Explore videos from the Discovery Channel.

Stay current by downloading new articles, web links, and multimedia.

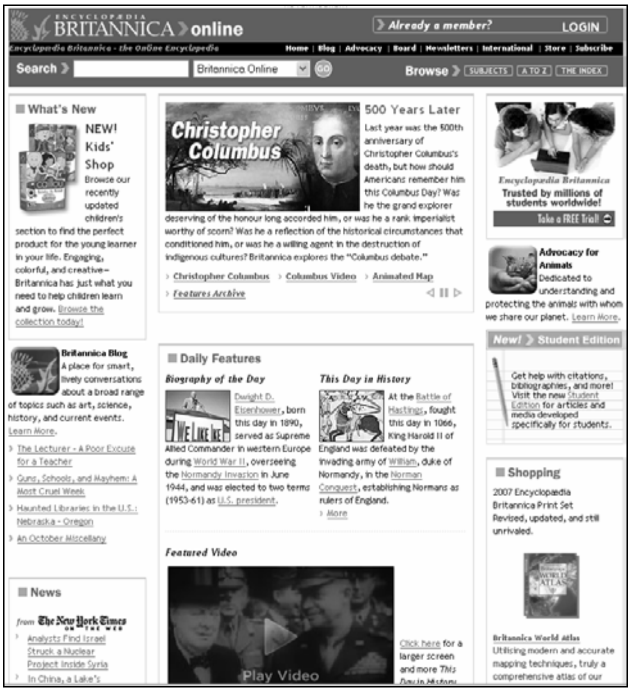
Fig. 4 *Encarta* Visual Browser (cd rom version)

Encyclopaedias published as multivolume sets of books for centuries, have been transformed in the 1990s into inexpensive cd roms or dvds integrating sound, pictures, animation and text.

In 1989, the first encyclopaedia in cd rom *Compton's Multimedia Encyclopaedia* was produced by Grolier. In 1993 the Microsoft Corporation released *Encarta Encyclopaedia,* the first general multimedia encyclopaedia on cd rom without a corresponding printed version. The first electronic version of *Encyclopædia Britannica* was also published in 1993, while *Canadian Encyclopaedia* appeared on cd rom in 1996.

In December 1997 *Encarta* became the first encyclopaedia to be published in dvd format. Dvds, storing much more information than cd roms, allowed greater use of complex multimedia features such as videos, animations and interactivities (Britannica, 2007).

Nowadays *Britannica* and *Encarta,* the two leading encyclopaedias, have captured the market. In the late 1990s, network technologies and the popularisation of the World Wide Web further provoked the evolution of encyclopaedias.


ENCYCLOPÆDIA BRITANNICA online
Already a member? LOGIN
Encyclopædia Britannica - the Online Encyclopedia
Home | Blog | Advocacy | Board | Newsletters | International | Store | Subscribe
Search
Britannica Online
GO
Browse
SUBJECTS
A TO Z
THE INDEX
Christopher Columbus
Take a FREE Trial!
New! Student Edition
Play Video
from The New York Times on the web

What's New

NEW! Kids' Shop Browse our recently updated children's section to find the perfect product for the young learner in your life. Engaging, colorful, and creative–Britannica has just what you need to help children learn and grow. Browse the collection today!

500 Years Later

Last year was the 500th anniversary of Christopher Columbus's death, but how should Americans remember him this Columbus Day? Was he the grand explorer deserving of the honour long accorded him, or was he a rank imperialist worthy of scorn? Was he a reflection of the historical circumstances that conditioned him, or was he a willing agent in the destruction of indigenous cultures? Britannica explores the "Columbus debate."

› Christopher Columbus › Columbus Video › Animated Map

› Features Archive

Encyclopædia Britannica

Trusted by millions of students worldwide!

Advocacy for Animals

Dedicated to understanding and protecting the animals with whom we share our planet. Learn More.

Britannica Blog

A place for smart, lively conversations about a broad range of topics such as art, science, history, and current events. Learn More.

› The Lecturer - A Poor Excuse for a Teacher

› Guns, Schools, and Mayhem: A Most Cruel Week

› Haunted Libraries in the U.S.: Nebraska - Oregon

› An October Miscellany

Daily Features

Biography of the Day

Dwight D. Eisenhower, born this day in 1890, served as Supreme Allied Commander in western Europe during World War II, overseeing the Normandy Invasion in June 1944, and was elected to two terms (1953-61) as U.S. president.

This Day in History

At the Battle of Hastings, fought this day in 1066, King Harold II of England was defeated by the invading army of William, duke of Normandy, in the Norman Conquest, establishing Normans as rulers of England.

› More

Get help with citations, bibliographies, and more! Visit the new Student Edition for articles and media developed specifically for students.

Shopping

2007 Encyclopædia Britannica Print Set Revised, updated, and still unrivaled.

Britannica World Atlas

Utilising modern and accurate mapping techniques, truly a comprehensive atlas of our

Featured Video

Click here for a larger screen and more This Day in History

News

› Analysts Find Israel Struck a Nuclear Project Inside Syria

› In China, a Lake's

Fig. 5 *Encyclopaedia Britannica* home page
http://www.britannica.com

The online versions normally include all the entries of the print and cd rom versions, as well as multimedia. They offer the advantage of freeing readers from the installation of cd roms or dvds. In the year 2000, the most important North American encyclopaedias, including *Encyclopædia Britannica, Encarta,* the *New Book of Knowledge* published by Groliers, and the *World Book Encyclopaedia* were accessible online (Britannica, 2007).

Although portals, search engines and web directories have progressively transformed the ways people search for information on the Web, that did not make encyclopaedias obsolete. On the contrary, reference works are needed more than ever to help the search and the filtering in the jungle of the information overload. Nowadays *Britannica* and *Encarta* dominate in the battle by promoting the value of authorial quality and editorial selectivity through new encyclopaedic products and services. Nonetheless, the latest strategies of *Britannica* and *Encarta* suggest more emphasis on a business model and less interest in the learning needs and demands of consumers (Panagiota, 2002).

3.4 The triumph of Online Encyclopaedias

In the early days of the World Wide Web, a number of writers predicted that the popularization of the Internet would lead to the death of every kind of printed publications. After several years that prediction doe not seem to have become reality. In fact electronic publications have not killed their printed counterparts since readers tend to treat printed and electronic versions as complimentary and not as competitors. An important exception is represented by encyclopaedias (Soojung-Kim Pang, 2000) as their hierarchical structure, or alphabetical arrangement, with their evolving nature is particularly adaptable to a disk-based or on-line computer format. These factors have caused the decline in popularity of printed encyclopaedias.

The advantages brought by online encyclopaedias are summarized below.

Low costs

Manufacturing costs have driven the development of electronic encyclopaedias. They have transformed the market in which encyclopaedia companies conducted their business. All major printed encyclopaedias have moved to the online method of delivery since it offers the advantage of being cheaply produced and can be consulted online from everywhere. Furthermore, freed from the expense of printing and binding more volumes, nowadays online encyclopaedias can offer a higher number of articles than their previous printed versions.

Multimediality

The most obvious advantage of online encyclopaedias is in their multimedia capabilities. Animated graphics, sound and video recordings have supplemented the text, photographs and drawings inherited from the printed medium. In this way, multimedia have enriched the content and the effectiveness of encyclopaedia’s pedagogical function.

Hypertextuality

Nowadays, online encyclopaedias make use of hypertext cross-references. The character of electronic texts encourages greater attention to content interconnection rather than to distinct individual articles. Unlike print, where encyclopaedic entries were essentially autonomous and self-contained objects, articles published online are joined to their kin by hyperlinks, which tend to encourage movement through interconnected texts. Thus, encyclopaedias are not a static and colossal collection of universal knowledge in one closed space but, as Neurath (1938) claims: *a living being and not a phantom, not a mausoleum or an herbarium, but a living intellectual force* or, as Selcer (2007) argues: *a vast, waving horizon, a net of multidimensional elements which can be connected according to multiple relationships*. In this dimension encyclopaedias convey a profound continuity of the unity of science, underlying its superficial discontinuity (Pombo *et al.*, 2006).

Dinamicity and up-to-date information

It is accepted that a printed volume slowly becomes obsolete. Articles and encyclopaedic volumes were once closed systems, since the opening of a page to introduce changes was economically very disadvantageous and if new articles or pictures were added, something had to be cut. Online encyclopaedias offer the advantage of being dynamic. Unlike paper or disk publications, new and frequently updated information can be presented almost immediately online, rather than waiting for the next release of a static format. Technology has transformed the nature of information which is now more up-to-date, temporary and permanently in progress. This is one of its main appeals (Soojung-Kim Pang, 2000).

Searching and Indexing

Electronic media offer previously unimaginable capabilities for searching and indexing as interactivity allows multiple methods of organization and retrievial of the same content. Articles are more accessible since, in addition to the alphabetical indexes compiled by editors for the print sets, online encyclopaedias employ high-speed search software that can retrieve an exhaustive set of files from their databases in response to specific queries.

Content specialization

The economic constraints imposed by the physical nature of the printed page, is no longer a concern. Early printed encyclopaedias required editors to be generalists, since a page could have anything on it. Thus, much of the work of producing an annual revision consisted of the craft-work of eliminating articles, words and counting lines, and rephrasing sentences to save (or add) a line or two.

Nowadays, the length of articles can reflect the importance of the subject rather than the space available on the page. Editors no longer need to be generalists, but they can be specialists working on articles that are not close to one another on a page but related by subject.

Reliability

Information reliability is another very important feature. Readers of printed encyclopaedias were less likely to notice incoherencies in the information provided. The physical separation of related articles in separate pages and often in separate volumes, made it more difficult to notice variations. Nowadays, thanks to the advantages offered by hyperlinking, contradictory information is just a mouse click away.

Author's commitment

Another emerging change relates to the new model of author-editor relationship which is shifting from one characterized by short periods of intense contact to one in which authors provide a continuous service, and from one that is focused only on individual writing to one defined by the sharing of expertise (Soojung-Kim Pang, 2000). Online encyclopaedias require people with different backgrounds, skills and interests to work closely together (e.g. designers, artists, authors, programmers, etc.). The idea of a finished article is obsolete since information is much more fluid and dynamic, with nothing fixed at the outset. In the past, keeping in constant touch with authors was not a high priority, since once an article was published, it might not be handled again for decades. Nowadays author's importance has become higher in electronic publishing and its presence more continuous (Soojung-Kim Pang, 2000). These developments carry out the predictions of many academic theorists who argue that hypertexts problematize the concept of the author. The freedom from print constraints has greatly affected editorial work from the intellectual skills required of editors, to the editing process itself. As will be shown in the next sections, the relationship with writing and reading has undergone deep changes which the specific case of Wikipedia testifies.

4. Expository Style of Encyclopaedias

Expository writing is a mode of writing in which the purpose of the author is to inform, explain, describe, or define the subject to the reader. According to Ball (1991) a well-written presentation remains focused on its topic and provides facts in order to inform its reader. It should be unbiased, accurate, and should use a scholarly third person tone. The text needs to encompass all aspects of the subject. Examples of expository writing can be found not only in encyclopaedias but in many other kind of informative writing such as magazine and newspaper articles, non-fiction books, travel brochures, business reports, memorandums, professional journal, etc.

In the creation of expository texts, writers cannot assume that readers have prior knowledge or former understanding of the topic that will be discussed. An important point authors have always to keep in mind is to use words that clearly show what he/she is talking about. Since clarity requires a strong organization, one of the most important mechanisms used to improve facts' presentation is to give the text a precise structure. Ball (1992) suggests four points:

Definition - Defining topics and subjects is particularly important in expository writing.

Description - Writing which intends to describe a person, place or thing is known as descriptive writing and is a form of expository writing.

Sequence - This structure is a form of expository writing that is used if the author intends to inform his or her readers by listing the order of steps in a process or listing events in chronological order.

Classification - It is an organizational strategy in which authors arrange groups of objects or ideas according to a common topic in detail. Placing different objects or ideas in categories is a type of classification.

The main peculiarity of the expository style used in the encyclopaedic genre is generally its formality, objectivity and impersonality. The voice of the author(s) disappears behind the presentation of facts and information. Articles are written in the third person, are unsigned, highly informational and abstract in content, explicit and context independent. Furthermore, as the central purpose of encyclopaedias is pedagogical, the degree of readability of the entries appears to be generally very high. Inviting readers to follow its own cursus, encyclopaedia is not a student' s manual. Its readers are an already lettered public, a *publique éclairé,* as Diderot and D'Alembert say, *a curious and intelligent reader* as stated in the *Preface* of the Britannica (Pombo *et al.*, 2006).

4.1 Stylistic Formality

Formality has been considered by many researchers the most important variation between styles or registers. Heylighen and Dewaele (1999:1) subdivide it in *deep formality* and *surface formality*. They define *deep formality* as:

> avoidance of ambiguity by minimizing the context dependence and fuzziness of expressions as unambiguous, context independent and without fuzzy expressions. This is achieved by explicit and precise description of the elements of the context needed to disambiguate the expression. A formal style is characterized by detachment, accuracy, rigidity and heaviness; an informal style is more flexible, direct, implicit and involved, but less informative.

The underlying assumption of most approaches is that a formal style is characterized by a special attention to form. The *Dictionary of language Teaching and Applied linguistics* (Richards *et al.* 1997: 144) defines formality as *the type of writing used in situations when the writer is very careful about choice of words and sentence structure*.

4.2 Decontextualization

One of the main features of encyclopaedic style, strictly associated to formality, is its context independence. Heylighen and Dewaele (1999:5) observe:

> Formality try to avoid ambiguity by including the information about the context that would disambiguate the expression into the expression itself, that is to say, by explicitly stating the necessary references, assumptions, and background knowledge which would have remained tacit in an informal expression of the same meaning. What really differentiate a formal style is that it achieves the same clarity without unstated assumptions.

By contrast, the language production is context-dependent when it is anchored to a spatio-temporal context to be meaningful and understandable; such anchor is called *deixis*. If this *anchor* does not exist information must be inferred from unstated background assumptions, or make reference to information expressed earlier (such as in the case of *anaphora*). In most written genres, and specifically in online encyclopaedias, what is written at one time and place is usually read at a different time and place by a multitude of people all over the world.

Consequently many deictic terms which refer to temporal or spatial contiguity should be less frequent in formal (offline/online) texts. Written texts and encyclopaedic entries should be more decontextualised with respect to the physical settings. Of course contextual assumptions, as will be further shown, are not absent from written discourse, but their occurrence is minimal when compared to face to face communication.

It stands to reason that the style used on the web has to be decontextualised as webpages, more than other genres, can be browsed by people speaking different languages, having different cultures and backgrounds and coming from all over the globe. It cannot be assumed that netsurfers share the specific high context culture with encyclopaedia's contributors, hence, it is essential that a context dependent terminology and ambiguous expressions be avoided and carefully replaced by explicit words fully comprehensible by the world wide web audience.

According to Heylighen and Heylighen (1999) the formal style is context independent, precise and not fuzzy and it allows a clear understandability which does not vary despite changes of reading context. Thus, formal writing is detached and impersonal. Its primary purpose is to transmit information and the words acquire an existence somewhat separate from their source. In the words of Florian Coulmas, this is the *reifying function* of writing which is particularly true of expository texts.

4.3 Exactness and Accuracy

Givon (1983) proposed that any discourse may be put along a continuum between two poles: the *syntactic mode* that is explicit, decontextualised, precise and stereotypically represented by expository prose and the *pragmatic mode* which is contextualized and loosely-structured.

According to Chafe (1987) a characteristic of formality is explicitness, that is avoidance of ambiguity. In most cases written discourse is unambiguous, it does not make use of fuzzy terms, and the exactness of writing is reflected in the avoidance of generalizations as well as in a greater use of deductive reasoning and supportive evidence. Formal style minimizes ambiguity by avoiding fuzzy and context dependent expressions. Fuzziness is avoided using precise and unequivocal expressions and providing information which has to be understandable independently of the original context of production.

Authors of encyclopaedic entries have no way of knowing who the potential readers will be and, thus, they can assume very little about them. This pushes the writer to word things very explicitly in order to be understood by any reader. Nevertheless, a totally unambiguous description is impossible also in formal style as a margin of indeterminacy is always foundable. Heylighen and Dewaele (1999:9) underline that an element of indeterminacy always remains, and a completely unambiguous description is practically impossible. They claim that the basic advantage of formality, which follows from their definition, is that:

> More formal messages have less chance to be misinterpreted by others who do not share the same context as the sender. This is clearly exemplified by written language, where there is no direct contact between sender and receiver, and hence a much smaller sharing of context than in speech […]. The definition also implies that validity or comprehensibility of formal messages will extend over wider contexts: (more people, longer time spans, more diverse circumstances, etc). This makes it easier for formally expressed knowledge to maintain and spread over many different persons, groups or cultures.

The concurrent disadvantage of invariance over contexts is that formal speech is more static and rigid, structurally more complex and not flexible. Therefore, formal style requires more time, attention and cognitive process to be produced and understood. Givon (1983: 1018) observes that:

> the absence of context forces the language user to code the necessary presuppositions within the message. The resulting 'syntactic mode' of expression involves a higher use of nouns that requires more lexical searching because of their relatively infrequent use.

By distancing themselves from the immediate context, formal texts will also be less direct than informal and involved texts. The latter, to imply meanings, can rely on a communicative and cultural context co-shared by participants.

4.4 Space, Time and Audience

According to Heylighen and Dewaele another implication of formality concerns audience size. This is a significant aspect to be considered also in web genres, since Internet readers come from all over the globe. Heylighen and Dewaele (1999: 25) point out:

> […] the larger the audience, in general, the more important it will be to secure accurate understanding. It is expected that speeches or texts directed to a large audience will be more formal than comments addressed to one or a few persons. This is confirmed by the higher formality score of speeches compared to conversations and publish texts compared to letters.

Heylighen and Dewaele also underline that formal textual production is strictly correlated to the concepts of space and time. The wider the spatial setting between sender and receiver is, the smaller the shared context will be, the higher the formality of the text produced. The same will happen when the time span between sending and receiving is long. In this case, the less will remain of the original context in which the discourse has been produced, the more an explicit, precise and context independent textual production will be needed. In conclusion, the above mentioned variables will influence the formality and the comprehensibility of encyclopaedic texts.

As shown in fig. 6, audience size, different writers/readers cultural background, settings of production and reception, time span and the need for understanding are factors which have to be absolutely taken into account if online encyclopaedias want to be effective, comprehensible and fulfil their main educational purpose.

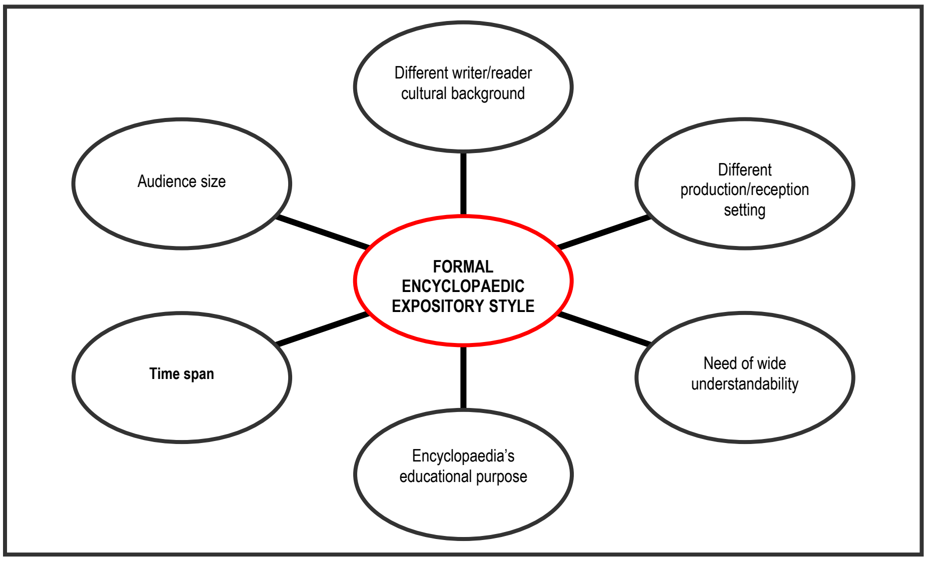

Different writer/reader cultural background
Audience size
Different production/reception setting
FORMAL ENCYCLOPAEDIC EXPOSITORY STYLE
Time span
Need of wide understandability
Encyclopaedia's educational purpose

Fig. 6 Formal encyclopaedic Expository Style

2. WIKIPEDIA AND WEB 2.0

Traditional encyclopaedias are written by a number of employed text writers, usually people with an academic degree, but the interactive nature of the Internet and the development of Web 2.0 has given birth to new collaborative projects such as *Nupedia, Everything2, Open Site, Wikipedia,* etc., which share the characteristic of being online alternatives to proprietary encyclopaedias. Nowadays these encyclopaedias can be seen as a collection of verbal and visual information arranged into a huge repository of hierarchical and associative lexias (Landow, 1997) with an hypertextual macrostructure combining a traditional and innovative approach to reading and writing (Elia, in press). Since the beginning of the new millennium Wikipedia, represents one of the online phenomenon more often under the spotlights. It is a freely available Web-based free-content co-authored encyclopaedia. It is a multilingual encyclopaedic project, operated by the *Wikimedia Foundation* [6] (Sloane, 2007). The name Wikipedia is a blended word made up of *wiki* (a type of collaborative website) and *encyclopaedia* (Wikipedia, 2006). A description of Wikipedia, which follows in the next sections, is crucial to the understanding of this new web emerging phenomenon.

1. Wikipedia: A General Overview

Wikipedia's English edition was launched by its co-founders Jimmy Wales and Larry Sanger on 15th January, 2001 as a complement to *Nupedia*[7], an English-language web-based encyclopaedia whose articles were written by experts and licensed as free content. Wales instead was the only creator of the *Wikimedia Foundation* in 2003. As of October 2007, Wikipedia with approximately more than eight million and half articles in 253 languages has been officially recognized as the largest international virtual community. The English edition being made up of 2,045,000 articles is the largest edition and it will very probably remain so in the future.

Looking at the recent statistics on the number of articles, the English edition of Wikipedia is over 20 times larger than Britannica's (Wikipedia, 2006). A key difference between the two encyclopaedias lies in article authorship. Britannica's articles are generally written by recognized

[6] *Wikimedia Foundation's* goals are to develop and maintain wiki-based projects and to freely provide their contents to the public. In addition to the multilingual general encyclopaedia *Wikipedia*, there is *Wiktionary* a multi-language dictionary, *Wikiquote* an encyclopaedia of quotations, *Wikisource* a repository of source texts in any language, and *Wikibooks* a collection of e-books for students.

[7] *Nupedia* was founded by Jimmy Wales with Larry Sanger as editor-in-chief (Marshall, 2006). *Nupedia* mostly known now as the predecessor of *Wikipedia*, lasted from March 2000 until September 2003. It was a Web-based encyclopaedia whose articles were written by experts and licensed as free content. It was characterized by an extensive peer-review process designed to make articles of a quality comparable to that of professional encyclopaedias. Nevertheless, it was not wiki based, and not publicly editable.

contributors, and are the product of an editorial staff and internal or external consultants. Most of Britannica's contributors are experts in their field and some of them are also Nobel laureates.

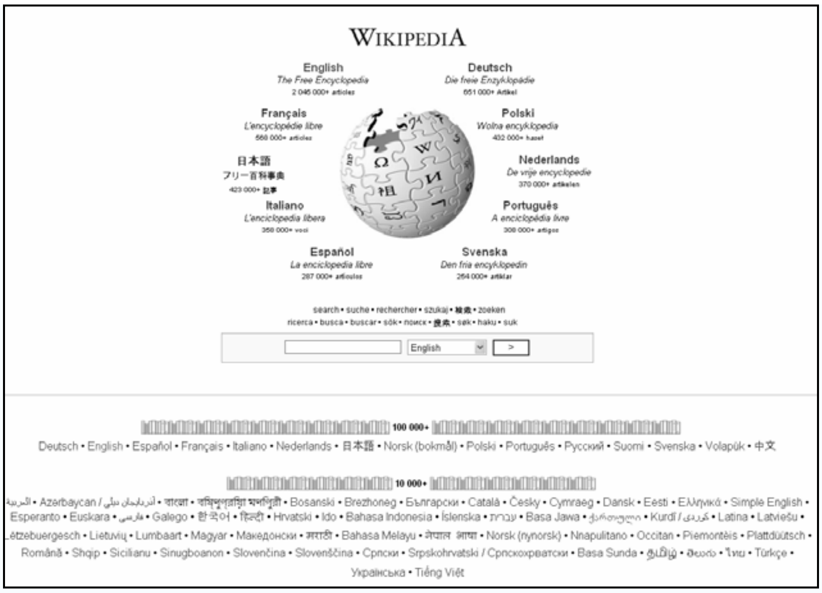

WIKIPEDIA
English
The Free Encyclopedia
2 046 000+ articles
Deutsch
Die freie Enzyklopädie
651 000+ Artikel
Français
L'encyclopédie libre
568 000+ articles
Polski
Wolna encyklopedia
432 000+ haseł
日本語
フリー百科事典
423 000+ 記事
Nederlands
De vrije encyclopedie
370 000+ artikelen
Italiano
L'enciclopedia libera
358 000+ voci
Português
A enciclopédia livre
308 000+ artigos
Español
La enciclopedia libre
287 000+ artículos
Svenska
Den fria encyklopedin
264 000+ artiklar
search • suche • rechercher • szukaj • 検索 • zoeken
ricerca • busca • buscar • sök • поиск • 搜索 • søk • haku • suk
English
>
100 000+
Deutsch • English • Español • Français • Italiano • Nederlands • 日本語 • Norsk (bokmål) • Polski • Português • Русский • Suomi • Svenska • Volapük • 中文
10 000+
العربية • Azərbaycan / آذربايجان ديلی • বাংলা • বিষ্ণুপ্রিয়া মণিপুরী • Bosanski • Brezhoneg • Български • Català • Česky • Cymraeg • Dansk • Eesti • Ελληνικά • Simple English •
Esperanto • Euskara • فارسی • Galego • 한국어 • हिन्दी • Hrvatski • Ido • Bahasa Indonesia • Íslenska • עברית • Basa Jawa • ქართული • Kurdî / كوردی • Latina • Latviešu •
Lëtzebuergesch • Lietuvių • Lumbaart • Magyar • Македонски • मराठी • Bahasa Melayu • नेपाल भाषा • Norsk (nynorsk) • Nnapulitano • Occitan • Piemontèis • Plattdüütsch •
Română • Shqip • Sicilianu • Sinugboanon • Slovenčina • Slovenščina • Српски • Srpskohrvatski / Српскохрватски • Basa Sunda • தமிழ் • తెలుగు • ไทย • Türkçe •
Українська • Tiếng Việt

Fig. 1 Wikipedia homepage
http://www.wikipedia.org

By contrast, the articles in Wikipedia are written by a community of editors with different levels of expertise: most editors do not claim any particular expertise;and many of them are anonymous and have no verifiable credentials. For this reason it has been argued (McHenry R., 2004) that Wikipedia cannot hope to compete with Britannica in accuracy.

Wikipedia relies on the authority of peer-reviewed publications rather than on the personal authority of experts. It does not force its contributors to give their names to establish their identity. Although some contributors are authorities in their field, Wikipedia only requires that information provided is supported by published and verifiable sources.

According to the statistics gathered by *Alexa* [8], Wikipedia in the first three months of 2007 is ranked among the first ten most clicked urls on the web, thus it can be considered one of the most popular reference websites. With around 50 million hits per day, it receives roughly 450 times more traffic than the online version of the *Britannica.* When *"YOU*" was awarded by the *Time Magazine* as the person of the Year in 2006 (you as user, creator and collaborator in all the community activities), this praise accelerated the success of online collaboration and interaction. Wikipedia was the first Web 2.0 service to be mentioned, followed by *YouTube* and *MySpace* (Grossman, 2006).

[8] *Alexa* http://www.alexa.com/

As Wikipedia is an open online authoring environment, anyone can add or improve text, images and sounds as contents are licensed under a free copyleft license, the *GFDL* (GNU Free Documentation License)[9].

Wikipedia's growth has been exponential in several of the major language editions. Its five largest editions are, in descending order, English, German, French, Polish, Japanese and Italian. Every language edition operates independently and translated articles represent only a small part of any edition[10].

Wikipedia has been described by its founder Jimmy Wales as *an effort to create and distribute a free encyclopaedia of the highest possible quality to every single person on the planet and in their own language* (Wales, 2005). It never considers any articles finished as they are subject to an everlasting editing process. Any visitor may edit Wikipedia's articles as a volunteer author and have their changes immediately displayed as wiki authorship is characterized by gradual and repetitive additions, or deletions of content over time. Wikipedia is not a form of one-way communication since, unlike other media, it has a strong collaborative imprinting. This phenomenon develops a sense of collective purpose and responsibility in the virtual community which further motivates public participation. People who write and edit articles for Wikipedia are defined as *Wikipedians*.

Wikipedia is built on the belief that cooperation among Wikipedians, thanks to the social software, will improve articles over time. Articles seem to become constantly better as contributors go back again and again to old articles adding new information, rewording ambiguous statements, correcting mistakes, etc. This means that over the years, the quality of the articles tends to improve, both in quality and accuracy. To paraphrase Linus Torvalds [11] *Given enough eyeballs, all typos factual errors and other errors of content are shallow* (Sanger, 2001). Wiki community's members define such a peculiar editing process as a collaborative work of art, a sort of *Darwinian-like evolutionary process* or an adversarial *battlefield of ideas* (Wikipedia, 2007).

Every contributor is intended to be of equal status when editing articles. The editing process is not controlled by any particular editorial group. However, maintenance tasks are performed by a group of volunteer administrators (*SysOps*) who, in accordance with the community policy, have the privilege of preventing articles from being edited or deleted.

Vandalism, which consists of a bad-faith addition, change or deletion, deliberately made to invalidate the encyclopaedia's integrity, is a problem for Wikipedia. Most acts of vandalism consist of replacing articles with obscenities or irrelevant content. Less important infractions may determine a

[9] The *GNU Free Documentation License* (GFDL) is a license for free content, designed by the Free Software Foundation (FSF) for the GNU project. The license stipulates that any copyleft of the material, even if modified, carries the same license. A copyleft license uses copyright law in order to ensure that every person who receives a copy, or derived version of a work, can use, modify, and also redistribute both the work and derived versions of the work. *Copyleft* is the opposite of *copyright*. Wikipedia is the largest documentation project to use this license.

[10] This is the list of major editions based on the number of articles up to October 2007: English (2,045,000), German (651,000), French (568,000), Polish (432,000), Japanese (423,000), Italian (358,000), Dutch (370,000), Portuguese (308,000), Spanish (287,000) and Swedish (254,000).

[11] Linus Benedict Torvalds (1969) is the original developer of Linux operating system.

temporary block, while long-term or permanent blocks caused by prolonged and serious infraction are given by an *Arbitration Committee* (ArbCom) which has the power of temporarily or permanently blocking users from editing.

2. The Literature on Wikipedia

The development of Wikipedia as a new web phenomen has recently attracted the attention of many scholars coming from different research areas: computing, sociology, linguistics, etc. Their positions are very different, since many criticisms as well as praises have been raised.

Wikipedia has been blamed for deficiencies in comprehensiveness because of its voluntary nature, for reflecting the systemic biases of its contributors and for inconsistency (Waldman, 2004).

Further critics argue that Wikipedia's open nature and lack of proper sources, for much of the information, make it unreliable (Schiff, 2006). Others suggest that Wikipedia is reliable most of the times, but it is not always clear to what extent is (Boyd, 2005). Editors of traditional reference works such as *Encyclopædia Britannica* have contested the project's utility and status as an encyclopaedia (McHenry, 2004). Concerns have also been raised on the lack of accountability resulting from users' anonymity, the vulnerability to vandalism and so forth. Other critics claim that Wikipedia's open structure makes it an easy target for advertisers (Sanger, 2006), (Torsten, 2005). Ahrens (2006) has noted the addition of news to articles by political organizations including the U.S. House of Representatives. The most visible and public criticism of *Wikipedia* has been conveyed by Lanier (2006) who criticizes *Wikipedia*'s growing importance in status and he sees this as a renaissance of the *idea that the collective is all-wise*. Lanier claims that the concept of *collective intelligence* can be a dangerous tool in the hands of any extreme ideology. Furthermore, it represents a risk for the future of individual minds as personal contributions will be lost in the *mare magnum* of the collective knowledge.

On the other hand, positive appreciations have been made in the article published by the journal *Nature* (Giles, 2006). Nature's scientists compared forty-two pairs of science articles from the Encyclopaedia Britannica Online to Wikipedia for factual errors, false statements and omissions and they discovered that the error rate among them was nearly the same. Experts found 162 errors in Wikipedia and 123 in Encyclopaedia Britannica. The results of the comparison were widely seen as a validation of Wikipedia's content and methods.

Lih (2004) studied Wikipedia's content construction and use processes from the perspective of participatory journalism. In addition to providing a rather comprehensive account of the Wikipedia project history, the author analyzed the change in the quality of Wikipedia articles before and after they had been cited in the press. Viegas et *al.* (2004) developed a tool for the *history flow-visualization*. This software allowed the analysis and display of the complex structure of the evolution of Wikipedia articles by visualizing the textual contributions of different authors at different times.

Resnick *et al.* (2005) highlighted the Wiki structure and its advantages in relation to other forms of online communication. Wikis with their new fundamental entity, *the editable node,* seem to establish a new form of editing pushing the boundaries of conventional online communication.

Other approaches stressed the productive power of Wiki discussions in the collaborative knowledge creation (Shah, 2005), (Lawler, 2005). In this perspective, the processes of contribution and discussion help to maintain a form of security that protects the data better than any other form of control.

Joseph Reagle (2006) explored the character of "mutual aid" and interdependent decision-making within Wikipedia. He focused on the wikiquette rules (e.g. good faith), which transforms community participation into a cooperative effort. He positively evaluated discussions in Wikipedia as tools which transform divergent into convergent controversy.

Holloway et *al.* (2006) reported a semantic analysis of *Wikipedia* covering a number of articles and categories of articles. The rapid growth of Wikipedia has also been a subject of this study. Capocci *et al.* (2006) for example, used social network modeling with Wikipedia to predict the growth patterns of Wikipedia. They found that this growth pattern is a *close analogy with that of the World Wide Web, despite the very different growth mechanism.*

Furthermore, discussion pages have been used to examine the information quality processes of Wikipedia articles. Stvilia *et al.* (2005) for example made use of article discussion pages to compile a list of ten information quality problems named by the authors such as: *accessibility, accuracy, authority, completeness, complexity, consistency, informativeness, relevance, verifiability and volatility.* On the basis of 60 randomly chosen articles they were able to show that discussions assured quality of information. Zlatic *et al*. (2006) performed a similar analysis but compared the linking between articles in different languages and found similarities pointing to a *unique growth process* across languages. Moreover, Pentzold and Seidenglanz (2006) explored the communicative functions in Wikipedia's community using Foucault' s discourse theory. They claim that discursive regularities named by Foucault lie in the Wikipedian collaborative writing process.

This doctoral thesis, as already claimed in the introduction, has been inspired by the article published by Hemigh and Herring (2005) where a linguistic comparison between two community-based encyclopaedias (*Wikipedia* and *Everything2*) and *Columbia Encyclopaedia* was presented. From their analysis they conclude that greater the degree of post-production editorial control afforded by the system, the more formal and standardized the language of the collaboratively-authored documents becomes. Their findings shed light on how users, acting through mechanisms provided by the system, can shape content in particular ways. The writing norms are constantly enforced through the permanent editing processes and the agency of socially-approved members (the *SysOps*) of the Wikipedia community.

3. What is Web 2.0 ?

In order to fully understand the Wikipedia phenomenon, a general overview of the web 2.0 revolution is provided in this section. The term Web 2.0 denotes a second generation of web-based communities and hosted services, such as social-networking sites, blogs, wikis and folksonomies[12], which aims at facilitating collaboration and sharing between users. The term Web 2.0 was coined by Tim O'Reilly, a guru in analysing macro trends in ICT and this definition became popular in 2004 after the first O'Reilly Media Web 2.0 conference (Graham, 2005). The core principle of this new web mode seems to lie in its participatory mechanisms. The Web becomes non hierarchical, democratic, open and non-authoritarian. On 30th September 2005, Tim O'Reilly wrote a paper summarizing the subject. The mind-map (Angermeier, 2005) in fig. 2 sums up the prompts of Web 2.0.

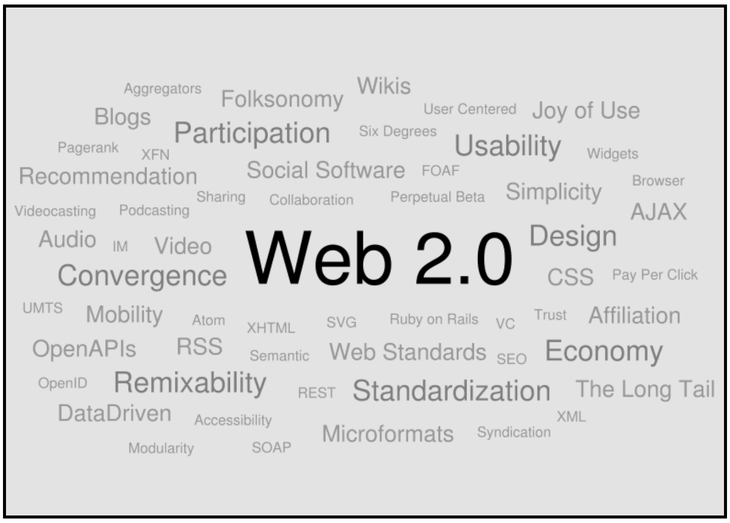


Fig. 2 Web 2.0 mind-map

Web 2.0 principles seem to be more related to political concepts than to Computer Science. This happens because there is an almost political principle at work in Web 2.0, that knowledge and information need to be free and not controlled, collectively created and mutually shared. Thus, the Internet becomes a platform to be shared with others (Prakash, 2007).

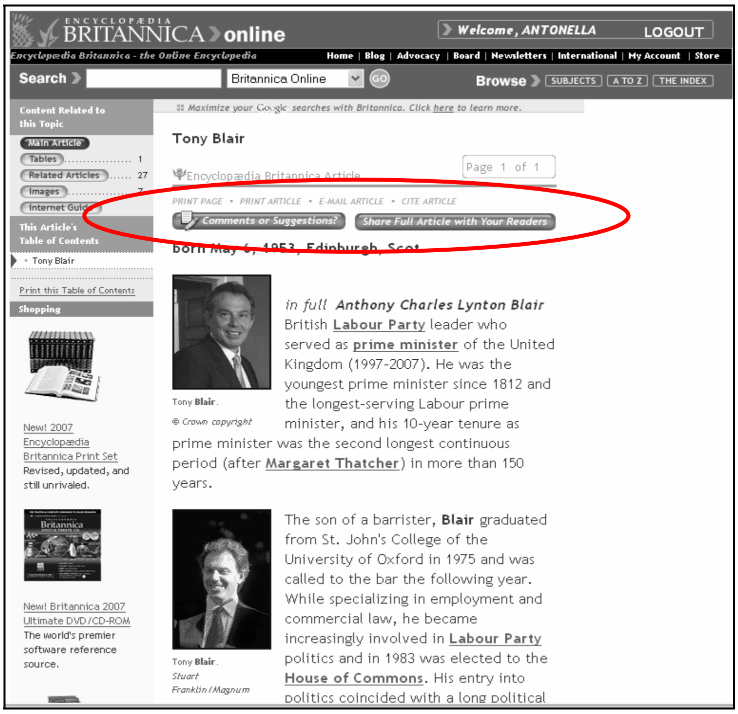
In a not too distant past, students and scholars went to dusty libraries to look at Encyclopaedia Britannica which was regarded as the fountainhead of all knowledge, as the last authorial word. When encyclopaedias appeared online in 1994 the same information could be accessed everywhere by laptops, but the nature of knowledge dissemination from an authoritative source continued and readers continued to be essentially passive participants.

[12] *Folksonomy* is a neologism which defines a practice of collaborative categorization using freely chosen keywords. It refers to a group of people cooperating spontaneously to organize information into categories. In other words *folksonomy* is a user generated taxonomy used to categorize and retrieve web content, using open-ended labels called *tags*. In contrast to formal classification methods, this phenomenon typically arises in non-hierarchical communities. The *folksonomic tagging* is intended to make a body of information increasingly easy to search, discover, and navigate over time. Apart from *Wikipedia*, two widely cited examples of websites using folksonomic tagging are *Flickr* (http://www.flickr.com) and *del.icio.us* (http://del.icio.us).

Then came Wikipedia in 2001. Its more significant difference from Britannica is that it is a collective project that can be built, revised, refined and changed by readers. Everyone has a voice in Web 2.0 applications and is free to participate, be it in *E- bay, Wikipedia*, *YouTube* or *Blogs* (Prakash, 2007). Nevertheless, the social and psychological ramifications of Web 2.0 are still to be seen and its implications are yet to be fully understood.

What is interesting to note is that although Britannica has never betrayed his proprietary origins, it has been clearly affected by Web 2.0 culture, being forced to introduce changes which have been personally defined as the product of the *wikification process*.

Since February 2007, a new button has been inserted at the top of each Britannica's article (fig. 3). Providing a *Comments or Suggestions* button (fig. 4), the chance has been given to Britannica readers to share their personal knowledge with a contributor panel which will evaluate the proposed content changes and will then publish all the approved variations on the original Britannica's article.


ENCYCLOPÆDIA BRITANNICA online
Welcome, ANTONELLA
LOGOUT
Encyclopædia Britannica - the Online Encyclopedia
Home | Blog | Advocacy | Board | Newsletters | International | My Account | Store
Search
Britannica Online
GO
Browse
SUBJECTS
A TO Z
THE INDEX
Content Related to this Topic
Main Article
Tables 1
Related Articles 27
Images 7
Internet Guid
This Article's Table of Contents
Tony Blair
Print this Table of Contents
Shopping
Page 1 of 1
Encyclopædia Britannica Article
PRINT PAGE • PRINT ARTICLE • E-MAIL ARTICLE • CITE ARTICLE
Comments or Suggestions?
Share Full Article with Your Readers

Maximize your Google searches with Britannica. Click here to learn more.

Tony Blair

born May 6, 1953, Edinburgh, Scot.

in full **Anthony Charles Lynton Blair**
British Labour Party leader who served as prime minister of the United Kingdom (1997-2007). He was the youngest prime minister since 1812 and the longest-serving Labour prime minister, and his 10-year tenure as prime minister was the second longest continuous period (after Margaret Thatcher) in more than 150 years.

Tony **Blair**.
© *Crown copyright*

The son of a barrister, **Blair** graduated from St. John's College of the University of Oxford in 1975 and was called to the bar the following year. While specializing in employment and commercial law, he became increasingly involved in Labour Party politics and in 1983 was elected to the House of Commons. His entry into politics coincided with a long political

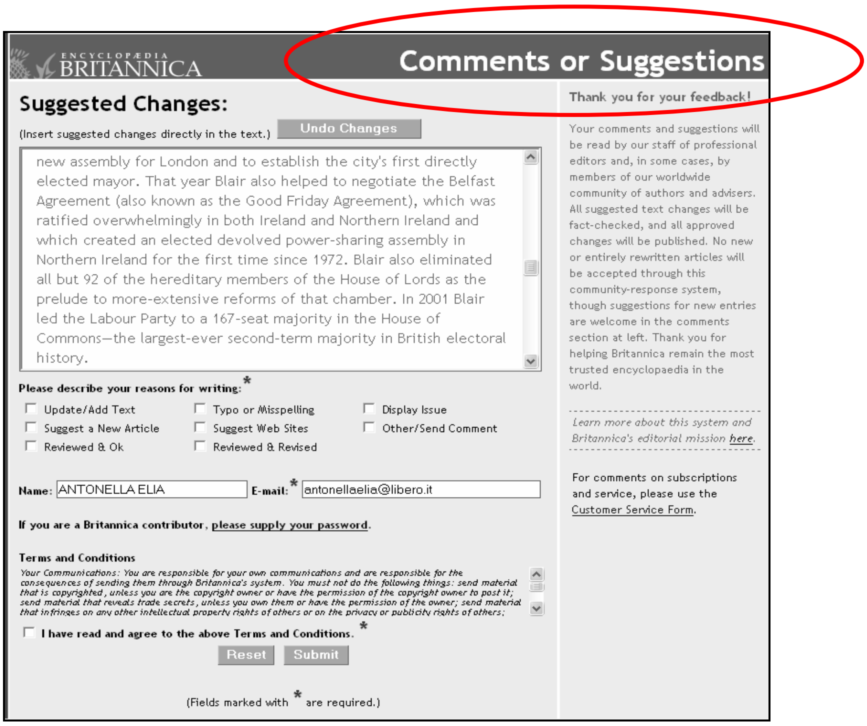
Tony **Blair**.
Stuart Franklin/Magnum

New! 2007 Encyclopædia Britannica Print Set
Revised, updated, and still unrivaled.

New! Britannica 2007 Ultimate DVD/CD-ROM
The world's premier software reference source.

Fig. 3 New buttons added on Britannica articles

In September 2007 a second button has been added on encyclopaedic articles. By clicking on *Share your articles with your reader* (fig. 5), Britannica subscribers can share the full text of an article with the readers of their personal website or blog, even if they are not subscribers to the online service.



Suggested Changes:

(Insert suggested changes directly in the text.) Undo Changes

new assembly for London and to establish the city's first directly elected mayor. That year Blair also helped to negotiate the Belfast Agreement (also known as the Good Friday Agreement), which was ratified overwhelmingly in both Ireland and Northern Ireland and which created an elected devolved power-sharing assembly in Northern Ireland for the first time since 1972. Blair also eliminated all but 92 of the hereditary members of the House of Lords as the prelude to more-extensive reforms of that chamber. In 2001 Blair led the Labour Party to a 167-seat majority in the House of Commons—the largest-ever second-term majority in British electoral history.

Please describe your reasons for writing:*

□ Update/Add Text □ Typo or Misspelling □ Display Issue
□ Suggest a New Article □ Suggest Web Sites □ Other/Send Comment
□ Reviewed & Ok □ Reviewed & Revised

Name: ANTONELLA ELIA E-mail:* antonellaelia@libero.it

If you are a Britannica contributor, please supply your password.

Terms and Conditions

Your Communications: You are responsible for your own communications and are responsible for the consequences of sending them through Britannica's system. You must not do the following things: send material that is copyrighted, unless you are the copyright owner or have the permission of the copyright owner to post it; send material that reveals trade secrets, unless you own them or have the permission of the owner; send material that infringes on any other intellectual property rights of others or on the privacy or publicity rights of others;

□ I have read and agree to the above Terms and Conditions.*

Reset Submit

(Fields marked with * are required.)

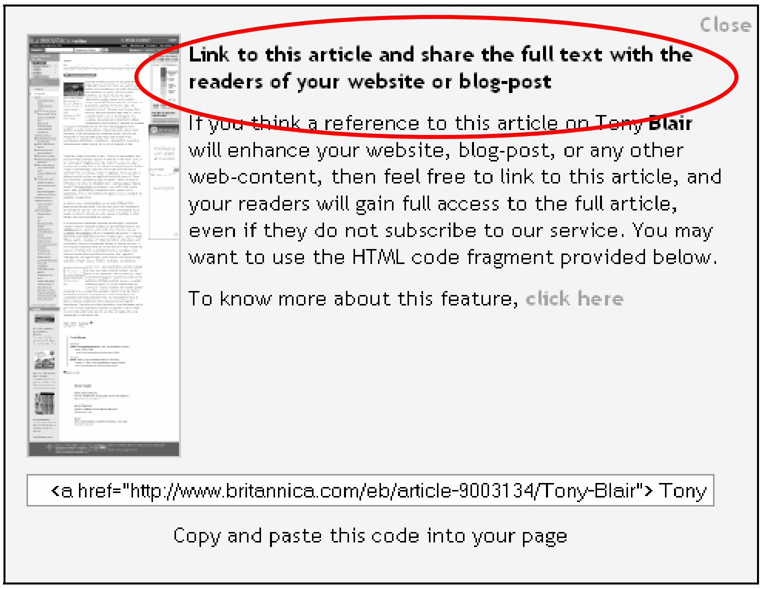
Thank you for your feedback!

Your comments and suggestions will be read by our staff of professional editors and, in some cases, by members of our worldwide community of authors and advisers. All suggested text changes will be fact-checked, and all approved changes will be published. No new or entirely rewritten articles will be accepted through this community-response system, though suggestions for new entries are welcome in the comments section at left. Thank you for helping Britannica remain the most trusted encyclopaedia in the world.

Learn more about this system and Britannica's editorial mission here.

For comments on subscriptions and service, please use the Customer Service Form.

Fig. 4 *Comments and suggestion* window
in Britannica Online



Link to this article and share the full text with the readers of your website or blog-post

If you think a reference to this article on **Tony Blair** will enhance your website, blog-post, or any other web-content, then feel free to link to this article, and your readers will gain full access to the full article, even if they do not subscribe to our service. You may want to use the HTML code fragment provided below.

To know more about this feature, click here

```
<a href="http://www.britannica.com/eb/article-9003134/Tony-Blair"> Tony
```

Copy and paste this code into your page

Fig. 5 *Share your article with your readers* window
in Britannica Online

Although the process is controlled, the variations made to the Britannica website clearly show to what extent Britannica, no longer indifferent to web cultural changes, had to include the audience and take into account Web 2.0 democratic values so as to preserve its own existence.

4. What is a Wiki?

Wiki is a virtual environment which is naturally suited for collaborative projects. While blogs can be highly personal, wikis are intensely communal (Read, 2005).

The *WikiWikiWeb* [13] founded by Ward Cunningham appeared on the web on 25th March 1995, two years before the birth of the first blog. Cunningham also invented the software name based on the Hawaiian term *wiki wiki*, meaning 'quick'. Moreover, he was the author with Leuf of the *Wiki Way* (2001), the first book dedicated to the subject.

Since the beginning of this new millennium, the use of *Peer to Peer* (P2P) technology has spread rapidly. A characteristic of wiki software is the ease with which pages can be created and updated. Most wikis are open to the general public without registration. Nevertheless, private wiki servers can sometimes be protected requiring user identification through login and password in order to be safeguarded against malicious behaviour (e.g. page deletion, vandalism and spamming).

While for years the standard was the *Wiki Markup Language* (original syntax of the *WikiWikiWeb)*, current formatting instructions vary considerably. Simple wikis allow only basic text formatting, whereas more complex ones support tables, images, formulas, and or even interactive elements.

Some wikis, such as Wikipedia, do not require the user to know wiki syntax, as they provide a *WYSIWYG*[14] editing, that translates graphically entered formatting instructions (e.g. **bold**, *italics, etc.*), into the corresponding html tags. Wiki software can be downloaded free of charge on the Internet. The reason for this is that the wiki code is available under the *GNU* General Public Licence, so the code is freely available to be reviewed and adjusted by developers. Wiki offers by default a search engine, and it is a true hypertext medium with non-linear navigational structure. This is made possible by *WikiLinks* which connect different wikipages. Links across different wiki communities are also possible using a special link pattern called *InterWiki*. New wiki pages are usually constructed by simply creating the appropriate links on a topically related page. A link opens an *edit* window, which allows users to enter the text for a new page. This simple editing mechanism generally ensures a high level of interlinking.

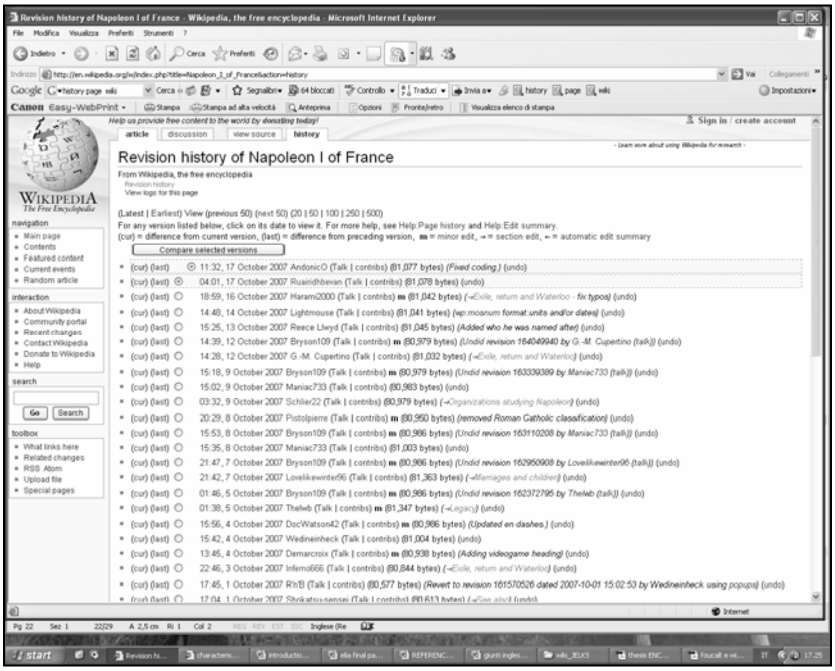
A further useful tool offered by the wiki software is the *history page* (fig. 6)*,* a space where all previous edited versions of articles and talk pages are listed with date, author and sometimes a comment. They are non-editable back up pages corresponding to past versions of wiki pages. Through this list all changes made to the page in reverse-chronological order are presented.

Wikis have two different writing modes and associated spaces. In the *document mode*, texts are collaboratively written by contributors, while in the second space, defined as *thread mode talk pages*,

[13] *WikiWikiWeb* http://c2.com/cgi/wiki

[14] *WYSIWYG* is the acronym for *What You See Is What You Get*. The term is used in computing to describe the likeness between the appearance of edited content and the final product.

contributors carry out discussions related to the subject of the main document [15] by posting signed messages (Cunningham & Leuf, 2001).


Revision history of Napoleon I of France - Wikipedia, the free encyclopedia - Microsoft Internet Explorer
File Modifica Visualizza Preferiti Strumenti ?
Indietro Cerca Preferiti
Indirizzo http://en.wikipedia.org/w/index.php?title=Napoleon_I_of_France&action=history
Vai Collegamenti
Google history page wiki Cerca Segnalibri 64 bloccati Controllo Traduci Invia a history page wiki Impostazioni
Canon Easy-WebPrint Stampa Stampa ad alta velocità Anteprima Opzioni Fronte/retro Visualizza elenco di stampa
Help us provide free content to the world by donating today!
Sign in / create account
article
discussion
view source
history
- Learn more about using Wikipedia for research -
WIKIPEDIA
The Free Encyclopedia
navigation
Main page
Contents
Featured content
Current events
Random article
interaction
About Wikipedia
Community portal
Recent changes
Contact Wikipedia
Donate to Wikipedia
Help
search
Go
Search
toolbox
What links here
Related changes
RSS Atom
Upload file
Special pages
Internet
Pg 22 Sez 1 22/29 A 2,5 cm Ri 1 Col 2 REG REV EST SSC Inglese (Re
start
Revision hi...
characteris...
introductio...
elia final pa...
REFERENC...
giunti ingles...
wiki_JELKS
thesis ENC...
foucalt e wi...
IT
17:25

Revision history of Napoleon I of France

From Wikipedia, the free encyclopedia

Revision history

View logs for this page

(Latest | Earliest) View (previous 50) (next 50) (20 | 50 | 100 | 250 | 500)

For any version listed below, click on its date to view it. For more help, see Help:Page history and Help:Edit summary.

(cur) = difference from current version, (last) = difference from preceding version, m = minor edit, → = section edit, ← = automatic edit summary

Compare selected versions

- (cur) (last) ◉ 11:32, 17 October 2007 AndonicO (Talk | contribs) (81,077 bytes) *(Fixed coding.)* (undo)
- (cur) (last) ◉ 04:01, 17 October 2007 Ruairidhbevan (Talk | contribs) (81,078 bytes) (undo)
- (cur) (last) ○ 18:59, 16 October 2007 Harami2000 (Talk | contribs) **m** (81,042 bytes) *(→Exile, return and Waterloo - fix typos)* (undo)
- (cur) (last) ○ 14:48, 14 October 2007 Lightmouse (Talk | contribs) (81,041 bytes) *(wp:mosnum format units and/or dates)* (undo)
- (cur) (last) ○ 15:25, 13 October 2007 Reece Llwyd (Talk | contribs) (81,045 bytes) *(Added who he was named after)* (undo)
- (cur) (last) ○ 14:39, 12 October 2007 Bryson109 (Talk | contribs) **m** (80,979 bytes) *(Undid revision 164049940 by G.-M. Cupertino (talk))* (undo)
- (cur) (last) ○ 14:28, 12 October 2007 G.-M. Cupertino (Talk | contribs) (81,032 bytes) *(→Exile, return and Waterloo)* (undo)
- (cur) (last) ○ 15:18, 9 October 2007 Bryson109 (Talk | contribs) **m** (80,979 bytes) *(Undid revision 163339389 by Maniac733 (talk))* (undo)
- (cur) (last) ○ 15:02, 9 October 2007 Maniac733 (Talk | contribs) (80,983 bytes) (undo)
- (cur) (last) ○ 03:32, 9 October 2007 Schlier22 (Talk | contribs) (80,979 bytes) *(→Organizations studying Napoleon)* (undo)
- (cur) (last) ○ 20:29, 8 October 2007 Pistolpierre (Talk | contribs) **m** (80,950 bytes) *(removed Roman Catholic classification)* (undo)
- (cur) (last) ○ 15:53, 8 October 2007 Bryson109 (Talk | contribs) **m** (80,986 bytes) *(Undid revision 163110208 by Maniac733 (talk))* (undo)
- (cur) (last) ○ 15:35, 8 October 2007 Maniac733 (Talk | contribs) (81,003 bytes) (undo)
- (cur) (last) ○ 21:47, 7 October 2007 Bryson109 (Talk | contribs) **m** (80,986 bytes) *(Undid revision 162960908 by Lovelikewinter96 (talk))* (undo)
- (cur) (last) ○ 21:42, 7 October 2007 Lovelikewinter96 (Talk | contribs) (81,363 bytes) *(→Marriages and children)* (undo)
- (cur) (last) ○ 01:46, 5 October 2007 Bryson109 (Talk | contribs) **m** (80,986 bytes) *(Undid revision 162372795 by Thelwb (talk))* (undo)
- (cur) (last) ○ 01:38, 5 October 2007 Thelwb (Talk | contribs) **m** (81,347 bytes) *(→Legacy)* (undo)
- (cur) (last) ○ 15:56, 4 October 2007 DocWatson42 (Talk | contribs) **m** (80,986 bytes) *(Updated en dashes.)* (undo)
- (cur) (last) ○ 15:42, 4 October 2007 Wedineinheck (Talk | contribs) (81,004 bytes) (undo)
- (cur) (last) ○ 13:45, 4 October 2007 Demarcroix (Talk | contribs) **m** (80,938 bytes) *(Adding videogame heading)* (undo)
- (cur) (last) ○ 22:46, 3 October 2007 Inferno666 (Talk | contribs) (80,844 bytes) *(→Exile, return and Waterloo)* (undo)
- (cur) (last) ○ 17:45, 1 October 2007 R'n'B (Talk | contribs) (80,577 bytes) *(Revert to revision 161570526 dated 2007-10-01 15:02:53 by Wedineinheck using popups)* (undo)
- (cur) (last) ○ 17:04, 1 October 2007 Shokatsu-sensei (Talk | contribs) (80,613 bytes) *(→See also)* (undo)

Fig. 6 Wikipedia's History Page

This research has focused its attention on the linguistic analyses of the two different writing modes and spaces offered by Wiki software and Wikipedia: the document mode encyclopaedic pages and its linguistic expression defined in this study as *WikiLanguage*, and thread mode talk pages and its *WikiSpeak*. An explanation in depth follows in the next two sections.

5. DocumentMode Wikipedia Articles

When using document mode pages, contributors create collaborative documents leaving their additions to wiki document, represented by encyclopaedic entries in Wikipedia. Multiple authors can edit and update the content of the document and gradually the content becomes a representation of shared knowledge or beliefs of the contributors (Cunningham & Leuf, 2001). The style used in encyclopaedic entries (defined in Wikipedia as *articles,* henceforth WAs) is explicitly promoted in an official *Manual of Style* [16] which is the official framework of reference for all Wikipedia's contributors. Here, authors can find rules on how to write article's titles, headings, notes, on how to use punctuation, spelling, national varieties of English, etc. According to Wikipedia *Manual of Style*, articles must firstly observe core principles of cooperation and objective writing based on three absolute and not negotiable principles, *Neutral Point of View* (NPOV) *Verifiability* (V) and *No*

[16] Manual of Style http://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style

Original Research (NOR). The first point is the most important. It states that articles should be written without bias. Thus, the presentation of facts necessarily requires that different types of prejudices such as: class, ethnic, racial, nationalistic, gender, linguistic, political and religious are avoided. Wikipedia requires that, where multiple perspectives exist within a topic, each should be fairly presented since readers should be allowed to freely form their own opinions. However, the meaning of the acronym NPOV has been often misunderstood. It does not mean *No Points Of View*, and it does not imply the absence or elimination of viewpoints.

With reference to the second point, *Verifiability* means that any reader should be able to check that material added to Wikipedia has already been published by a reliable source. Wikipedia *Manual of Style* states[17]:

> The threshold for inclusion in Wikipedia is verifiability, not truth. "Verifiable" in this context means that any reader should be able to check that material added to Wikipedia has already been published by a reliable source. Editors should provide a reliable source for quotations and for any material that is challenged or is likely to be challenged, or it may be removed (Wikipedia, 2007).

Furthermore, *No Original research* (NOR) [18] is a term used in Wikipedia to refer to:

> […] unpublished facts, arguments, concepts, statements, or theories. The term also applies to any unpublished analysis or synthesis of published material that appears to advance a position, or, in the words of Wikipedia's co-founder Jimmy Wales, would amount to a "novel narrative or historical interpretation (Wikipedia, 2007)..

Jointly, these three policies determine the type and quality of material acceptable in the encyclopaedic articles written in DocumentMode.

DocumentMode WA are coherent and self-contained. They reflect the result of the last update and are community property. They may have multiple and changing authors and are updated to reflect the community consensus. The collaborative writing process demonstrate that knowledge is collective and that ideas, not the writers, are the main focus (Elia, in press). The style expressed in WAs as will be shown in the next sections, is expository, extensive and monological.

Wikipedia's contributors strictly observe the Wikipedia *Manual of Style*. Articles are stylistically less innovative and original than ThreadMode talk pages. Encyclopaedic expository style is very formal, in that it never makes use of first and second personal pronouns, acronyms, jargon expressions, or neologisms.

The use of jargon is rigorously forbidden; nevertheless if its use is unavoidable, the 'banned' expressions must be hyperlinked to pages explaining accurately their meaning so that everyone can understand them. Texts are detached, accurate, rigid, refined and formal. Articles are impersonal since they are written in the third person. They are unsigned, highly informative, objective, and respectful of stylistic conventions (Elia, 2006).

[17] *Verifiability* http://en.wikipedia.org/wiki/WP:V

[18] *No Original Research* http://en.wikipedia.org/wiki/Wikipedia:No_original_research

According to *Wikipedia Assessment Department* to reach high quality standards articles should respect the attributes summarized below [19]:

THE PERFECT WIKIPEDIA ARTICLE...

- **fills a gap**; search for existing or related articles on the topic first.
- **has a good title** so it can be linked to and found easily and follows existing naming conventions.
- **starts with a clear description** of the subject; the lead introduces and explains the subject and its significance clearly and accurately, without going into excessive detail.
- **is understandable**; it is clearly expressed for both experts and non-experts in appropriate detail, and thoroughly explores and explains the subject.
- **is nearly self-contained**; it includes essential information and terminology, and is comprehensible by itself, without requiring significant reading of other articles.
- **branches out**; it contains wikilinks and sources to other articles and external information that add meaning to the subject.
- **and branches in**; editors have found and edited other significant wiki pages which make mention of the topic and link them to the article.
- **acknowledges and explores all aspects of the subject**; i.e., it covers every encyclopaedic angle of the subject.
- **is completely neutral and unbiased**; it has a neutral point of view, presenting competing views on controversies logically and fairly, and pointing out all sides without favoring particular viewpoints. The most factual and accepted views are emphasized, and minority views are given a lower priority; sufficient information and references are provided so that readers can learn more about particular views.
- **is of an appropriate length**; it is long enough to provide sufficient information, depth, and analysis on its subject, without including unnecessary detail or information that would be more suitable in "sub-articles", related articles, or sister projects.
- **reflects expert knowledge**; it is grounded in fact and on sound scholarly and logical principles.
- **is precise and explicit**; it is free of vague generalities and half-truths that may arise from an imperfect grasp of the subject.
- **is well-documented**; all facts are cited from reputable sources, preferably sources that are accessible and up-to-date.
- **is clear**; it is written to avoid ambiguity and misunderstanding, using logical structure, and plain, clear prose; it is free of redundant language.
- **is engaging**; the language is descriptive and has an interesting, encyclopaedic tone.
- **follows standard writing conventions** of modern English, including correct grammar, punctuation and spelling.
- **includes informative, relevant images**—including maps, portraits, photographs and artworks—that add to a reader's interest or understanding of the text, but not so many as to detract from it. Each image should have an explanatory caption.
- **is categorized.**

Surprisingly, many Wikipedia articles are of high quality. There is a page on Wikipedia named *Featured Articles* [20] in which particularly well-written and comprehensive articles are listed to exemplify the best works and professional standards of writing in encyclopaedic expository style.

The *Assessment Department* of the Wikipedia English Edition, has introduced a *Quality scale*[21], based on a set of rigorous criteria (e.g. *style, prose, completeness, accuracy, neutrality,* etc.), against which the quality of articles is judged. Nowadays, approximately 1500 articles have reached the highest status of *Featured Articles*. The six *class* parameters shown in fig. 7 are used:

[19] *The perfect article* http://en.wikipedia.org/wiki/Wikipedia:The_perfect_article
[20] *Featured articles* http://en.wikipedia.org/wiki/Wikipedia:Featured_articles
[21] *Quality scale* http://en.wikipedia.org/wiki/Wikipedia/Assessment

Quality Scale

Class	Criteria	Reader's experience	Editor's experience
FA	Reserved for articles that meet the *featured article* criteria and have received featured article status after community review.	Definitive. Outstanding, thorough article; a great source for encyclopaedic information.	No further editing necessary, unless new published information has come to light.
A	Provides a well-written, reasonably clear and complete description of the topic, as described in How to write a great article. It should be of a length suitable for the subject, with a well-written introduction and an appropriate series of headings to break up the content. It should have sufficient external literature references, preferably from the "hard" (peer-reviewed where appropriate) literature rather than websites. Should be well illustrated, with no copyright problems. At the stage where it could at least be considered for featured article status, corresponds to the "Wikipedia 1.0" standard.	Very useful to readers. A fairly complete treatment of the subject. A non-expert in the subject matter would typically find nothing wanting. May miss a few relevant points.	Minor edits and adjustments would improve the article, particularly if brought to bear by a subject-matter expert. In particular, issues of breadth, completeness, and balance may need work. Peer-review would be helpful at this stage.
GA	The article has passed through the *Good article* nomination process and been granted GA status, meeting the good article standards. This should be used for articles that still need some work to reach featured article standards, but that are otherwise good. Good articles that may succeed in FAC should be considered A-Class articles, but being a Good article is not a requirement for A-Class.	Useful to nearly all readers. A good treatment of the subject. No obvious problems, gaps, excessive information. Adequate for most purposes, but other encyclopaedias could do a better job.	Some editing will clearly be helpful, but not necessary for a good reader experience. If the article is not already fully wikified, now is the time.
B	Has several of the elements described in "start", usually a *majority* of the material needed for a completed article. Nonetheless, it has significant gaps or missing elements or references, needs substantial editing for English language usage and/or clarity, balance of content, or contains other policy problems such as copyright, NPOV or NOR. With NPOV a well written B-class may correspond to the "Wikipedia 0.5" or "usable" standard. Articles that are close to GA status but don't meet the Good article criteria should be B- or Start-class articles.	Useful to many, but not all, readers. A casual reader flipping through articles would feel that they generally understood the topic, but a serious student or researcher trying to use the material would have trouble doing so, or would risk error in derivative work.	Considerable editing is still needed, including filling in some important gaps or correcting significant policy errors. Articles for which cleanup is needed will typically have this designation to start with.
Start	The article has a meaningful amount of good content, but it is still weak in many areas, and may lack a table. For example an article on Africa might cover the geography well, but be weak on history and culture. Has at least one serious element of gathered materials, including any **one** of the following: • a particularly useful picture or graphic • multiple links that help explain or illustrate the topic • a subheading that fully treats an element of the topic • multiple subheadings that indicate material that could be added to complete the article	Not useless. Some readers will find what they are looking for, but most will not. Most articles in this category have the look of an article "under construction" and a reader genuinely interested in the topic is likely to seek additional information elsewhere.	Substantial/major editing is needed, most material for a complete article needs to be added. This article usually isn't even good enough for a cleanup tag: it still needs to be built.
Stub	The article is either a very short article or a rough collection of information that will need much work to bring it to A-Class level. It is usually very short, but can be of any length if the material is irrelevant or incomprehensible.	May be useless to a reader only passingly familiar with the term. Possibly useful to someone who has no idea what the term meant. At best a brief, informed dictionary definition.	Any editing or additional material can be helpful.

Fig. 7 Wikipedia’s Quality scale

Articles are also rated in Wikipedia in accordance with their *importance scale* [22] (fig. 8). Although a general and universal criterion to assess the importance of anything does not exist since it is based on subjective parameters, this is an interesting attempt to estimate the probability of the average reader of Wikipedia to look up the topic (and thus the immediate need to have a suitably well-written article on the subject).

Importance Scale

Status	Template	Meaning of Status
Top	**Top-Class**	This article is of the utmost importance to this project, as it forms the basis of all information.
High	**High-Class**	This article is fairly important to this project, as it covers a general area of knowledge.
Mid	**Mid-Class**	This article is relatively important to this project, as it fills in some more specific knowledge of certain areas.
Low	**Low-Class**	This article is of little importance to this project, but it covers a highly specific area of knowledge or an obscure piece of trivia.
None	**None**	This article is of unknown importance to this project. It remains to be analyzed.

Fig. 8 Wikipedia's Importance Scale

6. ThreadMode Talk Pages

*Talk page*s (henceforth TPs) are used by contributors to discuss how to improve the content of the official correspondent encyclopaedic pages. They use this space to post signed messages. TPs and WAs are intimately related, forming a combined whole, although for technical reasons they are separately stored in the web server's database.

The wiki writing mode conveyed in TPs has been defined as ThreadMode (Morgan, 2006). TPs are the most common area used by Wikipedians to communicate. They have multiple functions. For example, they are used to discuss the general direction of an article, its structure, scope and connection to other topics. In addition, they offer a place where authors can debate contributions or information quality issues.

A column in the *Wall Street Journal* (Gomes, 2007) points out that the best parts of Wikipedia is represented by the discussions related to entries themselves. Gomes claiming that the reading of TP is a rewarding experience, recommends to examine the discussion before reading the proper encyclopaedic article. With reference to the technique of collaborative writing, it is possible to see what the most controversial points in the construction of the article are, how people negotiate the facts

[22] *Importance Scale* http://en.wikipedia.org/wiki/Wikipedia:WikiProject/Assessment/Importance_scale

and the bias surrounding them, whom they agree with, who has access to talk pages and when the debate starts and ends. A ThreadMode TP looks like *conversational graffiti*. It is a written conversation among interested parties, as can be seen in this extract from a page in *Meatball Wiki* [23]:

> I think that hyperlinks should be red, not blue. –FredFlinstone Why do you think that, Fred?
> – BarneyRubble Because, it would make them stand out more. – ff

ThreadMode[24] talk pages are stylistically different from the bland, formal and neutral style expressed in the DocumentMode encyclopaedic pages. A threaded conversation shows many different points of view and lends energy to them, developing a significant sense of community. It has been claimed:

> This Wiki seems kinda dead to me. Is it because there are no ThreadMode discussions, which is a symptom of lack of controversy? There's not much fun in reading through a dictionary. --anon. [25]

My personal opinion is that it is easier to write in ThreadMode TPs than in DocumentMode WAs, since contributors are not obliged to abide any conventions or compulsory styles defined by any authoritative Manual of Style. Contributors have only to write, optionally sign and send their posts (anonymous contributions are also very frequent in Wikipedia). The style is here more original and often unconventional. It seems to work very well when individual contributions are concise, with a definite objective to a single subtopic. Longer segments which touch on several related topics can lead to a multi-threaded mode, in which people respond to multiple parts of a contribution, creating multiple threads which all progress simultaneously and sometimes erupt into a *ThreadMess* or a *ForestFire*[26]. ThreadMode TPs have a highly personal perspective as posts are written in the first person and are usually signed. As signatures are often redundant and conversational exchanges chaotic, threads are much more 'noisy' and 'fuzzy'[27] than in DocumentMode pages. The style in ThreadMode is dialogical, flexible, and direct. The register is free and informal. It presents multiple interacting positions and evolves without a predictive structure, since the discussion development is changeable and impulsive. The style in ThreadMode TP is exploratory, explicit, involving (Tannen, 1989), and rich in new terminology, although less informative than DocumentMode texts. ThreadMode discussions are the clear evidence to prove that knowledge is the result of constructivist collaboration and not a lonely production (Elia, in press).

[23] *Meatball Wiki* http://www.usemod.com/cgi-bin/mb.pl is a meta wiki dedicated to online communities. Thus, being 'a community about communities', has become the launching point for various other wiki-based projects. Nevertheless, its original goal was to focus on collaborative hypermedia, but current topics range from intellectual property to cyberpunk..

[24] ThreadMode http://www.usemod.com/cgi-bin/mb.pl?ThreadMode

[25] ThreadMode http://www.usemod.com/cgi-bin/mb.pl?ThreadMode

[26] See the *Glossary* in *Appendix*

[27] By *noisy* is meant the quality of lacking any predictable order or plan while *fuzzy* refers to a text which can appear to be confused and incoherent.

Stvilia *et al.* (2005) claim:

> A discussion page is an auxiliary wiki object which accompanies a Wikipedia article and, as the name indicates, is intended largely for the purposes of communication among the members of the Wikipedia community when constructing and maintaining the article content. Technically, a discussion page is the same wiki object as an article. Unless locked by Wikipedia administrators it can be updated by anyone. Updates to the article are logged and can be visualized through a history object. The difference between the article and its discussion page lies only in the role assigned to a discussion page in the Wikipedia infrastructure.

According to Schmidt (1996:155-200), TPs are *coordinative artifacts which help to negotiate and align member perspectives on the content and quality of the article*. He finds that TPs are often used by community's outsiders to ask questions related to the article's topic, and sometimes even soliciting assistance for other Wikipedia articles or projects outside of Wikipedia. Furthermore, Pentzold *et al.* (2006: 59) claim:

> The talk page presents itself as a comparatively unstructured forum without a predetermined topical framework. The authors develop its structure the moment they start a new line of argumentation. Their initial statement is followed by responses stating the author and the exact point of time. Sometimes, the talk on these pages outweighs the actual content with respect to its volume. For example, the talk pages of the article *Conspiracy theory* are approximately ten times as long as the associated article. Unsurprisingly, this effect seems to correlate with the importance or controversy of a topic.

A writing process, defined as *refactoring,* can often be noticed on Wikipedia pages, especially in TPs. The term *refactoring* [28] has its origins in computer programming. It refers to the process of rewriting, reorganizing, and shortening texts, while preserving content. It is not always an easy operation to perform, as its goals are to improve readability while preserving meaning and removing superfluous content without altering the basic information provided.

Refactoring TPs is necessary when there is an accumulation of previous unclear or irrelevant posts, whose effect is to discourage the involvement of potential contributors. It promotes productive discussions by improving clarity and accessibility. When participants have reached a consensus, someone will elaborate the new information and suggestions provided. By this process what has been written in TPs will be restructured and elaborated into a more formal and impersonal expository style in WAs (Morgan, 2006).

7. Cybergenre: A Theoretical Background

The web is a new communication medium that was invented only a few decades ago. It is also a large and heterogeneous community and a new virtual environment where interactions among web users and the possibility offered by technology modify existing genres and create new ones which

[28] *Refactoring* http://en.wikipedia.org/wiki/Refactor

better satisfy the new information and communication needs. As shown by Crowston and Williams (2000), who were among the first to study the development of genres on the web, the web has had a substantial impact on the genre repertoire. Most of webgenres come from previous traditions. When moving to a new medium, and before elaborating new formats, it is normal to use those formats available on existing media and then adapt them to the specific peculiarities of the new medium.

Shepherd and Watters (1998) coined the term *cybergenre*. Nowadays cybergenre (or *webgenre*) is characterized by *content*, *form* and *functionality*. The first two elements are common to traditional genres, while the third one refers exclusively to the capabilities offered by the web.

Before the new millennium, most genres on the web were still borrowed (reproduced genres) from other media, while a large proportion appeared to be adapted (variant genres) to the needs and capability of the new medium. Hence, traditional genres, such as encyclopaedias, newspapers and dictionaries have been influenced by cyberculture and the new functionalities offered by the web. Shepherd and Watters (1998: 1) claim:

> When an existing genre initially migrates to this new medium, it is usually as a faithful reproduction of the existing genre in both content and form with little new functionality. It may then evolve into a variant cybergenre as it incorporates functionality afforded by the computer and Internet. Cybergenres also include novel genres, either not based on previously existing genres or substantially different from existing genres on the basis of increased functionality.

Thus, cybergenres show different levels of functionality (defined in terms of *browsing, email facility, multimediality, search, discussion, interactivity, online ordering/enquiring, collaborative computing*, etc.). According to Shepherd and Watters (1998:1) a cybergenre is made up of two macro areas which they define as *extant* and *novel genres* (fig. 9).

Extant genres in which news articles, encyclopaedias and dictionaries are included, range from faithful replications of the original format, as they appear in their source media, to significant variants which fully exploit the new functionalities afforded by the Internet. *Replicated genres* include most digitalized text documents; they show very little innovative functionalities. On the other hand, *variant genres*, exploit the new technologies and represent an evolution of the original format.

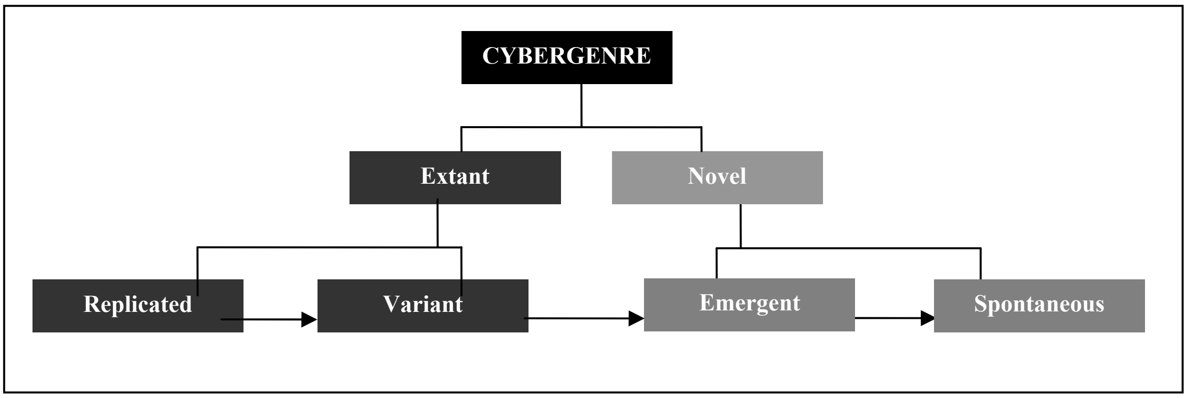


Fig. 9 Evolution of cybergenres (Shepherd and Watters, 1998)

As Erickson (1995:13-20) claims, Information and Communication Technology has the potential to greatly speed up the evolution of genres; thus, *novel* genres continuously emerge online. Some of them come out through an evolutionary pattern while others are spontaneous in nature. According to Shepherd and Watters (1998), novel cybergenres (fig. 9) are made up of two subclasses: *Emergent and* S*pontaneous genres.*

E*mergent genres* are considered those originally replicated in the new medium but which have evolved considerably from the original format, thanks to the new added functionalities. In brief, the fundamental evolutionary force is the progressive exploitation of the innovative functionalities afforded by the new medium. Thus, the typical evolutionary path is from simple replication through variant to emergent. S*pontaneous cybergenres* (e.g. home page, hotlists, interactive pages, virtual realities, etc.) have no counterpart in other media. They are almost totally based on the functionalities unique to the new medium, they represent the most advanced expression of genre evolution. This view is supported also by Haas and Grams (1998:489) as they claim:

> The Web, with its multimedia capabilities, has also spawned page types that have no equivalent in the print world, such as home page or a page containing audio or video clips, or interactive pages.

Together with the new functionalities suggested by Shepherd and Watters (1999), there are also other important attributes, I add, that characterize web genres: the use of hypertext, social software and new emerging techniques of collaborative writing, which have created a new way of reading and writing on the web. Introducing the concept of "modal shift" between reading, writing mode and navigating mode, Web 2.0 has introduced a new dimensional perspective on genre analysis.

8. Can Wikipedia be considered an Emergent Web Encyclopaedic Genre?

Yates and Sumner (1997:3-12) suggest that it is appropriate to rely on the progressive evolution of genre from replication to novel to maintain the notion of fixity in changing systems. The continuity in *content and form*, even if the functionalities change, provide the users with a familiar and strong metaphoric reference that transcends changes in functionality and evokes a natural progression of the genre. Such an approach provides continuity for the users. Rehm (2006) also suggests that the process of imitation maintains stability in the genre repertoire, while change is determined by the break of conventions. Emerging genres represent a transitional phase in genre evolution. They are genres, not fully standardized and not yet officially accepted by the academic community. Since the web is a recent phenomenon, fluid and evolving at a fast pace, the emergence of novel genres is much more rapid than in other media (Santini, 2007).

The concept of emerging genres has not been explicitly formulated in the genre literature as they convey textual patterns not yet classifiable in the official genre. For example, before the new millennium, blogs and wikis were already on the web, but they were just considered web pages. Only when the most active blog and wiki communities sprang up using their new label, blogs and wikis started spreading and being recognized as new web 2.0 genres. However, the emergence of a novel genre depends on social acceptance (Crowston and Williams, 2000) and this is probably the reason why Wikipedia is not yet officially recognized as an official web encyclopaedia, due to the several *querelles* on its open editing system which according to some, can negatively affect the quality of the information provided.

The fluidity and the dynamism of the web affect the web genre repertoire. Genre re-adjustments are not unusual in a transitional phase where the lack of any institutionalized control, as in the web, can stimulate the creation of new or hybrid genres (as it happens to the original project of *Nupedia* than transformed into *Wikipedia*).

Wikipedia, as an encyclopaedia, shows sets of standardized or conventional features which will be highlighted through the linguistic analysis which follows in the next chapter. The linguistic formality and coherence which it shows when compared to the encyclopaedic expository style of Britannica (see chapter 4), makes it clearly recognizable as an encyclopaedia, and this raises specific expectations. Nevertheless, Wikipedia is not exclusively an encyclopaedia, since it is also a free and egalitarian project as well as an online community. The phenomenon of hybridization, where two or more genres overlap, are not unusual in web environments. By applying Baktin's metaphor on interpretation of language to Wikipedia, genre conventions can be seen as the *centripetal* force that keeps stability in genre repertoire and allows continuity of communication (in this specific case the traditional encyclopaedic genre), while collaborative and technical innovations (wiki software) are the *centrifugal* forces that destabilize the system, allowing changes and genre evolution (Santini, 2007). This struggle between *stasis* and *change* (Yates and Sumner, 1997) gives rise to a transitional phase of emerging genres. As Santini (2007) claims *the web is a complex scenario where the lack of any institutionalized control stimulates creativity and hybridization among traditional offline and innovative online genres*.

In brief, Wikipedia as an hybrid genre cannot be classified using a single-genre label, since, as it will be shown, it is at the same time, an encyclopaedia, a collaborative project and a wiki community. A single genre classification scheme for Wikipedia seems to appear inappropriate. When dealing with a webgenre, important aspects which have always to be taken into account are its fast mutability and fluidity as Wikipedia and its evolving folksonomy and content of encyclopaedic entries clearly show.

As Orlikowski and Yates (1994) pointed out, genres are rarely homogeneous. Also traditional genres tend to overlap and mix; *tragicomedy* for example, blends aspects of two different genres. The only difference is that in an open communication space, like the web, where many communities meet, genre contaminations are likely to occur more easily (Santini, 2007).

My personal point of view is that offline and online electronic encyclopaedias have evolved progressively from the replication of the traditional paper format to a variant genre; in some cases (such as Wikipedia) they have acquired so many new technical functionalities (hypertextuality, search engines, multimedia, social and interactive functionalities, collaborative writing techniques, etc.) to be considered as a novel *emergent* genre.

Wikipedia is characterized by new functionalities which do not exist in the traditional paper form. Although it is an encyclopaedia, it is an online co-authored encyclopaedia. Thus, it acquires completely new distinctive features, as it exploits the new social networking and constructivist functionalities offered by the Web 2.0.

The cybergenre classification proposed by Shepherd and Watters (fig. 9) has proved to be very useful in analyzing Wikipedia. Online encyclopaedias (such as *Encyclopaedia Britannica, the Columbia Encyclopaedia, Encarta* etc.) are considered as expressions of a variant genre while hybrid genres which mix encyclopaedic projects with virtual communities (such as *Wikipedia, H2G2, Everything2)* are prototypes of the evolution of *variant* proprietary encyclopaedias into new *emergent* co-authored encyclopaedias.

But what is really new in Wikipedia and in online encyclopaedias? To what extent Wikipedia replicates an extant genre and how does it represent a novel genre? Is an emergent genre coming to life? What is replicated in online encyclopaedias compared to traditional reference works and what is emergent or completely new? Has the formal linguistic register of traditional encyclopaedias been maintained online or has it been affected by the informal and innovative values expressed by Web 2.0 culture in Wikipedia? Are the expository style of encyclopaedias, web usability and index of readability similar or different in Wikipedia and Britannica? The present research tries to give an answer to these several open questions.

9. Collaborative Writing: Strategies, Document Control Modes and Writing Roles

Although collaborative creation and organization have been in practice since biblical times, with scribes transcribing and at the same time often editing, updating, interpreting or reinterpreting original texts, open access large scale public collaborative content creation projects are relatively recent phenomena (Stvilia, 2005).

Lowry *et al.*’s taxonomy (2004) will be used as a framework of reference in order to define Collaborative Writing (henceforth CW) and to identify the typology of writing carried out in Wikipedia. CW, involving multiple people, increases the complexity of the writing process. CW’s focus on group work around a common objective is a critical definitional point as writing does not become collaborative just because multiple people are involved. One of the reasons for the amplified complexity is the need of coordination between multiple viewpoints and work efforts and the need to establish mutual consensus (Galegher & Kraut, 1990). Some researchers (Ede & Lunsford, 1990) have supported the importance of some group dynamics in CW process demonstrating that to become a real collaborative process, writers need first of all to build the group consensus. Furthermore, extra activities not involved in single-author writing such as communicating, negotiating, coordinating, monitoring, socializing, and so forth, are required. Writing tasks and group activities cannot be separated without negative repercussions. Thus, CW is not limited to writing and it has to be regarded from an holistic perspective. It can be defined as a social process that involves a group focused on a shared objective that is negotiated, coordinated, and communicated during the creation of a common document.

CW includes a variety of different writing strategies, activities, document control approaches, group roles, and work modes. First of all members of a CW group must agree upon basic strategies to successfully produce a collaboratively written document (Allen *et al.,* 1987).

A CW group can be structured around a *group single-author writing* (fig. 10) or a *sequential single writing* (fig. 11). In the former case the group works towards a coordinated consensus that is reflected in a document written only by one of the group members, while in the latter case each writer completes his or her task and then passes it on to the next person, who becomes the next single writer (Sharples, 1992). Such strategy reduces social interaction and can easily create a lack of group consensus. When the model of *parallel writing* (fig. 12) is adopted, the group divides CW work into discrete units and work in parallel (Sharples, 1993). This strategy conveys work in parallel by multiple writers. Some problems that occur include poor communication, stylistic differences, and informational overload (Ellis *et al.,* 1991).

Reactive writing (henceforth RW) (fig. 13) is defined by Lowry *et al.* (2004) as the strategy which occurs when writers create a document in real time, reacting and adjusting to each other's changes and additions without significant preplanning and explicit coordination. The term RW is used since written reaction may involve consensus or dispute, reflection, or spontaneous contributions. For example, while some authors write a section, others may simultaneously review the section and create new sections in response that may contradict or concur with the first author's point of view. Advantages of RW include the possibility of building consensus through free expression and the development of creativity. The primary drawback of this strategy is that it makes coordination difficult and can cause difficulties with version control.

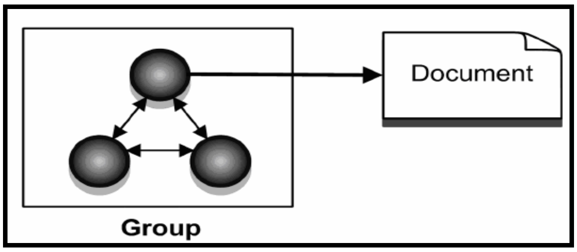


Fig. 10 Group single author writing (Lowry *et al.,* 2004)

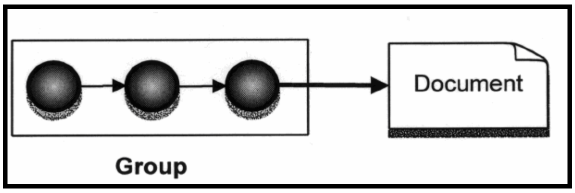


Fig. 11 Sequential writing (Lowry *et al.,* 2004)

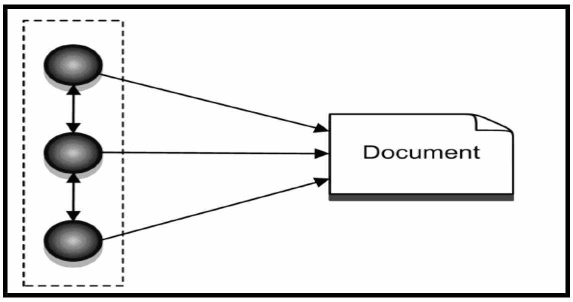


Fig. 12 Parallel writing (Lowry *et al.,* 2004)

Outlining, drafting, reviewing, revising and copyediting are some basic cognitive processes in CW familiar to most writers, which are involved in the actual production of a group document. These activities tend to occur in a dynamic and iterative way both in individual and collaborative writing (Lowry *et al.,* 2004).

In addition, other activities such as socialization, research, communication, negotiation, coordination etc., have a fundamental role in supporting the overall writing task. Again, these activities are not necessarily performed sequentially; they are carried out through iterative rounds of reading and review. Using these activities, it is possible to have a more comprehensive view of CW.

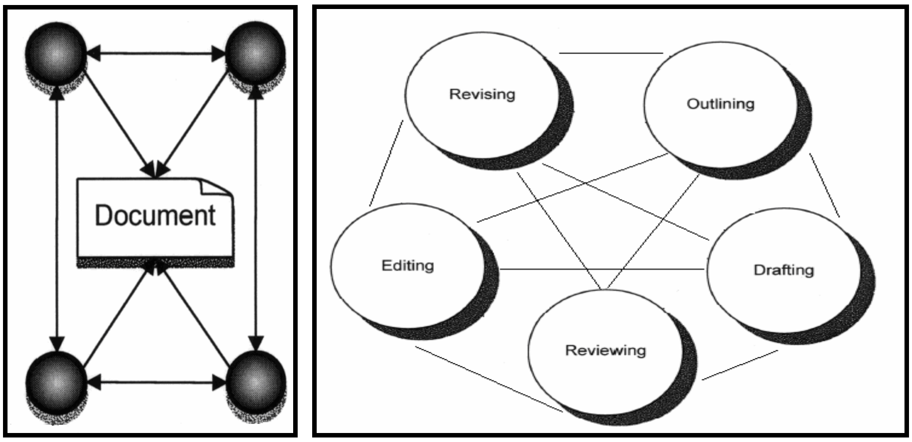


Fig. 13 Reactive Writing

Document control modes, which are the approaches chosen to manage a collaborative document can be centralized, relay, independent or shared[29] (Posner & Baecker, 1992). In the *shared* control mode all group members have simultaneous and equal access and writing privileges throughout the writing activity. This can be a highly effective, non threatening form of control in groups that work face-to-face, engage in frequent communication and have high levels of trust. Nevertheless, this mode can lead to conflict in groups working far away.

[29] In the *centralized control mode* one person controls the document throughout the writing activity. *Relay mode* happens when one person at a time controls changes within the group. This democratic technique is useful in groups that need to share power. In the case of *independent mode* each member works on a separate part of the document and maintains control of his or her portion throughout the writing process. It is a useful for groups working remotely on independent units of work (e.g. different chapters in a book). In this case each member works on a separate part of the document and maintains control of his or her portion throughout the writing process. It is a useful for groups working remotely on independent units of work.

In addition to the different document control modes, writers can also assume different roles in the CW process. The most common collaborative writing roles (e.g. editor , reviewer, group leader, facilitator, etc.) can be very strictly defined, interchangeable, or more than one role can be supported by anyone (Posner & Baecker, 1992).

10. Collaborative Writing in Wikipedia

According to Olga Pombo *et al.* (2006:252-265) encyclopaedias have historically been collective works, although some Medieval works and Renaissance and Baroque encyclopaedias, which today have been retrospectively included in the encyclopaedic genre, were written only by a single author. Many renowned experts together with various scholars and even unknown and anonymous authors contributed to the XVIII century's encyclopaedias. As Diderot claimed in the entry dedicated to "Encyclopédie"

> The Encyclopédie had the collaboration of first level science men, artists, musicians, writers like Quesnay, Rousseau, Voltaire, Du Marsais, Turgot, Montesquieu, Grimm or Duclos, side by side with craftsman, agricultures, gardeners, weavers, etc. and even many spontaneous and sometimes anonymous "colleagues", all united by a militant "intéret général du genre humain et par un sentiment de solidarité reciproque".

Thus, starting from the Enlightenment encyclopaedias became, as Neurath (1946:26) pointed out, a *polymorphic orchestra* since:

> in encyclopaedias scientists with different opinions will be given an opportunity to explain their individual ideals in their own formulation in such a way that encyclopaedia will become a platform for the discussion of all aspects of scientific enterprise.

10.1 From Individual to Collaborative Writing

With these assumptions in mind Wikipedia can be considered, in the scenario of the Web 2.0, the latest and more radical evolution from the original encyclopaedic model. Wikipedia, as a co-authored encyclopaedia, is not a generic sum of contributors' perspectives in independent encyclopaedic entries. Thanks to the new technical functionalities offered by wiki software, authors' voices, points of view and expertises are merged inside the same encyclopaedic articles.

Miller (2005) claims that for many generations, humans inscribed clay tablets and recorded information on papyrus but only rarely included their own names in the documents they produced. Only after the development of modern publishing methods, authorship acquired a legal and universal meaning. Copyright laws established the right of authors to control their publication.

Then came the Internet and the World Wide Web which began to challenge the concept of authorship and readership. This process began with electronic mail. Since the number of Internet users became wider, people started to look for ways to increase the sharing of the writing process. From these efforts has emerged the wiki, specifically designed to enable information sharing and collaborative writing and its most ambitious example: Wikipedia.

Miller (2005) claims that the idea of collaborative writing did not start with the Internet, of course, but this new form differs from the typical collaborations of the last century. The idea that any reader can also add, change or even delete another writer's document makes many writers uncomfortable, as Western laws have codified the rights of authors to own and control their personal works. Even when a work involves the efforts of several authors, the copyright prevails and every author's name appears on the work. Miller (2005:39) writes:

> Wikipedia has no such concerns. Just as Newton acknowledged that he stood on the shoulders of giants, so wiki authors understand that the recording of information by any one of us really only builds on the efforts of all the other thinkers, readers, and writers who have gone before. It embraces the process nature of reading and writing, preferring the constantly-evolving-but-never-finishing to the static and rapidly obsolescing "product." On a wiki site, anyone who reads a page can also edit it, borrow from it or even remove it. In fact, the wiki culture invites, almost compels readers to edit. Just because anyone can make changes doesn't free a writer from responsibility for what they write. The transition from the view of writing as a product to the understanding of writing and reading are different phases of the communication process. A single author doesn’t exist no longer. People periodically author, read, and share information.

Morgan (2006) also shows that in the wiki more than in other collective works, the main focus is the collective knowledge and not the single author, as author voices disappear behind the coral and objective writing. Each specific encyclopaedic article is coherent, self-contained and collectively written in a conventional way in DocumentMode, the main editing functionality offered by the wiki software. The style expressed in this ‘writing space’ proves to be expository, extensive and monological and it turns out to be rule oriented and stylistically formal since contributors strictly observe the ‘Manual of Style’. In conclusion, CW is a complex and dynamic group process in which many considerations and issues must be addressed.

According to Lowry *et al.* (2004)’s CW taxonomy, Wikipedia CW can be classified as *reactive writing* as contributors adjust to each other’s changes and work without strictly preplanned and explicitly coordinated activities. Being the most open writing system, *reactive writing* is also unpredictable and it needs a lot of supporting activities such as socialization, communication, negotiation and coordination. These specific activities explicitly take place in TPs in association with the main encyclopaedic articles, the back space where all Wikipedians according to the Wikiquette, discuss, agree and disagree, before writing, correcting, changing or editing official encyclopaedic articles.

Moreover, with reference to Lowry *et al.* (2004) taxonomy, Wikipedian document control mode is *shared*. This model can be very effective, although it involves frequent communication, high levels of reciprocal trust and unfortunately it can lead occasionally to conflict defined in Wikipedia as *edit war,* which sometimes needs the intervention of *mediation* or *arbitration committee* in order to be solved. The different collaborative writing roles previously mentioned (writer, consultant, editor, reviewer, group leader, etc.) could simultaneously be adopted by the same contributor. The choice depends on the attitude, commitment and involvement of each Wikipedian contributor.

Further to the definition of reactive writing, the term *massively distributed collaboration* (MDC) distinctively defines an emerging activity in content-creating virtual communities (e.g. mailing lists, blogs, wikis, etc.). For the first time Mitchell Kapor (2005) used this definition in a presentation at UC Berkeley on 11th September, 2005. In the introduction to his talk he claimed:

> The sudden and unexpected importance of the Wikipedia, a free online encyclopaedia created by tens of thousands of volunteers and coordinated in a deeply decentralized fashion, represents a radical new modality of content creation by *massively distributed collaboration*. This talk will discuss the unique principles and values which have enabled the Wikipedia community to succeed and will examine the intriguing prospects for application of these methods to a broad spectrum of intellectual endeavors.

MDC is nowadays applied in different domains such as education, research, music, corporations, political action, etc. Its central purpose is assembling a body of information which can be re-used later by the same contributors and by others.

10.2 CW in Wikipedia: Pros and Cons

Wikipedia is an emerging exciting online environment that is affecting and reshaping the way distributed contributors think, collaborate and work together. It offers the convenience of a shared online wiki workspace. Wiki software facilitates transparent online interactions and erases some of the boundaries that exist between author and reader. Using a wiki, members working on the collaborative production of an encyclopaedia, can more easily and frequently cross the borderlines between author and reader. Distributed contributors can interact with one another over the Internet by actively co-creating live Web content. This shift in writing methods challenges current thinking about effective Web design and enrich user experience. Wei *et. al.* (2005:206) claim:

> Wikis allow distributed teams to collaboratively write and edit documents through the Internet in a shared online workspace, without the need for special HTML knowledge or tools. The flexibility of wiki technology is a boon for increased cooperative work on large team projects. However, wiki technology also complicates notions of usable design as the information architecture of a wiki site may be created on the fly by all participants rather than by a dedicated technical communicator. Virtual work groups are becoming more common as the technology to support their work becomes increasingly more available. […] A sophisticated wiki such as Wikipedia has technical features which can easily support this massive documentation project. It is written, reviewed, and edited by volunteers worldwide and has features that support meta-conversation about the writing and editing of a page and allow users to easily compare past revisions of a page. Wikis that allow users to hold a stake in the community and develop a reputation ultimately can foster close, productive group work.

Since wikis give groups a shared online space to store documents, exchange information, and work collaboratively, they can be of great help to collaborative work. The only thing users need is access to a web browser. The centralization of a wiki can be useful for collaborative projects as it eliminates the difficulty of redistributing documents: there is only one document to work on rather than multiple copies circulating through the group members. The simplicity of a wiki also makes it less difficult to make small, spontaneous edits and minor changes which could seem picky or hypercritical if made on a Word document. Contributing authors can develop their ideas over a longer period of time and include more suggestions in the draft as a result of more frequent editing sessions. Furthermore, public wikis that provide information to larger audiences have the advantage of attracting more contributors (Wei *et al.* 2005).

Despite the vandalism and poor quality content, large public wiki projects such as Wikipedia have grown into mature projects with a high number of complete, well-written articles. Besides the productivity advantages, wikis are very useful as shared social spaces for group members working remotely since authors do not need to be in the same physical space, do not need to have a previous relationship with each other, and do not need to plan their actions.

Despite their benefits, wikis also present some disadvantages. Wei *et al. (*2005:206) claim:

> Chiefly, they require the users to learn wiki syntax in order to maximize the use of the formatting capabilities of the wiki. Adding plain text on a page is simple, but formatting headings, lists, or tables requires the knowledge of wiki syntax. Some of this syntax is easily learned though by novice users who can copy the syntax used by other wiki authors. Editing pages through a Web browser usually does not allow users to spell-check or have the same sophisticated editing functionality of a word processing program such as spelling and grammar checkers, thesaurus, synonyms etc. Wiki editing can also intimidate users new to the collaborative environment. If collaborative writers and editors are accustomed to the visual cues offered by Microsoft Word, wiki editing may be unsettling. It may take demonstrations to reassure the novice editor that edits are recorded, and can be compared in the *Revision History*.

Another disadvantage of Wikipedia is represented by the basic design which can look primitive, without graphics or exciting colors, like a relic of the early days of the World Wide Web.

Furthermore, wikis run the risk that some users may become invisible autocrats. Some dedicated users may enhance usability of the overall wiki for the entire group, but there is also the opposite risk of overpowering themselves.

Wikis are definitely challenging and they are redefining the concept of textuality and how it works. While hypertext has revolutionized the concept of textual linearity, wikis are developing the idea of *social textuality*. The wiki not only captures the content, but also the process; or rather, the wiki is the content and the process (Mejias, 2005). Wikis engender a new form of literacy: a social literacy which refers to the use of writing in social contexts (Lamb, 2004). This term refers to textual practices related to multiple and simultaneous authors. Wikis reflect the decisions not of a single individual, but of a community. Lamb (2004:42) summarizes some of the distinct traits of wiki writing as follows:

> […] content is ego-less, time-less, and never finished. Anonymity is not required but is common. With open editing, a page can have multiple contributors, and notions of page "authorship" and "ownership" can be radically altered. […] In wikis, the process becomes the product. What is important is not who changed a sentence in the text, but that the sentence has been changed and can be changed again, if someone doesn't like it. Wikis significantly alter our ideas about the ownership and stability of text to an extent that not even earlier forms of electronic text achieve: In a wiki, writing is open and ceases to be owned by any single individual. The surprising thing about wikis is that, although all the openness sounds like a recipe for disaster, committed communities seem to avoid chaos and actually manage to give shape to collectively shared meaning.

Crystal (2001:207), while not writing about wikis specifically, enumerated some of the problems of social literacy. He argues that, contrary to most traditional printed texts which have a single author, on the web:

> […] there are multi-authored pages where the style shifts unexpectedly from one part of a page to another. The more interactive a site becomes, the more likely it will contain language from different dialect backgrounds and operating at different stylistic levels—variations in formality are particularly common… People have more power to influence the language of the Web than in any other medium, because they operate on both sides of the communication divide, reception and production. They not only read a text, they can add to it.

Probably Wikipedia is a good gym where authors/readers are learning to 'filter out' the noise of multiple styles, and are becoming more comfortable with textual bricolage, and with the new web scenarios characterized by the concept of impermanence. Meanwhile authors/readers learning to interchange their roles are giving birth to a new social literacy and to a totally new virtual cultural scenario.

Web 2.0 conceptually refuses the idea of fixity. For those who believe in the knowledge immutability, this new paradigm is culturally unacceptable. To give an example The *Mississippi River bridge* in Minneapolis (Minnesota, United States) collapsed on August 1st, 2007, during the evening rush hour, falling into the river and onto its banks. Thirteen people died and approximately one hundred more were injured. Within 22 minutes from the event, the *Star Tribune* had updated its website with this news. Within 24 minutes, Wikipedia had added the information to its entry for the bridge[30]. The difference is that *The Star Tribune's News* site is run by a staff of professional journalists, while Wikipedia is not. The Interstate 35W bridge collapsed at about 6:05 p.m., at 6:29 p.m., a computer user in Lakeville added this sentence to Wikipedia's description of the bridge: *The bridge collapsed on August 1st, 2007, at approximately 6:00 pm. Several vehicles went into the water.* Three minutes later, John Warkel, a student at Henry Sibley High School in Eagan, added an update, citing KARE-11's website as a source.

Then further updates came. In the next 12 hours, people from around the world updated the encyclopaedic entry more than 450 times to reflect the changing news and to edit other peoples' work. Before the collapse, Wikipedia's short entry for the I-35W bridge was classified, according the

[30] *Mississippi River Bridge* http://www.en.wikipedia.org/wiki/I-35W_Mississippi_River_Bridge

Quality scale, as a *Stub*. The stub was created in May 2006 and edited only five times before 1st August 2007. During that night the entry became a full page with Wikipedia users adding information on the bridge's construction and history, as well as photos and updates about the collapse (Salas, 2007). In conclusion, as Mejias (2005) claims, the new writing modality embodied in Wikipedia can teach us about the responsibilities of social collaboration, the need for continuously updating information and the permanently unfinished state of human knowledge.

3. THEORETICAL FRAMEWORK OF REFERENCE AND METHODOLOGY

1. Framework of Reference

As the figure below shows, language can be thought in hierarchical terms so that morphemes form words, which form phrases, which form clauses, which form more complex clauses, which form discourse or text (fig. 1):

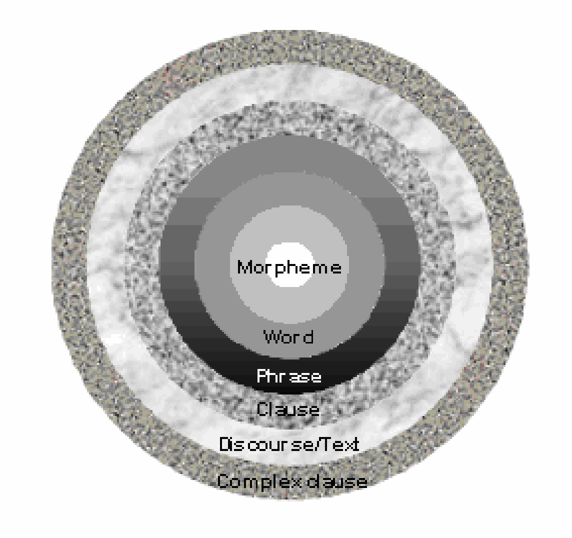


Fig. 1 Language hierarchical structure

Traditionally, grammatical description has been focused on phrase and clause level phenomena, but nowadays many linguists view grammar in the way Larsen-Freeman (1997) does as grammatical phrase or clause level choices are not independent of context and can properly be understood only in relation to it:

> Grammar does operate at the sentence level and governs the syntax or word orders that are permissible in the language. It also works at the subsentence level to govern such things as number and person agreement between subject and verb in a sentence. However, grammar rules also apply at the suprasentential or discourse level. For example, not every choice between the use of the past and the present perfect tense can be explained at the sentence level. Often, the speaker's choice to use one or the other can only be understood by examining the discourse context. Similarly, use of the definite article with a particular noun phrase after the noun phrase has been introduced in a text is a discourse-governed phenomenon. Much of the apparent arbitrariness of grammar disappears when it is viewed from a discourse-level perspective.

The approach which has been adopted, in trying to provide exhaustive answers to the research questions of this study, has been mainly based on a quantitative linguistic analysis. Its main purpose has been to find discrepancies and similarities inside the encyclopaedic genre and variations in the registers adopted in the different Wikipedia writing spaces. Thus, the frequency perspective has been adopted in order to verify whether or not the different grammatical choices are correlated to the original contexts of production (e.g. *Britannica* vs. *Wikipedia* - *Wikipedia* vs. *talkpages*) and whether they influence, or not, the nature and the quality of the discourse produced.

The descriptive approach which has been used in this study has combined corpus linguistics and a simplified factor analysis. It has associated quantitative analysis and computational techniques. It has also been complemented with some qualitative interpretations in order to confirm and enrich the empirical investigation and to verify the register variations and the respect of the web usability principles. The theoretical framework of reference has mainly been based on Douglas Biber's (1988, 1998, 2005) *Multidimensional Approach* (also known as *Factor Analysis)* and on the studies of Heylighen and Dewaele (1999) and Chafe (1987) on formality and variations within registers.

1.1 Biber's Multidimensional Approach to Register Variation

The Multidimensional Approach to register variation has been originally developed by Biber (1984) in his contrastive analysis of spoken and written registers in a variety of different English texts. The label *register* is used, also in this study, as a cover term for any change associated to a language variation.

Methodologically Biber's approach uses computer based text corpora[31] and sophisticated computational tools (such as automated *Tagging Programs*) in order to map the linguistic features characterizing the selected texts. In addition, statistical techniques (*Pearson Correlation Coefficients* and *Factor Analysis*) analyse the co-occurrence relations among the selected linguistic features, to identify the underlying dimensions of variation in the language. In brief, Biber's approach allows the identification of distinct groupings of linguistic features, that co-occur frequently in texts, which have been interpreted in terms of the communicative functions shared by the co-occurring features.

Specifically, Biber's approach is based on the investigation of a range of significant linguistic features counted in 500 spoken and written text samples taken from different registers: from telephone to face to face conversations, personal letters, fictions, broadcasts, biographies, prepared speech, academic papers, fiction, etc. Biber computes the *factor score* mapping the frequency of the linguistic features whose incidence (which he defines *loading*) will define and portray the specific linguistic dimension underlying the selected texts. Biber's analysis has defined six different dimensions (1998) to which he has attached the six interpretative labels which follow:

- Dimension 1 - Informational *vs.* Involved production
- Dimension 2 - Narrative *vs.* Non Narrative discourse
- Dimension 3 - Situation-dependent *vs.* Explicit reference
- Dimension 4 - Overt expression of persuasion
- Dimension 5 - Non abstract *vs.* Abstract style
- Dimension 6 - On-line Informational elaboration marking stance

[31] Texts are from the London-Lund corpus of Spoken English Corpus and the Lancaster-Oslo-Bergen Corpus of British English

Each dimension is based on the frequency variation of 67 linguistic classes (tenses, place and time adverbs, modals, contractions, negative forms, pronouns, nominalizations, gerunds, passives nouns, type/token ratio, word length, etc.) which have a positive or a negative loading in defining each specific dimension. In brief, this methodology is explicitly multidimensional, as it assumes that multiple parameters of variations are operative in any discourse domain. Biber's analysis shows *there is no single, absolute difference between speech and writing in English rather there are several dimensions of variation* (Biber 1988:199).

The comparative perspective has a keyrole in this approach since statistical data is interpreted in functional terms to determine the underlying communicative functions associated with each distributional pattern. The resulting dimensions are fundamental parameters of variation among English texts (Biber 1988: 200).

Biber intra/inter-genre analysis is based on different sampling techniques which allow different kinds of analysis. For instance, a sampling that extracts sections that are homogeneous with regard to purpose and topic allows a linguistic investigation on the distinctive characteristics of specific registers. When the comparison of frequency counts shows an extremely high degree of stability across samples, this result indicates that there is an internal linguistic consistency within registers adopted (as it will be shown in the specific case of *Britannica vs. Wikipedia*). On the other hand, an intergenre analysis uses sampling that disregards the changing textual purposes and topics and allows an overall characterization of the register variations (as it happens when *Wikipedia articles* and Wikipedia talk pages will be compared).

1.2 Written and Oral Discourse vs. *WikiLanguage* and *Wikispeak*

Texts, whether spoken, written or mediated, are produced in context. They have particular production circumstances that directly affect the register, that is to say the kind of language used. Circumstantial factors which have been identified as fundamental in many studies (Halliday, Swales, Biber et *al.*) are: the *participants* (their relationships and attitudes towards communication), the *setting* (including factors such as the extent to which time and place are shared by the participants and the level of formality), the *channel of communication*, the production and processing *time* (e.g. amount of time available), the communicative *purpose*, the *topic* (or subject matter), etc.

The two stereotypical and extreme modalities of oral and written discourse are, according to Biber, *involved* vs. *informational* production. As it has also been argued by Lakoff (1982, in Calude, 2005), the spoken/written dichotomy is not so strict, since the two modes cannot be separated by a rigid set of criteria. It has been found that, similarly to other linguistic phenomena, each of the two modes has further internal subdivisions.

John Gumperz, Wallace Chafe and Deborah Tannen, without taking into account electronic language case, had already claimed that this dichotomy was relative and in some cases unnecessary. They noted the existence of certain strategies common both to oral and written discourse, in which typical of written discourse strategies are applied to oral discourse and viceversa.

Biber confirms this approach to spoken and written discourse, claiming that the most frequently used clusters of grammar that represent functional dimensions can be thought of as forming a *continuum*[32]. He defines the polar ends of this continuum, as *involved vs. informational production.*The notion of *continuum* between oral and written discourse arises from the observation that some of the language types found in one mode share characteristics with language types found in the other mode[33].

Involved production mainly refers to a type of interactive communication. Traditionally it is more oral then written and informal when compared to informational production. The latter, on the other end, refers to a more detached, accurate and formal written communication. It is considered more literate when compared to the former. However, it is important to keep in mind that informational communication can be used in speech, and that involved communication can also be written.

The *continuum* from oral to written discourse is provided by online communication, where evidence of non spontaneous spoken discourse and spontaneous written discourse are conveyed in CMC. Spoken and written discourse appear to be in CMC not so much a dichotomy, but rather two aspects of the same phenomenon. It seems that online, the discourse producer moves itself and chooses the options which better conform to the means available for exchange purposes since strategies and resources of both forms are not different. What differs is mainly the medium, and the specific purpose of communication. In brief, the CMC drives this possibility to the extreme, endowing writing with the most typical features of the spoken discourse, as it will be shown in the next chapters. CMC and its numerous varieties spoken (written) in online communities (such as *WikiSpeak* in Wikipedia Community), represent a melting point which although in digital format, appropriately convey this *continuum* from informational to involved production.

This research will demonstrate that informational and involved production share similar core features whether they are delivered in their traditional oral or paper format or in online environments. In terms of its situational features, involved production is stereotypically represented by *face to face* conversation which is interactive, and dependent on shared space, time and background knowledge.

[32] According to Biber, within the domain of speech are included spontaneous conversations, talkback radio, monologues, public lectures, speeches, news broadcasts and so on. While within the domain of written language there are books and journals, newspapers, letters, notes diary entries, etc.

[33] According to Biber, for instance a diary entry, though typically shares a lot of characteristics with spoken language: it is unplanned, informal and it does not have the organization of a well-formed text, etc. Similarly, a prepared speech or lecture is very much like a written text since it is often first realized as a written text (people like to take notes of what they are about to say), it is planned and coherent, with an introduction, contents and conclusion and it often involves formal language.

In this research involved production is associated to *talk pages* where Wikipedians discuss how to improve content, quality and style of encyclopaedic articles. By contrast formal writing, stereotypically represented by the *informational production*, has been associated to Wikipedia *encyclopaedic articles* in this specific research.

As already mentioned, Biber's approach is based on the assumption that statistical co-occurrence patterns reflect underlying shared communicative functions. He claims that registers can be compared along the same dimension (e.g. intragenre analysis: *Britannica articles* vs. *Wikipedia articles* in this specific case) and frequency of linguistic features should co-occur if a similar register is used. On the other hand, Biber's statistical approach also allows a comparison between different registers (e.g. intergenre analysis: *Wikipedia articles* vs. *talkpages*). For this reason, the present study has mapped, recorded and compared linguistic variations between informational production, represented by encyclopaedic corpora, and involved production of talkpages.

Since the encyclopaedic genre is characterized by an informational and explicit style, it includes features belonging mostly to the linguistic classes identified by Biber's Factor analysis as Dimension 1 (Informational vs. Involved). Biber's Dimension 1 refers to discourse with highly informational purposes, carefully crafted and highly edited. This dimension seems to be dominant in the characterization of encyclopaedic genre. By contrast, involved production, seems to reflect linguistic peculiarities typical of talk pages.

This research is based on the frequency observation of a number of variables that have been considered linguistically significant in the textual units examined. Among the 67 classes identified by Biber, 23 linguistic features have been selected. The selection has been based on what has been considered distinctive in the definition of the two specific analyzed dimensions.

1.3 Other studies on Oral and Written Discourse

Register variations have also been investigated by other scholars. Heylighen and Dewaele (1999) divide the words of the lexicon into two classes, depending on whether they are used mainly to build more context dependent/or independent speech. Speakers in "context dependent speech" use a lot of words with a deictic function, referring to the spatio-temporal or communicative context (Heilighen and Dewaele, 1999). As a result of this concern, involved production often has a distinctly non-informational and fuzzy character (marked by hedges, and forms of reduced or generalized content). Furthermore, Chafe and Danielwicz (1987) propose the concept of *functional notions* (*integration, fragmentation, involvement and detachment)* that is particularly useful in the interpretation of different textual dimensions. Each of these functions can be applied to a particular aspect of the informational or involved production marked by peculiar linguistic features.

Chafe and Danielwicz claim that *integration* and *detachment* are the typical qualities of formal writing, whose textual production is marked by agentless passives and nominalizations, frequent nouns, adjectives, prepositional phrases, high lexical density, frequent long words, complex vocabulary and greater use of nominal structures. By contrast, they claim that *involvement* and *fragmentation* refer to those linguistic features which reflect the fact that actors involved in the verbal exchange typically interact with one another.

2. Methodology

This research which is mainly a frequency based study, uses a descriptive approach and empirically observes and compares three different subcorpora.

2.1 Corpus and Folksonomy

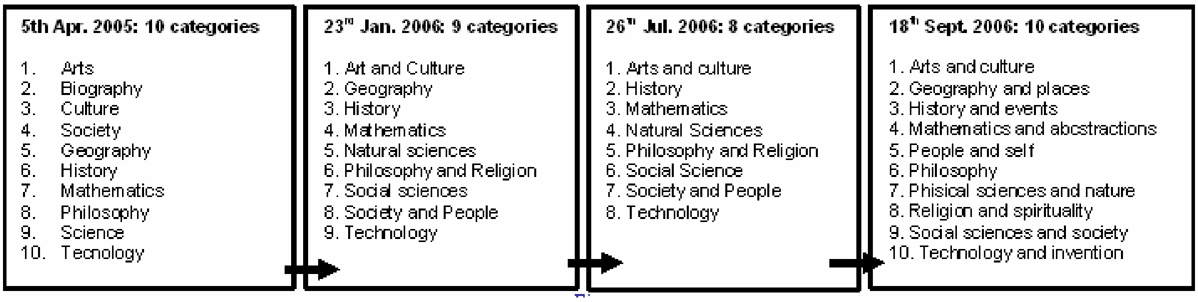
The overall corpus of reference is of 1,240,482 words (tokens). It is made up of three subcorpora: Britannica and Wikipedia encyclopaedic corpora with respectively, 247,103 and 391,637 tokens and Wikipedian talk pages with 601,742 tokens (fig. 2).

Corpus	
Britannica encyclopaedic articles	247,103 (tokens)
Wikipedia encyclopaedic articles	391,637 (tokens)
Wikipedian talk pages	601,742 (tokens)
Total Corpus	**1,240,482 (tokens)**

Fig. 2 Reference Corpus

The Encyclopaedic corpus is made up of an equal number of articles that appear in both Encyclopaedia Britannica and Wikipedia websites. The first two subcorpora include one hundred articles randomly selected from the ten categories of Wikipedia Folksonomy and by the one hundred equivalent articles found in Encyclopaedia Britannica Online. The selection includes encyclopaedic articles of different quality and at different evolution stages as testified by the identification label used by the Wikipedian department of the *Heraldry and Vexillology* (which assesses the quality of Wikipedia's articles). Some articles belong to the ★***FA** Class* (Featured Articles – the best ones), some to *A-class* (Articles with well written texts and contents), ⊕ ***GA Class*** (good article) and ***B-Class*** (articles to be improved). The selection has excluded articles belonging to the ***Start*** and ***Stub class*** (the former are articles too weak in many areas and lacking the key elements, while the latter are too short to provide encyclopaedic coverage of the subject).

The Wikipedian talk page corpus is made up of the back-office nodes associated to the 100 selected encyclopaedic articles where Wikipedians debate to improve the quality and the content of the related encyclopaedic pages. The folksonomy of Wikipedia, as well as its encyclopaedic articles and associated talk pages are dynamic and, thus, in constant evolution. This aspect has obviously represented a critical point in data collection and cataloguing.



5th Apr. 2005: 10 categories

1. Arts
2. Biography
3. Culture
4. Society
5. Geography
6. History
7. Mathematics
8. Philosophy
9. Science
10. Tecnology

23rd Jan. 2006: 9 categories

1. Art and Culture
2. Geography
3. History
4. Mathematics
5. Natural sciences
6. Philosophy and Religion
7. Social sciences
8. Society and People
9. Technology

26th Jul. 2006: 8 categories

1. Arts and culture
2. History
3. Mathematics
4. Natural Sciences
5. Philosophy and Religion
6. Social Science
7. Society and People
8. Technology

18th Sept. 2006: 10 categories

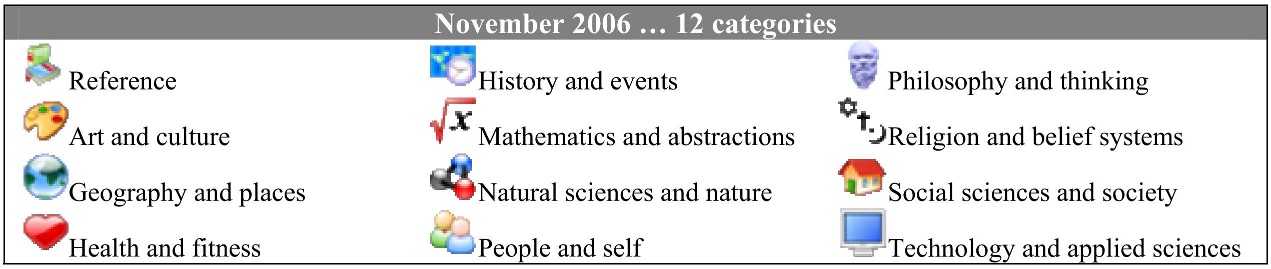
1. Arts and culture
2. Geography and places
3. History and events
4. Mathematics and abcstractions
5. People and self
6. Philosophy
7. Phisical sciences and nature
8. Religion and spirituality
9. Social sciences and society
10. Technology and invention

Fig. 3 Evolution of Wikipedia's categories

For example Wikipedia's Folksonomy (fig. 3) was made up of ten categories when the first survey was carried out (April 2005). Then, they became nine in January 2006 since the category 'Biography' disappeared. Some slight changes were also introduced in the name categories.

In July 2006, after the removal of "Geography", the categories became eight and, again they grew to ten on the 18th of September 2006.

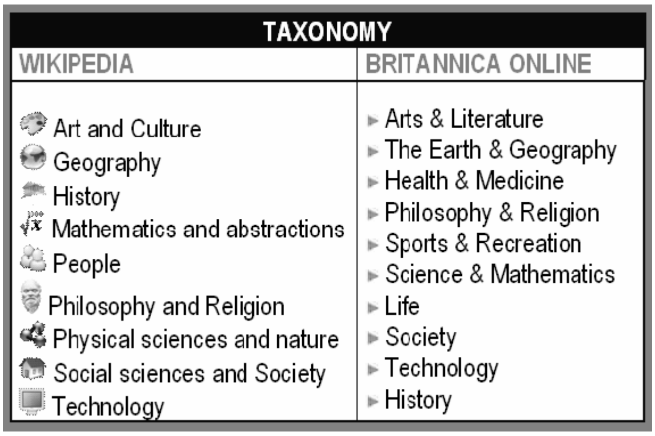
The category "Geography and places" was added and the category "Philosophy and Religion" was then subdivided in two new classes "Philosophy" and "Religion and Spirituality". The previous ten categories increased again to become twelve in November 2006 (fig. 4). The two new added categories were: 'Reference' and 'Health and fitness'. The previous categories "Philosophy" and "Religion" were renamed as "Philosophy and thinking" and "Religion and belief systems".



November 2006 … 12 categories		
Reference	History and events	Philosophy and thinking
Art and culture	Mathematics and abstractions	Religion and belief systems
Geography and places	Natural sciences and nature	Social sciences and society
Health and fitness	People and self	Technology and applied sciences

Fig. 4 November 2006 Wikipedia's categories

Despite the persistent changes in Wikipedia folksonomy, its basic taxonomy is not very dissimilar from Britannica's which has been, since the beginning of this research, consistently structured in ten steady 'subjects' and subdivided in more specific and equally stable subcategories (Fig. 5).



TAXONOMY

WIKIPEDIA	BRITANNICA ONLINE
Art and Culture	▸ Arts & Literature
Geography	▸ The Earth & Geography
History	▸ Health & Medicine
Mathematics and abstractions	▸ Philosophy & Religion
People	▸ Sports & Recreation
Philosophy and Religion	▸ Science & Mathematics
Physical sciences and nature	▸ Life
Social sciences and Society	▸ Society
Technology	▸ Technology
	▸ History

Fig. 5 Britannica and Wikipedia's taxonomy

Encyclopaedic articles and talk pages which are the reference corpus of this research have been mainly selected during the first semester of 2006. Thus, reference to that original classification system has been maintained during the present investigation for different reasons; first of all, because there was a natural numerical and topical matching with Britannica categories, secondly because the evolution of Wikipedia's Folksonomy is extremely fluid and too fast to be constantly followed and, above all, because it was not influential to the objectives of the present research.

Arts	Biography	Culture	Society	Geography
Cinemascope	Beatles	Diaspora	Alcoholism	Barcelona
Colosseum	Benjamin Franklin	Fairy tale	Euro	Bermuda triangle
Graffiti	Bill Gates	Flag	Feminism	Gobi desert
Holography	Albert Einstein	Geisha	Homosexuality	Hydrography
Proscenium	Fred Astaire	Jazz Dance	Women's suffrage	Himalaya
Jazz	James Dean	Pizza	Poverty	Klodzko
Madonna	Karl Marx	Romanticism	Racism	London
Polka	Adam Smith	Superstition	Tamil	Piccadilly Circus
U2	Vittorio Alfieri	Tea	Terrorism	San Josè
Wind rose	C. Columbus	Walt Disney	Zulu	Weather

History	Mathematics	Philosophy	Science	Technology
Anne Frank	Boolean algebra	Agnosticism	AIDS	Balloon
Aztec	Catastrophe theory	Aristotle	Big Bang	Gasoline
Silvio Berlusconi	Cryptography	Francis Bacon	Heart	Internet
Tony Blair	Graph theory	Epistemology	Neuron	Jet engine
British East India	Matrix	Michel Foucault	Nuclear weapon	Microprocessor
Wars of the Roses	Numerical analysis	Frankfurt school	Pneumonia	Microsoft
Ku Klux Klan	Pythagorean theorem	Philosophy of mind	Royal Astr.	Radar
Giuseppe	Quantum number	Skepticism	Sars	Typewriter
French Revolution	Real number	Thomas Huxley	Solar energy	Virtual Reality
George Bush	Vector space	Wittgenstein	Turquoise	World Wide Web

Fig. 6 Articles selected from Wikipedia and Britannica

As already mentioned, ten sample articles have been randomly selected from each category and analyzed. The advantage of representativeness and generalizability has been offered by the random technique. The choice of the same number of articles taken from the two encyclopaedias has given topical coherence to the present investigation. The two hundred articles which have been chosen are shown in fig. 6.

2.2 Working Phases

2.2.1 Research Area Definition

The first year of this doctoral research (2005) has been dedicated to a detailed and exhaustive literature review for the purpose of defining both the research area and the specific research questions. In addition to printed books and papers, most of the literature on the topic, has been found searching the web. The service offered by *Google Alerts* has been very useful to this purpose, since it provides e-mail notifications to the individual users about the latest pages published on the web on the chosen topic. During the first year, Wikipedia and Britannica websites have been analysed according to the principles of Web Usability and Readability. Sociolinguistcs dynamics within Wikipedia Community have also been observed.

2.2.2 Data collection, Rationalisation of Data and Corpus Definition

The second year of this research (2006) has been dedicated to the creation of a coherent and representative corpus made up of a collection of encyclopaedic articles (and talk pages in Wikipedia) available on the websites of the two encyclopaedias: Britannica (http://www.britannica.com) and Wikipedia (http://www.wikipedia.org). Articles in their original html format have been downloaded and saved. The Wikipedian talk pages associated to the main encyclopaedic articles have also been archived. Data has then been rationalised, according to content criteria. Thus, articles and talk pages have been cleaned out by removing information irrilevant to the specific content body (e.g. index of contents, graphs, photos and tables, references and extra links).

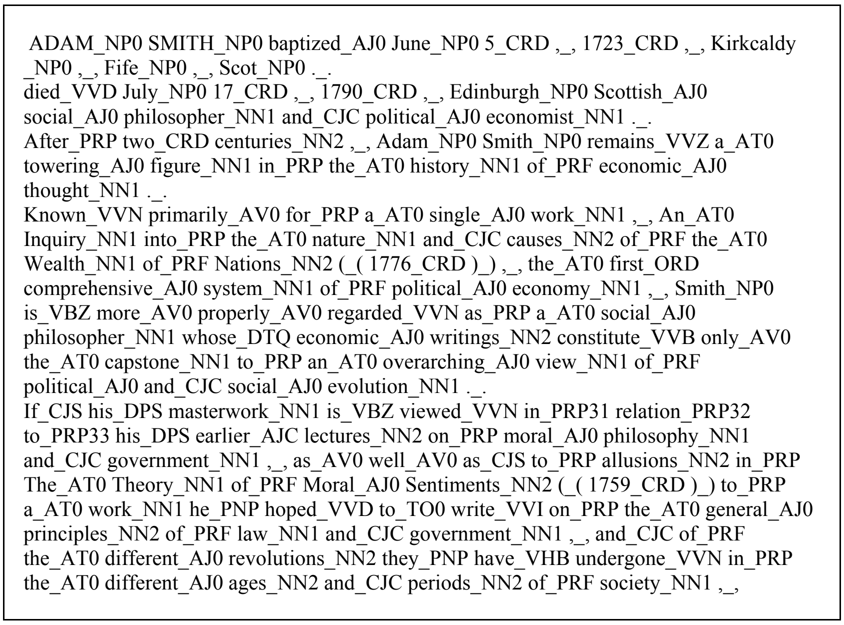
2.2.3 Conversion of texts and computational working tools

In order to allow a statistical quantitative analysis, encyclopaedic articles and talk pages in html format (.html) have been transformed in txt format (.txt) The program *HTMLAsText* (v1.05) has been used to convert HTML documents into simple text files.

Most of the statistical analyses have been carried out through the *Concordancer* program *AntConc*, developed by Laurence Anthony, at the School of Science and Engineering, Waseda University in Japan. *AntConc* is a lexical analysis tool, which can be used to search for keywords, perform concordance searches, etc. It has been mainly used to create word lists useful to compute the frequency of the linguistic classes considered functional for the purposes of this research.

Different concordancer programs, such as *Wordsmith tool* (a proprietary software developed by Mike Scott at the Oxford University Press) and *ConCapp* (an open source software developed by Chris Greaves at the Hong Kong Polytechnic University) have also been used, to verify the reliability of the findings and to make use of specific functions not available in other programs. The *TextAlyzer* program (hosted by *Lexicool.com website* http://www.lexicool.com) has also been employed for specific analysis such as *Index of Readability* and *Lexical Density.* In order to measure the hypertextual structures of the two encyclopaedias, the website *Link Analyzer Tool* has been used. In addition, the Microsoft program *Excel* has allowed the creation of dynamic data sheets and the automatic updating in the case of data variation. The mentioned softwares and websites have allowed the measurement of the selected linguistic features by means of a statistical quantitative approach to linguistic analysis not otherwise achievable through a traditional textual analysis methodology. Although concordancer software has been very useful to facilitate quantitative analysis, nevertheless, it has often been necessary to supplement the automatic analysis with a manual inspection to evaluate information in context.

CLAWS part-of-speech tagging software (developed at *Lancaster University)* has been another very useful tool employed in this research. CLAWS runs text files through a program which feeds them to a tagger assigning each word (or word combination) a particular part-of-speech. The tagset which has been chosen is *C5*; it includes over 60 tags and it is also used by *British National Corpus* (BNC). An excerpt of what a tagged portion of speech in the original version looks like is shown below (fig. 7).



```
 ADAM_NP0 SMITH_NP0 baptized_AJ0 June_NP0 5_CRD ,_, 1723_CRD ,_, Kirkcaldy
_NP0 ,_, Fife_NP0 ,_, Scot_NP0 ._.
died_VVD July_NP0 17_CRD ,_, 1790_CRD ,_, Edinburgh_NP0 Scottish_AJ0
social_AJ0 philosopher_NN1 and_CJC political_AJ0 economist_NN1 ._.
After_PRP two_CRD centuries_NN2 ,_, Adam_NP0 Smith_NP0 remains_VVZ a_AT0
towering_AJ0 figure_NN1 in_PRP the_AT0 history_NN1 of_PRF economic_AJ0
thought_NN1 ._.
Known_VVN primarily_AV0 for_PRP a_AT0 single_AJ0 work_NN1 ,_, An_AT0
Inquiry_NN1 into_PRP the_AT0 nature_NN1 and_CJC causes_NN2 of_PRF the_AT0
Wealth_NN1 of_PRF Nations_NN2 (_( 1776_CRD )_) ,_, the_AT0 first_ORD
comprehensive_AJ0 system_NN1 of_PRF political_AJ0 economy_NN1 ,_, Smith_NP0
is_VBZ more_AV0 properly_AV0 regarded_VVN as_PRP a_AT0 social_AJ0
philosopher_NN1 whose_DTQ economic_AJ0 writings_NN2 constitute_VVB only_AV0
the_AT0 capstone_NN1 to_PRP an_AT0 overarching_AJ0 view_NN1 of_PRF
political_AJ0 and_CJC social_AJ0 evolution_NN1 ._.
If_CJS his_DPS masterwork_NN1 is_VBZ viewed_VVN in_PRP31 relation_PRP32
to_PRP33 his_DPS earlier_AJC lectures_NN2 on_PRP moral_AJ0 philosophy_NN1
and_CJC government_NN1 ,_, as_AV0 well_AV0 as_CJS to_PRP allusions_NN2 in_PRP
The_AT0 Theory_NN1 of_PRF Moral_AJ0 Sentiments_NN2 (_( 1759_CRD )_) to_PRP
a_AT0 work_NN1 he_PNP hoped_VVD to_TO0 write_VVI on_PRP the_AT0 general_AJ0
principles_NN2 of_PRF law_NN1 and_CJC government_NN1 ,_, and_CJC of_PRF
the_AT0 different_AJ0 revolutions_NN2 they_PNP have_VHB undergone_VVN in_PRP
the_AT0 different_AJ0 ages_NN2 and_CJC periods_NN2 of_PRF society_NN1 ,_,
```

Fig. 7 Example of *CLAWS part-of-speech tagging*

CLAWS software has been very useful in searching the frequency of nouns and adjectives which could not be otherwise detected. Once samples have been tagged by CLAWS software, the *Antconc* progam has allowed to quantify specific occurrences of nouns and adjectives.

2.2.4 Classification and Analysis of Data

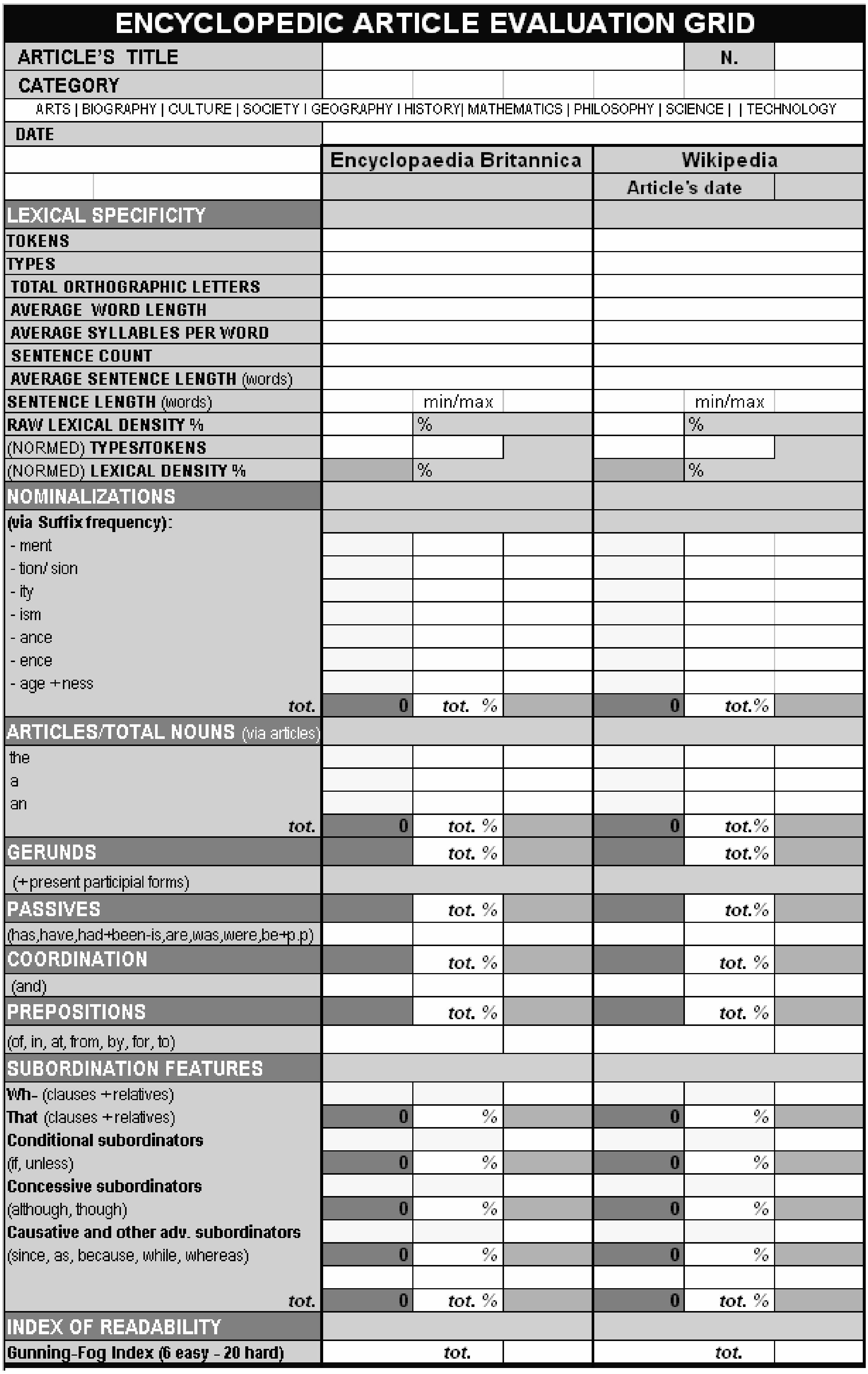
The main objective of this research phase has been the construction of a critical interpretative model and of a coherent system for the data classification and analysis.

Specifically, the co-occurrence of different linguistic features has been identified by a statistical technique based on frequency criteria mainly inspired to Biber's approach. Through this analysing technique, the correlations among the selected variables have been identified and grouped together. According to Biber, the group of salient linguistic variables is represented by the factor, which has been functionally interpreted in this specific research as "dimension" of register variation. Specifically, the internal register variation between the encyclopaedic expository style of Britannica and Wikipedia has been calculated. Furthermore, the cross linguistic comparison between the informational *WikiLanguage* and the involved and conversational *WikiSpeak* has allowed the measurement of register variation in the two different Wikipedian writing spaces (*document* mode vs. *thread mode).* The linguistic analysis has been carried out at micro/macroscopic levels.

2.2.5 Micro Analysis Phase

The micro analysis has been carried out only on encyclopaedic articles; talk pages have been excluded from this detailed investigation. In this first analysis, the frequency of a number of selected linguistic classes with a positive loading on formal encyclopedic expository style has been investigated for each article, and recorded in 100 specific *encyclopaedic Evaluation Sheets* (fig. 8). In this way, a detailed portrait of the micro textual distribution of the selected linguistic classes has been obtained.

The overall findings have then been gathered in a general "Overview table" where totals and averages have been juxtaposed to provide a broader reference frame (see *Appendix*). The analysis of some linguistic features has been sacrificed during this phase, or approximately calculated, since the reference corpus is not thoroughly annotated, an automatic or semiautomatic detection was impossible to run. As a consequence adjectives, typical of the formal expository style, have been excluded from the microscopic investigation and an indirect and approximate counting of nouns (rounded down) has been made through the occurrences of definite and indefinite articles.



ENCYCLOPEDIC ARTICLE EVALUATION GRID						
ARTICLE'S TITLE					N.	
CATEGORY						
ARTS | BIOGRAPHY | CULTURE | SOCIETY I GEOGRAPHY I HISTORY| MATHEMATICS | PHILOSOPHY | SCIENCE | | TECHNOLOGY						
DATE						
	Encyclopaedia Britannica			Wikipedia		
				Article's date		
LEXICAL SPECIFICITY						
TOKENS						
TYPES						
TOTAL ORTHOGRAPHIC LETTERS						
AVERAGE WORD LENGTH						
AVERAGE SYLLABLES PER WORD						
SENTENCE COUNT						
AVERAGE SENTENCE LENGTH (words)						
SENTENCE LENGTH (words)		min/max			min/max	
RAW LEXICAL DENSITY %		%			%	
(NORMED) TYPES/TOKENS						
(NORMED) LEXICAL DENSITY %		%			%	
NOMINALIZATIONS						
(via Suffix frequency):						
- ment						
- tion/ sion						
- ity						
- ism						
- ance						
- ence						
- age +ness						
tot.	0	tot. %		0	tot.%	
ARTICLES/TOTAL NOUNS (via articles)						
the						
a						
an						
tot.	0	tot. %		0	tot.%	
GERUNDS		tot. %			tot.%	
(+present participial forms)						
PASSIVES		tot. %			tot.%	
(has,have,had+been-is,are,was,were,be+p.p)						
COORDINATION		tot. %			tot. %	
(and)						
PREPOSITIONS		tot. %			tot. %	
(of, in, at, from, by, for, to)						
SUBORDINATION FEATURES						
Wh- (clauses +relatives)						
That (clauses +relatives)	0	%		0	%	
Conditional subordinators						
(if, unless)	0	%		0	%	
Concessive subordinators						
(although, though)	0	%		0	%	
Causative and other adv. subordinators						
(since, as, because, while, whereas)	0	%		0	%	
tot.	0	tot. %		0	tot. %	
INDEX OF READABILITY						
Gunning-Fog Index (6 easy - 20 hard)		tot.			tot.	

Fig. 8 Encyclopaedic Article Evaluation Grid

2.2.6 Macro Analysis Phase

In this phase, the macroscopic contrastive analysis carried out on the three subcorpora (*Britannica* vs. *Wikipedia*; *Wikipedia articles* vs. *Wikipedia talk pages*) has been carried out to measure the intra and inter genre register variation.

As already mentioned, two of the main research questions guiding the statistical analysis have been: *How much does the formality of the expository texts differ in the two encyclopaedias? What is the variation between the WikiSpeak and the WikiLanguage?* To this purpose, the linguistic features listed below (fig. 9) have been investigated in each of the three subcorpora.

LINGUISTIC CLASSES	
+	***Word length (characters)***
	Sentence length (words)
	Lexical density (tokens/types)
	Nominalizations
	Gerunds and present participles
	Definite/indefinite articles
	Nouns
	Adjectives
	Prepositional phrases
	Passives
	Subordination features
	Coordinating conjunctions
	Conjuncts
-	
	Place adverbials
	Time adverbials
	Person pronouns
	Demonstratives
	Infinitive pronouns
	Mitigating and Boostering devices
	Modals
	Lexical verbs
	Negative forms
	Interrogative sentences

Fig. 9 Linguistic Classes analyzed

Compared to the microanalysis phase, the queries made directly on the three subcorpora have allowed deeper investigations on additional and more specific linguistic classes.

Findings on total corpora (Britannica/Wikipedia/Talkpages), have been higlighted in tables and graphs, which have proved to be extremely useful tools to record, measure, compare and visualize more systematically data.

2.2.7 Sample Tagging

Since the reference subcorpora are not completely tagged, to compute the occurrence of nouns and adjectives in Britannica and Wikipedia, an inferential statistical approach has been used. The frequency of nouns and adjectives, has been calculated, using the *WWW CLAWS part-of-speech tagging software*. The Tagset *C5* has tagged nouns and adjectives (fig. 10) according to the following criteria:

NOUNS	
NN0	Noun (neutral for number) (e.g. *aircraft, data*)
NN1	Singular Noun (e.g. *pencil, goose)*
NN2	Plural Noun (e.g. *pencils, geese)*
NP0	Proper Noun (e.g. *London, Michael, Mars)*

ADJECTIVES	
AJ0	Adjective (unmarked) (e.g. *good, old)*
AJC	Comparative Adjective (e.g. *better, older)*
AJS	Superlative Adjective (e.g. *best, oldest)*

Fig. 10 Claws' Tags for nouns and adjectives

Applying the random sampling technique, the initial 10,000 tokens of the first 10 articles of each encyclopaedic category have been extracted and tagged online by the *WWW CLAWS tagging software*. In a second phase, *AntConc* concordancing program has allowed to compute specific frequencies.

2.2.8 Normalization of Frequency Count

Since the linguistic investigation is mainly frequency based, the count of occurrences has been normalized to make the quantitative findings comparable (*relative frequency*).

Normalization of frequency count has been made following Biber's theory which demonstrates that raw frequency counts are not directly comparable when textual units have different lengths.

Biber (1998:263) points out to this end:

> When corpus based studies examine the frequency of features across text and registers, it is important to make sure that the counts are comparable. Normalization is a way to adjust raw frequency count from texts of different lengths so that they can be compared accurately. […] the total number of words in each text must be taken into consideration when norming frequency counts. Specifically the raw frequency count should be divided by the number of words in the text, and then multiplied by whatever basis is chosen for norming.

In this study normalization has been made on a basis of 100 words per text. The choice of 100 words has been determined by the length of the shortest article found. To this purpose the following formula has been applied:

$$\textbf{Count of Occurrences : Tokens = X : 100} \rightarrow \textbf{X} = \frac{\textbf{Count of Occurrences * 100}}{\textbf{Tokens}}$$

Thus, *absolute frequencies* (total occurrences) have been multiply for the basis chosen for normalization (*relative frequency* per 100 words) and then divided by the total number of words in the text (total tokens).

2.2.9 Chi-Square Test

All the data resulting from the linguistic analyses has been submitted to *Chi-Square* test to assess its specific reliability.

Inferential statistics provides an important tool for assessing whether observed patterns are meaningful. There are many different statistical techniques (e.g. *T-test, Chi Square, Anova, Log Likelihood,* etc.) which can be applied depending on the types of variable under observation (Biber, 1998: 275).

The Chi-Square (X^2) test has been chosen among the different significance testings. It has been calculated online by the *Interactive Calculation Tool for Chi Square Test* [34] by Preacher (2001). It has been a precious tool since it has allowed to automatically chi-square the findings of this research.

The test has allowed to determine whether or not the quantitative difference, in the frequency of the linguistic classes analysed, is the result of a genuine variation between two or more items, or whether it is just due to chance (Baroni, 2006). Thus, the *Null Hypothesis,* that is to say the hypothesis we want to test (the variation is due to chance), can be confirmed or rejected by the Chi-Square test.

Chi-Square test is based on absolute and not on relative (normed) frequencies. In order to chi quare all the linguistic classes analyzed, the online *Calculation Tool* has been filled in with the data shown in fig. 11a. A specific example is provided for nominalizations (fig. 11b).

Interactive Calculation Tool for Chi Square Test (University of Kansas)
http://www.psych.ku.edu/preacher/chisq/chisq.htm

	Corpus 1	**Corpus 2**	**Total**
Frequency of search item(s)	**A**	**B**	**A+B**
Frequency of other words	**C-A**	**D-B**	**C+D -A-B**
Total number of words in corpus	**C**	**D**	**C+D**

Fig.. 11a Absolute frequency's calculation

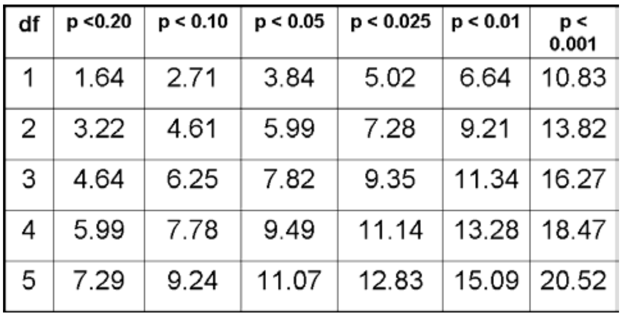
	Britannica	**Wikipedia**	**Total**
Frequency of nominalizations	13014	18110	**A+B**
Frequency of other words	247103 – 13014 = 234089	39163 – 18110 = 373527	**C+D -A-B**
Total number of words in corpus	247103	391637	**C+D**

Fig.. 11b Example of nominalizations' absolute frequency

Chi Square test is based on a *Contingency* table (fig. 12) which has been consulted to find the exact cause of significance. Thus, the *degree of freedom* (df) has been verified (R stands for *Raw* while C for *Column).*

$$df = (r-1)(c-1)$$

Since each selected linguistic category has been individually assessed, the degree of freedom always turns out to be "1", thus the first line has been taken into account. Then, the P-value (p), which indicates the probability of error, has been checked.



df	p <0.20	p < 0.10	p < 0.05	p < 0.025	p < 0.01	p < 0.001
1	1.64	2.71	3.84	5.02	6.64	10.83
2	3.22	4.61	5.99	7.28	9.21	13.82
3	4.64	6.25	7.82	9.35	11.34	16.27
4	5.99	7.78	9.49	11.14	13.28	18.47
5	7.29	9.24	11.07	12.83	15.09	20.52

Fig. 12 Contingency table (–up to df 5)

When $p < 0.05$ it means that there is 5% probability that the difference detected is due to a random variation. In other words, there is a 95% probability that the difference is a true reflection of variation in the two corpora, thus it is significant and not due to chance. In this case the *Null Hypothesis* is rejected.

In corpus linguistics, probability values of less than 0.05 (written as p> 0.05) are assumed to be significant, whereas those greater then 0.05 are not.

To give an example, the Chi-Square in fig. 13 shows the frequency of nominalizations in the two encyclopaedic corpora. Since the X^2 resulting value is 134,909 df 1 has been consulted and the P-value is < 0,0001. This data means that with a probability of 99.99% the difference reflects a significant variation in the two corpora, thus it is not due to chance.

	Britannica	**Wikipedia**	**Total**
Frequency of nominalizations	13014	18110	**A+B**
Frequency of other words	247103 – 13014 = 234089	39163 – 18110 = 373527	**C+D -A-B**
Total number of words in corpus	247103	391637	**C+D**
X^2	**134,909**		

Fig. 13 Frequency of nominalizations in BAs vs. WAs

Specific findings of this research are shown in the next sections.

4. WIKIPEDIA AS ENCYCLOPAEDIA: FINDINGS

1. Linguistic classes with a positive loading on informational production

For their linguistic peculiarities and communicative purpose, encyclopaedias can be fully included in the category of informational production, as their main aim is to inform, to educate and to present facts and information in specific entries. Biber through the multidimensional analysis has mapped linguistic feature patterns in different typologies of spoken and written English texts.

He claims (Biber, 1988:155) that expository texts are informational, detached, elaborated, highly explicit and context independent. They are characterized by the need for precise and dense packaging of information. He states that the following linguistic features are typically recurrent in informational texts:

> A high frequency of noun, word length, prepositional phrases, lexical density and attribute adjectives can be associated with an high informational focus and a careful integration of information in a text, and a high frequency of nouns, thus indicates great density of information. Prepositional phrases also serve to integrate high amounts of information into a text. Word length and type token ratio similarly mark high density of information, but they further mark very precise lexical choice resulting in an exact presentation of information content. A high token-type ratio results from the use of many different lexical items in a text, and this more varied vocabulary reflects extensive use of words that have very specific meanings. Attribute adjectives are used to further elaborate nominal information. […] Together these 5 elements are used to integrate high amounts of information into a text, to present information as precisely as possible. These features are associated with communicative situations that require a high informational focus.

As informational texts, the presentation of information in encyclopaedias is packed in textual units which make use of an explicit formal expository style. Thus, sharing the traits of informational production, its specific peculiarities have been analyzed in the sections which follow.

1.1 Lexical Specificity

Lexical density (type/token ratio[35]) *word length and sentence length* are the three elements which have been investigated to define *lexical specificity* of Britannica and Wikipedia encyclopaedic corpora. Spoken discourse is produced *on the fly* and is intended to be consumed, heard, in the same rapid and dynamic manner. Written discouse on the other hand is static; it is produced at the pace set by the writer alone and can be consumed at any speed that the reader chooses. The effects of such diversity seems to generate differences in the linguistic production. One of the main aspect concerns vocabulary use.

[35] For *type/token ratio* is meant the number of different words (types) divided by the total number of words (tokens) in a specific textual sample.

Chafe and Danielewicz (1987) claim that as a consequence of these differences, speakers tend to operate with a narrower range of lexical choices than writers. Producing language on the fly, they hardly have time to go through all of the possible choices they might make, and may typically settle on the first words that occur to them.

Biber (1988) also gets to similar conclusions. He claims that lexical specificity seems to be correlated with the production of differences between speaking and writing. An higher *lexical specificity* seems to be associated to formal written genre, marking a high density of information, by reflecting precise word choice and an exact presentation of informational content. The result is that the vocabulary of spoken language is more limited in variety. In order to empirically examine the different use of vocabulary, Chafe and Danielewicz calculate the type/token ratio. Furthermore, Yates study (1996) indicates that CMC is more akin to writing than speech in terms of range of vocabulary used. The most obvious conclusion is to consider lexical specificity as the product of the medium itself, and the opportunity it brings for longer gestation over the content of utterances.

1.1.1 Lexical Density

A high type/token ratio reflects the use of many different words in a text (vs. extensive repetition of relatively few words), representing a more careful word choice and a more precise presentation of informational content. Halliday (1985) also considers a high lexical density typical of formal writing.

In the present research, lexical density has been measured through the type/token ratio. For example, the article *Graffiti* in Encyclopaedia Britannica contains 406 tokens and 224 types. The type/token ratio is 224/406 and the raw lexical density is 55.2 %. On the other hand, the same article in Wikipedia contains 4141 tokens and 1488 types. The type/token ratio is 1488 /4141, and the resulting raw lexical density is 35.9% (fig. 1).

Graffiti	Britannica		Wikipedia	
Tokens	406		4141	
Types	224		1488	
Raw types/tokens ratio	224	406	1488	4141
Raw lexical density %	55.2		35.9	
Normed types/tokens ratio	224	406	224	427
Normed lexical density %	55.2	(100 words)	52.5	(100 words)

Fig. 1 Graffiti article's Lexical Density

It should be noted that the ratio decreases as the number of words in a sample increases, therefore the ratio of text with different length is not comparable (Chafe 1987, Biber 1988).

According to Biber (1988), many of the different words used in the first 100 words of a text will be repeated, consequently in each additional 100 words the number of new types decreases since, as mentioned above, the relationship between text length and unique words (tokens/types) is not proportional.

In fact, when the length of encyclopaedic articles varies widely, as it frequently happens in the specific case studies analyzed, the raw lexical density will appear to be much higher in the shorter text.

Thus, when calculating lexical density in the microanalysis phase, and in order to have authentic and comparable data, the length of the longer article has been reduced to the length of the shorter one. Furthermore, to homogenise and standardize the results of the two macro encyclopaedic corpora analysis, a further normalization has been made on the basis of 100 words. In order to do this, the following formula has been applied:

$$\textbf{Types : Tokens = X : 100} \quad \rightarrow \quad \textbf{X} = \frac{\textbf{Types * 100}}{\textbf{Tokens}}$$

The results show that when lexical density has been calculated on similar samples, the standardized types/tokens ratio tends to be very similar in the two encyclopaedias. The example provided (fig. 1) proves that the normed lexical density in the article *Graffiti* is 55.2 % in Britannica and 52,5 % in Wikipedia vs. the previous row lexical density being 35.9% in Wikipedia.

The micro analysis has highlighted the fact that the difference between the lexical density of each pair of encyclopaedic articles (fig. 3) is similar in most of the cases [36]. Except for 15 articles (on 100) lexical density is always slightly higher in Britannica.

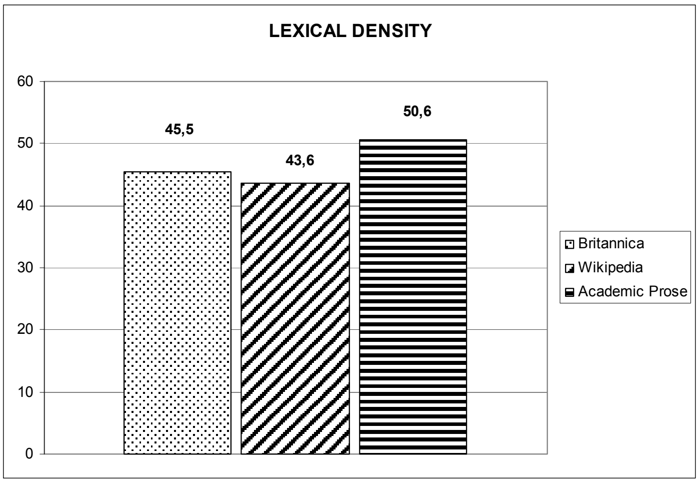


Fig. 2 Britannica, Wikipedia, Academic Prose: Lexical Density

The macro analysis has confirmed the previous microscopic findings since normalized lexical density proves to be similar in the two encyclopaedic corpora.

Specifically, total lexical density is 45.5 % in Britannica and 43.6 % in Wikipedia corpus. According to Biber, Halliday and Chafe's theories, this means that the lexical variety in the two encyclopaedias is very similar. However, as the percentage is higher in Britannica, this data confirms

[36] Specific data is shown in the table in *Appendix*

the slight predominance of a formal register in Britannica when compared to Wikipedia. Nevertheless, it is lower than in Academic Prose (fig. 2).

Biber's analysis shows that Academic Prose, which he considers the most formal genre in the scale of informational production, has a lexical density of 50.6%. Hence, formality of the encyclopaedic genre although lower than Academic Prose, is not very significantly distant.

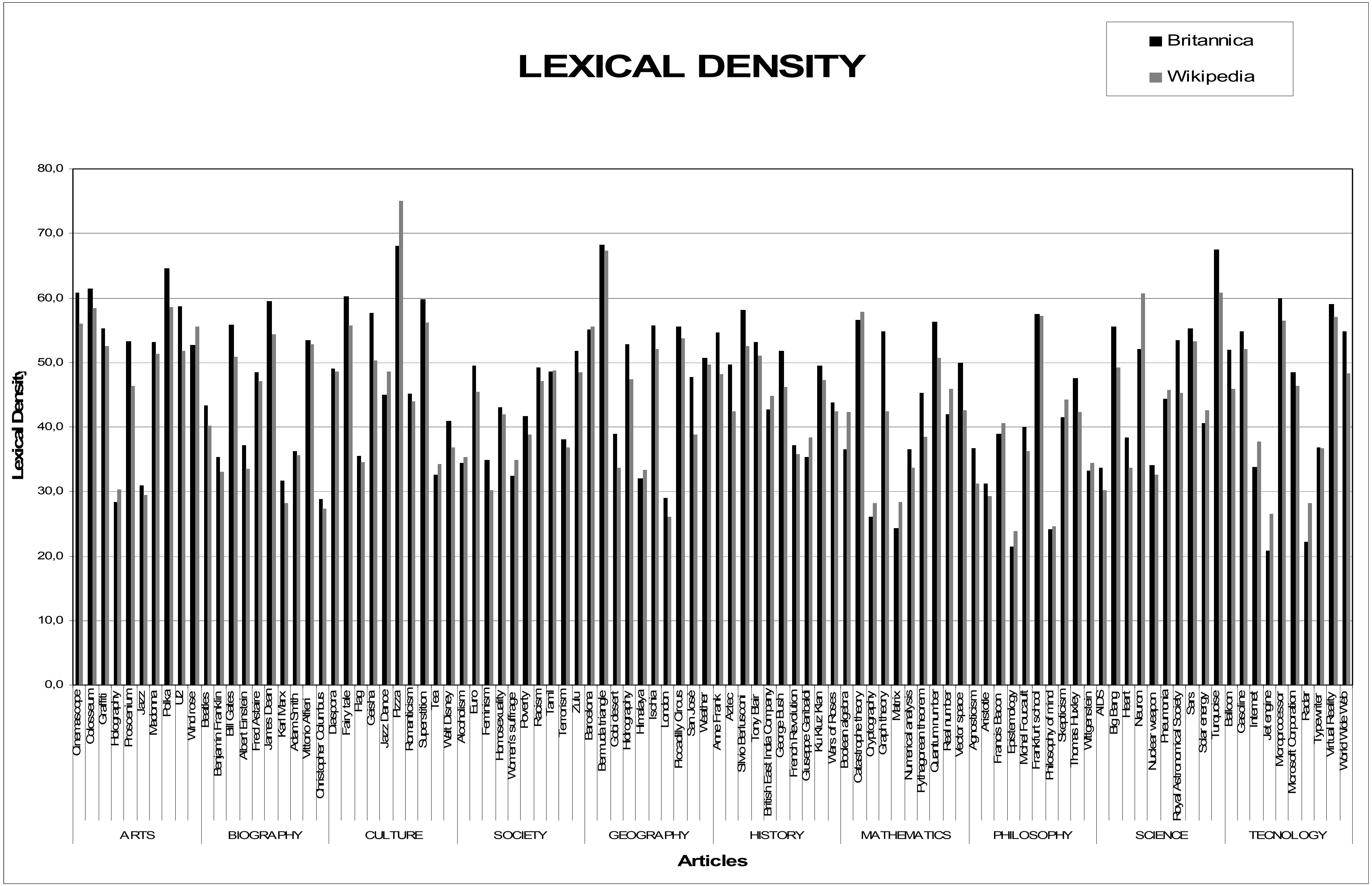


Fig. 3 Lexical density per article

1.1.2 Average Word Length

Biber (1988) claims that longer words and a high lexical density frequently co-occur in formal written genres. He states that longer words convey a more specific and specialized meaning than the shorter ones. Zipf (1949), considered one of the first pioneers in the linguistic quantitative analysis, shows that words become shorter when they are more general in meaning and more frequently used.

On the basis of these theoretical assumptions, a difference in the formal expository style has been detected through the measurement of word length in Britannica and Wikipedia. The measurement of the average word length has revealed that words in the two corpora have an equal average number of characters. The range goes from a minimum value of 3.9 (in *Matrix* article) to a maximum of 6.7 (*Microprocessor* article) in Britannica, and from 4.4 (*Vector Space* article) to 6.1 (*Hydrography* article) in Wikipedia. Most of the articles (92/100 in Britannica and 81/100 in Wikipedia) have an average word length of 5 characters which corresponds to two/three syllables per word. In detail, the average word length is of 5.3 characters in Britannica and of 5.2 in Wikipedia. This data reveals again that the findings in the two corpora are very close.

In his multidimensional analysis Biber (1988:255) finds the average word length of Academic Prose to be of 4.8 characters per word. Similar average word length in Wikipedia, Britannica and Academic Prose has been detected.

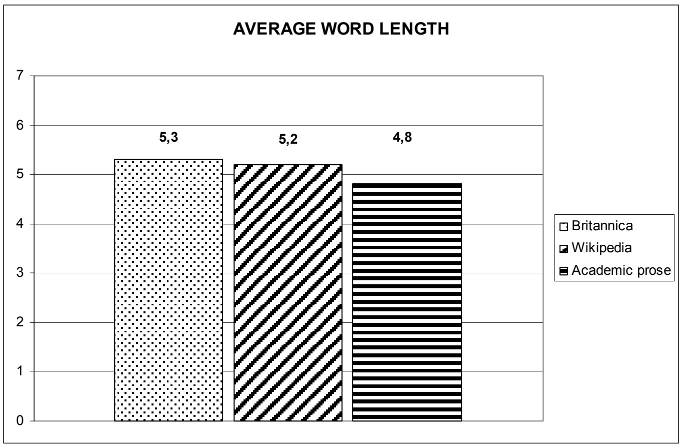


Fig. 4 Average Word Length in Bas, Was, Academic Prose

Specifically, average word length proves to be slightly higher in encyclopaedias than in Academic Prose. I consider that the main reason for longer words (although minimal) found in encyclopaedias is probably due to the pedagogical need of clarity, exactness and precision in the information delivery. Fig. 4 compares the average word length in Britannica, Wikipedia and Academic Prose, while fig. 5 shows average word length in each pair of encyclopaedic articles.

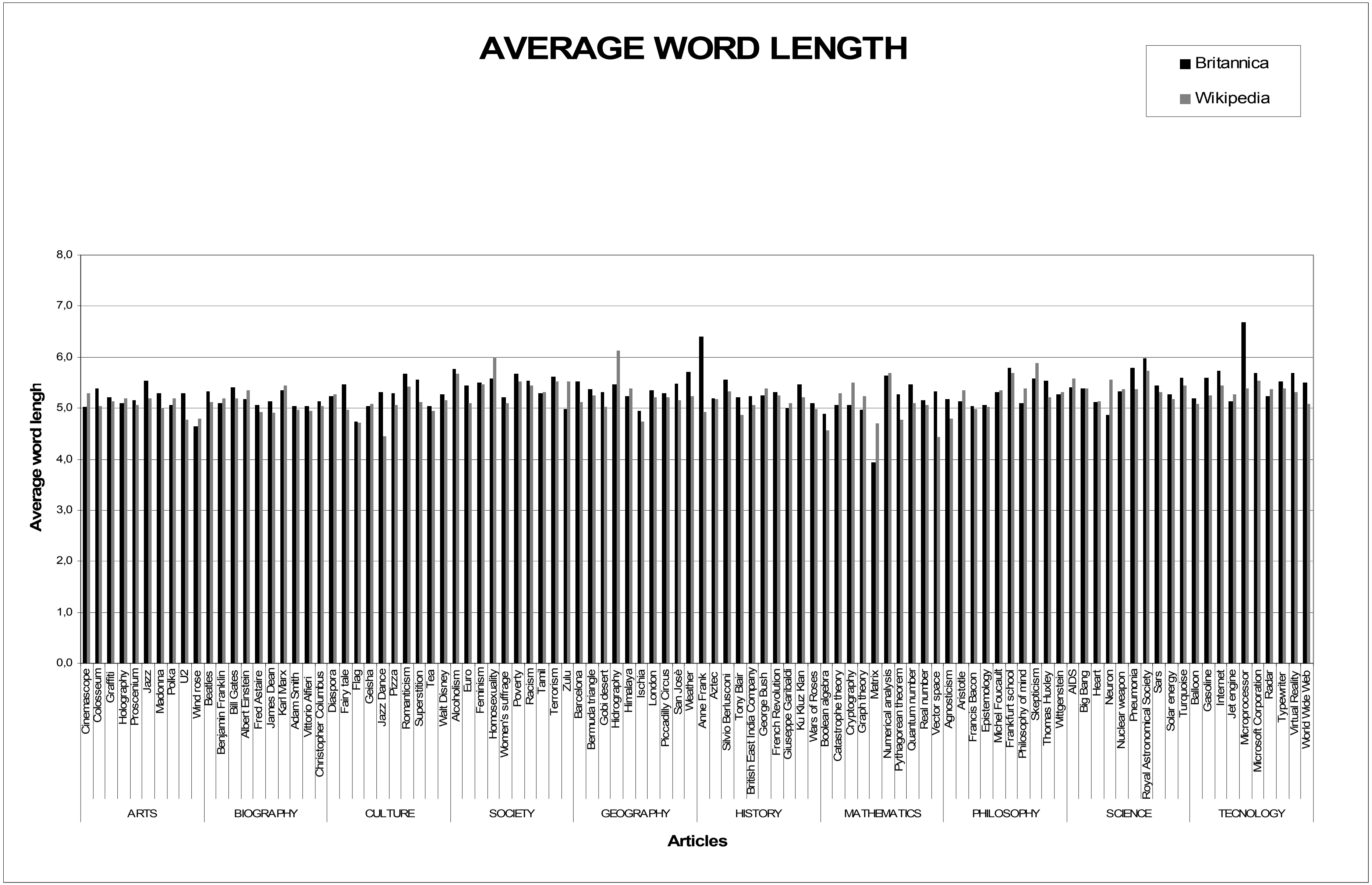


Fig. 5 Average Word Length in BAs vs. WAs

1.1.3 Sentence Length

One of the most noticeable and consistent properties of formal and academic production is that it is produced in longer clauses. This happens because formal written texts can go through planning and editing. By contrast, involved production, mainly made up of informal, interactive and oral texts, should be characterized by shorter units, consisting in simple clauses which should be syntactically less complex than informational written texts. Chafe and Danielewics (1987:5) define clauses in spoken texts as *intonation units*. With regard to written discourse they define intonation units as stretches of language between two punctuation marks:

> moving from conversation to academic writing, there will be an increase in the intonation unit size, because writers do not have to produce language on the fly, and are so freed from constraints.

They claim that writers connect clauses in a complex way, sculpting them into long planned sentences. Writers, unlike speakers, have the time and leisure to perfect the complex and coherent sentence structures. There are many linguistic devices whose effect is to increase the size of written intonation units, such as prepositional phrases, nominalizations and attributive adjectives.

According to Chafe and Danielewics, academic writing shows a relatively normal distribution of sentence lengths centred around an average of 24 words as if writers possessed an intuitive concept of *normal sentence* length. On the other hand, the average length of spoken sentences is of 18 words.

Nevertheless, not all the linguists agree with the assumptions of Chafe and Danielewics. Some researchers have interpreted the distributional pattern which conveys maximum content in the fewest word as marking a highly exact presentation of information. For example Ong (1982) claims that there is greater redundancy in speech than in writing and that there are different reasons for that: first of all, because speech is ephemeral and cannot be returned to, and secondly, because redundancy helps the listener's understanding. According to Tannen (1989), writing requires greater effort and for this reason written textual units tend to be more concise. It has also been suggested by Sheperd and Watters (1998) that longer sentences in spoken discourse are due to repetitions which enhance coherence and involvement whereas availability of cohesive devices in writing tends to diminish reliance on repetition and is the main cause of shorter sentences.

On the basis of these divergent theoretical assumptions, Britannica vs. Wikipedia sentence length has been measured. Apart from the different theoretical positions, what has been interesting for the purpose of this research is that, sentence length appears to be very close in the two encyclopaedias and not so far from the average sentence length of academic written texts, which Chafe, as previously mentioned, found to be typically of 24 words. As shown in fig. 6 the average sentence length has proved to be very similar in the two corpora, although slightly longer sentences have been found in Wikipedia (22.09 words per sentence) than in Britannica's (22.05 words per sentence). However, the difference can be considered insignificant and the formality of the expository style is confirmed in the two encyclopaedias by similar average sentence length.

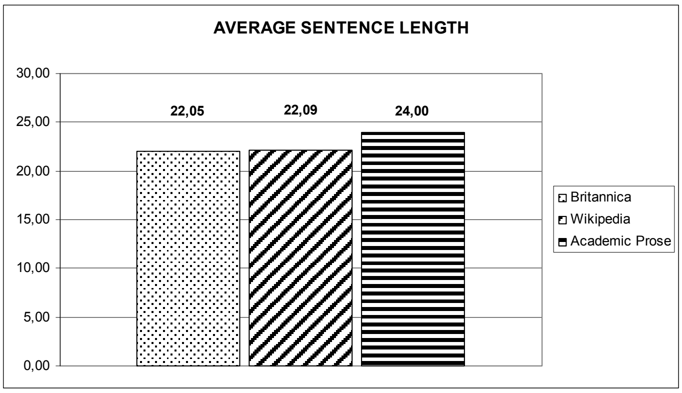


Fig. 6 Average sentence length in in BAs, WAs, Academic Prose

The micro analysis (fig. 7) proves that the range from the minimum and the maximum number of words per sentence is very close. In fact, the shortest sentence has 14.4 words (*Vittorio Alfieri*), whereas the longest one 32.8 (*Racism*) in Britannica corpus. By contrast, the shortest sentence has 14,7 words (*Geisha*) and the longest 35.8 (*Microsoft Corporation*) in Wikipedia.

In conclusion, the contrastive analysis has shown that there is an average difference of just two words (24 vs. 22 words) per sentence, thus encyclopaedias make use of shorter sentences if compared to academic texts. The micro/macroscopic analysis seems to confirm, once again, that the formality of the two corpora is very similar and not far from the formal style conveyed in academic texts.

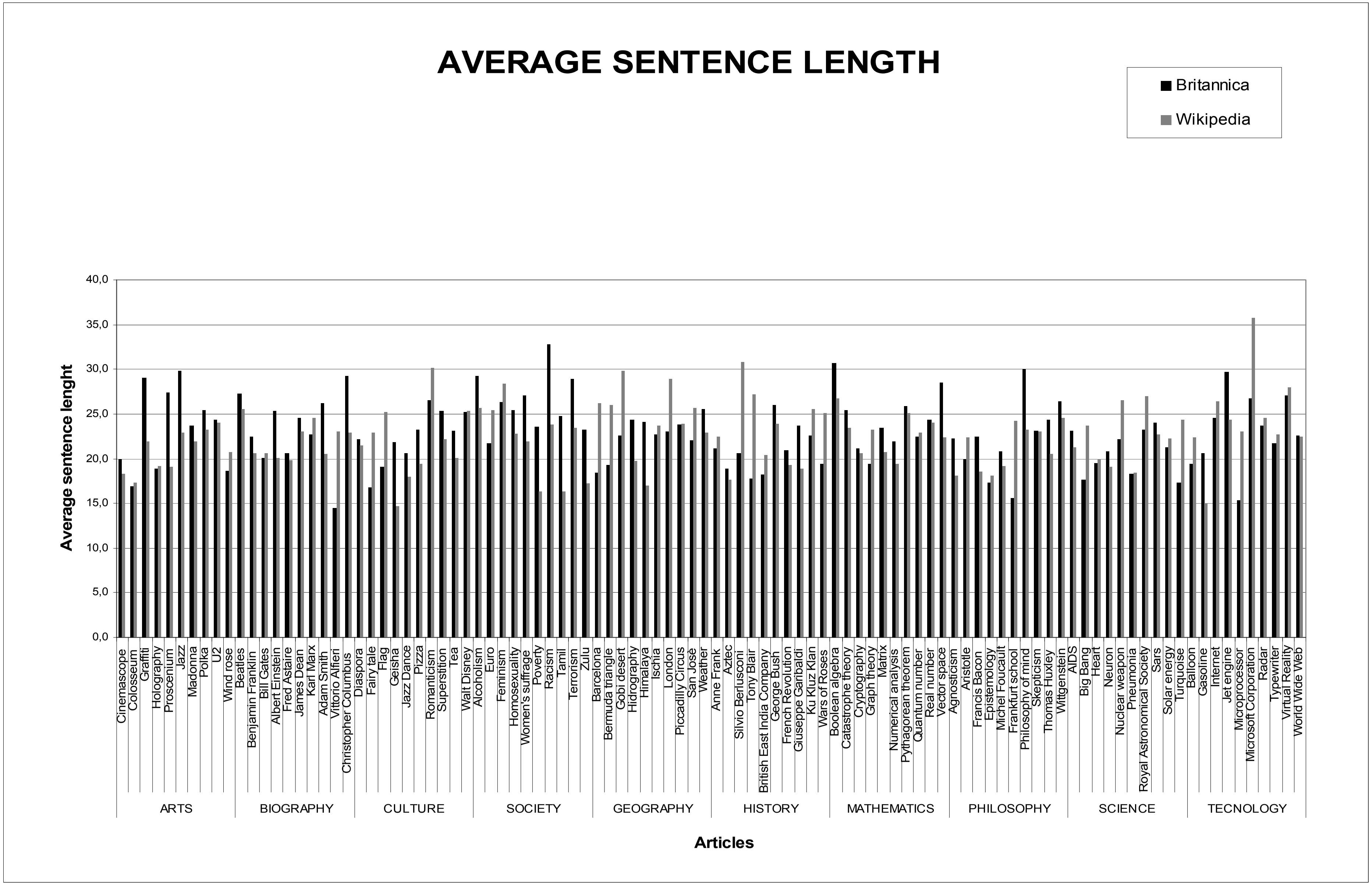


Fig. 7 .Average sentence length in in BAs vs. WAs

1.2 Nominal Forms

The overall nominal characterization of a text and the distinction between nominal and verbal style are identified as the fundamental peculiarity of written discourse by Biber (1988:227) who claims:

> A high nominal content in a text indicates a high abstract informational focus, as opposed to primarily interpersonal or narrative foci. Nominalizations, including gerunds, have particularly been taken as markers of conceptual abstractness.

To assess the incidence of nominal forms on the formal encyclopaedic expository style, the present research has investigated the frequency of nominalizations, gerunds, participial forms, articles and nouns in the two encyclopaedic corpora.

1.2.1 Nominalizations

Nominalizations (the formation of a noun from a verb or an adjective) includes all words with Latin origin ending with the suffixes *-age, -ment, -ance, ence, -tion, -ity, -ism –ness* (and their plural forms). Nominalizations have been used in many register studies. Chafe and Danielewicz (1987) focus on their use and note that they expand the idea of textual units, integrate information in fewer words and tend to co-occur with passive constructions and prepositions. According to Heylighen and Dewaele (1999:17) *nominalizations are a principle means whereby a single clause can be constructed from what might otherwise have been several clauses*. It is a device which shortens the sentence length. Many words which originated as nominalizations have become standard items of the academic vocabulary.

Thus, as the high occurrence of nominalizations is considered typical of formal written genres, their frequency has been investigated to assess their incidence on the expository style of our encyclopaedic corpora. The occurrence of nominalizations has given results which are not so dissimilar in the two encyclopaedias. Frequency is higher in Britannica (5.26 %) than in Wikipedia (4.62 %). The specific distribution of nominalizations in the two corpora is shown in fig. 8.

As a consequence, even if slightly different, the high occurrence of this linguistic class has positively influenced the formal encyclopaedic style of both encylopedic corpora.

Nominalizations				
	Britannica	**%**	**Wikipedia**	**%**
- tion	5504	2.23	7371	1.88
- ity	1710	0.69	2469	0.63
- ment	1364	0.55	2001	0.51
- ence	1187	0.48	1457	0.37
- age	765	0.31	1328	0.34
- ism	794	0.32	1135	0.29
- ance	674	0.27	1031	0.26
- sion	760	0.31	1016	0.26
- ness	256	0.10	302	0.08
Total	**13014**	**5.26**	**18110**	**4.62**

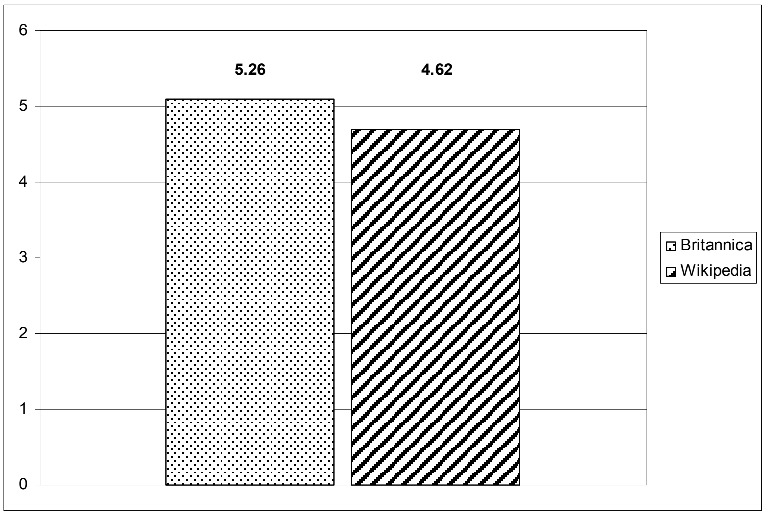


Fig. 8 Nominalizations in BAs vs. WAs

Some concordances of the query made for the nominalizations ending with the suffix *–tion* follow.

```
but also as a partial   exposition of a much larger scheme of his
ger scheme of historical evolution.  Early life Much more is k
t  Kirkcaldy, a  small (population 1,500) but thriving fishing vi
cholarship  (the  Snell Exhibition) and traveled on horseback to
re spent largely in self-education,  from which Smith obtained a
s in Edinburghùa form of education  then much in vogue in the pr
Latin, the level of sophistication for so young an audience today
 acquired the detailed information concerning trade and business
ays the psychological   foundation on which The Wealth of Nations
ing passions for self-preservation and self-interest.  Smith's a
Sentiments  the famous observation that  he was to repeat later
and the largely amoral explication of the economic system  in th
n also be seen as  an  explanation of the manner in which individ
sible for the measures of taxation that  ultimately provoked the
ons of Hume and his own admiration for The  Theory of Moral Sent
```

The analysis of the two encyclopaedic corpora, has brought to light a similar distribution of the nominalizations' typologies. As can be observed (fig. 8) the percentage of nominalizations ending with the suffix '*–tion'* is in first rank (BAs 2.2 vs. WAs 1.9), (e.g. *eruption, protection, frustration, legislation, recognition, etc*.). It is followed by words ending with the suffix '*-ity'* (BAs 0.69 vs. WAs 0.63) (e.g. *inability, creativity, etc.*), *'- ment'* (BAs 0.55 vs. WAs 0.51) (e.g. *improvement, employment, etc*.) and *'–ence'* (BAs 0.48 vs. WAs 0.37) (e.g. *experience, abstinence, interdependence,* *obedience*, *etc*.). The frequency of words ending with the suffixes ‘*-age, -ism, -ance, -sion’* (e.g. h*eritage, materialism, perseverance, impression,* etc.) is of about 0.3 in both corpora. The words ending with the suffix ‘*–ness’* (e.g. *drunkness, effectiveness, etc.*) have the lowest frequency (BAs 0.10 vs. WAs 0.08). The microanalysis of Britannica vs. Wikipedia articles has shown that, in most cases, the frequency of the nominalizations is higher in the Britannica corpus. However, this value is not absolute, given that, as can be observed in fig. 9, the nominalization frequency is higher in the following 36 Wikipedia articles.

Nominalizations		
Articles	**Britannica**	**Wikipedia**
Skepticism	5.8	**8.4**
Numerical analysis	6.9	**7.7**
Polka	4.7	**7.6**
AIDS	5.7	**7.2**
Homosexuality	6.5	**7.1**
Frankfurt school	4.6	**7.1**
Poverty	6.7	**6.8**
Big Bang	5.4	**6.3**
Proscenium	5.5	**5.9**
Karl Marx	4.7	**5.9**
Hidrography	2.7	**5.7**
Philosophy of mind	4.7	**5.4**
French Revolution	4.7	**5.2**
Holography	4.1	**5.0**
Vittorio Alfieri	2.0	**4.8**
San Josè	4.6	**4.7**
Cryptography	4.4	**4.6**
Sars	4.4	**4.6**
Bermuda triangle	3.9	**4.5**
Neuron	1.2	**4.4**
Vector space	3.9	**4.3**
Catastrophe theory	2.5	**4.0**
Real number	2.9	**4.0**
Tamil	3.7	**3.9**
Piccadilly Circus	3.3	**3.6**
Benjamin Franklin	3.1	**3.5**
Turquoise	1.4	**3.5**
Bill Gates	2.8	**3.3**
Quantum number	2.2	**3.3**
Colosseum	2.0	**2.9**
Anne Frank	2.1	**2.9**
Beatles	2.6	**2.8**
Wars of Roses	2.6	**2.8**
Graph theory	1.8	**2.8**
U2	2.5	**2.6**
Pythagorean theorem	1.6	**1.8**

Fig. 9 Nominalization frequency in BAs vs. WAs

The microanalysis has shown that the nominalization distribution has always been very similar both in each pair of articles and in the overall corpora. The only article which has shown a marked discrepancy in the nominalization occurrences is the '*Pizza'* article where the frequency is of 6.9 % in Britannica and of 1.6 % in Wikipedia.

Lowest and Highest Nominalizations		
Article	**Britannica**	**Wikipedia**
Pizza	6.9%	1.6%
Neuron	1.2%	4.4%
Jazz dance	10.7%	8.6%

Fig. 10 Lowest and highest Nominalization frequency in BAs vs. WAs

By contrast, in Britannica corpus, the encyclopaedic article which has shown to have the lowest frequency of nominalizations is *Neuron*, with a frequency of 1.2 % (correspondent article in Wikipedia 4.4 %) while the highest frequency has been found in the *Jazz Dance* article (10.7 %). The same article has also recorded a high number of nominalizations in Wikipedia corpus (8.6 %), thus, it is not a coincidence that this article has been assessed by Wikipedia department of *Heraldry and Vexillology* as a *featured* article. Fig. 11 shows the value of nominalizations in each specific pair of articles.

In conclusion, the findings have not shown a very dissimilar quantitative and qualitative (average) distribution of nominalization typologies in the two encyclopaedic corpora (5.26% BAs vs. 4.62% WAs). Nonetheless, total occurrence of nominalizations is higher in Britannica, and this data has definitely affected the more formal register of its encyclopaedic expository style. Furthermore, the microanalysis has shown that although the average frequency of nominalizations is not very different, some differences in their distribution have been detected in the two corpora. This proves that whichever writing technique is adopted, either individual or collaborative, it cannot ensure an homogeneous distribution of nominalizations in the encyclopaedic corpora.

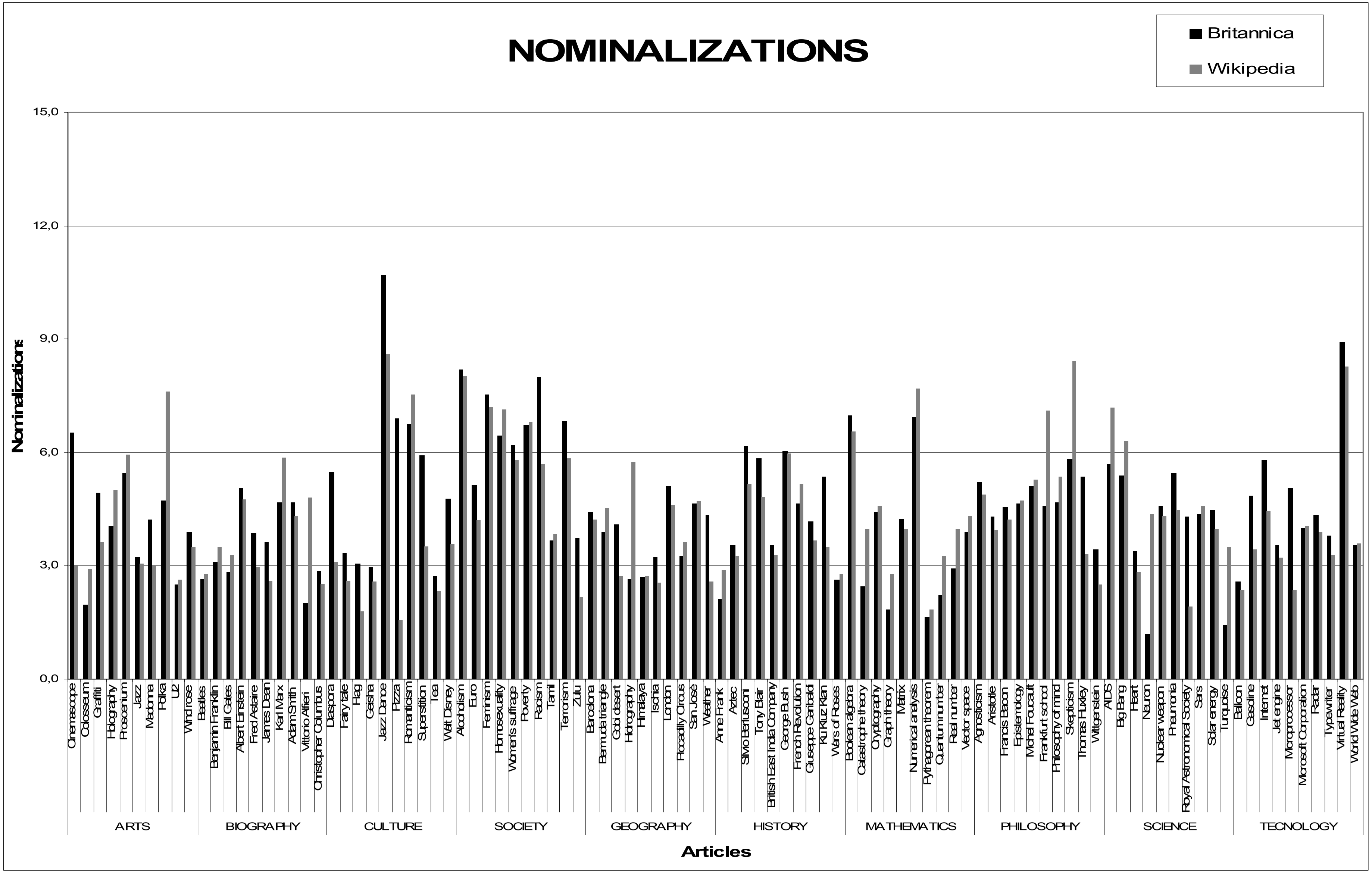


Fig. 11 Nominalizations in BAs vs. WAs

1.2.2 Gerunds and Present Participial Forms

All participial forms and gerunds and verbal nouns are closely related to nominalizations in their function. The typical interpretation associated with their distribution is that participles and gerunds are used for integrating information or for discourse structural elaboration.

Studies that consider the occurrence of participles and gerunds typically find that they occur more frequently in formal writing than in speech (Biber, 1988). O'Donnell (1974), Chafe and Danielewicz (1986) have considered gerunds and participial forms as a distinguishing marker of register.

Statistically, gerunds and participial forms are among the most difficult forms to analyze, since they can function as nouns, adjectives or verbs and, within their use as verbs, they can function as main verbs (present progressive, perfect or passive), complement clauses, adjectival clauses, or adverbial clauses, as the examples below (from Britannica's AIDS article) show:

> *(GERUND)*
> HIV slowly attacks and destroys the immune system, the body's defence against infection, *leaving* an individual vulnerable to a variety of other infections and certain malignancies that eventually cause death.
>
> *(PRESENT PROGRESSIVE*)
> According to the United Nations 2004 report on AIDS, some 38 million *people are living* with HIV, approximately 5 million people become infected annually, and about 3 million people die each year from AIDS.
>
> *(PRESENT PARTICIPLE)*
> and the World Health Organization estimates that 9 out of 10 people *needing* treatment will not receive it.
>
> *(NOUN)*
> Attempts to reduce intravenous drug use and to discourage the *sharing* of needles have also led to a reduction in infection rates in some areas.
>
> *(ADJECTIVE*)
> most DNA *synthesizing* enzymes have, many mutations arise as the virus replicates

In this research, as in other works (Chafe,1982; Beaman,1984) a specific distinction has not been made among the different functions thus, gerunds and all participial forms have been grouped in a single class. Their micro and macro occurrence has been investigated in the two corpora.

As fig. 12 clearly shows, the average occurrence of gerunds and participial forms is practically the same in Britannica and Wikipedia (2.38 % BAs vs. vs. 2.41% WAs).

Gerunds and Present Participial Forms				
	Britannica	%	Wikipedia	%
Total	5870	2.38	9456	2.41

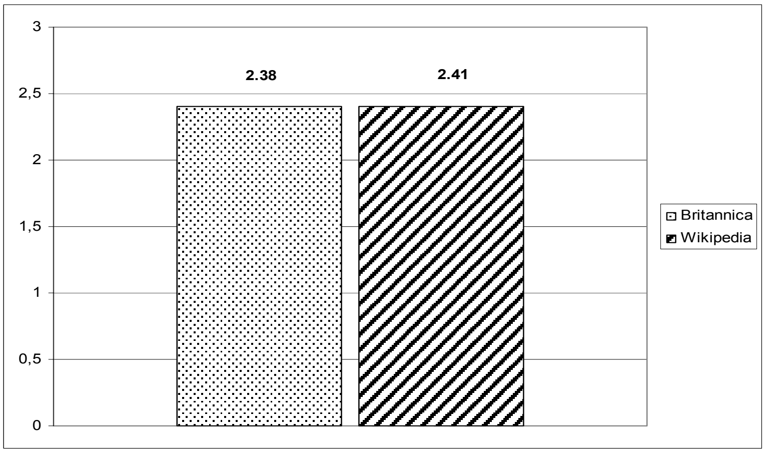


Fig. 12 Gerunds and Present Participial Forms in BAs vs. WAs

Some concordances are shown below.

```
es, Adam Smith remains a towering figure in the history of  eco
ly the capstone to an overarching view of  political and social
l (population 1,500) but thriving fishing village near Edinburgh
ation 1,500) but thriving fishing village near Edinburgh, and
received his elementary schooling in  Kirkcaldy and that at the
seems to have been a main shaping  force in Smith's development
n Smith's development. Graduating in 1740, Smith won a scholarsh
lege. Compared to the stimulating atmosphere of Glasgow, Oxford
mporary    philosophy.  Returning to his home after an absence o
n much in vogue in the prevailing spirit of ôimprovement.ö  The
Ne  as  extraordinarily demanding. Afternoons were occupied wit
mith played an active role, being elected dean of faculty in 175
ngs were spent in the stimulating company of Glasgow society.
great merchants who were carrying on the colonial trade that  h
e detailed information concerning trade and business that was to
with  Hume and the  other leading philosophers of his time, he t
ook as a universal and unchanging  datum from which social inst
tator,ö approving  or  condemning our own and others' actions wi
lth of Nations: that self-seeking men are often  ôled by an inv
isible hand . . . without knowing it, without intending it, [to]
out knowing it, without intending it, [to]  advance the interes
 At one level there is  a seeming clash between the theme of soc
Ently  married and  was searching for a tutor for his stepson an
ual salary of ú300 plus traveling  expenses and a pension of ú3
ulouse, where Smith began working on a book (eventually to be Th
s an antidote to the excruciating boredom of the provinces.  Af
y to have  considered  dedicating The Wealth of Nations to him,
```

Minimum and maximum values of gerunds and all participial forms do not coincide. As can be seen in fig. 13 and in the table in *Appendix*[37], the range shows a variable distribution of gerunds and present participles in each pair of encyclopaedic articles.

[37] All the data related to the linguistic analyses are shown in the general table reported in Appendix where the overall findings of the microanalysis have been summariezed in a unique prospectus.

Whereas the minimum occurrence of gerunds and present participial forms is 0.5 % in Britannica (*Boolean Algebra*) and the highest value is 5.5 (*Virtual Reality*), their frequency ranges from 0.0 (*Pizza*) to 3.8 (*Poverty*) in Wikipedia's articles.

Thus, the analysis has shown that although the distribution of gerunds and participial forms is different in the two macro corpora and in each pair of encyclopaedic articles (see *Appendix*) in terms of minimum and maximum value, their average is almost coincident and so is, consequently, their overall incidence on the index of formality of encyclopaedic expository style.

Furthermore, the microscopic analysis has demonstrated that the frequency of gerunds and participial present forms is slightly lower in Britannica than in Wikipedia in 53 articles (more than 50%). Although quantitatively insignificant, this data reveals an opposite trend as, until now, all the linguistic classes considered have shown a slightly higher incidence in Britannica's formal style (fig. 14).

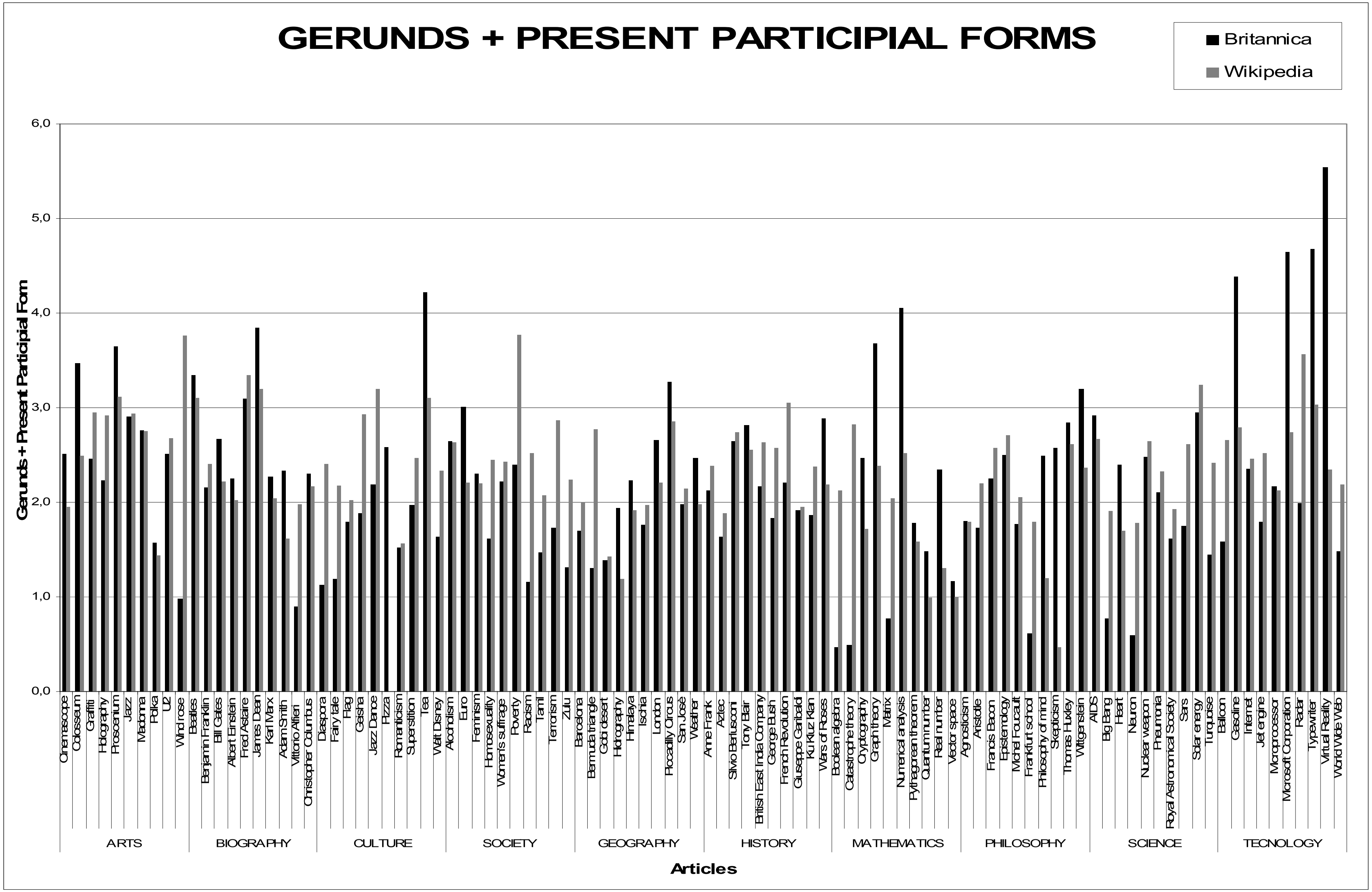


Fig. 13 Gerunds and present participial forms in BAs vs. WAs

Articles	Britannica	Wikipedia
Boolean algebra	**0.5**	2.1
Catastrophe theory	**0.5**	2.8
Neuron	**0.6**	1.8
Frankfurt school	**0.6**	1.8
Big Bang	**0.8**	1.9
Matrix	**0.8**	2.0
Vittorio Alfieri	**0.9**	2,0
Wind rose	**1.0**	3.8
Diaspora	**1.1**	2.4
Racism	**1.2**	2.5
Fairy tale	**1.2**	2.2
Bermuda triangle	**1.3**	2.8
Zulu	**1.3**	2.2
Turquoise	**1.4**	2.4
Tamil	**1.5**	2.1
World Wide Web	**1.5**	2.2
Romanticism	**1.5**	1.6
Balloon	**1.6**	2.7
Homosexuality	**1.6**	2.4
Royal Astronomical Society	**1.6**	1.9
Aztec	**1.6**	1.9
Walt Disney	**1.6**	2.3
Barcelona	**1.7**	2.0
Terrorism	**1.7**	2.9
Aristotle	**1.7**	2,2
Sars	**1.8**	2.6
Ischia	**1.8**	2.0
Michel Foucault	**1.8**	2.1
Flag	**1.8**	2.0
Jet engine	**1.8**	2.5
George Bush	**1.8**	2.6
Ku Kluz Klan	**1.9**	2.4
Geisha	**1.9**	2.9
Giuseppe Garibaldi	**1.9**	2.0
Superstition	**2.0**	2.5
San Josè	**2.0**	2.2
Radar	**2.0**	3.6
Pneumonia	**2.1**	2.3
Anne Frank	**2.1**	2.4
Benjamin Franklin	**2.2**	2.4
British East India Company	**2.2**	2.6
Jazz Dance	**2.2**	3.2
French Revolution	**2.2**	3.1
Women's suffrage	**2.2**	2.4
Holography	**2.2**	2.9
Francis Bacon	**2.2**	2.6
Internet	**2.4**	2.5
Poverty	**2.4**	3.8
Nuclear weapon	**2.5**	2.6
Epistemology	**2.5**	2.7
U2	**2.5**	2.7
Silvio Berlusconi	**2.6**	2.7
Solar energy	**2.9**	3.2

Fig. 14 Ordered frequency of Gerunds and present participials in BAs vs. WAs

1.2.3 Definite / Indefinite Articles

According to Heylighen and Dewaele (1999), the frequency of articles, nouns, adjectives and prepositions is expected to increase with the formality of a text.

Following this assumption, the frequency of definite (*the*) and indefinite articles (*a, an*) has been calculated. Also in this case, their occurrence is very similar in the two corpora although slightly more significant in Britannica (10.02% vs. 9.68%) as fig. 15 shows.

Definite / Indefinite Articles				
	Britannica	**%**	**Wikipedia**	**%**
the	18.147	7.34	27.780	7.09
a	5.506	2.23	8.573	2.19
an	1.109	0.45	1.576	0.40
Total	**24.762**	**10.02**	**37.929**	**9.68**

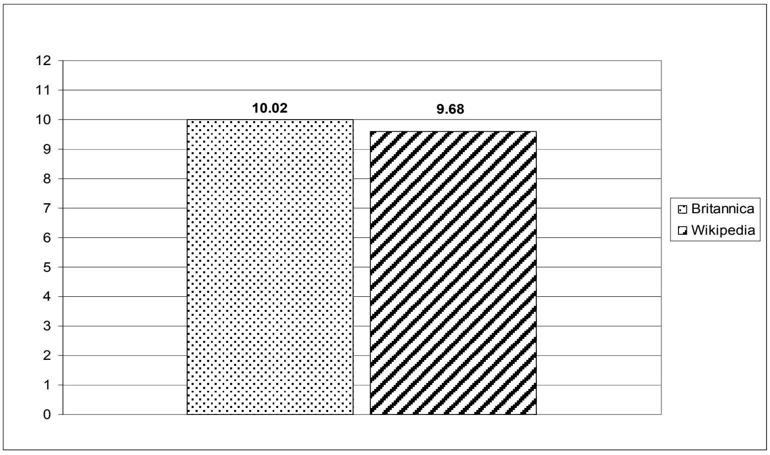


Fig. 15 Definite/indefinite articles in BAs vs. WAs

Of course the higher occurrence of articles is strictly associated with the more significant frequency of nouns, as will be shown in the next section. However, this data has not been confirmed in the 33 articles shown in fig. 16, where a constant and slightly lower frequency of definite and indefinite articles has been detected.

In Britannica the lowest frequency of definite and indefinite articles has been found in the article *Superstition* and *Turquoise* with a percentage of 5.4, whereas the highest frequency has been found in the article *Wind Rose* (16.1). On the other hand, in the Wikipedia corpus the range goes from a minimum occurrence of 6.7 in *Feminism* to a maximum value of 21.8 in *Wind rose*. The frequency of definite/indefinite articles in each specific encyclopaedic article (see table in *Appendix*) has been microscopically portrayed and shown in fig. 17.

Definite/Indefinite Articles		
Articles	**Britannica**	**Wikipedia**
Superstition	**5.4**	9.9
Turquoise	**5.4**	7.7
Racism	**6.1**	8.3
Fairy tale	**6.9**	9.3
Pneumonia	**6.9**	8.2
Microprocessor	**7.2**	10.8
Homosexuality	**7.4**	7.9
Balloon	**7.5**	9.8
Solar energy	**7.5**	9.3
Jazz	**7.6**	8.8
Wittgenstein	**7.9**	8.2
Numerical analysis	**8.2**	10.0
Madonna	**8.3**	9.5
Bermuda triangle	**8.4**	10.7
Frankfurt school	**8.6**	11.4
Weather	**8.6**	9.7
Microsoft Corporation	**8.6**	9.2
Silvio Berlusconi	**8.8**	9.6
Giuseppe Garibaldi	**9.0**	10.8
Feminism	**9.0**	6.7
Thomas Huxley	**9.4**	9.6
Royal Astronomical Society	**9.4**	10.7
Anne Frank	**9.4**	10.2
Polka	**9.4**	10.0
Virtual Reality	**9.8**	10.2
U2	**10.7**	11.0
Ischia	**10.9**	12.5
Colosseum	**10.9**	13.0
Aztec	**11.4**	11.6
Gobi desert	**11.6**	12.6
Pythagorean theorem	**12.3**	13.8
Himalaya	**12.5**	13.2
Holography	**12.6**	13.0
Proscenium	**14.2**	14.3
Wind rose	**16.1**	21.8

Fig. 16 Frequency of definite/indefinite articles in ascendant order

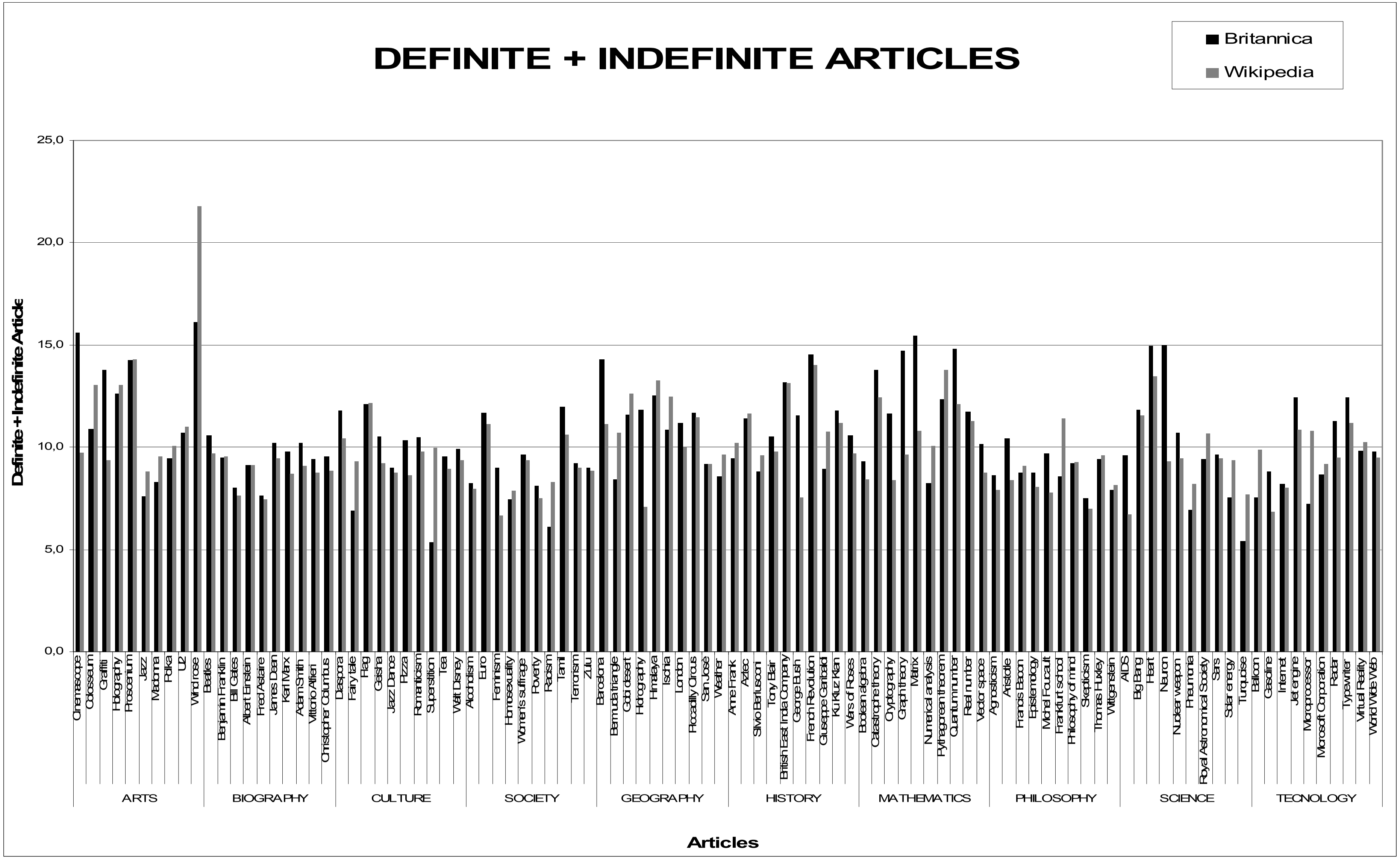


Fig. 17 Definite and indefinite articles in BAs vs. WAs

1.2.4 Nouns

Biber (1988), Heylighen and Dewaele (1999) have investigated the overall noun occurrence in written and spoken discourse, showing that their frequency is higher in written academic prose than in oral speech.

Using the *CLAWS part of speech tagging software* and an inferential statistical approach (see section 2.2.7), specific occurrences of neutral (for number), singular, plural and proper nouns on a sample corpus of 10.000 tokens has been calculated. Then, the overall nominal value of the two encyclopaedic corpora has been deduced. Findings are shown in fig. 18.

Noun Frequency				
	Britannica	%	Wikipedia	%
Total	73883	29.90	114671	29.28

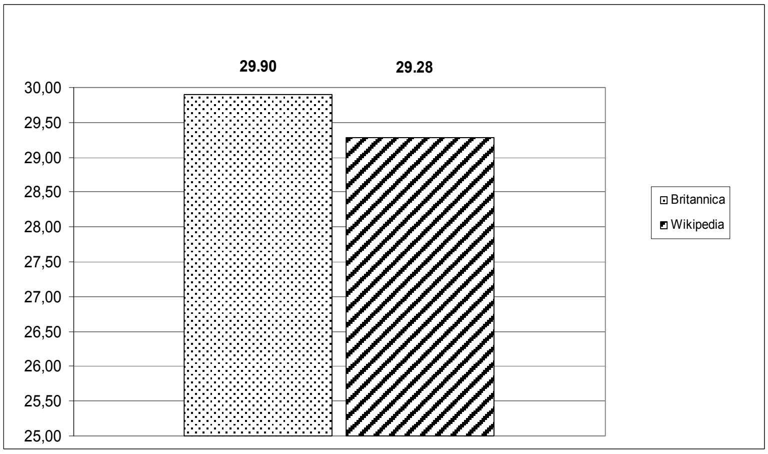


Fig. 18 Noun frequency in WAs vs. BAs

The high frequency of nouns seems to be dominant compared to other linguistic classes and to significantly influence the level of formality of the encyclopaedic style. Moreover, also in the case of nouns, findings confirm a very similar frequency in Britannica and Wikipedia's corpora although the frequency is slightly higher in Britannica (29.90% BAs vs. 29.28% WAs). Two concordance excerpts from the query made on singular and plural nouns on the tagged corpora follow:

```
  AJ0  social_AJ0 philosopher_NN1 and_CJC political_AJ0 economis
d_CJC political_AJ0 economist_NN1 ._.  After_PRP two_CRD centur
Z a_AT0  towering_AJ0  figure_NN1 in_PRP the_AT0 history_NN1 of_
re_NN1 in_PRP the_AT0 history_NN1 of_PRF economic_AJ0  thought_
of_PRF economic _AJ0  thought_NN1 ._.  Known_VVN primarily_AV0
for_PRP a_AT0 single_AJ0 work_NN1 ,_, An_AT0  Inquiry_NN1 into_
work_NN1 ,_, An_AT0  Inquiry _NN1 into_PRP the_AT0 nature_NN1 an
y_NN1 into_PRP the_AT0 nature_NN1 and_CJC causes_NN2 of_PRF the_
s_NN2 of_PRF the_AT0  Wealth_ NN1 of_PRF Nations_NN2 (_( 1776_CR
RD  comprehensive_AJ0 system_ NN1 of_PRF political_AJ0 economy_N
 of_PRF political_AJ0 economy NN1 ,_, Smith_NP0  is_VBZ more_AV
_AT0 social_AJ0  philosopher _NN1 whose_DTQ economic_AJ0 writing
B only_AV0  the_AT0 capstone _NN1 to_PRP an_AT0 overarching_AJ0
0 network_NN1 of_PRF networks_NN2  ,_, the_AT0 Internet_NP0 eme
_DPS constituent_NN1 networks_NN2 ._.  It_PNP supports_VVZ huma
il_NN1 )_) ,_, chat_VVB rooms_NN2 ,_, newsgroups_NN2 ,_,  and_C
_VVB rooms_NN2 ,_, newsgroups_NN2 ,_,  and_CJC audio_AJ0 and_CJ
DT0 different_AJ0  locations_ NN2 ._.  It_PNP supports_VVZ acce
y_PRP  many_DT0 applications_ NN2 ,_, including_PRP the_AT0 Worl
ber_NN1 of_PRF  e-businesses _NN2 (_( including_PRP subsidiaries
_( including_PRP subsidiaries_NN2 of_PRF traditional_AJ0  brick
rick-and-mortar_AJ0 companies_NN2 )_) that_CJT carry_VVB out_PRP
PS sales_NN0 and_CJC services_NN2 over_PRP the_AT0 Internet_NN1
N1 ._. )_)  Many_DT0 experts _NN2 believe_VVB that_CJT the_AT0 I
_NN1 ._.  Early_AJ0 networks _NN2 The_AT0 first_ORD computer_NN1
rst_ORD computer_NN1 networks_NN2 were_VBD  dedicated_VVN speci
```

1.3 Adjectives

Adjectives seem to expand and elaborate the information presented in a text. Chafe and Danielwicz (1987) group adjectives together with prepositional phrases and subordinating constructions as devices used for idea unit integration and expansion. Biber (1998) also finds that the frequency of adjectives is higher in formal and academic written genres. As for nouns, the same inferential statistical methodology has been used for determining the overall frequency of adjectives. Basic adjectives, comparative and superlative adjectives have been distinguished (see section 2.2.7). Some concordances are shown below.

```
he_AT0 most_AV0  destructive _AJ0 epidemics_NN2 in_PRP recorded_
epidemics_NN2 in_PRP recorded_AJ0 history_NN1 ._.  In_PRP 2005_
1 ._.  In_PRP 2005_CRD alone _AJ0 ,_, AIDS_NN1 claimed_VVN betwe
etween_PRP an_AT0  estimated_ AJ0 2.8_CRD and_CJC 3.6_CRD millio
s_NN1 to_PRP  antiretroviral _AJ0 treatment_NN1 ,_, both_AV0 mor
e_NN1  in_PRP cardiovascular _AJ0 risks_NN2 &lsqb;_( 5_CRD &rsqb
e_AT0 rise_NN1 of_PRF  viral _AJ0 escape_NN1 and_CJC resistance_
RD &rsqb;_) ._.  The_AT0 Red _AJ0 Ribbon_NN1 is_VBZ the_AT0 glob
bon_NN1 is_VBZ the_AT0 global_AJ0 symbol_NN1 for_PRP  solidarit
ity_NN1 with_PRP HIV-positive_AJ0 people_NN0 and_CJC those_DT0
```

```
ly_AV0 lost_VVD its_DPS wider_AJC_popularity_NN1 as_CJS it_PNP wa
 perplexed_VVN many_DT0 older_AJC  musicians_NN2 and_CJC fans_NN
esponses_NN2 from_PRP younger_AJC_ones_NN2 (_( ranging_VVG from_P
J0 artform_NN1 with_PRP wider_AJC_appeal_NN1 ,_, to_PRP a_AT0  s
nd_CJC recreate_VVI  earlier_ AJC_styles_NN2 of_PRF jazz_NN1 )_)
tle_VVI for_PRP a_AT0 smaller_AJC_audience_NN1 of_PRF  aficionad
also_AV0 embraces_VVZ greater_AJC_opportunity_NN1 for_PRP men_NN2
ts_NN2 in_PRP the_AT0 broader_AJC  sense_NN1 do_VDB not_XX0 clai
a_AT0 type_NN1 of_PRF lighter_AJC_than_CJS  air_NN1 aircraft_NN0
 AJ0  flammability_NN1 Rozier_AJC_balloons_NN2 use_VVB both_AV0 h
```

As for nouns, the overall frequency of adjectives has also shown a high incidence when compared to other linguistic classes. A low quantitative variation has been detected in the two corpora, as fig. 19 shows (10.54% BAs vs. 10.06% WAs)

Adjectives Frequency				
	Britannica	%	Wikipedia	%
Total	26045	10.54	39398	10.06

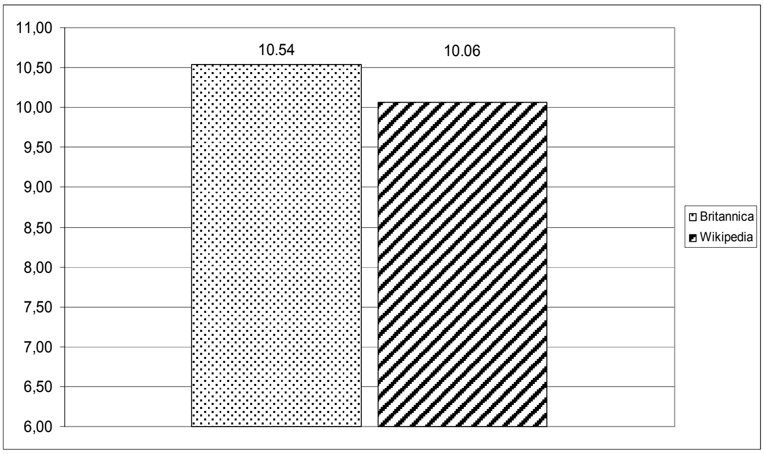


Fig. 19 Adjectives frequency in BAs vs. WAs

1.4 Prepositions

Prepositions are important devices for packing high amounts of information in the discourse. Chafe and Danielewicz (1987) as mentioned in the previous section, describe prepositions as a device for integrating information into idea units and for expanding the amount of information they contain.

Biber (1988:237) also claims that *prepositions tend to co-occur frequently with nominalizations and passives in academic prose, official documents and other informational types of formal written discourse.* Furthermore, Heylighen and Dewaele (1999) state:

> Within the 'formal' categories prepositions perform the best […] prepositions are typically used to start a further specification, or simply adding precise information on the circumstances in which something happens.

As can be noticed in the sample of *Encyclopaedic Article Evaluation Grid* (chapter 3, fig. 8), only the frequency of some basic prepositions (*of, in, to, by, for, from, at*) has been calculated for each pair of articles during the microscopic analysis phase. By contrast, a more detailed analysis on prepositions has been carried out on the overall Britannica and Wikipedia encyclopaedic corpora during the macroscopic analysis. Findings, ordered in ascendant frequency sort, are shown in fig. 20.

The data confirms the general trend emerging from the analysis of other linguistic features, that is to say, a very similar frequency of prepositions although its number is slightly higher in the Britannica corpus (14.23% BAs vs. 13.42% WAs). As fig. 20 shows, the analysis of prepositions has shown a proportional incidence and distribution in each encyclopaedic corpus.

Prepositions				
	Britannica	%	Wikipedia	%
Of	11354	4.59	15084	3.85
In	6343	2.57	10026	2.56
To	5820	2.36	8915	2.28
By	1991	0.81	2906	0.74
For	1897	0.77	3148	0.80
With	1566	0.63	2666	0.68
On	1421	0.58	2562	0.65
From	1222	0.49	1924	0.49
At	1014	0.41	1466	0.37
Than	385	0.16	560	0.14
Into	335	0.14	539	0.14
Between	299	0.12	415	0.11
Through	261	0.11	307	0.08
Out	180	0.07	280	0.07
During	175	0.07	442	0.11
Against	152	0.06	260	0.07
Among	138	0.06	173	0.04
Without	130	0.05	169	0.04
Within	95	0.04	164	0.04
Upon	79	0.03	86	0.02
Toward/s	78	0.03	126	0.03
Througho	59	0.02	74	0.02
Off	54	0.02	87	0.02
Per	52	0.02	77	0.02
Except	34	0.01	34	0.01
Opposite	20	0.01	20	0.01
Onto	9	0.00	30	0.01
Versus	4	0.00	10	0.00
Plus	3	0.00	13	0.00
Total	35170	14.23	52566	13.42

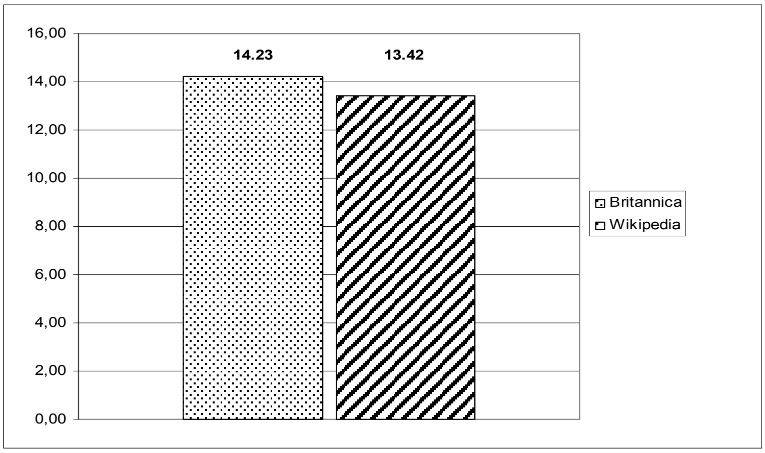


Fig. 20 Prepositions in BAs vs. WAs

To provide an example, a concordances' excerpt of the most used preposition (*of*) is shown below.

```
with live groups. It was one  of the first cellar clubs in  Li
pposed to the strict  policy  of jazz for venues such as The Ca
 Coombs. The Cavern  was one  of the more well-known spots wher
en went through a progression of names: Johnny and The Moondogs
ng at The Beatles. The origin of the name  "The Beatles" with
ory he wrote about the naming of the group for the  Liverpool
  met and befriended a group  of German art students who called
 particular the introduction  of the famous Beatle haircut, whi
he hair and clothing styles   of the band). While in Hamburg, T
 his backing band on a series of recordings for the German  Po
```

Just glancing at the excerpt taken from *Ischia* Britannica article, it is evident to what extent the text is crowded with prepositions.

> Italian Isola D'ischia, Latin Aenaria, island *at* the northwest entrance *to* the Bay *of* Naples, *opposite* Capo (cape) Miseno, Napoli province, Campania region, southern Italy, just west-southwest *of* Naples. Oblong *in* shape, *with* a circumference *of* 21 mi (34 km) and an area *of* 18 sq mi (47 sq km), the island consists almost entirely *of* volcanic rock and rises *to* 2,585 ft (788 m) *at* Monte Epomeo, an extinct volcano. The date *of* the first eruption is estimated *to* have been about 2200 BC; an eruption *of* the 7th century BC, according *to* the Roman scholar Pliny the Elder, drove away the first Greek settlers, and another *in* 470BC put a Syracusan garrison *to* flight. There were several eruptions *in* Roman times. The last *on* record occurred *in* 1301–02, when the population fled *to* Baia *on* the mainland and did not return *for* four years. There have been destructive earthquakes more recently, the last *in* 1883 when the entire town *of* Casamicciola was destroyed. The island was known *to* the Greeks as Pithecusa (probably meaning "island *of* monkeys") and *to* the Romans as Aenaria. *From* the Middle Ages it was subjected *to* frequent attacks and invasions, usually related *to* the struggles *for* supremacy *on* the mainland. Its volcanic soils are fertile, and the wine, called Epomeo, that is produced *on* Ischia is famous. Wheat, olive oil, and citrus fruits are also economically important. The clay *of* Ischia is believed *to* have been used *by* the ancient potteries *of* Cumae and Puteoli (Pozzuoli). Well known *for* its mild climate, picturesque scenery, and numerous thermal mineral springs, Ischia is much frequented as a health and vacation resort. The more important towns are *in* the north *of* the island: Ischia, the administrative centre and seat *of* a bishop, consisting *of* the fishing village *of* Ischia Ponte *with* a medieval castle, and Ischia Porto;

PREPOSITIONS		
Article	**Britannica %**	**Wikipedia %**
U2	8.4	10.0
Barcelona	8.5	10.1
San Josè	8.6	9.6
Matrix	9.1	9.5
Real number	9.1	9.7
Pizza	9.5	9.6
Fairy tale	9.5	10.4
Cinemascope	9.5	10.9
Graph theory	9.6	9.8
Microsoft Corporation	9.9	10.8
Jazz	10.1	10.5
Bill Gates	10.3	11.7
Bermuda triangle	10.4	10.7
Holography	10.4	11.2
Gobi desert	10.5	11.3
Euro	10.6	11.1
Silvio Berlusconi	11.0	11.7
Tamil	11.0	11.3
Nuclear weapon	11.0	11.5
Thomas Huxley	11.3	13.4
Big Bang	11.3	11.4
Philosophy of mind	11.3	11.8
Racism	11.3	11.7
Sars	11.8	12.1
Romanticism	12.0	15.0
Solar energy	12.5	16.4

Fig. 21 Prepositions: a comparison in BAs vs. WAs

Fig. 21 shows that only 26 Britannica encyclopaedic articles out of 100 have a lower frequency of prepositions once compared to Wikipedia.

In the remaining 74 Britannica's articles, a higher occurrence has been recorded (see *Appendix*) consequently this linguistic class has a positive influence on the formality of the expository encyclopaedic style. The frequency of prepositions in each pair of encyclopaedic articles is pointed out in fig. 22. The specific data can be read in the table reported in *Appendix*.

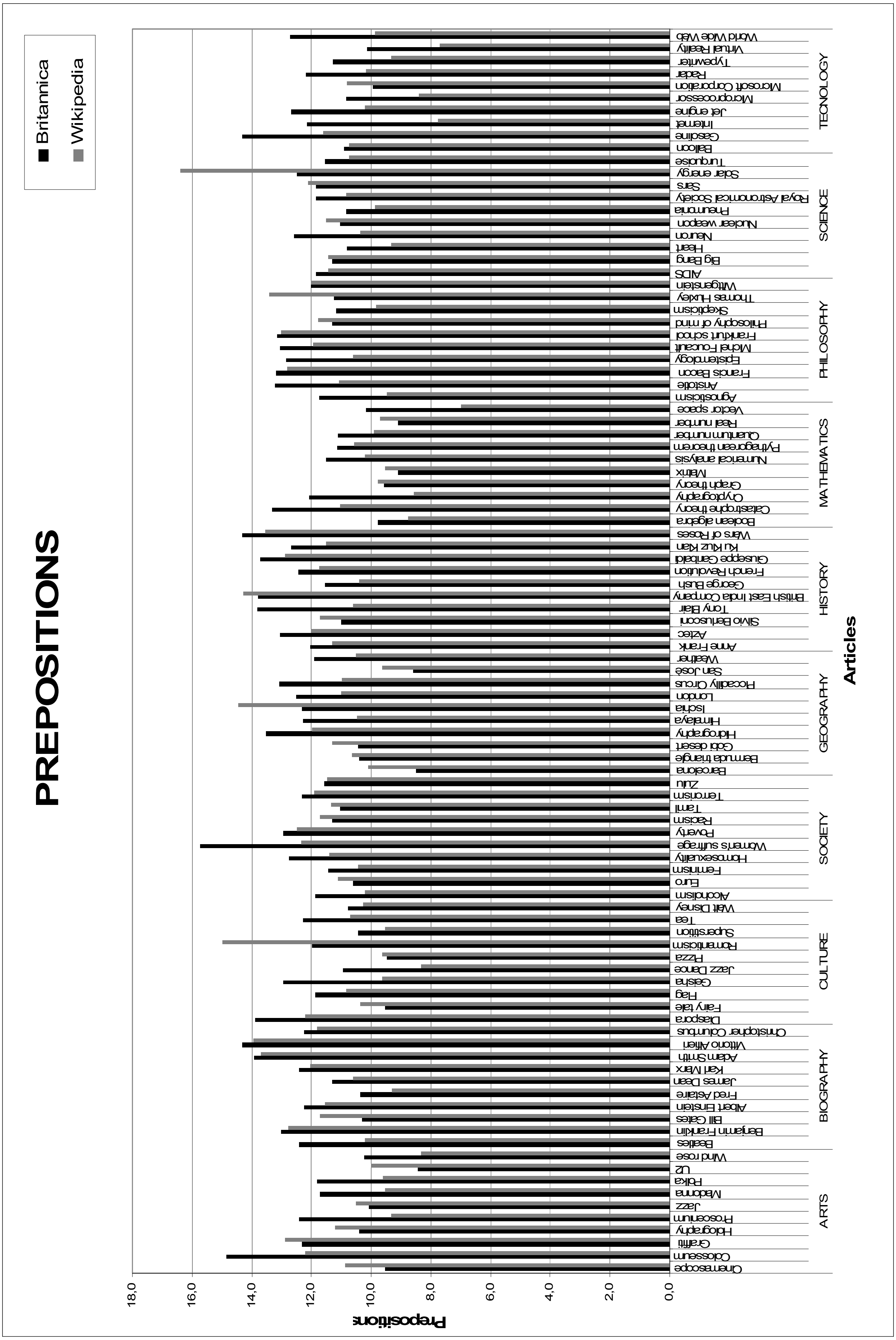


Fig. 22 Prepositions in BAs vs. WAs

1.5 Passives

A discourse with very frequent passive constructions is typically abstract in content and formal in style, therefore passives have been considered as one of the most important surface markers of the decontextualized or detached style that stereotypically characterizes formal writing. In passive constructions, the agent is demoted or eliminated altogether, resulting in a static and more abstract presentation of information. *Agentless* passives are used when the agent does not have a salient role in the discourse, while *by-* passives are used when the agent is very closely related to the discourse topic.

Biber’s (1998:163) multidimensional analysis also confirms that passives are associated with a static, nominal and impersonal style. Typical informational academic production makes use of several passive constructions such as *by passives, agentless passives, past particles reduced to relatives,* etc.

In addition to Biber’s analysis, further studies which have used passives for register comparison include Brown and Yule (1983), Chafe (1982), Chafe and Danielewicz (1987).

In the present research, the frequency of *agentless* passives and *by-passives* has been investigated by the occurrences of the following verbal constructions:

- has been (adv) + past participle (by)
- have been (adv) + past participle (by)
- had been (adv) + past participle (by)
- is (adv) + past participle (by)
- are (adv) + past participle (by)
- was (adv) + past participle (by)
- were (adv) + past participle (by)
- be (adv) + past participle (by)

Some occurrences of *agentless passives* and *by- passives (+ adv.)* from the Britannica corpus are shown in the two concordances’excerpts reported below.

```
ewish poet and  philosopher,  has been  authoritatively described as ô
AIDS Memorial  Quilt, which   has been  displayed worldwide both to ra
and popular culture, HIV/AIDS has been  double-edged. On the  one han
professions. Thus, alcoholism has been  thought to be caused by defe
he U.S. A rate of 3.5 percent has been  reported from Sweden and  1.1
percent. The rate in France   has been  estimated at as high as 15 per
s. None of  these treatments  has been  shown in controlled studies to
y good at heart.ö  The diary  has been  translated into more than 50 l
rified by Auguste Comte.  It   has been  suggested that Bacon's thought
Cardona     Valley, salt      has been  exploited since Roman times; a
istian household; however,it  has been  claimed that he was a conver
Taino chieftain's settlement  has been  identified nearby. Concepci¾
ase pencil. Manual  tracking  has been  largely replaced by automatic
about the echo signal. Colour has been  employed, for  example, to in
adar cross section of a man   has been  measured at microwave frequenc
```

```
of  political economy, Smith  is more properly regarded as a so
evolution. If his masterwork  is viewed in relation to his  ea
ion.  Early life Much more    is known about Adam Smith's thoug
f Smith's childhood  nothing  is known other than that he recei
nd less naive if the question is reformulated to ask how  inst
ts exerted on  Smith, but it  is known that he thought sufficie
ization through which society is impelled, unless blocked by
ted that each of these stages is accompanied by institutions
ropical evergreen rain forest is confined to the humid foothill
ness. Mesua ferrea (ironwood) is found  on porous soils at alt
st Alpine vegetation. Juniper is widely distributed, preferring
ier areas; on Nanga Parbat it is found even at an  altitude of
life of the eastern Himalayas is derived mainly from that of th
k has been  domesticated and  is used as a beast of burden in L
exported to India, where oil  is extracted from them. Bhutan al
 Of the plantation crops, tea is grown mainly on the hills and
rict. Tea in limited quantity is  also grown in the Kangra Val
sonal migration of livestock) is widely practiced during  the
nerals, although exploitation is restricted to the  more acces
ountains, and  alluvial gold  is recovered in the nearby bed of
```

Fig. 23 shows the overall occurrences of passives forms in Britannica and Wikipedia corpora. The frequency of passives is 0.96 % in both corpora, thus the analysis has revealed an identical average incidence of this verbal construction in determining the expository style of Britannica and Wikipedia.

Passives				
	Britannic	**%**	**Wikipedia**	**%**
be + (adv) + p.p.	**576**	0.23	**730**	0.19
is + (adv) + p.p.	**574**	0.23	**939**	0.24
was + (adv) + p.p.	**529**	0.21	**1034**	0.26
are + (adv) + p.p.	272	0.11	409	0.10
were + (adv) + p.p.	263	0.11	388	0.10
has been + (adv) + p.p.	58	0.02	117	0.03
had been + (adv) + p.p.	58	0.02	58	0.01
have been + (adv)+ p.p.	45	0.02	93	0.02
Total	**2375**	**0.96**	**3768**	**0.96**

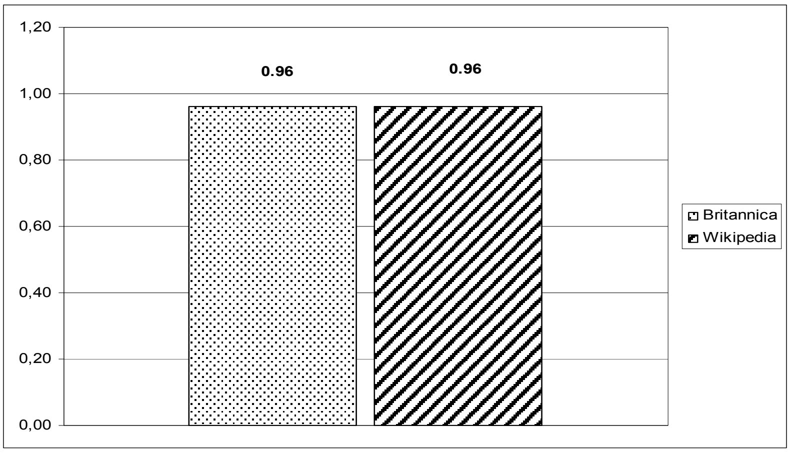


Fig. 23 Passives in BAs vs. WAs

Unlike academic written discourse, where the use of passives prevails, the quantitative incidence of passives on the encyclopaedic expository style is clearly not very high. In order to increase the textual Index of Readability and Web Usability, active verbal structures are preferred, as the main purpose of encyclopaedias is popular and educational and the largest target audience is made up of school and university students.

Despite the low occurrence, the passive form which is preferred in both encyclopaedias is the impersonal one [*be, is, was* + *(adv)* + *p.p.*]. Furthermore*,* Wikipedia makes a more extensive use of past tense impersonal forms [*was* + *(adv)* + *p.p.*]. The *agentless* passive is the most used construction as it makes the presentation of the information more static and abstract. Specific data and some examples of the most recurrent constructions are provided below.

BE + (ADV) P.P.
Chief among the lost works are: Eudemus, in the tradition of Plato's Phaedo; On Philosophy, a type of philosophical program containing themes *to be developed* later in his Metaphysics; the Protrepticus, or exhortation to the life of Aristotele *(from Aristotele – Britannica)*

IS + (ADV) + P.P.
Smith *is more properly regarded* as a social philosopher whose economic writings constitute only the capstone to an overarching view of political and social evolution.
(from Adam Smith – Britannica)

WAS + (ADV) + P.P.
Much of his later work *was done* on the west coast of Ireland in the rural isolation he preferred.
(from Wittgenstein – Wikipedia)

By- Passives				
	Britannica	**%**	**Wikipedia**	**%**
was + p.p. + by	91	0.04	157	0.04
is + p.p. + by	86	0.03	132	0.03
be + p.p. + by	67	0.03	74	0.02
were + p.p. + by	32	0.01	47	0.01
are + p.p. + by	24	0.01	39	0.01
had been + p.p. + by	12	0.00	19	0.00
has been + p.p. + by	11	0.00	10	0.00
have been + p.p. + by	7	0.00	16	0.00
Total	**330**	**0.13**	**494**	**0.13**

Fig. 24 By-Passives in BAs vs. WAs

The average frequency of *by*-passives is 0.13 % in both corpora. Thus, its frequency is statistically very limited, as fig. 24 shows. This construction is used only when the agent is very closely related to the discourse theme.

Some concordances are shown below

vely courtship dance of Bohemian folk origin. It	**is characterized by**	three quick steps and a hop and is danced to
region where targets are expected. When a target	**s illuminated by**	the beam, it intercepts some of the radiated ener
range and angular direction. Range, or distance,	**is determined by**	measuring the total time it takes for the radar s
cond (the speed of light). The range to a target	**is determined by**	measuring the time that a radar signal takes to t
. The ultimate range accuracy of the best radars	**is limited by**	the known accuracy of the velocity at which
transmitter generates the high-power signal that	**is radiated by**	the antenna. In a sense, an antenna acts as a tr
s radiated as a narrow beam. A paraboloid, which	**is generated by**	rotating a parabola about its axis, forms a symme
applications without the phase shifters. The beam	**is steered by**	the mechanical movement of the entire antenna.
In most cases the sensitivity of a radar receiver	**is determined by**	the noise generated internally at its input. Beca
eceiver noise The sensitivity of a radar receiver	**is determined by**	the unavoidable noise that appears at its input.
ed microwave air-surveillance radar, whose range	**is limited by**	the curvature of the Earth. Besides detection and
l trade with the ancient Greeks and Romans	**is verified by**	literary, linguistic, and archaeological ev
of the time allowed for black tea. Fermentation	**is stopped by**	heating in iron pans, and the leaf is subjected to
urposes. Delicate veining, caused by impurities,	**is desired by**	some collectors as proof of a natural stone. Turq
nting position, and the imprint of type on paper	**is produced by**	a trigger action. The type-wheel machines offer a

Although the overall average frequency of the selected passive forms is the same in the two corpora (0.96%), fig. 25 shows that there is not a proportional distribution of this verbal structure in the pairs of encyclopaedic articles. A dissimilar distribution can also be observed in the specific data reported in *Appendix*.

Thus, it can be assumed that the quantitative and qualitative distribution of passive forms mostly depends on the personal writing style of contributors. The investigation carried out on some samples has indicated an alternate use of passive and active forms in the same article. Neverthless, a definitely broader inclination toward the use of active forms has been noted in some sample articles analyzed in Wikipedia. This is probably due to the general recommendation clearly expressed in the *Wikipedian Manual of Style* which suggests avoiding frequent passives and complex verbal structures in order to improve the index of textual readability.

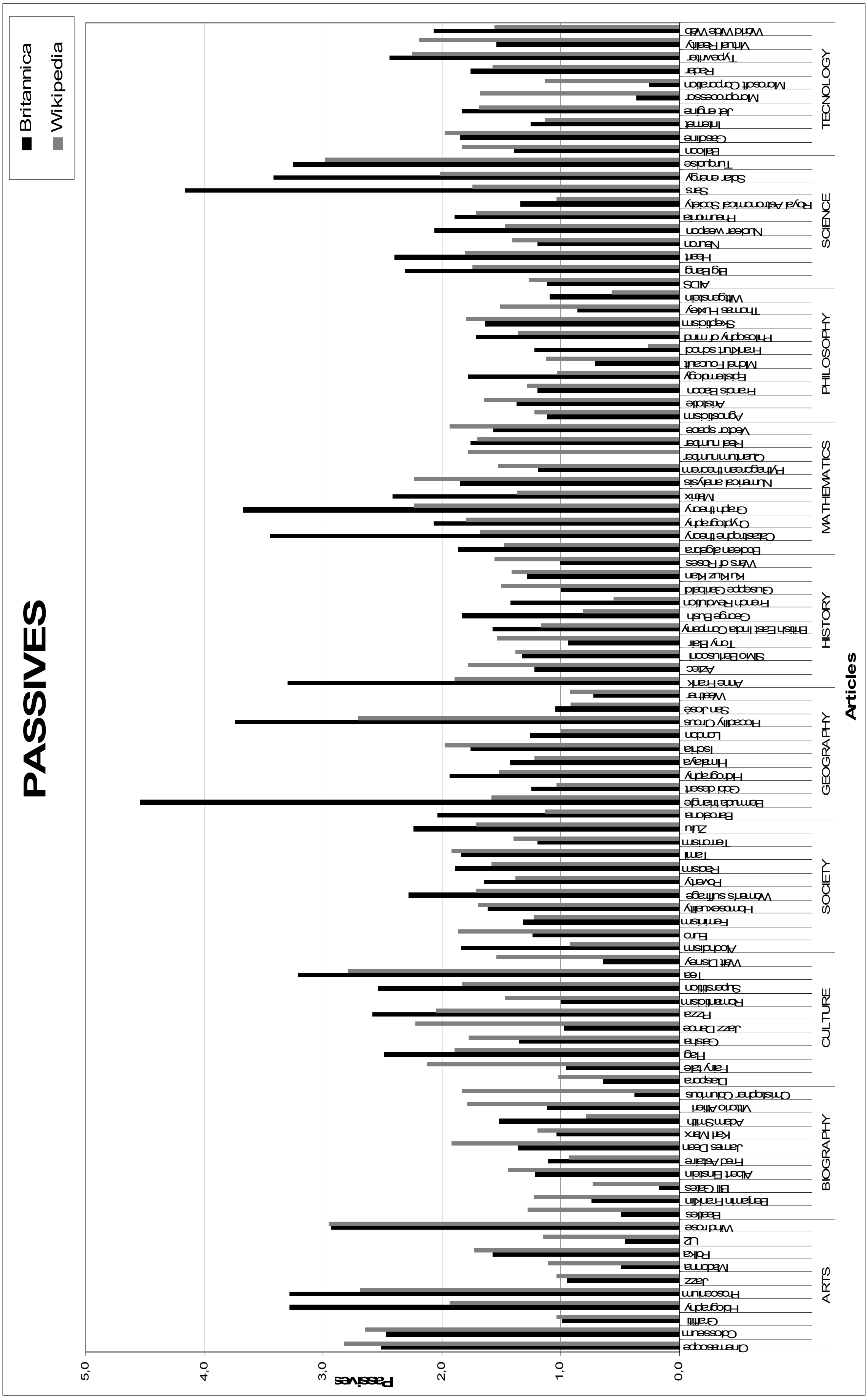


Fig. 25 Passives in BAs vs. WAs

1.6 Subordination Features

Subordination is a clause linkage device which enables the language user to express several related events. The debate regarding complexity of language is relevant to the study of subordination because they are often considered linked, as an increase in the number of subordinate structures is generally associated with an increase in the degree of complexity.

Within reference grammars, subordination and coordination are defined in terms of two binary features, namely *embeddeness* and *dependency*. According to Foley and Van Valin (1984:239) *embeddedness* refers to whether a clause is a constituent of (embedded inside) another clause. *Dependency* has to do with whether the clauses stand in a *whole-whole equivalence relation* or in a *part-whole relationship*. Thus, subordination involves embedded clauses and the notion of dependency. By contrast, coordination does not present either embedded clauses or dependency relations between them (Calude, 2005).

Subordination is the most recurrent linguistic feature used for register comparisons. Some studies have concentrated on the linguistic differences between spoken and written discourse and the conclusions have been different and even contradictory because of the problem of selection of data representing the two media. Due to the importance of the subject, a simplified synopsis of the main studies which have investigated "complexity and subordination" in spoken and written English texts in the last 30 years has been reported in fig. 26 (Calude, 2005).

There are several observations which can be drawn from the body of work summarized (fig. 26). First of all, there is no overall consensus regarding the degrees of complexity found in oral and written discourse, (defined in this specific study as *involved* vs. *informational* production). Latter studies have shown that both utterances exhibit some kind of complexity. However, even among those who agree with this hypothesis, further differences are still found, in terms of the ways in which syntax complexity manifests itself. Some linguists, such as Beaman (1984) and Halliday (1979), believe that speech is structurally more complex (in that it contains more embedded clauses) and written discourse is lexically denser. Halliday (in Biber, 1988: 229), for example, claims:

> Conversational speech has more subordination than written styles, because the two modes have different kind of complexities: spoken language, because it is created and perceived as an ongoing process, is characterized by 'an intricacy of movement complex sentence structures with low lexical density (more clauses, but fewer high-content words per sentence), written language, in which the text is created and perceived as an object, is characterized by a denseness of matter, simple sentence structures with high lexical density (more high content words per clause, but fewer clauses).

Others claim that grammatical complexity can be found in both oral and written discourse, and that differences derive from the types of syntactic constructions found in each medium.

1970s	
1974	**O'DONNELL**'s work indicates that writing is overall more diverse and complex than speech, having more gerunds, participles, attributive adjectives, passives, modal and perfective auxiliaries, but less noun clauses, infinitives and progressive auxiliaries.
1976	**POOLE and FIELD** report more embedding, more adverbials and more personal pronouns in SL, but fewer adjectives and complex verbal structures. These results, they claim, are a consequence of the increased amount of time available to writers (and not to speakers), leading to simpler structures in writing (but not speaking)
1977	**KROLL**'s work suggests that WL has more instances of subordination than SL.
1979	**CHAFE** argues that WL is more reliable on subordination than SL due to the "detached" relationship between writer and audience. Furthermore, in general, written language tends to be planned rather than unplanned, whereas speech is exactly the opposite – hence writing tends to contain more subordination than speech. A third view comes to light, introduced by **HALLIDAY,** who argues that both speech and writing are complex, but in different ways. Speech is complex in terms of grammatical structure, whereas writing is complex in terms of lexical items. In the same year, **LAKOFF** writes that the dichotomy between speech and writing is altogether misguided and that the two language mediums form a continuum. She suggests (nine) binary features could be used to separate the different language modes situated in-between the two extremes of the continuum.
1980s	
1980	**PRICE and GRAVES** claim that writing has more adjectival and adverbial phrases.
1982	**CHAFE** consolidates his earlier findings that WL is more complex and more integrated (e.g., less fragmentary) than SL, in that it has more nominalizations, prepositional phrases, more present and past participials, more attributive adjectives, more THAT- and TO- complement clauses and more relative clauses.
1984	A new study by **BEAMAN** agrees with Halliday's earlier claims, that SL is at least as complex as WL, if not more complex in some cases. SL is found to have more dependent clauses, whereas WL has more lexical density.
1988	The dichotomy between speech and writing is again under attack by a detailed study carried out by **BIBER**, who shows that the two language mediums cannot be successfully separated using a multi-feature analysis. Interestingly, he also picks up on some of the discrepancies mentioned by Beaman, and adds to their source problems in defining the variables investigated (such as the notions of *sentence* and *subordination*).
1990s	
1994	Halliday's results are further supported by **MILLER**, who proposes that speech and writing are different language mediums, not only containing different kinds of structures, but also containing structures which may "look" and "behave" differently when used in the two different modes. However, the overall findings seem agree with the general trend that speech is less complex than writing. One important factor in the analysis of speech is found to be the level of education obtained by the participants involved. That is, the higher their level of education, the higher their exposure to written material and the more similarity can be observed between their speech and their writing.
1995	Two studies carried out by **GREENBAUM and NELSON** report no real differences between speech andwriting. However conversational data is found to be an exception to this pattern. Conversations stand out as a language type, in that they exhibit decidedly less instances of subordination than any written text.
1997	In accord with Halliday's and in part with Miller's work, **KIRK** also finds SL and WL to be equally complex. According to Kirk, some subordinate clauses are more common in SL (THAT-complements and WH-complements), while others are more frequent in WL (infinitive clauses, ING-clauses and ED-participial clauses).
1998	Finally, **MILLER and WEINERT's** book on *Spontaneous spoken language: syntax and discourse* examines conversations (impromptu, as well as narratives) and task-related dialogue. The results obtained confirm earlier findings by Kirk, Miller and Halliday that (1) speech is different from writing, (2) some constructions may occur in medium but not the other, and finally, (3) one and the same construction may appear and function very differently in the two mediums (relative clauses being a prime example).

Fig. 26 Subordination in Spoken and Written English texts
(freely adapted from Calude, 2005)

Thus, a group of researchers, with Halliday as main representative, claims that spoken has more subordination than written discourse. On the other hand, Beaman (1984) and Biber (1989) find that different subordination forms are distributed differently. Beaman for example observes that there are more *that complement clauses* in interactional texts which often mark the stance of the speaker or writer (e.g. with the verbs *think, wish, hope,* etc.), while Biber (1998:74) argues that *that complement clauses* and causative adverbial subordinators (*since, as, because*) are not very frequent in informational written texts. He finds that they co-occur more frequently with first and second person pronouns and reduced and interrogative forms.

The above mentioned divergent theoretical positions have stimulated the present study. At this stage of the research the question searching for an answer is: *What is the incidence of subordination features on the expository style of the two encyclopaedias?* In trying to answer this question, the occurrences of a selected number of subordination features in the encyclopaedic articles and in the overall corpora have been measured and compared. The distribution of the selected subordinated clauses listed below has been investigated.

- Relative clauses (e.g. what, which, who, whom, whose, that)
- Conditional subordinators (e.g. if, unless)
- Concessive subordinators (e.g. although, though)
- Causative subordinators (e.g. because, since, as)
- Other subordinators (e.g. while, whereas)

1.7 Relative Clauses

Relative clauses have been frequently used as markers of register variation. According to Beaman (1984), relative clauses provide a way to talk about nouns, either for identification or simply to provide additional information. Chafe (1987) considers both *wh- clauses* and *that clauses*, devices to expand and integrate the information provided in a text.

Most of the studies generally find that relative clauses occur more frequently in writing than in speech. Other scholars, however, do not treat all relative clauses as a single feature and, not finding an homogeneous distribution of this pattern, formulate different interpretations. Some analyses suggest that these constructions are considerably more frequent in spoken than in written discourse. Winter (1982) for istance, claims that the most important function carried out by relative sentences is to express attitudinal comments, whereas wh-clauses provide a way to talk about questions and often indicate the speaker's evaluation or attitude. Frequency of *wh-relatives* clauses and *that clauses* has been investigated in the two encyclopaedic corpora.

1.7.1 Wh-Clauses

Biber claims (1988:55):

> Wh-relative clauses are used to specify the identity of referents within a text in an explicit and elaborated manner. The co-occurrence of conjuncts, passive constructions, and past participial clauses marks informational discourse that is impersonal and formal in style.

Different types of *wh-clauses* introduced by *what, which, who, whom* and *whose* have been investigated. Some random examples are reported below.

There is such a thing as divine philosophy *what* was later called rational, or natural, theology […]
These include B lymphocytes, *which* produce antibodies needed to fight infection […]
Anarcho-feminists, *who* found a larger audience in Europe than in the Unite States
the woman *who* married J.F. Kennedy […]
The emperor of Cathay, *whom* Europeans referred to as the Great Khan […]
Nicholas of Autrecourt (1300-50), *whose* views anticipated the radical skepticism of Hume […]

The excerpt below shows some concordances for the relative pronoun *who*.

```
ompany of the great merchants who   were carrying on the colonial
s headed by Franþois Quesnay, who   called themselves les  Úconom
r of the  Duke of Buccleuch,  who   had joined them in Toulouse, t
against the poor, or of those who   have some property against  t
some property against  those  who   have none at all.ö Finally, Sm
gs problems. The manufacturer who    accumulates stock needs more
erchants and  manufacturers,  who   neither are, nor ought to be,
æGnostic' of Church history   who   professed to know so much abou
says (1893), reproached those who   pretended to delineate  ôthe
y applicable to many of those who   nowadays adopt the more  comf
calö (the atheist is thus one who   is simply  without a belief i
s stronger than that of those who   simply  confess that they
```

As can be observed in fig. 27, a similar average occurrence of the total number of clauses introduced by *wh-words* has been found both in Britannica and Wikipedia corpora (0.65% BAs vs. 0.69% WAs). The most used wh-pronoun is *which*, followed by *who, what, whose* and *whom*. As the data (fig. 27) shows, their descendent occurrence is constantly slightly higher in Wikipedia and always proportional and coherent in the two encyclopaedic corpora.

Wh- Clauses				
	Britannica	**%**	**Wikipedia**	**%**
Which	883	0.36	1552	0.40
Who	316	0.13	699	0.18
What	292	0.12	297	0.08
Whose	70	0.03	85	0.02
Whom	41	0.02	61	0.02
Total	**160**	**0.65**	**269**	**0.69**

Fig. 27 *Wh-clauses* in Ba vs. WAs

1.7.2 That as Subordination Feature

Since *that* can assume many different grammatical functions, its investigation has required a careful manual selection of its specific functions in the context. The occurrence of *that* with the following grammatical functions has been measured.

➔ That as relative clause *(subject position)* e.g.:
[…] in best start with the treatment of those problems *that* are relevant, interesting, or important to him.

➔ That as relative clause *(object position)* e.g.:
[…] the painting *that* the artist created […]

➔ That as adjective complement *(adj + that)* e.g.:
[…] If it is already true *that* there will be a sea battle tomorrow, the […]

➔ That as verb complement *(verb+ that)* e.g.:
[…] he states *that* poetry is more philosophic than history and thus […]

Some concordances of *that* are reported.

```
blished posthumously in 1969, that  “'Knowledge' and certainty be
n of genocide.” They believed that  “a super bomb should never be
s like these, Moore contended that  “a thing can't be certain unl
ence as her guide to proclaim that   “all men and women [had been]
 the  3 floods recommended    that  “apart from erecting further w
theorist, quoting his dictum  that  “by the position which women h
urther postulate: he supposes that  “exhalations,” some moist and
(1953), Wittgenstein states   that  “explanation must be replaced
x in debate in 1846 recalled  that  “he spoke only in the imperati
as he said to his sovereigns, that  “my hard and troublesome voya
General Council declared      that  “on the German side the war wa
 became a common saying       that  “One man's terrorist is anothe
Some of the major factors     that  affect performance are discus
apples have only one property that  affects each sense organ diffe
te religious revival meetings that  African Americans in many part
ergy. This produced a schism  that  aggravated the violence of the
brings out both the fact      that  agnosticism has something to d
```

A comparison between the Britannica and the Wikipedia corpora has revealed a higher occurrence of *that* – clauses in Britannica (1.11% BAs vs. 0.59 %WAs).

That Clauses				
	Britannica	**%**	**Wikipedia**	**%**
Total	2751	1.11	2322	0.59

Fig. 28 *That-clauses* in BAs vs. WAs

1.8 Adverbial Clauses

Adverbial clauses are important devices which should mark greater textual elaboration and convey informational relation in texts. They seem to characterize formal informational texts.

Nevertheless, this assumption is not universally accepted. Thomson (1984) and others, in fact, consider speech richer than writing in adverbial clauses.

There are several subclasses of adverbial clauses. The most common are those introduced by *causative*, *concessive*, and *conditional* adverbs. They have been easily and automatically identified in an unambiguous way by the concordancer software. Causative and adverbial subordinators are considered by Biber (1988:236) as:

> Markers of affect or stance, that is justification for actions or beliefs (because) or conditions for actions or beliefs (if, unless). These subordination features thus, seem to be associated with a relatively loose presentation of information, and they seem to mark a range of affective functions relating to the elaboration of personal attitudes or feelings.

Thus, causative and conditional adverbial subordination seem to be related to affect or stance in that they set discourse frame for particular propositions and present justification for actions or beliefs.

Total findings (fig. 29,30,31) and some random examples of conditional, concessive and causative subordinators detected in the two corpora are outlined.

Both total and relative frequencies (in percentage) are shown. The occurrence of conditional subordinators (*if, unless*) (fig. 29) has proved to be similar in the two corpora (0.14 % BAs vs. 0.10 % WAs), although slightly higher in Britannica.

However, from the statistical point of view, their frequency seems to be very low in the encyclopaedic expository style. Some examples in context are provided below.

> *If* this learning process is not interrupted and especially *if* the social surroundings respond encouragingly or permissively or ambivalently to heavy drinking and intoxication, then the vulnerable personality will become conditioned to react *(from Alcoholism – Britannica)*
>
> *Unless* such lowered rates eventually result in women bearing fewer children, the result is a sharp acceleration in population growth, which can reach rates of 3-4 percent annually in some cases. (*from Poverty – Britannica)*
>
> For example, *if* alcoholism is not considered a disease, third party payments to physicians and hospitals for its treatment would cease. (*from Alcoholism – Wikipedia)*
>
> Blair refused to renegotiate the rebate *unless* the proposals included a compensating overhaul of EU spending, particularly on the Common Agricultural Policy which takes 40% of the EU budget. (*from Blair – Wikipedia)*

Conditional Subordinators				
	Britannica	**%**	**Wikipedia**	**%**
If	328	0.13	383	0.10
Unless	15	0.01	15	0.00
Total	**343**	**0.14**	**398**	**0.10**

Fig. 29 Conditional Subordinators in BAs vs. WAs

The frequency of concessive subordinators (*although, though)* has also given similar results in the two encyclopaedias being 0.12 % in Britannica and 0.10% in Wikipedia (fig. 30).

Concessive Subordinators				
	Britannica	**%**	**Wikipedia**	**%**
Although	160	0.06	267	0.07
Though	137	0.06	158	0.04
Total	**297**	**0.12**	**425**	**0.11**

Fig. 30 Concessive Subordinators in BAs vs. WAs

Some random examples of their use in context are provided below.

Although HAART does not appear to eradicate HIV, it largely halts viral replication, thereby allowing the immune system to reconstitute itself. (*from AIDS - Britannica)*

Secrecy, though still an important function in cryptology, is often no longer the main purpose of using a transformation, and the resulting transformation may be only loosely considered a cipher. (*from Cryptography - Britannica)*

Although there is no blood test specific for alcohol abuse or alcohol dependence (alcoholism), prolonged heavy alcohol consumption may lead to several abnormalities. (*from Alcoholism – Wikipedia)*

Though an enemy of kings, the aristocratic feeling of Alfieri rendered him also a decided foe to the principles and leaders of the French Revolution. *(from Vittorio Alfieri- Britannica)*

As fig. 31 shows, the overall occurrence of causative subordinators (*since, as, because)* has given again not very different frequencies in the two corpora (0.22 % BAs vs. 0.26 %WAs).

Causative Subordinators				
	Britannica	**%**	**Wikipedia**	**%**
Since	115	0.05	331	0.08
As	220	0.09	374	0.10
Because	216	0.09	315	0.08
Total	**551**	**0.22**	**1020**	**0.26**

Fig. 31 Causative Subordinators in BAs vs. WAs

Some random examples in context are provided below:

Since all participants must possess the same secret key, if they are physically separated as is usually the case there is the problem of how they get the key in the first place. (*from Cryptography – Britannica)*

Because medicine was a traditional occupation in certain families, being handed down from father to son, Aristotle in all likelihood learned at home the fundamentals of that practical skill (*from Aristotele –Britannica)*

This is where frequent confusion arises since physiologic dependence does not imply the existence of the disease state which psychiatrists call dependence. *(from Alcoholism – Wikipedia)*

In this case, the thrust is developed in the propulsor as it energizes and accelerates the airflow through the propulsor, e.g., an airstream separate from that flowing through the prime mover. *(from Jet engine – Britannica)*

> Mountain streams are confined to the Gobi's fringes and even then quickly dry up as they disappear into the loose soil or the salty, enclosed depressions *(from Gobi desert – Britannica)*

> The shock appears to have been twofold because Bacon, who was casual about the incoming and outgoing of his wealth, was unaware of any vulnerability and was not mindful of the resentment of two men whose cases had gone against them in spite of gifts they had made with the intent of bribing the judge. (*from Francis Bacon – Britannica)*

> Edward and Warwick next marched north, gathering a large army as they went, and met an equally impressive Lancastrian army at Towton (*from Wars of the Roses – Britannica)*

> Because of the redundancy of the English language, only about 25 symbols of ciphertext are required to permit the cryptanalysis of monoalphabetic substitution ciphers, which makes them a popular source for recreational cryptograms. (*from Cryptography – Britannica)*

> It was also called the Peripatetic School because Aristotle preferred to discuss problems of philosophy with his pupils while walking around -- peripateo -- the shaded walks -- peripatoi -- around the gymnasium). (*from Aristotele – Wikipedia)*

As is a multi functional grammatical item, which can perform different linguistic functions. For example it can be an adverb (*He is as intelligent as his brother* adverb in comparison), a conjunction (*as he is ill, he cannot go out,* as = because), a preposition (*portrayed as a victim,* as = like), etc.

Only *as* in the function of causative subordinator has been taken into account in this analysis, that is to say when it introduces the reasons or causes for an action or event, thus indicating a cause-effect relationship with other information within the same sentence. This search has implied a very careful semiautomatic selection to clean the original "polluted" results which were more or less five times higher. The findings have shown that causative subordinators are the third most used subordinators both in Wikipedia and Britannica corpora. This is probably due to the main intrinsic educational purpose of reference works which provide cause/effect explanation for most of the information provided through a clear, precise and sequential presentation of facts and events. The counting of other subordinators, such as *while, whereas, whereby, as long as*, has confirmed again not very dissimilar average frequencies (0.06% BAs vs. 0.11% WAs) (fig. 32). Nevertheless, the frequency is this time, lower in Britannica. W*hile* and *whereas*, seem to be the most employed adversative subordinators in both encyclopaedic corpora.

Other Subordinators				
	Britannica	**%**	**Wikipedia**	**%**
While	122	0.05	366	0.09
Whereas	30	0.01	28	0.01
Whereby	4	0.00	9	0.00
As long as	4	0.00	10	0.00
Total	**160**	**0.06**	**413**	**0.11**

Fig. 32 Other Subordinators in BAs vs. Was

Some random examples in context are provided.

While a printing apprentice, he wrote under the pseudonym of 'Silence Dogood' who was ostensibly a middle-aged widow. (*from Benjamin Franklin – Wikipedia)*

At low flight speeds the streamtube approaching the lip is larger in cross-section than the lip flow area, whereas at the intake design flight Mach number the two flow areas are equal. (*Jet engine - Wikipedia)*

Boosting referred to a process whereby thermonuclear reactions were used as a source of neutrons for inducing fissions at a much higher rate than could be achieved with neutrons from fission chain reactions alone. (*from Nuclear weapon – Britannica)*

the dread of the consequences of drinking acts as a chemical fence to prevent the patient from drinking as long as he continues taking the drug. (*from Alcoholism – Britannica)*

Total Subordination Features				
	Britannica	**%**	**Wikipedia**	**%**
That clauses	2751	1.11	2322	0.59
Wh- clauses	1602	0.65	2694	0.69
Causative subordinators	551	0.22	1020	0.26
Conditional subordinators	343	0.14	398	0.10
Concessive subordinators	297	0.12	425	0.11
Other subordinators	160	0.06	413	0.11
Total	**5707**	**2.31**	**7270**	**1.86**

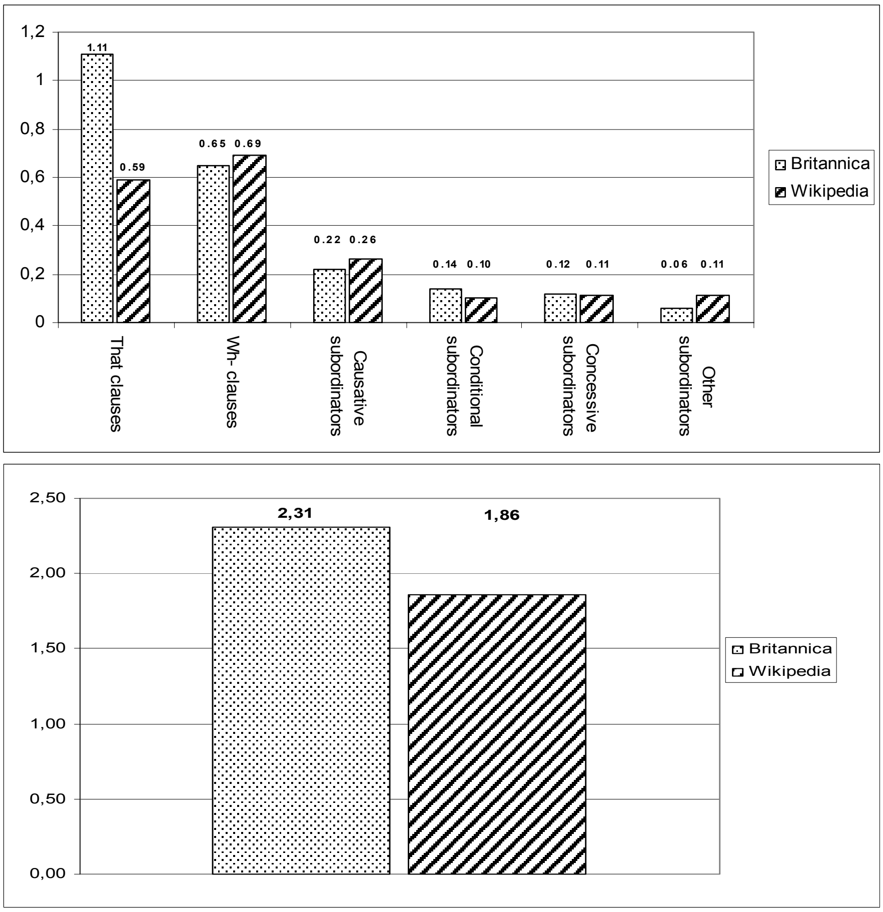


Fig. 33 Specific and overall subordination features in BAs vs. WAs

In conclusion, the microscopic analysis has revealed a dishomogeneous distribution of the selected subordination features in the 100 pair of analyzed articles in both encyclopaedic corpora (see Appendix) and, furthermore, their overall occurrence shows a dissimilar average frequency in the two encyclopaedic corpora.

As fig. 33 shows, the loading of this linguistic class on encyclopaedic expository style is 2.31 % in Britannica vs. 1.86 % in Wikipedia. Although their frequency is not very far, the data shows an higher loading of subordination in Britannica.

If subordination structures involve embedded clauses and the notion of dependency, it should result in a higher structural elaboration and textual complexity. Based on the density of such features it is possible to conclude that with reference to this linguistic class, Britannica encyclopaedic expository register is more elaborate and complex than Wikipedia. The difference in the frequency is slightly more marked when compared to other linguistic classes analysed until now. As the first graph in fig. 33 shows, in both corpora the highest frequency is held, in descendent order, by *that* clauses, followed by *wh-clauses* and causative subordinators. The loading of conditional and concessive subordinators and of the miscellaneous class (other subordinators) is considerable lower. The two bars in the second graph (fig. 33) show the total number of subordination features in the two encyclopaedic corpora.

1.9 Coordination Features

Coordinating conjunctions, which link different grammatical units, are grammatical patterns expressing very basic grammatical relationships. They can link grammatical units of almost any size such as single words, phrases, clauses, morphemes, whole sentences, nouns, verbs adjectives, adverbs, prepositions, pronouns, and determiners. Chafe and Danielewicz (1986: 17) claim:

> Although it is not particularly difficult to accomplish conjoining of clauses and phrases, speakers do not do it as often as writers, in fact academic writers use it three times as often a conversationalists.

Coordination, which contrasts with subordination, is a clause linkage device, which is used to link related clauses which are not involved in a dependency relation. They are easy to use and generally frequent in both oral and written discourse.

Conjunction	Relationship	Example
And	**Addition**	[…] During this period the virus continues to replicate, *and* there is a slow decrease in the CD4 count (the number of helper T cells) (*from AIDS - Wikipedia)*
Nor	**Alternative (negative)**	[…] The two-volume work suggested that women inclined neither toward marriage *nor* a religious vocation should set up secular convents where they might live, study, and teach. (*from Feminism - Britannica)*
But	**Contrast**	[…] HIV-2 can cause AIDS, *but* it does so more slowly than HIV-1. (*from AIDS - Britannica)*
Or	**Alternative**	[…] Moral Sentiments complemented *or* was in conflict with The Wealth of Nations, which followed it. (*from Adam Smith - Britannica)*
Yet	**Contrast**	[…] *Yet* by emphasizing education and political rights that were the privileges of the upper classes (*from Feminism - Britannica)*

Fig. 34 Conjunctions and their functions

The five selected coordinating conjunctions, shown in fig. 34, clarify a specific relationship between equally important ideas. Due to the importance of their grammatical function, their frequency has been investigated in the encyclopaedic corpora.

And is the most used coordinating conjunction in written texts. This data is confirmed by the findings shown in fig. 35. *And* can solve different functions as it can be either a phrase or a clause coordinator; both having a complementary functions. *And* as an independent clause coordinator is represented by clauses linked by an initial *and*, as the example below shows.

> He did this in a lecture at the Royal Institution in February 1860, *and* spoke in favour of Darwin's theory of natural selection in the debate at the British Association [...]

According to Chafe (1982, 1985) *and,* as phrasal coordinator has an integrative function and is used for idea unit expansion, increasing the sentence length. It can join two adverbs, adjectives, verbs or nouns [X^1 + *and* + X^2] (X are both adv/adj/v/n)]. Some examples found in the two encyclopaedic corpora are provided below:

> (ADVERB *AND* ADVERB) [...] *macrophages and dendritic cells. It also directly and indirectly destroys CD4+ T cells. As CD4+ T cell*
>
> (ADJECTIVE *AND* ADJECTIVE) [...] *is common infecting bone marrow, bone, urinary and gastrointestinal tracts, liver, regional lymph of same-sex relationships in the temperate and sub-tropical zone stretching from Northern India*
>
> (VERB *AND* VERB) [...] *It can also infect and cause disease in the eyes and lungs. Progressive multifocal leukoence* [...]
>
> (NOUN *AND* NOUN) [...] *winning the gold medal for anatomy and physiology. In 1845 he published his first* [...]

Some concordances of *and* are shown below.

```
es, such as automatic weapons and  compact, electrically detonate
ave terrorists a new mobility and  lethality, and the growth of a
a new mobility and lethality, and  the growth of air  travel pro
 travel provided new methods  and  opportunities. Terrorism was v
 Germany under Adolf  Hitler  and  the Soviet Union under Stalin.
est, imprisonment,  torture,  and  execution were carried out wit
to  create a climate of fear  and  to encourage adherence to the
ence to the national ideology and  the declared economic, socia
he declared economic, social, and  political goals of the state.
al conflicts (e.g.,  Ireland  and  the United Kingdom, Algeria an
d the United Kingdom, Algeria and  France, and Vietnam and France
 Kingdom, Algeria and France, and  Vietnam and France and  the U
geria and France, and Vietnam and  France and  the United States
rance, and Vietnam and France and  the United States), in dispu
```

The frequency of other important coordinating conjunctions such as *or, but, nor, yet* has also been detected. The overall frequency has shown similar occurrences in the two encyclopaedic corpora (fig. 35). Nevertheless, the total occurrence of this linguistic class is also in this case higher in Britannica than in Wikipedia (4.11% BAs vs. 3.64% WAs). It proves again the slightly higher quantitative incidence of this linguistic feature on Britannica encyclopaedic expository style. All the analysed coordinating conjunctions, have a coherent, proportional and descendent loading in both encyclopaedic corpora, as the data and the first graph in fig. 35 show. The overall representation of coordinating conjunctions in the two corpora is shown in the second graph (fig. 35).

Coordination Features				
	Britannica	**%**	**Wikipedia**	**%**
And	7948	3.22	11255	2.87
Or	1340	0.54	1945	0.50
But	766	0.31	928	0.24
Nor	56	0.02	49	0.01
Yet	54	0.02	62	0.02
Total	**10164**	**4.11**	**14239**	**3.64**

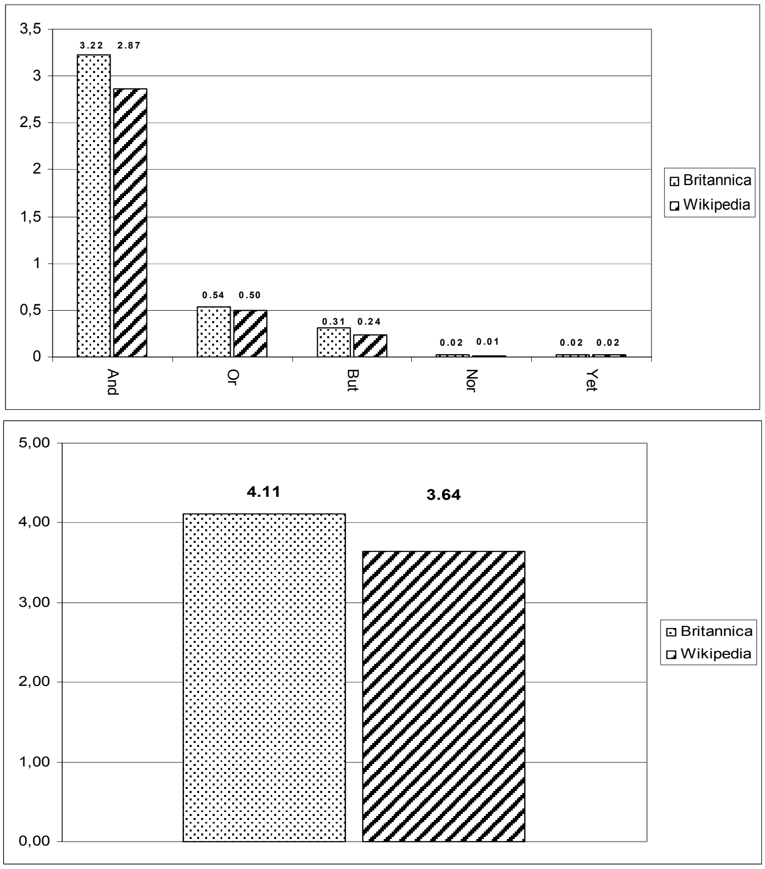


Fig. 35 Specific and overall coordination features in BAs vs. WAs

Comparing the occurrences of subordination and coordination features, it is evident that their frequency is not balanced in the two encyclopaedic corpora, as the data proves (fig. 36).

In particular, Britannica makes a more generous use of both subordination (2.31 % BAs vs 1.86 % WAs) ad coordination devises (4.11% BAs vs 3.64% WAs) than Wikipedia (fig. 36). Its more redundant use integrates the information provided in a better way and produces a higher structural elaboration and textual complexity in Britannica expository style.

Coordination and Subordination				
	Britannica	**%**	**Wikipedia**	**%**
Coordinat	10054	4.11	11255	3.64
Subordina	5707	2.31	7270	1.86

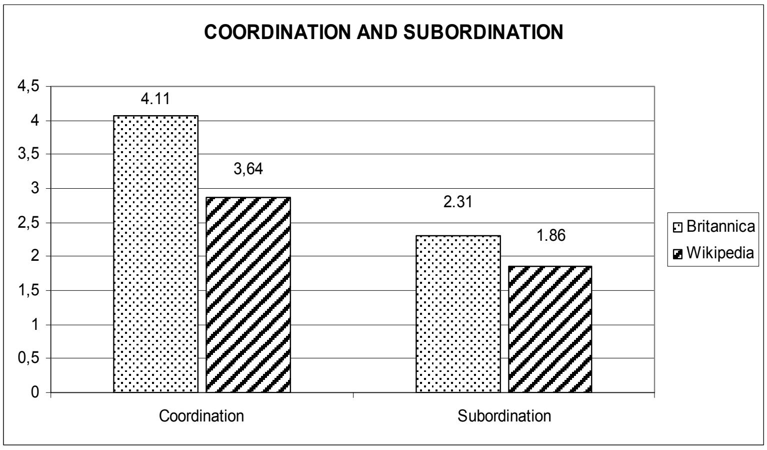


Fig. 36 and Subordination: a comparison

1.10 Conjuncts

Conjuncts are adverbs through which further information is added to the sentence. They connect the sentences with previous parts of the text. They explicitly mark logical relations between clauses, and as such have a keyrole in texts with a highly informational focus. Despite their importance in marking logical relations, few studies have analyzed their distribution. Ochs (1979) notes that conjuncts are commonly found in formal written discourse. Altenberg (1986) looks at concessive and antithetic conjuncts and finds that they are generally more common in writing than in speech. Impersonal and formal expository style is characterized by a recurrent use of them, as Biber claims (1998:163):

> Conjuncts occur frequently with prepositions, passives and nominalizations in highly informational genre such as academic prose, official documents and professional letters.

The queries made in the two encyclopaedic corpora has made possible the measurement of the conjuncts identified by Biber in his Multidimentional Analysis, and to quantify their incidence on the encyclopaedic formal expository style. Findings have revealed very similar frequencies in the two corpora. Following the general trend recorded until now, the frequency of conjuncts has confirmed a slightly higher numerical incidence of this linguistic class in Britannica (0.47%) than in Wikipedia (0.37), as well as a proportional loading on the formal register of both encyclopaedias. As fig. 37 highlights, the 18 selected conjuncts have a different distribution in the two corpora. Except for *however*, *therefore* and *hence,* which occupy the first, sixth and tenth place, the other conjuncts occupy a different position in the scale of occurrences.

Conjuncts					
	Britannica	%		Wikipedia	%
However	254	0.10	However	445	0.11
Thus	188	0.08	For example	166	0.04
For example	121	0.05	Rather	139	0.04
That is	116	0.05	Thus	126	0.03
Rather	97	0.04	Instead	98	0.03
Therefore	63	0.03	Therefore	97	0.02
In addition	53	0.02	That is	91	0.02
Instead	43	0.02	In addition	51	0.01
Moreover	33	0.01	As a result	44	0.01
Hence	30	0.01	Hence	37	0.01
Never	29	0.01	Otherwise	30	0.01
Similarly	26	0.01	Never	22	0.01
On the other hand	25	0.01	Similarly	22	0.01
As a result	24	0.01	On the other hand	22	0.01
Furthermore	18	0.01	Nonetheless	20	0.01
Nonetheless	18	0.01	Consequently	20	0.01
Otherwise	16	0.01	Furthermore	18	0.00
Consequently	9	0.00	Moreover	16	0.00
Total	1260	0.47		1464	0.37

Fig. 37 Conjuncts in BAs vs. WAs

Definitely, the most recurrent conjunct in the two corpora is *however*. It can be used in a number of different ways. When used as a conjunctive adverb it joins two simple sentences to make a compound sentence. It indicates that the relationship between two independent clauses is of contrast or opposition. When *however* is used to write a compound sentence it should be preceeded by a semi-colon or a comma, and followed by a comma. Some concordances of *hower* are shown below.

```
the blood become undetectable  ; however, the virus is still present in
round  pine Lycopodium selago  ; however, except for the latter, which ha
cord of his titles and claims) ; however, he died a  disappointed man.
Genoa to a Christian household ; however, it has been  claimed that he w
ng to the shrine of the Virgin ; however, hostile Portuguese  authoritie
in order and formed the cipher ; however, when the strip was wrapped arou
on, also consists of 64  bits  ; however, only 56 of these can be chosen
ettlement was called Lundenwic ; however, virtually nothing is known abou
few cycles of  the sine wave   ; however, in a radar system having the va
 Honolulu, on December 7, 1941 ; however, the significance of the radar
```

```
red. Any  purported criterion   , however, would appear to be based on a
d be led to suspend  judgment   , however, they would find peace of mind
nsical  living. This provided   , however, neither a theoretical basis f
beyond all possible experience  , however, leads into contradictions  a
eativity is the basis of truth  , however, men make  interpretations by
Scepticism, has sought to show  , however, that, on the standards offere
s of individual cells together  , however, as in  modern solar batterie
; the statutes that do  exist   , however, generally share some common e
ö The element of  criminality   , however, is problematic, because it do
 this  definition is flexible   , however, and on occasion it has been e
```

However can also be used to begin a sentence as the examples which follow show.

> *However*, wreckage has not been found, and some of the theories advanced to explain the repeated mysteries have been fanciful. (*from Bermuda –Britannica)*

> *However,* the test was clearly disappointing and in a 1933 studio memo David O. Selznick, who had signed Astaire to RKO and commissioned the test, described it as "wretched". However the test was clearly disappointing and in a 1933 studio memo. (*from Fred Astaire –Wikipedia)*

> various charges including corruption and undue appropriation. *However*, no definitive conviction sentence has ever been issued on Silvio Berlusconi himself for any of the trials which have concluded so far; (*from Berlusconi –Wikipedia)*

In this case *however* is followed by a comma and what follows is a complete sentence. Sentences beginning with *however*, are closely related to the sentences which precede them. Although conservative grammarians generally insist on the fact that *however* should not be used to begin a sentence, this rule has been often ignored also by reputable writers and by encylopedists as the concordances below show.

```
nished when flying in the area  . However, wreckage has not been found,
mock naval     engagements      . However, it is uncertain whether the a
be charged  the agreed amount   . However, there is a whole gamut of new
ography in the open literature  . However, Admiral Bobby Inman, while
c group least affected by AIDS  . However, most shared  with gay men th
mputer network became feasible  . However, time-sharing systems were the
f located in the United States  . However,  most of these ISPs provided
obbligatos of their own making  . However, these  explorations remained
London   County Council (LCC)   . However, the City Corporation successf
 redevelopment area in Poplar)  . However, severe  air pollution from c
I, was restored  in the 1950s   . However, between 1968 and 1981 the cit
rather than move  through it    . However, road congestion remained a ma
cannot  be false or in error    . However, the view that first-person, p
ous powers of the human  mind   . However, it would be most unwise at pr
om the surface of the  ground   . However, at the lower frequencies (bel
 requires a 500-MHz bandwidth)  . However, since the  energy is directe
```

Although the occurrence of *however* is the highest among the selected conjuncts, its usage is differently distributed in the two corpora. As shown in fig. 38, Britannica largely uses it to make compound sentences indicating that the relationship between the two independent clauses is of contrast or opposition. On the other hand, Wikipedia prefers to use *however* to begin a sentence. Unlike Wikipedia, Britannica's choice seems to demonstrate respect towards the more formal and orthodox style featured by prescriptive grammar rules. Furthermore, unlike Britannica, Wikipedia does not often employ punctuation before and after *however.* In this way, it expresses its preference towards a more unconventional way of writing.

However and Punctuation				
	Britannica	**%**	**Wikipedia**	**%**
. However ,	13	0.005	**198**	0.05
; (or) , however	**201**	0.08	110	0.02

Fig. 38 However and Punctuation in BAs vs. WAs

1.11 Punctuation Marks

A survey on the use of the most common punctuation marks (*commas, full stops* and *semicolons*), has proved that Britannica uses them more extensively than Wikipedia (8.73% BAs vs. 7.77% WAs) (fig. 39).

Punctuation Marks				
	Britannica	**%**	**Wikipedia**	**%**
Commas (,)	14045	5.68	20066	5.12
Full stops (.)	6797	2.75	9721	2.48
Semicolons (;)	742	0.30	667	0.17
Total	**21584**	**8.73**	**30454**	**7.77**

Fig. 39 Punctuation Marks in BAs vs. WAs

For example *commas*, used to indicate separation of different elements within the grammatical structure of a sentence, occur 5.68 % in BA vs. 5.12 % in WA. The frequency of *commas* is so wide because the functions they can perform are numerous; usually they separate one element of a locution from another. In the analysed corpora commas separate independent clauses joined by coordinating conjunctions such as *and* (*commas* followed by the conjunction *and* occur 1809 times in BAs; 0.73%) and 2405 times in WAs 0.61%). Furthermore, commas can set off appositives, and other parenthetical elements and can also separate adverbial clauses and phrases from the main clause they precede and join words in series and a string of adjectives modifying a noun.

F*ull stops,* commonly placed at the end of several different types of sentences are also more recurrent in Britannica than in Wikipedia, as fig. 39 shows (2.75 % BAs vs. 2.48 % WAs).

Semicolons (used in both corpora) coordinate two independent clauses not joined by a coordinating conjunction, particularly when they are joined by conjuncts such as *however* and *furthermore*. They are also used to separate clauses or phrases in series constructions when these already contain commas. As was expected, semicolons frequency is lower than commas and full stops in both corpora; and again their use is more recurrent in Britannica than in Wikipedia (0.30 % BAs vs. 0.17 % WAs).

The frequency of punctuation marks is proportional in the two corpora according to the following order: *full stops, commas* and *semicolons.* Their overall frequency is higher in Britannica than in Wikipedia (8.73 BAs vs. 7.77 WAs). This data could confirm, once again, a slightly higher respect in Britannica towards the prescriptive rules on stylistic formality which recommend the usage of punctuation marks as they can graphically suggest what only intonation can make clear in speech. By contrast, as it has already been observed, Wikipedia sometimes does not pay attention to a meticulous use of punctuation marks. Nevertheless, *Wikipedia Manual of Style* is sensitive to its correct use, as the quotation below from the *Italian* Wikipedia community shows:

> Ancora, le virgole scandiscono il ritmo delle frasi. Una frase, anche corta, che abbia tante, troppe virgole, messe vicine, diventa lenta, pesante, faticosa, per chi, come voi, ora legge. Le frasi con poche virgole invece scorrono via molto veloci e senza intoppi ma spesso diventano molto più lunghe e piene di subordinate tanto che può diventare difficile per la mente del lettore capire cosa sta leggendo ora dal momento che non si è ancora potuta fermare un attimo da quando è iniziata la frase per tirare le somme e ricapitolare tutto quello che gli è passato sotto il naso perché ancora non ha trovato uno straccio di virgola o magari un bel punto. Insomma, ecco un esempio illuminante di come non dovete scrivere e non dovete abusare di subordinate a catena e giri di parole. (http://it.wikipedia.org/wiki/Aiuto:Manuale_di_stile)

1.12 Comments and Remarks

To conclude, Biber (1988, 1995, 2005) has developed a "multidimensional analysis" of register variation in order to map linguistic features pattern in different typologies of spoken and written English texts.

According to Biber (1988:155) expository texts are informational, detached, elaborated, highly explicit and context independent. Expository texts are characterized by the need for precise and dense packaging of information. If compared to other written or spoken registers, informational texts contain more high content words and phrases. In particular Biber (1988:104-5) claims that the following linguistic features are typically used in informational texts:

> […] a high frequency of noun, word length, prepositional phrases, lexical density and attribute adjectives can be associated with an high informational focus and a careful integration of information in a text, and a high frequency of nouns, thus indicates great density of information. Prepositional phrases also serve to integrate high amounts of information into a text. Word length and type token ratio similarly mark high density of information, but they further mark very precise lexical choice resulting in an exact presentation of information content. A high token-type ratio results from the use of many different lexical items in a text, and this more varied vocabulary reflects extensive use of words that have very specific meanings. Attribute adjectives are used to further elaborate nominal information. […] Together these 5 elements are used to integrate high amounts of information into a text, to present information as precisely as possible. These features are associated with communicative situations that require a high informational focus.

This is the reason why the linguisitic classes which he identifies as typical of informational production, and I add of encyclopaedias, have been analysed in this research. The main aim of encyclopaedias is to inform, to educate and to present facts and information in specific entries. As informational texts, the presentation of encyclopaedic information is packed with textual units which make use of an explicit formal expository style.

The purpose of my research inspired by Biber's statistical approach has been the identification of the underlying linguistic parameters of variation, and to specify the linguistic similarities and differences between Britannica and Wikipedia encyclopaedic expository style with respect to *informational* vs. *involved* dimensions. Whereas Biber analysis has produced results that show systematic differences in a range of different registers, from conversation to academic writing, the aim of the first part of this research has been to map intra-genre register variations (Britannica vs.

Wikipedia), and in the second part to map inter-genre variations between informational *WikiLanguage* and involved *Wikispeak*, as will be shown in the next chapter.

To summarize, a selected number of linguistic classes which according to Biber have a positive loading in defining the informational production, has been investigated in this section in order to quantitatively map the formal encyclopaedic expository style of Britannica vs. Wikipedia. A micro and a macroscopic contrastive analysis has been carried out for the purpose of defining, through a frequency criterion, the positive incidence of the selected linguistic features on the formal register of the encyclopaedic expository style, highlighting similarities and differences in the two corpora through the examples and the concordances' excerpts provided.

The data reported in fig. 40, clearly shows that all the findings are not very dissimilar, although most of the times they are slightly higher in Britannica except for sentence length (22.05 Words BAs vs. 22.09 WAs) and gerunds and participial forms (2.38% BAs vs. 2.41% WAs).

(+) Linguistic Features		
	BRITANNICA	**WIKIPEDIA**
Word length (characters)	5.30	5.20
Sentence length (tokens)	**22.05**	22.09
Lexical density (tokens/types)	45.5	43.6
Nominalizations	5.26	4.62
Gerunds and present participles	**2.38**	2.41
Definite/Indefinite Articles	10.02	9.68
Nouns	29.90	29.28
Adjectives	10.54	10.06
Prepositions	14.23	13.42
Passives	0.96	0.96
Subordination features	2.31	1.86
Coordination features	4.11	3.64
Conjuncts	0.47	0.37
Punctuation marks	8.73	7.77
TOTAL (+)	**88.53**	**83.87**

Fig. 40 (+) Linguistic Features in BA vs. WAs

As already pointed out in this section, the above mentioned linguistic classes are also very frequent in academic writing, considered by Biber as the most typical and extreme formal expression of the informational production.

As previously shown, one of the main peculiarities of informational production which, according to frequency criteria makes formal encyclopaedic expository style very close to academic papers, is associated with the use of longer words and sentences, as well as with a higher lexical density. Sentences are expanded through a variety of devices, some of the most frequent ones, being nominalizations, gerunds and present participial forms, prepositions, definite and indefinite articles, nouns, adjectives, prepositions, an extensive use of subordination and coordination devices and punctuation marks. Furthermore, differently from academic writing, a reduced number of passive constructions and conjuncts has been detected in encyclopaedic expository genre of both Britannica and Wikipedia corpora.

Fig. 40 outlines the final macro data related to the linguistic classes which have a positive loading on the formality of the encyclopaedic expository style. The final score is just an orientative data (in which word length, sentence length and lexical density as are not included as they are not grammatical categories). The total frequency of the linguistic classes analysed in the two macro corpora is shown in fig. 41, while the microscopic frequency of each specific pair of encyclopaedic articles is mirrored in fig. 42 . Specific data related to the microscopic analysis is shown in Appendix.

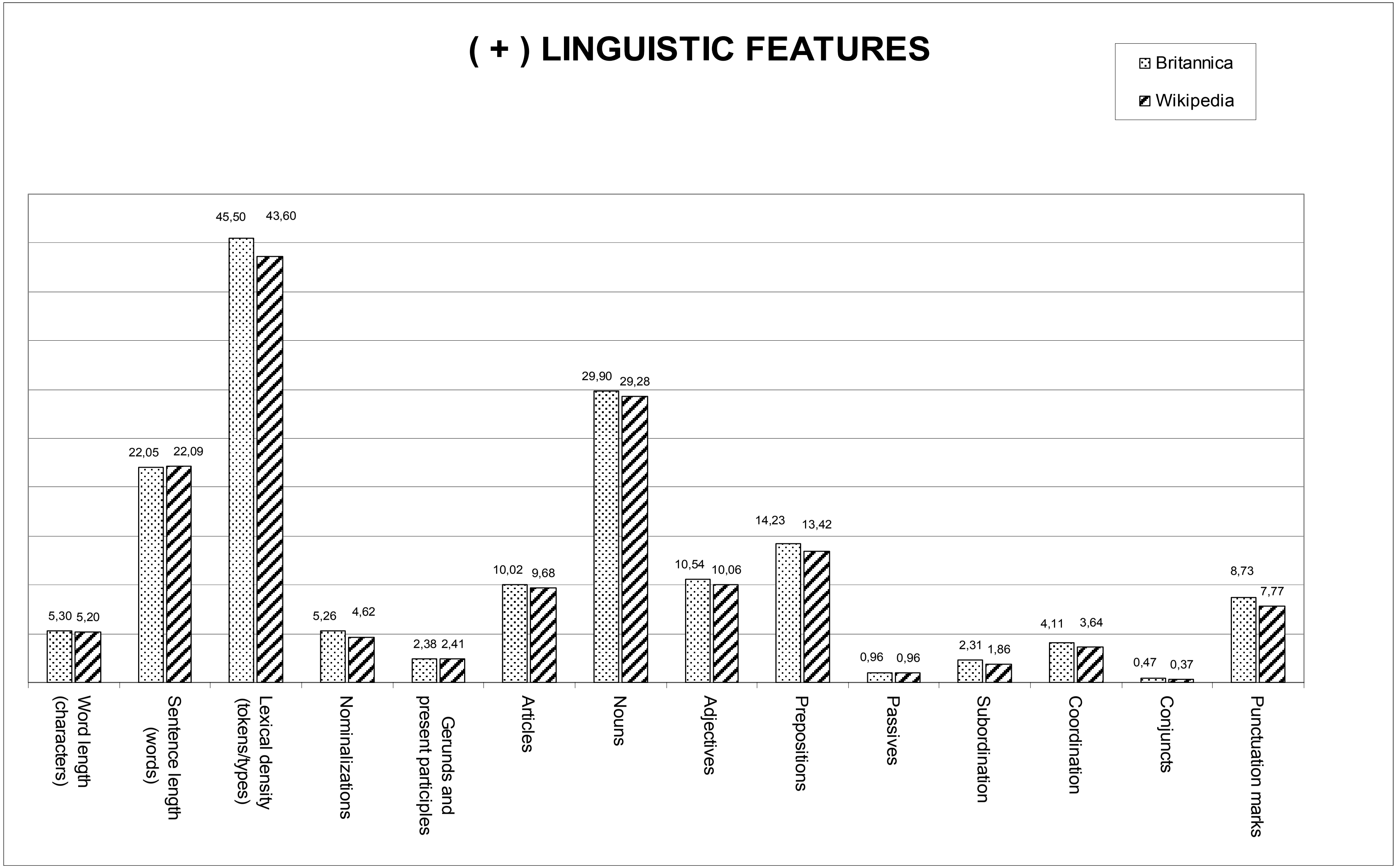


Fig. 41 (+) Linguistic features in BAs vs. WAs

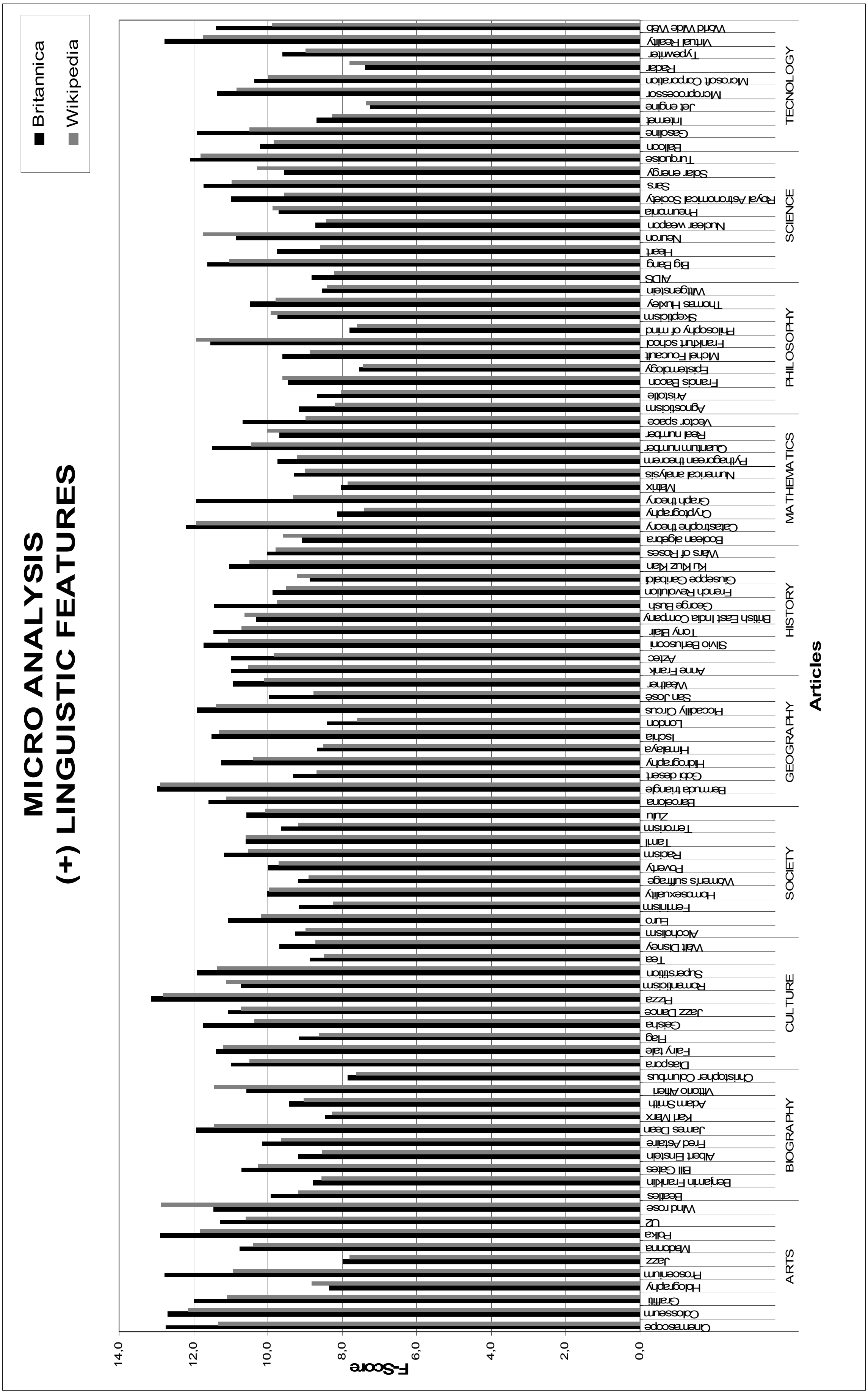


Fig. 42 Micro analysis (+) linguistic features in BAs vs. WAs

2. Linguistic Classes with a Negative Loading on Informational Production

After having analysed the linguistic classes with a positive loading on the formality of the encyclopaedic expository style, the analysis of those which are typical of the opposite dimension (informational vs. *involved* production) has been taken into account. According to Biber's multidimensional approach, if the total frequency of those linguistic classes having a positive loading on formal informational production is summed, and the total frequency of those elements with a negative loading on this dimension subtracted from it, the final value will portray the specific dimension of the linguistic register analysed. Thus, following this approach, linguistic classes with a negative loading on informational production (related to dimension 1) have been subtracted from the total amount of the linguistic features having a positive loading on it, for the purpose of obtaining the degree of the formal informational production of our specific encyclopaedic corpora. To quantify and map our specific encyclopaedic style, the occurrences of the following negative linguistic classes will be explained and investigated:

- Place adverbials
- Time adverbials
- Personal pronouns
- Demonstratives
- Infinitive pronouns
- Mitigating and Boostering devices
- Modals
- Lexical verbs
- Negative forms
- Interrogative sentence
- Reduced forms

2.1 Place and Time Adverbials

Place and time adverbials mark direct reference to the physical and temporal context of the text, or to the external physical and temporal world. Chafe and Danielwicz (1987) consider place and time adverbials as markers of involvement and Biber (1988) interprets their higher distribution as marking situated vs. abstract textual content. Place and time adverbials, which have been analyzed, are listed in fig. 43/44. The selection has been taken from Quirk *et al.* (1985).

Place Adverbials				
	Britannica	**%**	**Wikipedia**	**%**
Under	146	0.06	267	0.07
Toward(s)	78	0.03	126 (39)	0.03
Around	73	0.03	177	0.05
Above	72	0.03	108	0.03
Near	51	0.02	71	0.02
Here	43	0.02	57	0.01
Outside	31	0.01	84	0.02
Below	29	0.01	65	0.02
Behind	17	0.01	55	0.01
Nearby	16	0.01	24	0.01
Inside	5	0.00	35	0.01
Ahead	4	0.00	9	0.00
Next to	4	0.00	10	0.00
On top of	2	0.00	2	0.00
Nowhere	-	-	3	0.00
Total	**571**	**0.23**	**119**	**0.28**

Fig. 43 Place Adverbials in BAs vs. WAs

Some random examples extracted from the two corpora are provided below.

Under the influence of a larger labour supply, the wage rise is moderated and profits are maintained. (*from Adam Smith - Britannica)*

In 1748 he began delivering public lectures in Edinburgh under the patronage of Lord Kames. *(Adam Smith - Wikipedia)*

Still, around the world, women are advancing their interests, although often in fits and starts. (*from Feminism - Britannica)*

men's contribution to child care and domestic labour are typically centred around the idea that it is unfair for the woman to be expected to perform more than half of a household's domestic work. (*from Feminism – Wikipedia)*

Jews lived outside Palestine, about four-fifths of them within the Roman Empire, but they looked to Palestine as the centre of their religious and cultural life. (*from Diaspora - Encyclopædia Britannica)*

In modern use, the 'Diaspora' refers to Jews living outside of the Jewish state of Israel today. (*from Diaspora – Wikipedia)*

Friends who had searched the family's hiding place after their capture later gave Otto Frank the papers left behind by the Gestapo. (*from Anne Frank –Britannica).*

More than 8,000 women, including Anne and Margot Frank and Auguste van Pels, were transported, but Edith Frank was left behind . (*from Anne Frank – Wikipedia)*

Concerning contiguity, people are inclined to think of things that are next to each other in space and time. (*from Epistemology - Encyclopædia Britannica)*

Lillywhites is a major retailer of sporting goods located on the south side, next to the Shaftesbury fountain. (*from Piccadilly Circus – Wikipedia)*

As can be seen in fig. 43, the most used place adverbial is *under*. As the following concordances clearly shows, it is extensively used, in most of the cases, with a figurative value.

```
of perfect liberty, operating under the drives and constraints of
 Rather, it was to show that, under the impetus of the  acquisiti
in lessening child mortality. Under the  influence of a larger la
traditional Christian beliefs under the impact of modern scientifi
ding (first published in 1748 under another title),  which attemp
viral genes by the host cell. Under appropriate conditions these
at an alcoholic is not always under internal pressure to drink and
daily for a few  days; then,  under carefully controlled condition
f attempting to  drink while  under disulfiram medication. A small
le's  death that the school,  under Theophrastus, acquired extensi
as the  school may have been  under Aristotle, it was very importa
ations from different periods under the same title, the editors
 was still at the Academy and under the immediate influence of  P
onogrßfico, Barcelona, Espa±a Under the ruler Itzc¾atl  (1428û40)
ken part in the  examination  under torture of Peacham, which turn
ive work that was to appear   under the title of Instauratio Magna
sence except for the property under investigation. Any property
re than 50,000 ac (20,200 ha) under intensive cultivation. Tourism
chival recordings  assembled  under the supervision of the band an
```

The second most frequently used place adverbial is *toward/s.* The *American Heritage Dictionary of English Usage* (2000) claims that *toward* is more often used in American English, while *toward***s** is widely used in British English. The terms are seldom intermingled in offline texts. This difference has also been noted by Quirk (1985) in *A Comprehensive Grammar of the English Language*. Whereas in Britannica 100% of the times only the American English expression (*toward*) has been employed, a mixed use of this adverb can be noticed in Wikipedia (*toward* 39 times vs. *toward***s** 87 times).

This mixed usage confirms the global and collaborative production of the English edition of Wikipedia, which does not succeed, in this specific case, in imposing a unique spelling choice. The British English spelling is preferred, in spite of the higher number of American contributors, and this choice is probably due to the fact that the British English is linguistically considered the most traditional and formal point of reference.

The occurrence of some 'place adverbials', such as *here* and *above*, is also largely due to textual internal deixis (e.g. *it is shown here, it was shown above, etc.*) as the two concordances' excerpts of *here/above* clearly prove.

```
two points should be noted    here: first, the issue is closely r
g, or be amused by something. Here it   is very difficult to cite
of the concept, not covered   here, that are stressed by Phenomen
of   the senses. The emphasis here is on the way of knowing rathe
Finally,  it  is  significant here, as it was in the discussion o
that   people can have of it. Here the difference between the men
lled the peak power, is taken here to be 1 megawatt. Since a puls
```

```
e AIDS-defining tumors listed above,     HIV-infected patients are
o cause colitis, as described above, and   CMV retinitis can cause
about 1 in 150     (see table above). Post-exposure prophylaxis wi
ve, if they meet the criteria above, the process   is likely the s
mid of Tenochtitlan rose 60 m above the city.   Houses were made o
the most famous of these. See above for   a description of the Ram
```

The overall occurrence of the selected *place adverbials* (fig. 4) is similar, although lower in Britannica than in Wikipedia corpus (0.23% BAs vs. 0.28% WAs). This lower occurrence proves, once again, the use of a slightly more decontextualized and abstract style in Britannica. As shown in fig. 44 the frequency of time adverbials is 0.77% in BAs vs. 0.88% in WAs.

Time Adverbials				
	Britannica	**%**	**Wikipedia**	**%**
When	365	0.15	531	0.14
After	283	0.11	537	0.14
Early	208	0.08	323	0.08
Later	182	0.07	344	0.09
While	139	0.06	366	0.09
Before	118	0.05	214	0.05
Now	111	0.04	230	0.06
Until/til	111	0.04	166	0.04
Late	105	0.04	118	0.03
Earlier	55	0.02	59	0.02
Once	51	0.02	105	0.03
Again	49	0.02	79	0.02
Today	39	0.02	155	0.04
Immediately	31	0.01	34	0.01
Initially	23	0.01	40	0.01
Recently	14	0.01	60	0.02
Formerly	8	0.00	16	0.00
The first time	6	0.00	31	0.01
By the time	5	0.00	9	0.00
Whenever	4	0.00	2	0.00
Tomorrow	4	0.00	5	0.00
As soon as	3	0.00	2	0.00
Yesterday	2	0.00	6	0.00
Everytime	0	0.00	0.00	0.00
Next/last time	0	0.00	0.00	0.00
Afterwards	0	0.00	10	0.00
Tonight	0	0.00	2	0.00
Lately	0	0.00	1	0.00
Total	**191**	**0.77**	**344**	**0.88**

Fig. 44 Time adverbials in BAs vs. WAs.

Although the quantitative variation is as usual minimal it proves to be higher in Wikipedia also in the frequency of time adverbials.

In Wikipedia, the overall higher occurrence of place and time adverbials, which have a negative loading on formal expository style, is again evidence of the slightly lower formality detected in Wikipedia. Of course, both presentation of facts and events in encyclopaedic articles and discussions in talk pages, need clear references to a temporal and spatial setting.

This is the main reason why the frequency variation, as is shown in the next chapter, is not very significant also when Wikipedian articles are compared with the associated talk pages. Some random examples of time adverbials, in their original context of use, are provided below.

> Meanwhile, sometime *before* July 1591, Bacon had become acquainted with Robert Devereux, the young earl of Essex, who was a favourite of the Queen, although still in some disgrace with her for his unauthorized marriage to the widow of Sir Philip Sidney. *(from Bacon Francis – Britannica)*
>
> *Before* beginning this induction, the inquirer is to free his mind from certain false notions or tendencies which distort the truth. (*from Bacon - Wikipedia*)
>
> Aristotle composed the work, *now* lost, On Kingship, in which he clearly distinguishes the function of the philosopher from that of the king. (*from Aristotele -Britannica)*
>
> but *now*, following Plato's example, he gave regular instruction in philosophy in a gymnasium dedicated to Apollo Lyceios, from which his school has come to be known as the Lyceum. (*from* Aristotele - *Wikipedia*)
>
> It was *the first time* in almost 1,500 years that live performances had been held in the amphitheatre.
> (*from* Colosseum - Britannica)
>
> Like a Virgin was also *the first time* Madonna used her most enduring career strategy:
> (Madonna – *Wikipedia*)
>
> Francis' cousin through his mother was Robert Cecil, *later* earl of Salisbury and chief minister of the crown at the end of Elizabeth I's reign and the beginning of James I's. (*from* Bacon Francis - Britannica)
>
> On June 27, 1576, he and Anthony were entered de societate magistrorum at Gray's Inn, and a few months *later* they went abroad with Sir Amias Paulet. *(from Bacon Francis – Wikipedia)*

2.2 Personal Pronouns

One of the main differences between speech and writing, which many researchers focus upon, is in the use of personal pronouns. Chafe and Danielewicz (1987) have considered differences in the relationship between writer/reader and speaker/listener analyzing the use of personal pronouns in various genres representing formal and informal, written and spoken varieties of American English.

Chafe (1982:45) describes the difference as evidence of diverse levels of *involvement* and *detachment.* He argues that the involvement of speakers with their audiences arises from the fact that:

> it is typically the case that a speaker has face to face contact with the person to whom he or she is speaking. That means, for one thing, that the speaker and listener share a considerable amount of knowledge concerning the environment of the conversation. It also means that the speaker can monitor the effect of what he or she is saying on the listener, and that the listener is able to signal the understanding and ask for clarification ...to have less concern for consistency than for experiential involvement.

Yates (1996), Fowler and Kress (1979:201) also examine the usage of pronouns. The latter claim that the omission of subjectivity from written texts is mainly due to conventional social practices rather than direct effect of the specific medium in use. They claim:

> Removal of the pronoun associated with personal speech is felt to be appropriate to the impersonal, generalisingtone of newspapers, textbooks, scientific articles. It is not the medium of writing that creates the impersonality but rather the "appropriate" attendant social practices.

Moreover, Fowler and Kress note that the use of *I* for example is rare in the text of the *Observer* newspaper. It appears most frequently in self-centered articles by people of note, in investigative reporting and in eye-witness accounts. Other several studies have used *first person pronouns* for register comparisons Their use has always been associated with ego-involvement. According to Biber (1988:225) some studies have grouped all pronominal forms together as a single category which is interpreted as marking relatively low informational load, lesser precision in referential identification or a less formal style. This category has been generally interpreted as marking interpersonal focus (Poole and Field, 1976), interactional (Chafe 1992) and involved communication. According to Chafe (1992) second person pronouns, both in the singular and plural form, require a specific addressee and they indicate a high degree of involvement with the interlocutor. Chafe and Danielewicz (1987) and Biber (1986) consider a frequent use of the third singular person pronoun *it* as marking a relatively inexplicit lexical content due to a non informational focus. Biber (1988: 226) claims:

> the personal pronoun "it" is the most generalized pronoun since it can stand for referents ranging from animate beings to abstract concepts. This pronoun can be substituted for nouns, phrases, or whole clauses.

The frequency of the above mentioned personal pronouns (and associated object and reflexive personal pronouns and possessives) has been quantified in both encyclopaedic corpora. It is expected from the mentioned teories to detect low occurences of first, second and third person pronouns in encyclopaedic articles. This is indeed the case.

As already claimed, this linguistic class has a negative loading on the formality of the informational production as its expository style has to be impersonal and objective. In fact, first and second person pronouns have only been found in reported speech or direct quotations in our encyclopaedic corpora. Their use is officially banned. It is declared in Wikipedia *Manual of style* [38]:

> *Avoid first-person pronouns and one*
> Wikipedia articles must not be based on one person's opinions or experiences. Thus, *I* can never be used except when it appears in a quotation. For similar reasons, avoid the use of *we* and *one*. A sentence such as "We should note that some critics have argued in favor of the proposal" sounds more personal than encyclopaedic.
> Nevertheless, it is sometimes appropriate to use *we* or *one* when referring to an experience that *anyone*, any reader, would be expected to have, such as general perceptual experiences. For example, although it might be best to write, "When most people open their eyes, they see something", it is still legitimate to write, "When we open our eyes, we see something", and it is certainly better than using the passive voice: "When the eyes are opened, something is seen."
> It is also acceptable to use *we* in mathematical derivations; for example: "To normalize the wavefunction, we need to find the value of the arbitrary constant *A*."
>
> *Avoid second-person pronouns*
> Use of the second person (*you*), which is often ambiguous and contrary to the tone of an encyclopaedia, is discouraged. Instead, refer to the subject of the sentence or use the passive voice, for example:
> (use) When a player moves past "Go", that player collects $200.
> (use) Players passing "Go" collect $200.
> (use) $200 is collected when passing "Go".
> (don't use) When you move past "Go", you collect $200.
>
> This guideline does not apply to quoted text, which should be quoted exactly.The guideline also does not apply to the Wikipedia namespace, where you refers to the writers to whom articles in the namespace are addressed.

Some random examples in context are provided below.

He would have made, *I* fear, a poor gypsy, commented his principal biographer. (*From Adam Smith - Britannica)*

However, later in the same lecture, discussing modern non-anthropomorphic concepts of God, Russell states: That sort of God is, *I* think, not one that can actually be disproved, as *I* think the omnipotent and benevolent creator can. (*From Agnosticism – Wikipedia)*

There is a famous inscription by Wren's son in St. Paul's Cathedral, addressing the visitor in the following words: "Lector, si monumentum requiris, circumspice" ("Reader, if *you* seek a monument, look about you") (*From London - Britannica)*

The questionnaire asks the following questions: Have *you* ever felt you needed to Cut down on your drinking? Have people Annoyed *you* by criticising your drinking? Have you ever felt Guilty about drinking? (*From Alcoholism – Wikipedia)*

It should be noted that each of these stages is accompanied by institutions suited to its needs. (*From Adam Smith - Britannica)*

In all of this, *it* is notable that Smith was writing in an age of preindustrial capitalism. (*From Adam Smith – Wikipedia)*

[38] *Manual of Style* http://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style

The data in fig. 45 shows that occurrences of first and second person pronouns (in subject and object position + object and reflexive personal pronouns and possessives) are slightly more frequent in Wikipedia, whereas the occurrence of third singular person pronoun *it* is slightly higher in Britannica. The different distribution of personal pronouns may be used as indicator of the scale of personal involvement of author(s) and readers in the text (*I, you*) and of the degree of impersonal style, objectivity or fuzziness of the information provided (*it*). However, the overall variation in the frequency of personal pronouns is as usual very low in the two encyclopaedic corpora (1.05% BAs vs. 0.84% WAs).

I/WE (+ associated object/reflexive p. p. and possessives)				
	Britannica	**%**	**Wikipedia**	**%**
I	152	0.06	238	0.06
We	62	0.03	183	0.05
My	22	0.01	74	0.02
Mine	3	0.00	8	0.00
Me	16	0.01	65	0.02
Us	16	0.01	18	0.00
Our	48	0.02	63	0.02
Ours	0	0.00	1	0.00
Myself	3	0.00	5	0.00
Ourselves	5	0.00	1	0.00
Subotal	**327**	**0.13**	**656**	**0.17**

YOU (+ associated object/reflexive p. p. and possessives)				
	Britannica	**%**	**Wikipedia**	**%**
You (subject + object)	14	0.01	113	0.03
Yourself	2	0.00	2	0.00
Your	12	0.00	44	0.01
Yours	0	0.00	2	0.00
Yourselves	1	0.00	1	0.00
Subtotal	**29**	**0.01**	**162**	**0.04**

IT (+ associated object/reflexive p. p. and possessives)				
	Britannica	**%**	**Wikipedia**	**%**
It (subject +object)	1498	0.61	1695	0.43
Itself	109	0.04	70	0.02
Its	623	0.25	690	0.18
Subtotal	**223**	**0.90**	**245**	**0.63**

TOTAL	**258**	**1.05**	**327**	**0.84**

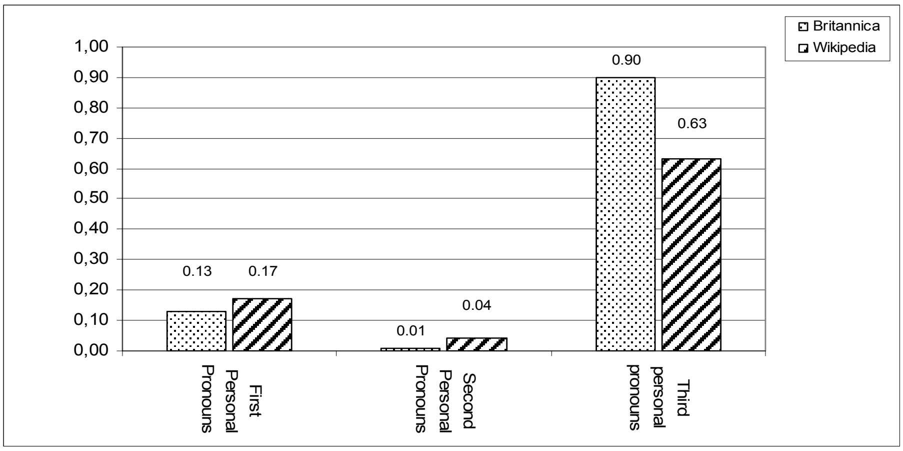


Fig. 45 Personal Pronouns in BAs vs. WAs

2.3 Demonstratives

Demonstratives have also been extensively used for register comparison. They are markers of generalized pronominal reference and are important devices for marking referential cohesion in a text (Halliday *et al.,* 1976). They can have a deictic function or can refer to a specific nominal entity or to an explicit, often abstract concept (e.g. *this shows…*). Demonstratives are used for both text-internal deixis and for exophoric, text-external reference.

Demonstratives				
	Britannica	**%**	**Wikipedia**	**%**
That	387	0.16	1210	0.31
This	856	0.35	1754	0.45
These	419	0.17	639	0.16
Those	196	0.08	227	0.06
Total	**185**	**0.75**	**383**	**0.98**

Fig. 46 Demonstratives

As fig. 46 shows, the total occurrence of demonstratives turns out to be slightly higher in Wikipedia than in Britannica. A careful and semiautomatic investigation has, of course, excluded the occurrence of *that* with relative, complementizer or subordinator functions. Findings confirm, once again, the general trend recorded until now, which shows similar frequencies in both encyclopaedic corpora.

Thus, the negative loading of this linguistic class on formality is coherently lower in Britannica than in Wikipedia. Some random examples in context are provided below.

> Late *that* year he returned to Kirkcaldy, where the next six years were spent dictating and reworking The Wealth of Nations, followed by another stay of three years in London (*from Adam Smith - Britannica*)

> *That* work helped to create the modern academic discipline of economics and provided one of the best-known intellectual rationales for free trade, capitalism and libertarianism. (*from Adam Smith -Wikipedia*)

> The answer to *this* problem enters in Book V, in which Smith outlines the four main stages of organization through which society is impelled, unless blocked by deficiencies of resources, wars, or bad policies of government (*from Adam Smith – Britannica*)

> *This* work, which established Smith's reputation in his day, was concerned with how human communication depends on sympathy between agent and spectator. (*from Adam Smith – Wikipedia*)

> None of *these* treatments has been shown in controlled studies to be more effective than others. (*from Alcoholism*- Britannica)

> Each of *these* symptoms may be continuous or periodic. (*from Alcoholism – Wikipedia*)

> In *those* years the Beatles effectively reinvented the meaning of rock and roll as a cultural form. (*from Beatles* - Britannica)

> After the conquest *those* roads were no longer subject to maintenance and were tragically lost to the test of time. (*from Aztec - Wikipedia*)

2.4 Indefinite Pronouns

Indefinite pronouns, not referring to a specific person, place or thing, add a value of fuzziness to the formal tone of the encyclopaedic texts which should be written in a formal and informational style. Vagueness should be avoided, as the dominant style of this style should be accurate and precise. For this reason indefinite pronouns have a negative incidence in defining the formality of the encyclopaedic style. Some examples in context are provided below:

> Solar energy is also used on a small scale for *other* purposes besides those described heretofore. (*from Solar energy- Britannica*)
>
> The ANC, and *other* movements, were banned.) This acceptance extended to the covert provision of funds and guerilla warfare training to Inkatha by the government. (*from Zulu - Wikipedia*)
>
> In *some* countries of the region the prevalence of HIV infection of inhabitants exceeded 10 percent of the population. (*from Aids- Britannica*)
>
> However, *some* may improve feelings of well-being in people who believe in their value. (*from Aid s- Wikipedia*)
>
> He had a low profile as a musician while acting as the producer of *several* successful films. (*from Beatles- Britannica*)
>
> In the modern districts of the city are *several* avenues on which most of the international merchants offering clothing, jewelry, leather goods and *other* items have their stores. (*from Barcelona – Wikipedia*)
>
> The doctrine that humans cannot know of the existence of *anything* beyond the phenomena of their experience. (*from Agnosticism - Britannica*)
>
> Their objective was nothing less than complete destruction of the state. *Anything* that contributed to this goal was regarded as moral. (*from Terrorism – Wikipedia*)

As fig. 47 shows, the overall occurrence of indefinite pronouns is the same in the two corpora. Nevertheless, unlike other linguistic classes, their incidence is not proportional in the two corpora.

Whereas the most frequent indefinite pronoun is *other* (0.22% BAs vs. 0.23% WAs) in both corpora, Wikipedia makes a slightly more intensive use of the pronoun *many* (0.15% BAs vs. 0.20% WAs) compared to Britannica.

Indefinite Pronouns				
	Britannica	%	Wikipedia	%
Other	553	0.22	902	0.23
All	497	0.20	712	0.18
More	483	0.20	806	0.21
Some	476	0.19	805	0.21
Most	409	0.17	707	0.18
Many	366	0.15	780	0.20
Any	230	0.09	302	0.08
Each	204	0.08	201	0.05
Another	177	0.07	235	0.06
Much	164	0.07	287	0.07
Several	142	0.06	235	0.06
Either	97	0.04	100	0.03
Others	97	0.04	175	0.04
Something	80	0.03	72	0.02
a/few	73	0.03	125	0.03
Little	64	0.03	103	0.03
Nothing	57	0.02	35	0.01
Anything	38	0.02	22	0.01
Someone	25	0.01	36	0.01
Everything	22	0.01	19	0.00
Anyone	19	0.01	32	0.01
No one	15	0.01	12	0.00
None	11	0.00	15	0.00
Everyone	4	0.00	12	0.00
Nobody	4	0.00	3	0.00
Somebody	1	0.00	0	0.00
Anybody	1	0.00	2	0.00
Everybody	1	0.00	6	0.00
Total	431	1.74	674	1.72

Fig. 47 Indefinite Pronouns in BAs vs. WAs

2.5 Mitigating and Boostering Devices

2.5.1 Downtoners

Downtoners, which immediately precede adjectives are a group of adverbs that scale down the effect of the modified item, giving some indication of the degree of uncertainty or probability of the information provided (Biber *et al.,* 2005:178). According to Quirk *et al.* (1985:597-602), they have a *general lowering effect on the force of the verb.* Chafe and Danielewicz (1987) claims that they are commonly used in formal academic writing to indicate reliability of the information, marking uncertainty towards a proposition. Stubbs (1983:185) claims that downtoners, typically recur in spoken discourse (and I add, in any kind of involved production, CMC included) for the purpose of *facilitating cooperation between the partners by avoiding threatening the hearer.* Some random examples in context are provided below.

> Faced with conflict of interest and other charges, he resigned after *only* seven months in office. (*from Berlusconi - Britannica*)

> his *only* gratification, in the absence of freedom among the continental states, came from contemplating the wild and sterile regions of the north of Sweden, where gloomy forests, lakes and precipices encouraged his sublime and melancholy ideas (*from Vittorio Alfieri – Wikipedia*)

> The control mode of the turboprop also is *somewhat* different from that of a helicopter's turboshaft engine. (*from Jet engine - Britannica*)

> a gas turbine engine is used as powerplant to drive (propeller) shafhigh efficiency at lower subsonic airspeeds(300 knots plus), high shaft power to weight Limited top speed (aeroplanes), *somewhat* noisy, complexity of propeller drive. (*from Jet engine – Wikipedia*)

> to affirm, as Spencer did, the existence of a being about whom absolutely nothing else can be said is a *rather* comical hypostatization (taking of an abstraction as real), which is surely indiscernible from affirming no being at all. (*from Agnosticism- Britannica*)

> Silvio Berlusconi undoubtedly has a *rather* long record of judicial trials, as several crimes have been alleged to him or his firms (see also the following subsection on Berlusconi's trials), including false accounting, tax fraud, corruption and bribery of police officers and judges. (*from Berlusconi- Wikipedia*)

> The *relatively* small effect of the weight flow of fuel in creating a difference between the weight flow of the inlet and exhaust streams is intentionally disregarded. (*from Jet engine- Britannica*)

> Turbojet engines take a *relatively* small mass of air and accelerate it by a large amount, whereas a propeller takes a large mass of air and accelerates it by a small amount. (*from Jet engine - Wikipedia*)

Biber (2005) analyzing both informational and involved production, points out the different use of downtoners in spoken and academic written discourse as fig. 48 shows. He claims that while the downtoner *pretty*, is commonly used in the AE conversations, it is never found in academic prose. By contrast, the dowtoner *relatively* is commonly found in academic prose, but rarely in conversation.

The present investigation has reached similar results. The downtoner *pretty* commonly used in conversations, as it will be shown, is also very recurrent in Wikipedian talk pages, but not in encyclopaedic articles, as fig. 49 shows.

	Academic prose (c. 5 million words)	Conversation (AE; c. 2.5 million words)
Pretty	-	******** [39]
Relatively	****	-
Rather	**	-
Fairly	**	-
Slightly	**	-
Almost	*	-
Somewhat	*	-
Nearly	-	-

Fig. 48 Distribution of downtoners (Biber & Conrad 2005: 178)

By contrast, *relatively* is more often recurrent in encyclopaedic articles than in talk pages. The frequency variation of downtoners in the two encyclopaedic corpora does not seem to be, as usual, very striking. It is slightly higher in Britannica (0.25%) than in Wikipedia (0.23%). This data marks the lightly superior degree of uncertainty, and lack of neutrality of the former, since downtoners, although apparently insignificant devices, convey the author's point of view (fig. 49). As will be shown in the next chapter, the frequency of downtoners is higher in talk pages, confirming the necessity to use, also in CMC, mitigating devices to facilitate cooperation inside the working community.

Downtoners				
	Britannica	**%**	**Wikipedia**	**%**
Only	379	0.15	524	0.13
Rather	97	0.04	139	0.04
Relatively	**34**	**0.01**	**54**	**0.01**
Merely	29	0.01	23	0.01
Nearly	27	0.01	53	0.01
Partly	23	0.01	22	0.01
Somewhat	22	0.01	28	0.01
Slightly	20	0.01	30	0.01
Partially	14	0.01	16	0.00
Practically	10	0.00	5	0.00
Fairly	7	0.00	19	0.00
Hardly	7	0.00	5	0.00
Barely	3	0.00	1	0.00
Scarcely	1	0.00	2	0.00
Pretty	**0**	**0.00**	**1**	**0.00**
Mildly	0	0.00	0	0.00
Total	**633**	**0.26**	**922**	**0.23**

Fig. 49 Downtoners in BAs vs. WAs

[39] Each * = 50 occurrences per million words; = less than 20 occurrences per million words

2.5.2 Hedges

Hedges are also mitigating devices which mark propositions as probable or uncertain. They are employed in both spoken and written discourse, involved or informational production and have a keyrole in communication. Through their use, writers and speakers (I suggest *writing speakers* in CMC) distinguish their opinions from facts and evaluate the (un)certainty of their assertions. Differently from downtoners which give some indication of the degree of uncertainty, hedges simply mark a proposition as uncertain.

Biber (1988:240) finds hedges co-occuring more frequently with interactive features and with reduced or generalized lexical content (e.g. first, second and third person pronouns, interrogative sentences, contractions, emphatics, etc.) and Chafe (1982) claims that hedges *mark fuzziness in involved discourse*, convey doubt and point out the author's tentative assessment of information provided balancing conviction with caution. Through hedges, academic writers seek to modify their assertions, toning down potentially risky claims, and what they believe to be correct.

Hedges				
	Britannica	**%**	**Wikipedia**	**%**
Almost	98	0.04	91	0.02
Kind of	57	0.02	23	0.01
Sort of	21	0.01	15	0.00
More or less	8	0.00	10	0.00
Something like	4	0.00	1	0.00
At about	4	0.00	12	0.00
Maybe	1	0.00	4	0.00
Total	**189**	**0.07**	**156**	**0.03**

Fig. 50 Hedges in BAs vs. WAs

With these premises in mind, the occurrence of the most common hedges, listed in fig. 50, has been measured in the two encyclopaedic corpora.

As can be noticed, their frequency is very similar in the two corpora, although it proves to be slightly higher in Britannica than in Wikipedia (0.07% BAs vs. 0.03% WAs).

Hence, Britannica contributors more than Wikipedians convey their *Personal Point of View* marking their statements as uncertain or probable. As was expected, their incidence is very low in both encyclopaedic corpora, since mitigating devices are not typical of informational production, but of the opposite involved dimension (see chapter 5).

Some examples of *hedges* in context are provided below.

he form, titles, and order of Aristotle's texts that are studied today were given to them by Andronicus *almost* three centuries after the philosopher's death, (*from Aristotele – Britannica*)

People from Bosnia can be found *almost* anywhere in the world. (*from Diaspora - Wikipedia*)

the most successful one being the exemplary account in Novum Organum of how his inductive tables show heat to be *a kind of* motion of particles. *Bacon* (*from Francis–Britannica*)

the German Criminal Court Laboratory, the Bundeskriminalamt (BKA) was asked to examine *the kind of* paper and the types of ink used in the manuscript of the diary. (*from Anne Frank - Wikipedia*)

In non industrial societies (present and past), this *sort of* inability to provide for one's basic needs rests mainly upon temporary food shortages caused by natural phenomena or poor agricultural planning. (*from Poverty–Britannica*)

He later explained that he "was joking", and he meant to create a relaxed climate, that this *sort of* meeting were meant to "create friendship, cordiality, simpatia and kind relationships" between the participants (*Berlusconi - Wikipedia*)

Ragtime differs substantially from jazz in that it was a through-composed, fully notated music intended to be played in *more or less* the same manner each time, (*from Jazz – Britannica*

this treatment is *more or less* acceptable by tradition, and because such material is usually of a higher grade to begin with. (*from Turquoise – Wikipedia*)

2.5.3 Amplifiers

Amplifiers (e.g. *clearly, obviously*, of *course*) have the opposite effect of downtoners, since they boost the force of the verb (Quirk e*t al.* 1985:590-7). Through amplifiers academic writers (and I suggest, also encyclopaedia's contributors) emphasize what they believe to be correct, amplifying their certainties. Amplifiers can communicate both interpersonal and ideational (or conceptual) information. They convey authorial participation and are central aspects of the rhetorical or interactive character of academic writing marking certainty or conviction in the proposition and writer's involvement and solidarity with the audience (Hyland, 2000). According to Chafe (1985) amplifiers are used to indicate the reliability of propositions.

As downtoners and hedges, amplifiers are mainly typical features of involved production, conveying the point of view and the evaluative position of the author. Thus, their frequency is expected to be low in the objective and neutral informational encyclopaedic production, and actually it is.

The quantitative analysis which follows has shown an almost identical frequency of amplifers being 0.15 % in both encyclopaedic corpora (fig. 51). Thus, the negative influence of this linguistic feature on the formality of the encyclopaedic production is very low.

Amplifiers				
	Britannica	**%**	**Wikipedia**	**%**
Very	138	0.06	275	0.07
Highly	37	0.01	60	0.02
Clearly	35	0.01	32	0.01
completely	26	0.01	34	0.01
Extremely	23	0.01	29	0.01
Fully	18	0.01	23	0.01
Greatly	16	0.01	38	0.01
Of course	15	0.01	13	0.00
Obviously	10	0.00	6	0.00
Strongly	10	0.00	21	0.01
Totally	8	0.00	10	0.00
Altogether	7	0.00	8	0.00
Absolutely	5	0.00	7	0.00
Intensely	5	0.00	2	0.00
Enormously	5	0.00	4	0.00
Thoroughly	4	0.00	6	0.00
Perfectly	3	0.00	4	0.00
Utterly	2	0.00	2	0.00
Entirely	-	-	31	0.01
Total	**367**	**0.15**	**605**	**0.15**

Fig. 51 Amplifiers in BAs vs. WAs

Some examples of their use in the original encyclopaedic context are provided below.

There are no possible or conceivable conditions in which this proposition is not true (on the assumption, *of course*, that the words, husband and married are taken to mean what they ordinarily mean). (*from Epistemology – Britannica)*

But *of course*, it might turn out that he was mistaken, and that what he thought was true was actually false. This is not the case with knowledge. (*from Epistemology - Wikipedia)*

A whole generation of low - and medium - bypass engines has *completely* supplanted the first generation of aircraft powered by (zero-bypass) turbojet engines. (*from Jet engine – Britannica)*

Estimates of the prevalence of alcoholism vary *greatly*, depending on how it is defined as well as on the methods of estimation. (*from Alcoholism - Wikipedia)*

The course persuaded the inquirer that reason cannot attain truth; yet certainty in true religious belief was still thought *absolutely* necessary for salvation. (*from Agnosticism - Britannica)*

The current trend in VR is actually to merge the two user interfaces to create a *fully* immersive and integrated experience. (*from Virtual reality - Wikipedia)*

2.5.4 Emphatics

While amplifiers indicate the degree of certainty in a proposition, emphatics simply mark the presence of certainty. Labov (1984) discusses forms of this type under the label of “intensity” as they convey emotions and personal view towards the linguistic proposition.

According to Biber (1988) *emphatics* mark involvement with the topic and frequently occur in conversations. As can be noticed in fig. 52 their negative incidence on the formality and neutrality of the encyclopaedic production is similar and their occurrence is very low in both encyclopaedic corpora.

Emphatics				
	Britannica	**%**	**Wikipedia**	**%**
Most	409	0.17	707	0.18
Such a + adj	81	0.03	49	0.01
Just	74	0.03	106	0.03
Real + adj	54	0.02	147	0.04
Really	41	0.02	37	0.01
A lot + adj	2	0.00	20	0.01
For sure + adj	-	-	1	0.00
Total	**661**	**0.27**	**106**	**0.28**

Fig. 52 Emphatics in BAs vs. WAs

Some examples in the original encyclopaedic context are provided below.

> Although these anomalies may seem simple and unproblematic at first, deeper consideration of them shows that just the opposite is true. (*from Epistemology – Britannica)*

> In the second sense of belief, to believe something just means to think that it is true. That is, to believe P is to do no more than to think, for whatever reason, that P is the case. (*Epistemology – Wikipedia)*

> Perhaps never before in history had there been such a large spontaneous gathering as the one that cheered him through the streets of London. (*from Garibaldi-Britannica)*

> Since the church was such a big part of the lives of South Carolinians (91% of church-goers in South Carolina in 1888 were either Methodist or Baptist), they were discouraged from joining the suffrage movement. (*from Women's suffrage – Wikipedia)*

> how many women really wanted equality? The debate was not limited to the United States. *(from Feminism- Britannica)*

> However, it was only really considered an amusing curiosity of no obvious value. *(From Jet engine – Wikipedia*

To summarize, the total occurrence of the above mentioned mitigating or boostering devices such as downtoners, hedges, amplifiers and emphatics, has a negative loading in defining the formality of the encyclopaedic expository style as, intentionally or untentionally, they convey the writer’s point of view, disregarding one of the “golden rules” of the informational production which claims the fundamental value of neutrality and objectivity in the exposition of facts and events. The total value of mitigating and boostering devices show to be higher, although slightly, in Britannica corpus (0.75% BAs vs. 0.69% WAs). Neverthless, their occurrence is minimal, thus they do not compromise the formality of the encyclopaedic informational production (fig. 53).

Total Mitigating and Boostering devices				
	Britannica	**%**	**Wikipedia**	**%**
Downtoners	633	0.26	922	0.23
Hedges	189	0.07	156	0.03
Total mitigating devices	**822**	**0.33**	**1078**	**0.26**
Amplifiers	367	0.15	605	0.15
Emphatics	661	0.27	1067	0.28
Total boostering devices	**1028**	**0.42**	**1672**	**0.43**
Total	**1850**	**0.75**	**2750**	**0.69**

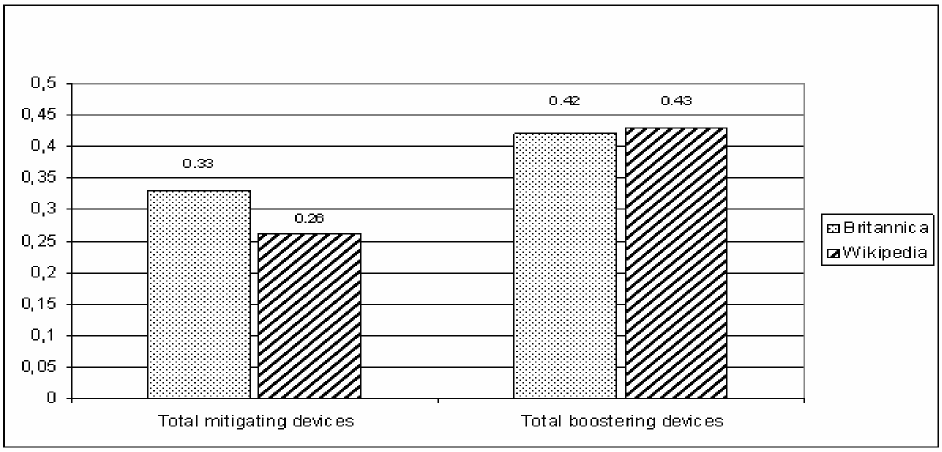


Fig. 53 Mitigating and boostering devises in BAs vs. WAs

In conclusion, both mitigating and boostering devices have a keyrole in detecting evaluative position of the writer and his/her emotional participation in the presentation of information. By breaking the hypothetical neutrality, impersonality and objectivity of encyclopaedic expository prose they convey the authorial judgement and contributor's personal point of view in a veiled way.

2.6 Modals

Hodge and Kress (1988 :121) begin their discussion of modality in language by noting that:

> In everyday communication it manifestly matters a great deal what weight we attach to an utterance. A statement may be said emphatically, without qualifications, and we know that we are being asked to believe that it is true. Or it may be hedged with 'I think', 'it may be that'. Perhaps it is spoken with rising intonation like a question, and we know that the speaker is offering the statement more tentatively. Or it may be said with a laugh or an ironic sarcastic tone, and we know the speaker does not believe the statement at all.

These methods of encoding attitude towards a statement or the content of an utterance are described by Hodge and Kress (1988) as the *modality system of language.* Though they use this term to cover many aspects of communication, they note that the system manifests itself most notably in the use of *modal auxiliaries*. Their analysis provides interesting results. Another definition of modality is given by Kiefer (1994:2514) as follows:

> the relativization of the validity of sentence meanings to a set of possible worlds. Talk about possible worlds can thus be construed as talk about the ways in which people could conceive the world to be different.

According to Quirk *et al.* (1985) the use of modal verbs does not imply a simple declaration of facts as it includes the assertion or denial of any degree or manner of affect, belief, certainty, desire, obligation, possibility, or probability on the part of the utterer. They claim that it is possible to distinguish three functional classes of modals: (1) those marking permission, possibility, or ability; (2) obligation or necessity; and (3) volition or prediction.

According to Biber (1988:107), the use of modal verbs and semi-modals (*have to*) is very common in conversation. In particular, possibility modal are used to flag uncertainty or lack of precision in the presentation of information. Finally, Yates (1996) has also investigated the usage of modals in CMC, speech or writing arguing that their frequency is significantly higher in CMC, with writing having the lowest usage of all three.

As can be noted, *possibility* modals have the highest occurrence in both encyclopaedic corpora (fig. 54), although their frequency is greater in Britannica than in Wikipedia. (0.48% BAs vs. 0.40% WAs). Possibility modals are followed by *predictive* and *necessity* modals. In particular, the occurrence of possibility and necessity modals is slightly higher in Britannica, while the opposite trend has been recorded for predictive modals.

The analysis has revealed once again, a similar overall occurrence of this linguistic class in the two encyclopaedias although in Britannica the total frequency is slightly higher than in Wikipedia (0.82% BAs vs. 0.72% WAs). Consequently, modals have a similar negative incidence in defining the formal informational production in both encyclopaedic corpora, but they have a slightly higher negative loading in Britannica, unveiling in a disguised way, especially through the use of possibility and necessity modals, the position of the writer(s) and the lack of precision in the presentation of information.

The occurrence of possibility, predictive and necessity modals in the two encyclopaedic corpora is shown in fig. 54.

Modals				
	Britannica	**%**	**Wikipedia**	**%**
Possibility Modals				
Can	541	0.21	753	0.19
May	291	0.12	547	0.14
Could	240	0.10	208	0.05
Might	118	0.05	79	0.02
Total	***1190***	***0.48***	***1587***	***0.40***
Predictive Modals				
Would	364	0.15	561	0.14
Will	169	0.07	328	0.08
Shall	3	0.00	16	0.00
Total	***536***	***0.22***	***905***	***0.23***
Necessity Modals				
Must	169	0.07	141	0.04
Should	85	0.03	144	0.04
Have to	29	0.01	24	0.01
Ought	10	0.00	5	0.00
Total	***293***	***0.11***	***314***	***0.08***
Total	**201**	**0.82**	**280**	**0.72**

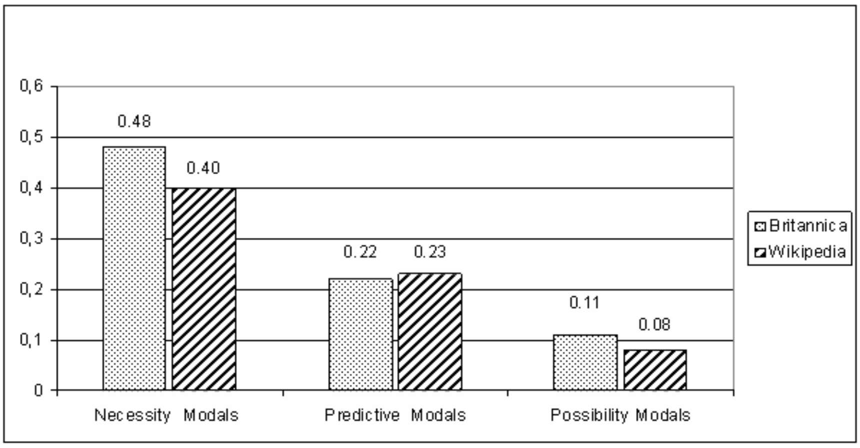


Fig. 54 Modals in BAs vs. WAs

Some random examples are provided below.

Since this image is aerial, the microscope *can* be positioned in such a way that it *can* focus on the required region. In the same way, a camera also *can* be focused at the required depth and *can* photograph objects inside a deep transparent chamber. (*from Photography - Britannica)*

It *can* only lightly brighten or darken zones of the
hologram. This does not prevent the creation of the half-spherical wave fronts when the hologram is illuminated. (*from Holography – Wikipedia*)

It is clear that communications connectivity *will* be an important function of a future Internet as more machines and devices are interconnected. (*from Internet - Britannica)*

Some commercial organizations encourage staff to fill them with advice on their areas of specialization in the hope that visitors *will* be impressed by the expert knowledge and free information, and be attracted to the corporation as a result. (*from Internet- Wikipedia)* he *would* have studied the role in therapy of diet, drugs, and exercise; he *would* have learned how to check the flow of blood, apply bandages, fit splints to broken limbs, reset dislocations, and make poultices of flour, oil, and wine. (*from Aristotele – Britannica)*

As such, Aristotle's early education *would* probably have consisted of instruction in medicine and biology from his father. Little is known about his mother, Phaestis. It is known that she died early in Aristotle's life. (*from Aristotele- Wikipedia)*

all persons such as hospital workers or family members who come into close contact with a patient *must* follow strict routines of cleanliness (*from SARS– Britannica)*

The ceremonies *must* be held on sacred ground at sacred times, with all actors in special costumes. All actors *must* assume an attitude of solemn respect toward the proceedings. (*from Terrorism- Wikipedia)*

2.7 Lexical Verbs

Lexical verbs form the primary verbs of a language. According to Biber, lthe most common lexical verbs (*get, go, say, know, think, see, want, come, give, mean, take, make)* are very frequent in conversation.

Their high occurrence has been recorded by many linguists in conversations and in more general involved production. By contrast, they should not be very frequent in informational production as a more elaborate lexical choice should be preferred as graphs from Biber's presentation reveal (fig. 55a/b).

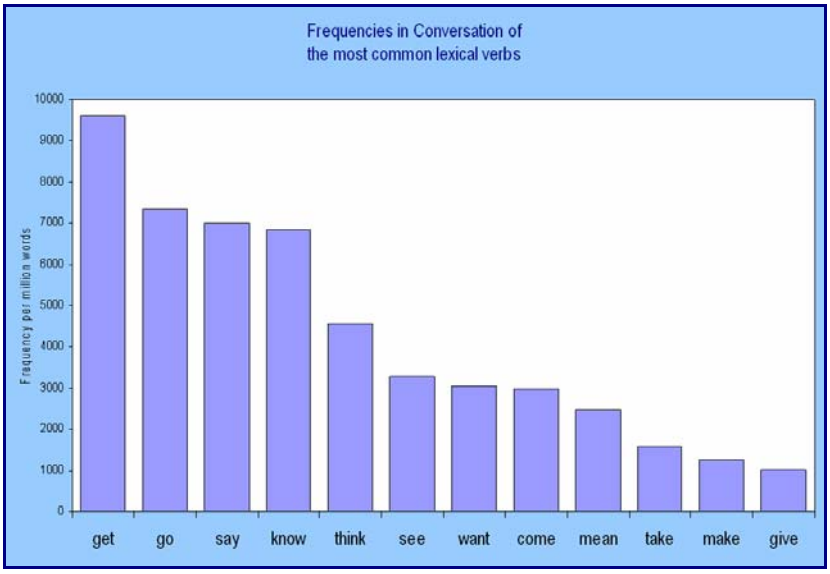


Fig. 55a What can corpus linguistics tell us about English grammar?
Biber (2006)

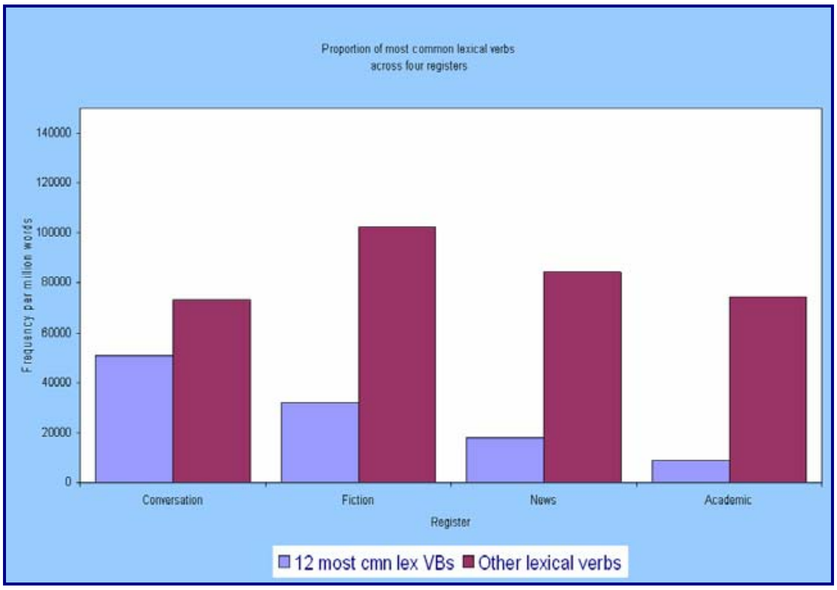


Fig. 55b What can corpus linguistics tell us about English grammar?
Biber (2006)

Thus, a high frequency of lexical verbs has a negative loading on the encyclopaedic genre since it conveys a poverty of language, due to the restricted variety in use. Following the general trend observed until now, frequency of lexical verb, is low and very similar in the two encyclopaedic corpora, although slightly higher in Wikipedia (0.84% BAs vs. 0.97% WAs).

Lexical Verbs *				
	Britannica	**%**	**Wikipedia**	**%**
Know	409	0.17	508	0.13
Make	317	0.13	503	0.13
Give	223	0.09	326	0.08
ee	203	0.08	577	0.15
Say	188	0.08	235	0.06
Think	180	0.07	185	0.05
Take	177	0.07	575	0.15
Mean	126	0.05	210	0.05
Come	123	0.05	192	0.05
Go	82	0.03	183	0.05
Want	25	0.01	46	0.01
Get	23	0.01	251	0.06
Total	**207**	**0.84**	**379**	**0.97**

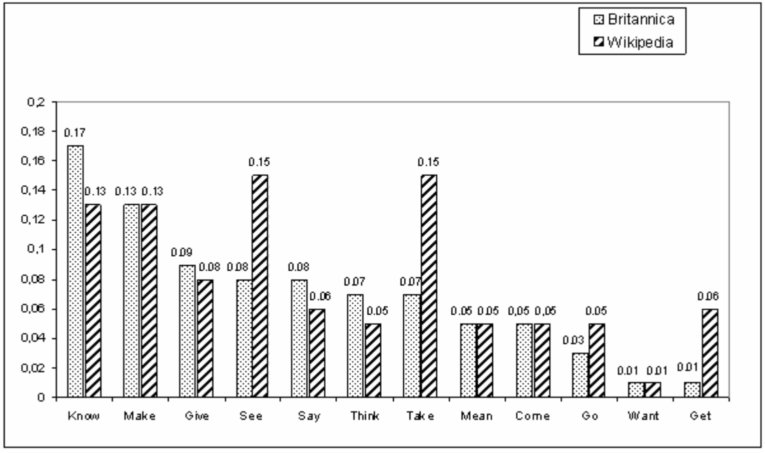


Fig. 56 Lexical verbs in BAs vs. WAs

Each figure in fig. 56 represents the sum of the total occurrences found for the listed verbs in the simple present (included third person), past and perfect tenses.

As can be noted, the occurrence of lexical verbs is not homogenously distributed in the two corpora since Wikipedia makes a more extensive use of the verbs *see, take* and *get* than Britannica.

2.8 Negative Forms

The use of negative forms is not significantly present in formal and expository style, as fig. 57 shows. There is twice as much negation overall in speech as in writing, a distribution that Tottie (1983) attributes to the greater frequency of repetitions, denials, rejections, questions and mental verbs in speech. Tottie distinguishes between synthetic and analytic negation. Synthetic negation (*no, neither)* is more literary and seemingly more integrated; by contrast, analytic negation (*not*) is more colloquial and seems to be more associated with a fragmented presentation of information and with text of low informational density.

Negative Forms				
	Britannica	**%**	**Wikipedia**	**%**
Not [40]	904	0.37	1209	0.31
No	290	0.12	356	0.09
Nor	56	0.02	49	0.01
Neither	36	0.01	22	0.01
Total	**1286**	**0.52**	**1636**	**0.42**

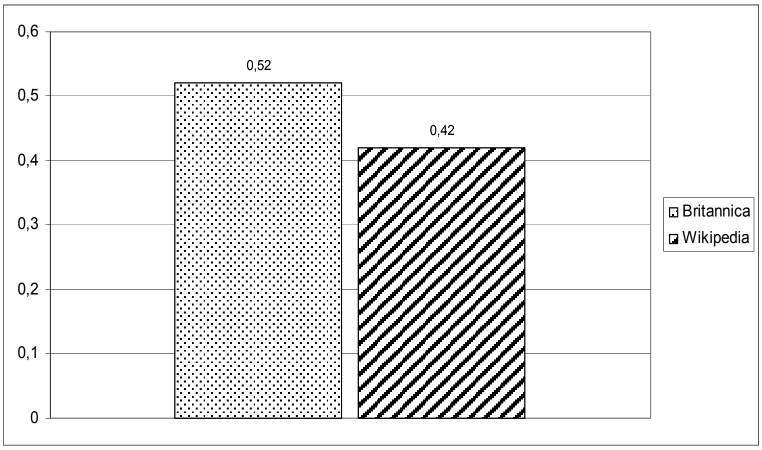


Fig. 57 Negative forms in BAs vs. WAs

According to Biber (1988:107) analytic negation is an alternative to the more integrative synthetic negation. As can be seen, data in fig. 57 shows a general low incidence of negative forms in both encyclopaedic corpora. Unexpectedly, a lower occurrence of synthetic negations

[40] *does not, do not, did not* is included in the frequency of *not.*

compared to analytic negations has been detected in both corpora. Furthermore, the occurrence of analytic negation is higher in Britannica than in Wikipedia (0.37% BAs vs. 0.31% WAs) (fig. 57).

The resulting overall occurrence of negative forms is slightly higher in Britannica (0.52% BAs vs. 0.42% WAs) than in Wikipedia. Nevertheless, the overall frequencies are not, as usual, very dissimilar in the two corpora. Some random examples in context follow.

> Another Sophist, Gorgias, advanced the skeptical-nihilist thesis that nothing exists; and if something did exist, it could *not* be known; and if it could be known, it could not be communicated. (*from Skepticism – Britannica*)
>
> Because debunkers often attack popular ideas, many are *not* strangers to controversy. (*from Skepticism - Wikipedia*)
> I have *no* desire to prove anything by it. I just dance. (*from Fred Astaire - Britannica*)
>
> The latter observation will be *no* news to the professsion, which has long admitted that Astaire starts dancing where the others stop hoofing". (*from Fred Astaire - Wikipedia*)
>
> He considered simple expressions *neither* true *nor* false and held that they may signify things in one or another of the following categories: substance, quantity, quality, relation, place, time, position, state, action, and affection.(*from Aristotele - Britannica*)
>
> Berlusconi himself claims to have resolved his conflict of interest: for example, he cites the fact that he is *neither* longer president of Mediaset, nor 100% owner. (*from Berlusconi - Wikipedia*)
> Unlike most of the national currencies that they replaced, euro banknotes *do not* display famous national figures. (*from Euro - Britannica*)
>
> via which they *do not* offer cross-border payments. In this way, banks in France continue to charge more for cross-border transfers than for domestic transfers. (*from Euro - Wikipedia*)

2.9 Interrogative Sentences

Interrogative sentences and particularly those which make use of second person pronouns, indicate a concern with interpersonal functions and involvement with the addressee (Biber, 1988). This is the reason why the frequency of interrogative sentences, as shown in fig. 58, is very low in both corpora. Nevertheless, their overall occurrence is higher in Britannica than in Wikipedia (0.035% BAs vs. 0.04% WAs). Most of them have been found in direct quotations or in impersonal open questions to introduce a subject. Nevertheless, the rhetorical device of open question, often found in Britannica has never been detected in Wikipedia.

Interrogative Sentences				
	Britannica	%	Wikipedia	%
Total	87	0.035	17	0.004

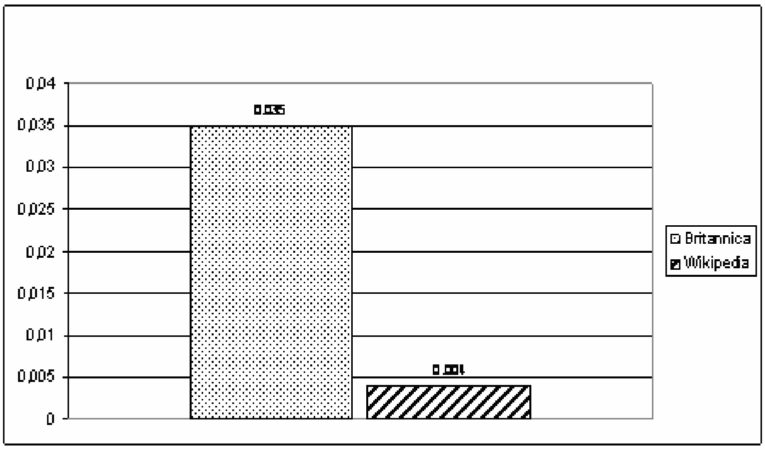


Fig. 58 Interrogative sentences in BAs vs. WAs

Some examples follow:

What types of human beings are there? What is their essence? What is the essence of human history ? Of humankind ? Contrary to so many of his intellectual predecessors, Foucault sought not to answer these traditional and seemingly straightforward questions but to critically examine them and the responses they had inspired. (*from Foucault- Britannica)*

In discussions after dinner Darwin asked his guests, "Why do you call yourselves Atheists?" (*from Agnosticism - Wikipedia*)

Wittgenstein once put the question this way: “And the problem arises: what is left over if I subtract the fact that my arm goes up from the fact that I raise my arm?” (*from Aristotele – Britannica)*

In his 1953 essay, What “Is An Agnostic? ” Russell states: An agnostic thinks it impossible to know the truth in matters such as God and the future life with which Christianity and other religions are concerned. (*from Agnosticism - Wikipedia*)

Yet questions remain: How will Western feminism deal with the dissension in its ranks, from women who believe the movement has gone too far and grown too Radical ? How uniform and successful can feminism be at the global level ? Can the problems confronting women in the mountains of Pakistan or the deserts of the Middle East be addressed in isolation, or must such issues be pursued through international forums ? (*from Feminism- Britannica*)

"Why do you call yourselves Atheists?" saying that he preferred the word "Agnostic." Aveling replied that "Agnostic was but Atheist writ respectable, and Atheist was only Agnostic writ aggressive." Darwin responded by asking, "Why should you be so aggressive?" wondering what was to be gained from forcing new ideas on people when free thought was "all very well" for the educated, but were ordinary people "ripe for it ?" Aveling then asked what if "the revolutionary truths of Natural and Sexual Selection" had been confined to the "judicious few" and he had delayed publication of the Origin of Species, where would the world be? Surely "his own illustrious example" encouraged freethinkers to proclaim truth "abroad from the house-tops […] Robert G. Ingersoll (*from Agnosticism – Wikipedia*)

2.10 Reduced Forms

This linguistic construction which involves a surface reduction, is completely absent in formal expository writing. Linguists have traditionally explained their frequent use in conversation as being a consequence of fast and easy production. The use of contractions seems to be tied to appropriateness considerations as much as the differing production circumstances of oral and written discourse. Biber (1988) finds that this feature tends to co-occur frequently with interactive features (such as first and second person pronouns) and with certain types of subordination; in addition, it seems to be preferred in American rather than in British English written texts, apparently because of greater attention to grammatical prescription by the latter.

As was expected, these reduced forms have not been found either in Britannica or in Wikipedia as they are officially forbidden in every kind of formal writing. Wikipedia Manual of Style declares [41]:

> *Contractions*
> In general, formal writing is preferred; therefore, the use of contractions, such as "don't", "can't" and "won't", is avoided unless they occur in a quotation.

Other linguistic features typical of CMC discourse, such as the use of interjections or the unconventional use of punctuation marks, acronyms and emoticons in CMC are totally absent both in Britannica and Wikipedia corpora.

Their use is considered stylistically inappropriate in any kind of formal written text. Nevertheless, as it will be shown in the next chapter, they are widely used in the WikiSpeak written(spoken) in Wikipedian talk pages.

[41] *Manual of Style* http://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style#Contractions

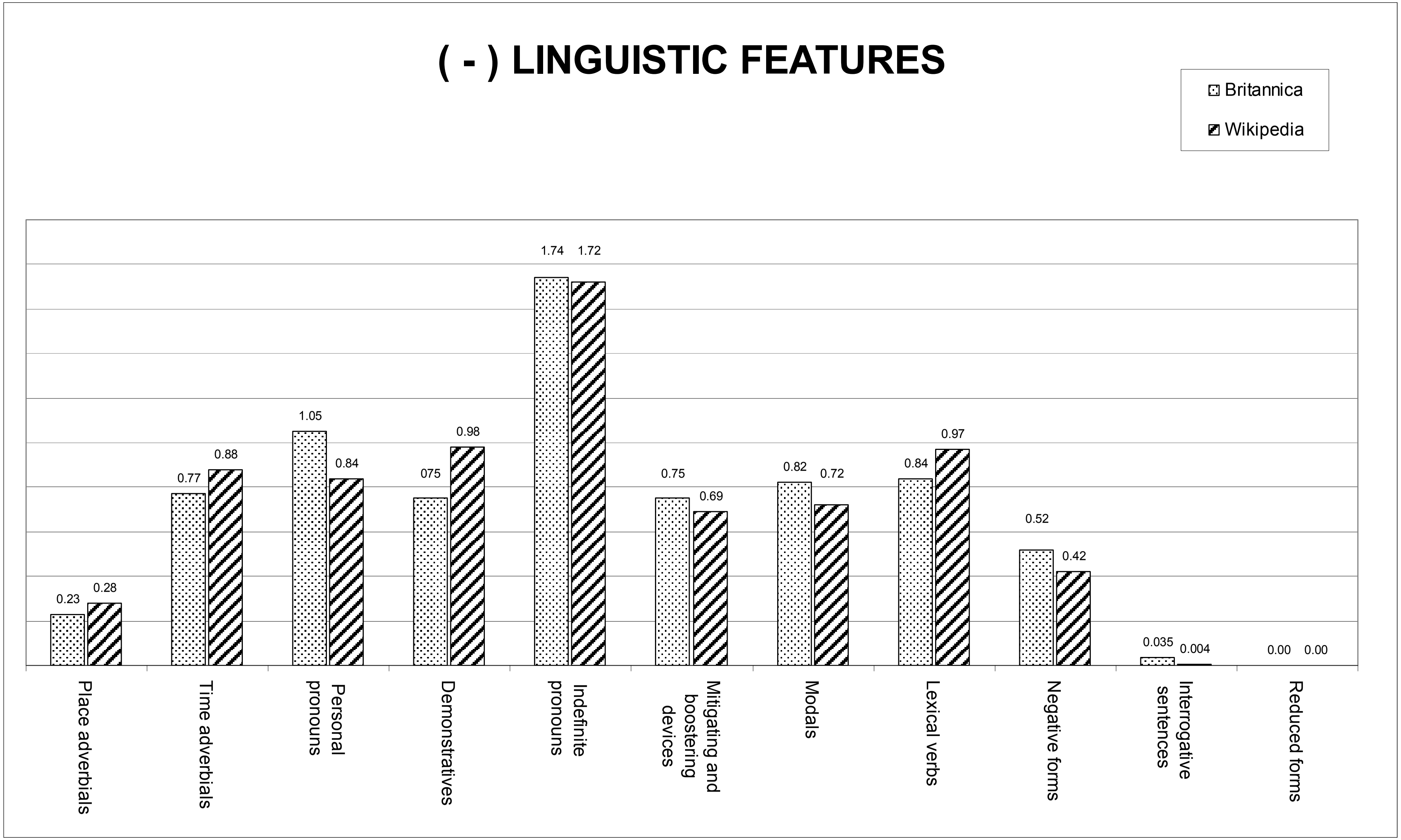


Fig. 59 (-) Linguistic Features in BAs vs. WAs

3. Comments and Remarks

The linguistic classes with a negative loading on informational production which have been investigated up to now in the two encyclopaedic corpora are outlined in fig. 59/60.

In particular, place and time adverbials, demonstratives and lexical verbs have a slightly higher frequency in Wikipedia, whereas personal pronouns, mitigating and boostering devices, modals, negative and interrogative forms and indefinite pronouns are more extensively used in Britannica. However, the little quantitative variation recorded is statistically irrelevant.

Britannica vs. Wikipedia Encyclopaedic Expository Style			
		Britannica %	**Wikipedia %**
	Word length (characters)	*5.3*	*5.2*
	Sentence length (words)	*22.5*	*22.9*
	Lexical density (tokens/types)	*45.5*	*43.6*
⊕	**Nominalizations %**	**5.26**	**4.62**
	Gerunds and present participles	**2.38**	**2.41**
	Articles	**10.02**	**9.68**
	Nouns	**29.90**	**29.28**
	Adjectives	**10.54**	**10.06**
	Prepositions	**14.23**	**13.42**
	Passives %	**0.96**	**0.96**
	Subordination features	**2.31**	**1.86**
	Coordination features	**4.11**	**3.64**
	Conjuncts	**0.47**	**0.37**
	(+)TOTAL	**+79.80**	**+76.10**
⊖	**Place adverbials**	**0.23**	**0.28**
	Time adverbials	**0.77**	**0.88**
	Personal pronouns (I, you, it)	**1.05**	**0.84**
	Demonstratives	**0.75**	**0.98**
	Indefinite pronouns	**1.74**	**1.72**
	Mitigating and Boostering devices	**0.75**	**0.69**
	Modals	**0.82**	**0.72**
	Lexical verbs	**0.84**	**0.97**
	Negative forms	**0.52**	**0.42**
	Interrogative sentences	**0.035**	**0.004**
	(-)TOTAL	**- 7.50**	**-7.50**
	FINAL SCORE	***72.30***	***68.60***

Fig. 60 Britannica vs. Wikipedia Encyclopaedic Expository Style

The total of all the negative linguistic features shows to be practically the same in the two corpora (7.50%) (fig. 61). This result proves to what extent both Britannica authors and Wikipedia contributors are very careful in avoiding the use of those linguistic elements which can invalidate the objectivity, neutrality and formality of the encyclopaedic expository style.

Nevertheless, the *Final Score* resulting from the difference of positive and negative linguistic features, shows a higher formality of Britannica expository style (72.30%). Thus, the minor final score in Wikipedia (68.60%) demonstrates a lower conformity of Wikipedians in following the linguistic rules which determine a formal encyclopaedic expository style (fig. 62).

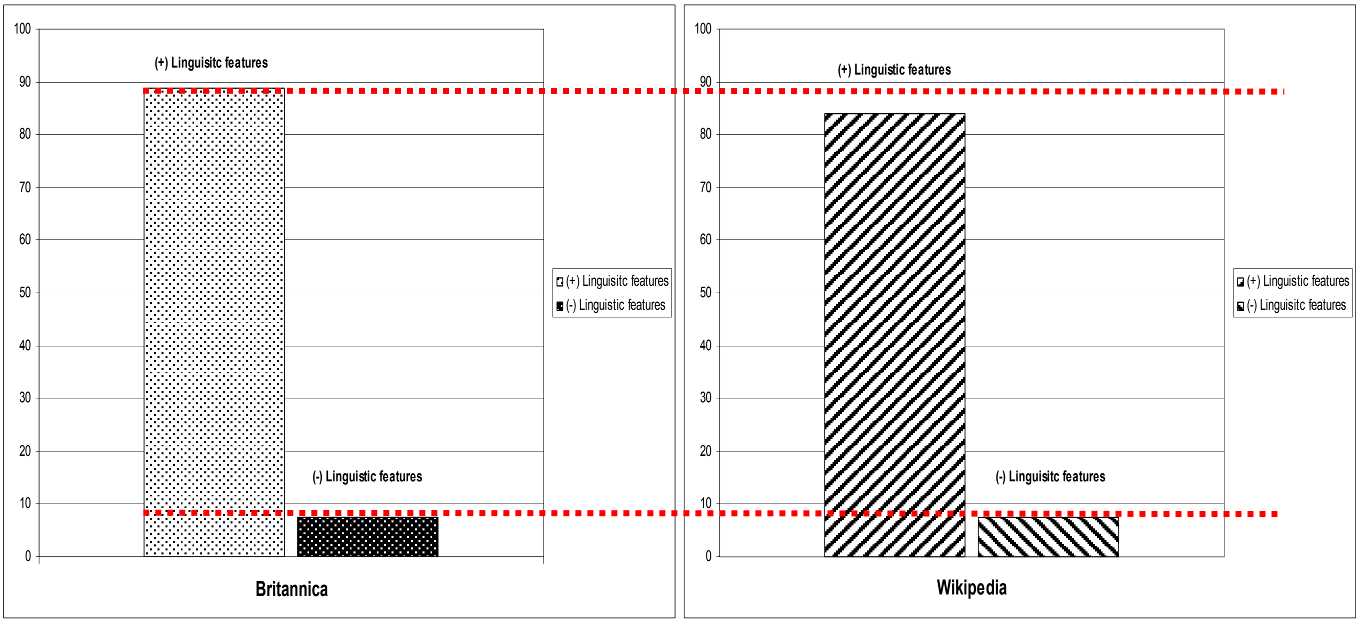


fig. 61 (+/-) Linguistic features in Britannica vs. Wikipedia

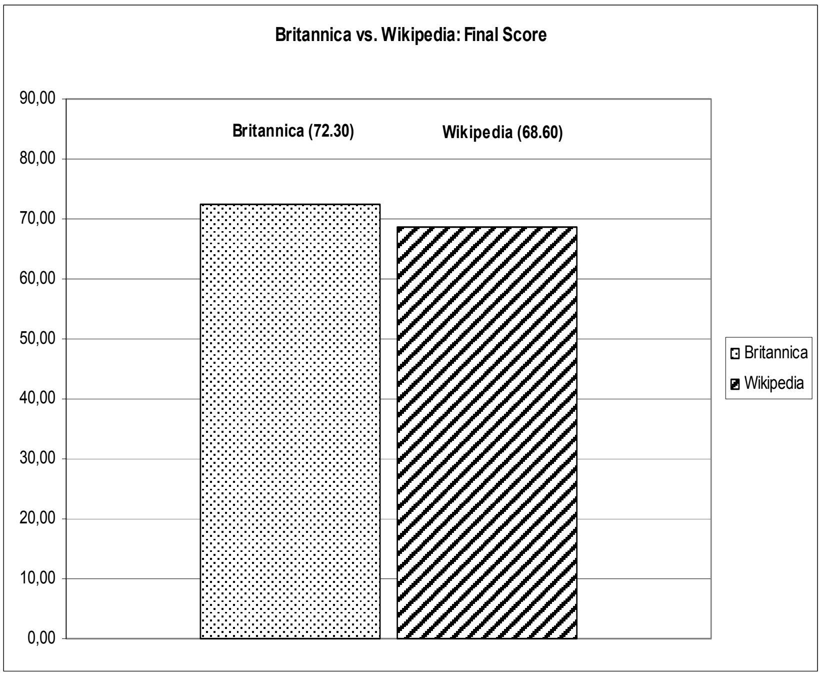


Fig. 62 Final Expository Style of Britannica and Wikipedia

As can be noted the first three elements (*word length, sentence length* and *lexical density*) in fig. 60 have not been included in the computation of the *Final Score* since they are not classes linguistically homogeneous with the other elements listed. Subtotals and final score have a merely indicative function. The purpose is to convey the different value of encyclopaedic expository style of Wikipedia and Britannica in single and more easily identifiable figures. Quantitatively speaking, the stylistic variation in the two encyclopaedic corpora is only 3.70 %.

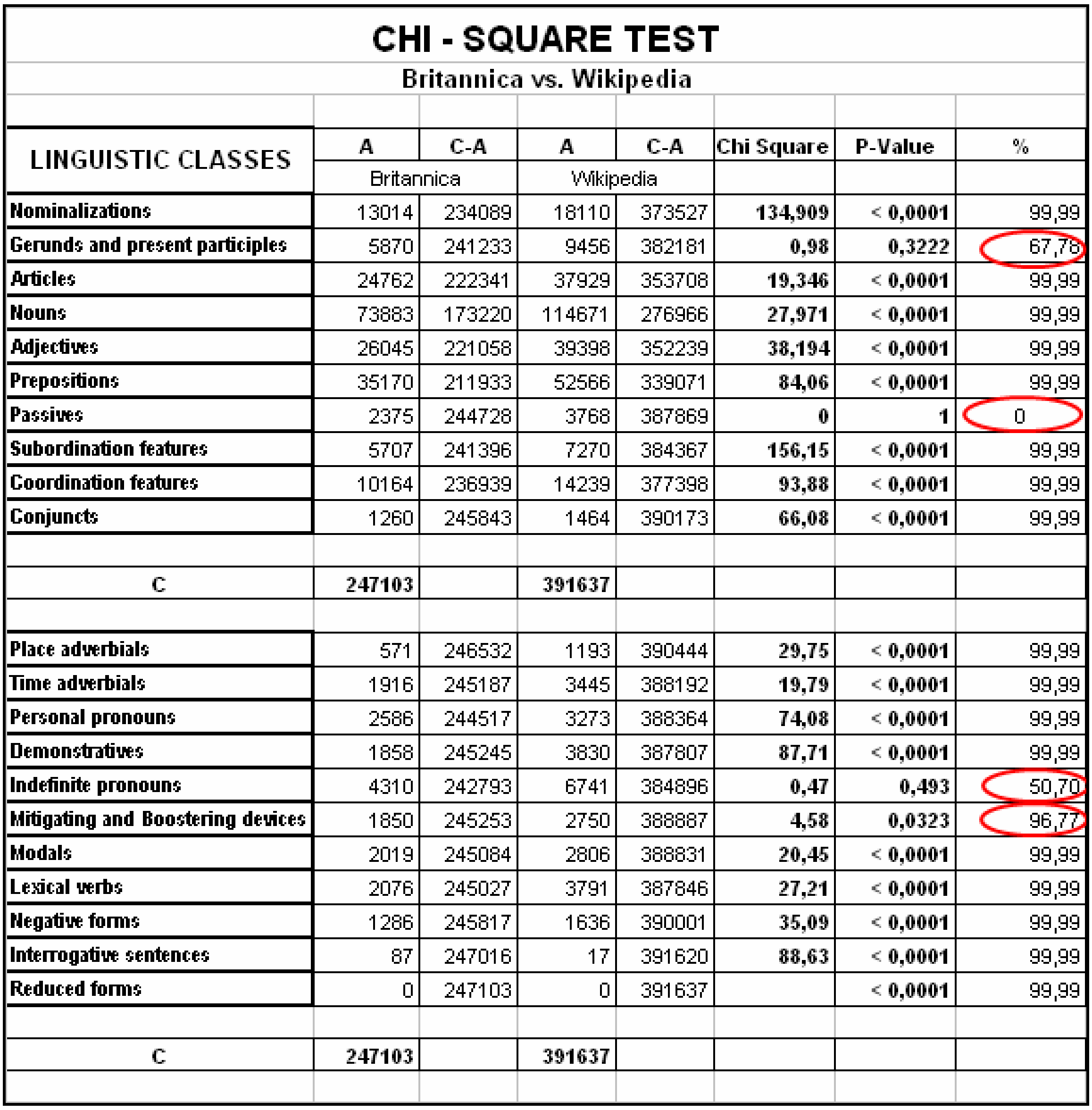
4. Chi-Square Test

In 17 out of 20 linguistic classes analyzed, the Chi Square test shows that all the data coming from the comparison of Britannica and Wikipedia has a 99.99 % reliability. Thus, the results are highly consistent, since they are a reflection of a significant variation in the two corpora, and not due to random variation. This means that the *Null hypothesis* (the difference is due to chance) of the test is rejected [42].

Specifically, the data in fig. 63 shows a slightly lower P-value for *mitigating and boostering devices.* The P-value, that is to say the probability of certainty, is reduced to 96.77%. Furthermore, the relative frequency of *indefinite pronouns* is very close in the two encyclopaedic corpora (1.74% BAs vs 1.72% WAs) consequently the P-value proves to be reliable at 50.70%. The similarity in the frequency of gerunds and present participles (2.38% BAs vs. 2.41% WAs) shows that the probability of data reliability is of 67.78%. Moreover, with the relative frequency of passives being identical (0.96 % in both corpora), there is 100% probability that the total coincidence in the number of passives is due to pure accident, thus in this specific case we cannot reject the *Null hypothesis.*

In conclusion, the overall data related to the comparison of the two encyclopaedic corpora appears to have a very high degree of reliability, being the data and the resulting P-value a symptom of a true underlying difference and thus not the result of a random variation in 85% of the linguistic classes analyzed.

[42] *see* chapter 2, section 2.2.9



CHI - SQUARE TEST

Britannica vs. Wikipedia

LINGUISTIC CLASSES	A	C-A	A	C-A	Chi Square	P-Value	%
	Britannica		Wikipedia				
Nominalizations	13014	234089	18110	373527	**134,909**	**< 0,0001**	99,99
Gerunds and present participles	5870	241233	9456	382181	**0,98**	**0,3222**	67,78
Articles	24762	222341	37929	353708	**19,346**	**< 0,0001**	99,99
Nouns	73883	173220	114671	276966	**27,971**	**< 0,0001**	99,99
Adjectives	26045	221058	39398	352239	**38,194**	**< 0,0001**	99,99
Prepositions	35170	211933	52566	339071	**84,06**	**< 0,0001**	99,99
Passives	2375	244728	3768	387869	**0**	**1**	0
Subordination features	5707	241396	7270	384367	**156,15**	**< 0,0001**	99,99
Coordination features	10164	236939	14239	377398	**93,88**	**< 0,0001**	99,99
Conjuncts	1260	245843	1464	390173	**66,08**	**< 0,0001**	99,99
C	**247103**		**391637**				
Place adverbials	571	246532	1193	390444	**29,75**	**< 0,0001**	99,99
Time adverbials	1916	245187	3445	388192	**19,79**	**< 0,0001**	99,99
Personal pronouns	2586	244517	3273	388364	**74,08**	**< 0,0001**	99,99
Demonstratives	1858	245245	3830	387807	**87,71**	**< 0,0001**	99,99
Indefinite pronouns	4310	242793	6741	384896	**0,47**	**0,493**	50,70
Mitigating and Boostering devices	1850	245253	2750	388887	**4,58**	**0,0323**	96,77
Modals	2019	245084	2806	388831	**20,45**	**< 0,0001**	99,99
Lexical verbs	2076	245027	3791	387846	**27,21**	**< 0,0001**	99,99
Negative forms	1286	245817	1636	390001	**35,09**	**< 0,0001**	99,99
Interrogative sentences	87	247016	17	391620	**88,63**	**< 0,0001**	99,99
Reduced forms	0	247103	0	391637		**< 0,0001**	99,99
C	**247103**		**391637**				

Fig. 63 Chi-Square Test: Britannica vs. Wikipedia

5.Encyclopaedias and Web Writing

Webpages are complex objects. Even when taken individually, they appear to be a composite type of document, with a visual organization of the space where different communicative purposes and different functionalities are included at the same time. The intertwining of visual and verbal is not new. What is new is the frequency of use of such a solution. While the linear organization of most paper documents was reflected in an initial evolutionary phase of web pages having an organization similar to that of printed pages, the latest evolution shows a visual organization that allows the inclusion of several functionalities and contents with different communicative purposes in a single document (Shepherd *et al.,* 1998). For example, the space on a web page can be divided into different sections, organized around the main body of the document. Navigational buttons, menus, search boxes, table of contents and links are all elements visually located in different areas of a single page (Haas, 2000:186-187).

The use of images and other graphical elements such as fonts of different types, sizes and colours, as well as the use of formatting devices, such as columns, section breaks, pictures, etc., is not a recent phenomenon. Nevertheless, their use and the effect of multimediality, hyperlinking, interactivity and multi-functionality have a crucial influence on webpages.

Both in traditional paper documents and webpages, readability, clarity, order, and reliability of information are fundamental aspects. The spatial organization of graphics and text on the webpages can direct reader's attention and make the interaction with the website more enjoyable and effective.

A good graphic design creates a visual logic and a positive optical impact. Pages which are not graphically interesting do not motivate the viewer. For example, dense text documents are hard to be read, particularly on the low-resolution screens of personal computers. Visual and functional continuity in website organization, graphic design, and typography are essential to convince the audience that a website offers accurate and useful information. A good page design simplifies navigation and makes it easier for readers to take advantage of the information provided. On the other hand, without good and readable contents, highly graphical pages disappoint the user.

Thus, in the analysis of web encyclopaedic pages, their visual organization, functionality, web usability, hypertextuality and index of readability are factors which cannot be ignored without losing important information on the text. A web page can be considered as a sort of container of multiple texts made up of intrinsically associated components. Artificially separating what is considered to be the main textual body from the rest is an arbitrary operation and it would not make sense in many cases.

In brief, in a webpage all the elements contribute to form a whole. Compared to the graphical complexity of many webpages, the reader is immediately struck by the minimalist layout of Wikipedia encyclopaedic pages which can be explained by the flexibility provided by the wiki software, the simplicity of its syntax, which allows everybody to contribute to the collaborative development of the project, and the creation of complex texts without much effort or expertise. The textual and multi-modal changes and the upgrading of the extant encyclopaedic genre (Shepherd *et al.,* 1998) has created fresh conventions whose introduction has been spurred by new rapid communication needs and by the evolution brought by Web 2.0.

5.1 Index of Readability

The key function of the encyclopaedic genre is educational, thus articles should provide a general overview on a specific subject through an understandable and popular expository style. Since most of the encyclopaedic readers are school learners, it is essential that texts be written in a clear, linear and comprehensible way in order to be easily understood to fulfill their primary pedagogical purpose.

MacCormick *et al.* (1982) submitted encyclopaedias to readability tests. They proved that encyclopaedias written by experts require high levels of reading skills in order to be easily understood. It is expected that Encyclopaedia Britannica Online continues this tradition since a group of experts controls the content. But, to what extent Wikipedian collaborative writing affects readability? When articles are edited in Wikipedia, their Index of Readability is not automatically tested, and this factor could generate articles of mixed readability levels, and extremely different from what is assumed to be the monitored system of Encyclopaedia Britannica. The comparison of Index of Readability of Britannica vs. Wikipedia will show to what extent the collaborative writing process of Wikipedia matches up or not with the Index of Readability of Encyclopaedia Britannica.

Reading a text is mainly a left-brain activity. It demands focus, word recognition, decoding linear processing, and prediction of outcomes. Readability formulas offer the opportunity to assess only the surface characteristics of texts. They evaluate features that can be subjected to mathematical computation such as semantic (the difficulty of words) and syntactic factors (the difficulty of sentences). As already pointed out in the previous sections, word and sentence length influence stylistic formality. Complex texts often contain difficult and long words because they discuss abstract ideas, whereas easy texts use common and short words as they are focused on concrete experiences. Only sentence and word lengths and complexity of linguistic structures can be measured by *Readability* formulas.

Webster's dictionary defines *readable* a text *easy to be read, interesting, agreeable, attractive in style and enjoyable*. It is clear that most of these features cannot be measured mathematically as qualitative factors such as tone, complexity of ideas, page design, textual comprehensibility or obscurity, textual cohesion and coherence, interest, appeal and enjoyment aroused in the reader, are elements which cannot be evaluated through mathematical formulas.

Nevertheless, my personal point of view (and not only) is that semantic and syntactic factors measured through sentence and word length definitely affect text readability in a significant way. For this reason, *Gunning Fox* Index has been chosen, among the numerous available readability formulas (e.g. *Dale-Chall, Flesch-Kincaid, Fry,* etc.), in order to assess the Index of Readability of Britannica and Wikipedia corpora.

5.1.1 Gunning’s Fog Index of Readability

In 1952 Robert Gunning created one of the most popular readability formulas. It predicted, with an 80% accuracy, the difficulty of a written passage. This formula indicates the reading skill (based on grade level) necessary to understand a text on the first reading (of course, the lower the number, the more understandable the content will be to the reader).

Gunning's Fog formula is easy to apply. It is based on the calculation of (1) average sentence length and (2) on the percentage of the polysyllabic words contained in a text. 1 and 2 have to be added and the sum must be multiplied by 0.4. The formula to be applied is the following:

$$0.4 * \left(\left(\frac{\text{words}}{\text{sentence}}\right) + 100\left(\frac{\text{complex words}}{\text{words}}\right)\right)$$

The formula is an objective tool for measuring readability and it predicts quite satisfactorily the difficulty of a text. The Fog Index scores of some American resources are shown in fig. 64. They have been provided by Philip Chalmers[43] to help establish and assess the textual readability of documents.

Fog Index Scores	
Score	**Resources**
6	TV guides, The Bible, Comic books
8	Reader's Digest, Ladies' Home Journal
8 - 10	Most popular novels
9	Reader's Digest
10	Time, Newsweek
11	Wall Street Journal
12	Atlantic Monthly
14	The Times, The Guardian
15 - 20	Academic papers

Fig. 64 Gunning’s Fog Index Scores

[43] Philip Chalmers in *Lines from a Floating Life* http://ninglun.wordpress.com/

For example if a text has a score of 12, it means that it has the reading level of a U.S. high school senior. Texts designed for a wide audience generally require a *Fog Index* of less than 12. Score 17, for example, indicates a level of textual difficulty at post-graduate level.

5.1.2 Britannica vs. Wikipedia: Index of Readability

With these premises in mind, the *Index of Readability* of Britannica and Wikipedia articles has been calculated in order to understand the readability of the two encyclopaedias. Purpose of this specific investigation has been to check if the typical web reader of online encyclopaedias, mainly made up of high school learners and university students, have the reading skills necessary to easily understand its content. The measurement of course has also fulfilled the purpose of comparing the two resulting scores in order to verify discrepancies or similarities. The analysis has been carried out through the online text analysis tool *Textalyser* hosted on the *lexicool.com* website[44] which has automatically computed Gunning Fog's Index of Readability.

As data in Appendix shows, the microanalysis on 200 encyclopaedic articles has revealed, in most cases, very similar average Indexes of Readability in the two corpora (fig. 67). The final average score is 11.5 in Britannica and 11 in Wikipedia. According to Robert Gunning, texts designed for a wide audience and with a popular purpose in mind generally require a Fog Index below 12.

Britannica and Wikipedia have a final average score (fig. 65) very close to the Index of Readability of some of the most popular American magazines such as: *Time, Newsweek* (10), *Wall Street Journal* (11) and *Atlantic Monthly* (12) (fig. 64). Thus, the interpretation of the final results has demonstrated that the two encyclopaedias have successfully passed the test since the respective scores demonstrate that they should be comprehensible by a wide audience.

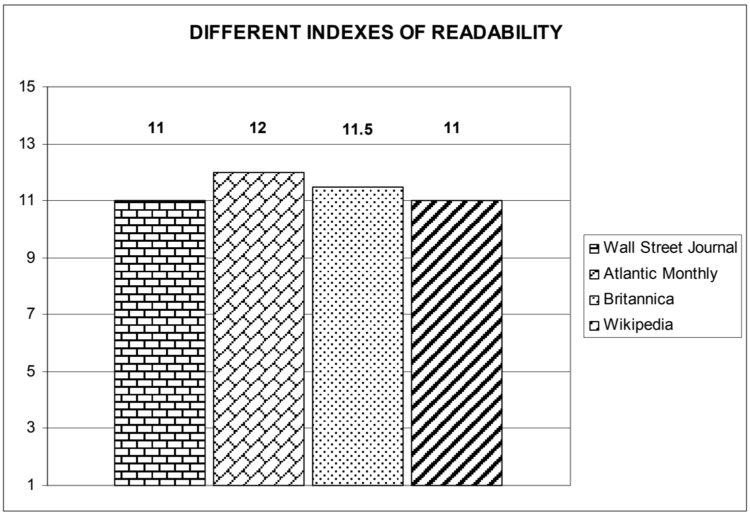
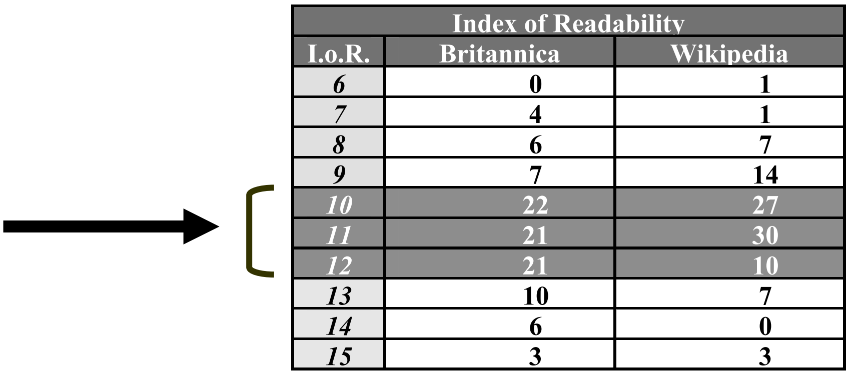


Fig. 65 Different Indexes of Readability

[44] *Textalyser* http://www.lexicool.com

Fig. 66 shows the numerical distribution of Indexes of Readability in Britannica and Wikipedia articles. As the data proves, about 65% articles in the two encyclopaedic corpora have an Index of Readability between 10-12, consequently this range can be considered the most significant one.



Index of Readability		
I.o.R.	Britannica	Wikipedia
6	0	1
7	4	1
8	6	7
9	7	14
10	22	27
11	21	30
12	21	10
13	10	7
14	6	0
15	3	3

Fig. 66 Index of Readability

In particular, 17% of articles in Britannica vs. 23% in Wikipedia are easily readable and understandable having an Index of Readability from 6 to 9, whereas 19% of articles in Britannica vs. 10% in Wikipedia have a more complex Index of Readability (from 13 to 15). This data shows, first of all, independently from the individual or collaborative writing technique adopted by Britannica encyclopaedists or Wikipedians, that Index of Readability is not homogeneously distributed in the two corpora. Secondly, the average Index of Readability of the two encyclopaedic corpora (11,5 BAs vs. 11 WAs) proves that Wikipedian articles should be slightly simpler to be read and understood.

In detail, the highest score (15.8) has been found in the article *Racism* in Britannica, the lowest in the article *Graffiti* (6.6) in Wikipedia. This last article has recorded a score of 12.2 in Britannica. This is the only case in which a marked score variation has been detected. Excluding these exceptions, all the remaining articles have recorded very similar Indexes of Readability.

In conclusion, according to Gunning Fog's index of readability, the average score of 11.5 in Britannica and 11 in Wikipedia indicate the years of formal education that a person requires to easily understand the content of encyclopaedic articles on the first reading.

The Fog Index required to have a reading level of a U.S. high school senior is of 12, hence Britannica and Wikipedia articles are easily understandable also by a less educated audience (11-11.5) and both encyclopaedias should succeed in fulfilling their primary educational and popular purpose.

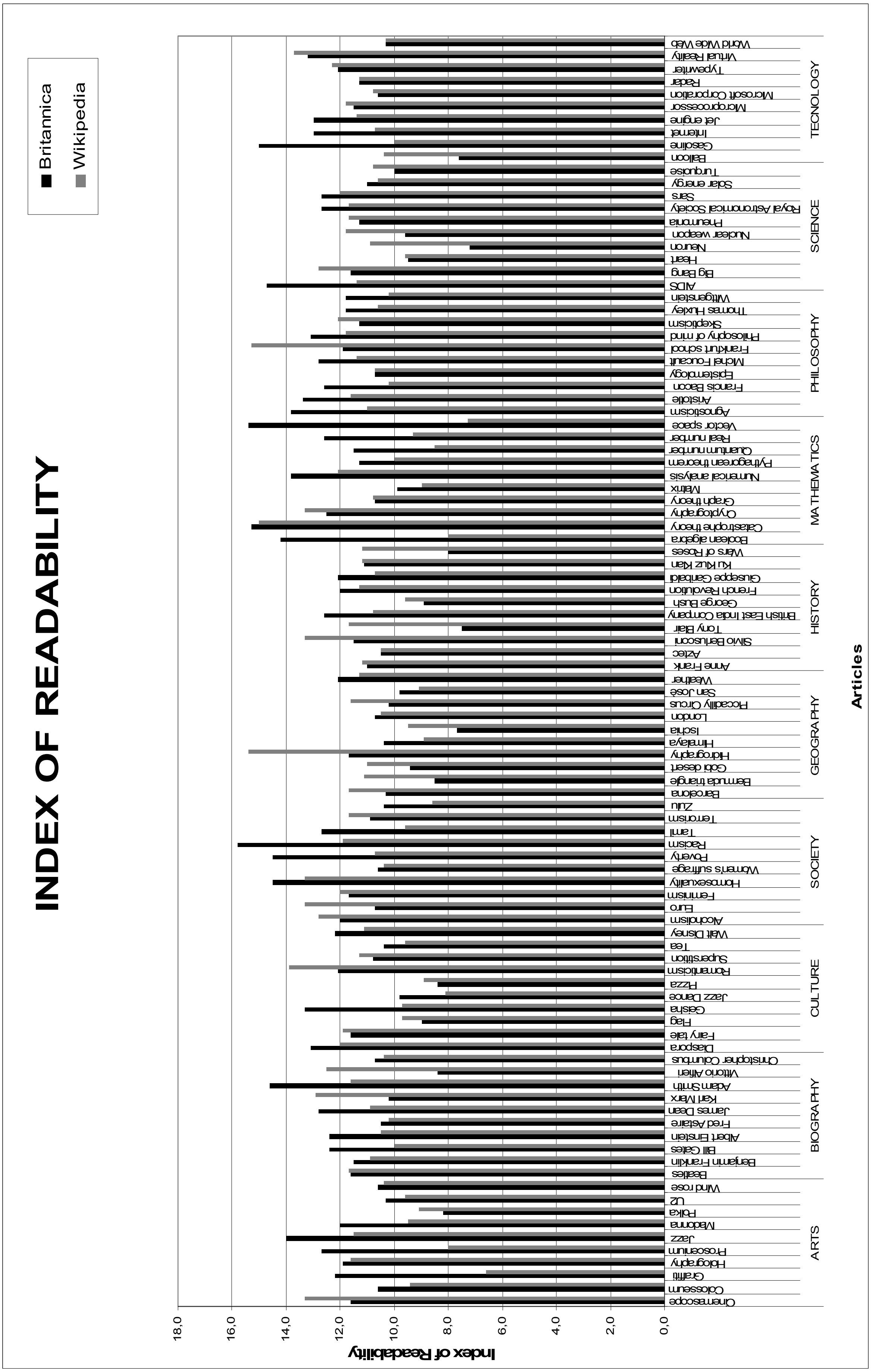


Fig. 67 Index of Readability BAs vs. WAs

5.2 Web Usability

As shown in the previous section, the content of texts with a pedagogical and popular purpose, should be written with readability in mind. The early research on readability was conducted only on traditional printed texts. Nowadays, there are new and additional elements which need to be taken into account when considering digital genres. When websites and, in this specific case, online encyclopaedias are assessed, it is essential to consider online readability in terms of both online content readability and web usability.

Reading a text is mainly a left-brain activity. During the last thirty years, as well as traditional readability formulas connected to writing content, there has been a great interest in the graphic aspects of writing that appeal also to the right side of the brain. They include editorial design, layout, symmetry, the generous use of illustrations, colours, blank spaces, graphs, bulleted lists, etc. They are essential factors which have to be taken into account in order to globally understand the nature of a text. All the aspects which co-occur in improving the readability and usability, fruition and comprehensibility of online texts will be discussed in the next sections.

5.2.1 Explicit and Unambiguous Language

Most of online readers surf the web because they are looking for specific information, and they do not find it by reading a Web page word by word but rather by scanning the page for relevant items. For this reason it is important to take into account some basic stylistic conventions when writing and structuring webpages. The *Web Style Guide* site (Lynch*,* 2005) claims:

- *Be frugal.* Don't use the first paragraph of each page to tell users what information they'll find there. Instead, start with the information, written in the concise and factual prose style shown above.
- *Stick to the point.* Write in easily understood sentences. Steer clear of clever headings and catchy but meaningless phrases that users must think about and explore further to understand.
- *Think globally.* Remember that you are designing documents for the *World Wide* Web and that your audience may not understand conventions specific to your little corner of the world. Also, avoid metaphors and puns that may make sense only in the context of your language and culture.

Furthermore, reading from computer screens is more tiring for the eyes and about 25% slower than reading from printed papers. Thus, the clearer the style of writing is, the easier it will be for the site visitors to absorb what has been written on the webpage. Some techniques for using clear and simple language include, for example, the avoidance of slang or jargon expressions, the use of shorter words where possible, the avoidance of complex and ambiguous sentence structures, omission of needless words, inclusion of just one idea or concept per sentence, the use of active instead of passive verbs, and the organization and structuring of information in an orderly and logical way. Tailoring texts in concise sentences, using an objective language, and at an appropriate reading level for the target audience improves textual readability.

The linguistic analysis carried out on Wikipedia and Britannica, following a statistical approach based on Biber's Factor analysis, has shown the use of an encyclopaedic expository style and an Index of Readability comprehensible by a large audience. The language used proves to be unambiguous, explicit and context independent. It does not make use of jargon expressions, and the average length of words is not very long, being of 2-3 syllables (5.3 characters per word BAs vs. 5.2 WAs). However, sentence length does not appear to be very short (22.05 words BAs vs. 22.09 words WAs), being very similar to the sentence length of academic writing (24 words). Encyclopaedic texts rely on a balanced use of coordination and subordination features. Furthermore, the use of passive verbal structures is very low and identical in the two encyclopaedias (0.96%).

Although visual and audio media are included on encyclopaedic web pages, its primary mode of communication is through written text. The analysis has confirmed that both Britannica and Wikipedia make use of an expository style immediately comprehensible also to non-specialist readers. Furthermore, writing for the web has a number of important implications. First of all, webpage layout affects the reading strategies, hyperlinking allows multiple entry points to information provided on the website and the page organization in separate blocks for search, navigation and content definitely affects reading and navigation.

In the following sections, web usability and the application of its principles in the analysis of the two encyclopaedias will be presented.

5.2.2 Front Load Content

Traditionally a printed page contains (in the following order) the introduction, the main content, and the conclusions. Unfortunately, when scanning through web content the readers do not tend to read all the text and neither do they read all the way to the bottom of the screen. Consequently, conclusions are easily missed if left at the end of the page. For this reason, front-loading is applied to web pages**.** According to Web usability principles (Nielsen, 1979) the opening paragraph on every page should always contain a summary of the main content of that page and its conclusions. In this way, the reader can instantly gain an understanding of what the page is about and decide whether they want to read it or not. A great example of front-loaded content is conveyed in newspaper articles, where the opening paragraph always presents the conclusion of the article.

Analyzing Wikipedia corpus, it has been noted that each encyclopaedic article is introduced by a *front load* section which summarizes and briefly defines the main topic developed in the article. This happens because Wikipedians rigorously follow the *Manual of Style* according to which:

> the purpose of an encyclopaedia is to codify human knowledge in a way that is most accessible to the most people and this demands clear descriptions of what the subject matter is about. So the reader is not dropped into the middle of the subject from the first word, but he/she is eased into it.

As the two excerpts below from Britannica and Wikipedia's *Euro* articles show, the attention towards web usability writing techniques has also been found in Britannica where articles are always opened by an introductory paragraph where brief explanation on the specific topic is provided.

Encyclopædia Britannica Article

monetary unit and currency of the **Euro**pean Union (EU). It was introduced as a non cash monetary unit in 1999, and currency notes and coins appeared in participating countries on January 1, 2002. After February 28, 2002, the **euro** became the sole currency of member states, and their national currencies ceased to be legal tender. The **euro** is represented by the symbol €.

Euro

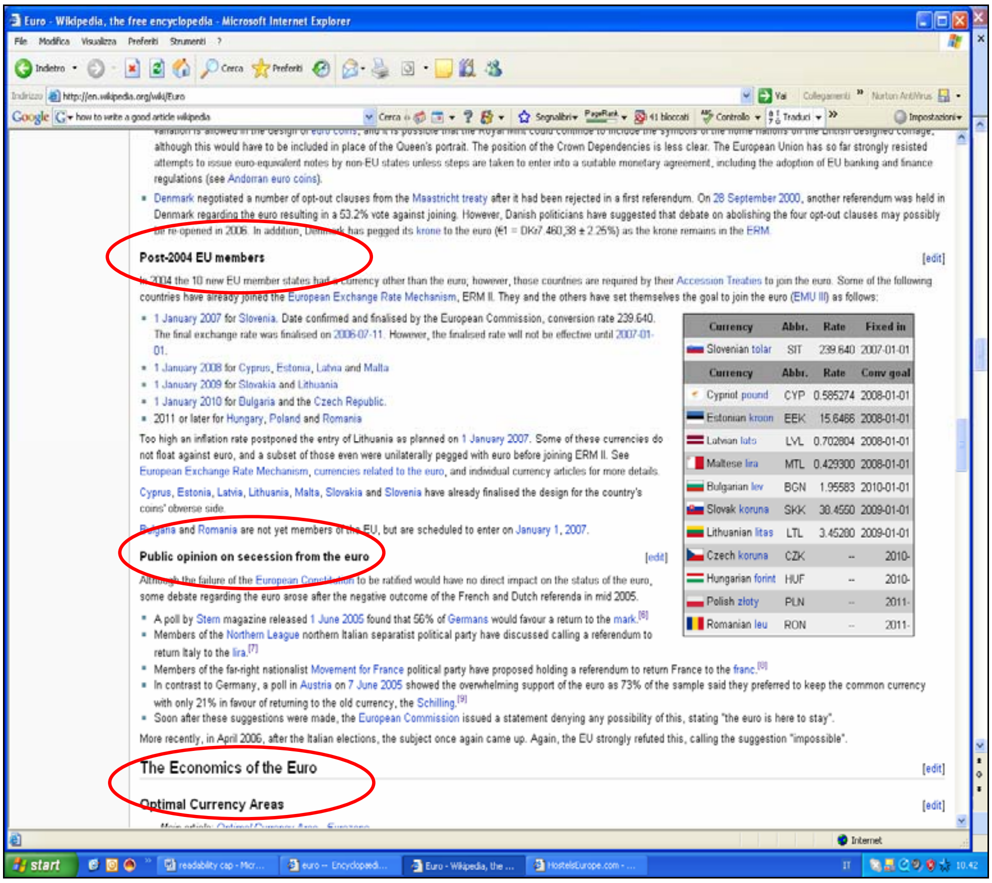
From Wikipedia, the free encyclopaedia

The **euro** (currency sign: **€**; banking code: **EUR**) is the official currency of the European Union member states of Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal, and Spain - also known as the Eurozone - and is the single currency for more than 300 million people in Europe. Including areas using currencies pegged to the euro, the euro affects more than 480 million people worldwide.[1]

5.2.3 Sections and Descriptive Subheadings

According to Nielsen (1999) the main heading on the page should provide an overall view of what the page is about. In addition, Web Usability principles recommend to organize webpages into a system of hierarchical sections and subsections which should be quickly retrieved through the corresponding sub-headings written in bold. The main function carried out by descriptive headings and sub-headings is to be a sort of keyword which allows site visitors to easily identify what each section of the page is about and to quickly retrieve and access the desired information. Each sub-section should contain, on average, from two to four paragraphs.

As fig. 68a shows, Wikipedia fully respects web usability standards. It rigorously structures its web article strictly following Nielsen's principles. It makes use of frequent sections and subsections easily retrieved through the corresponding headings and subheadings. The segments of information are broken up by horizontal dividers which make the visual organization of information in blocks more evident. Compared to Wikipedia, Britannica does not follow such a marked logical structure since the layout is absent in its web articles which reproduce the format of a more traditional printed page. In Britannica, the content is organized in very long and sequential series of text grouped into few sections structured in long paragraphs, split only by a double line spacing and distributed on several webpages (fig.68b).


Euro - Wikipedia, the free encyclopedia - Microsoft Internet Explorer
File Modifica Visualizza Preferiti Strumenti ?
Indietro Cerca Preferiti
Indirizzo http://en.wikipedia.org/wiki/Euro
Vai Collegamenti Norton AntiVirus
Google how to write a good article wikipedia Cerca Segnalibri PageRank 41 bloccati Controllo Traduci Impostazioni
Internet
start readability cap - Micr... euro -- Encyclopedi... Euro - Wikipedia, the ... HostelsEurope.com - ... IT 10.42

variation is allowed in the design of euro coins, and it is possible that the Royal Mint could continue to include the symbols of the home nations on the British designed coinage, although this would have to be included in place of the Queen's portrait. The position of the Crown Dependencies is less clear. The European Union has so far strongly resisted attempts to issue euro-equivalent notes by non-EU states unless steps are taken to enter into a suitable monetary agreement, including the adoption of EU banking and finance regulations (see Andorran euro coins).

- Denmark negotiated a number of opt-out clauses from the Maastricht treaty after it had been rejected in a first referendum. On 28 September 2000, another referendum was held in Denmark regarding the euro resulting in a 53.2% vote against joining. However, Danish politicians have suggested that debate on abolishing the four opt-out clauses may possibly be re-opened in 2006. In addition, Denmark has pegged its krone to the euro (€1 = DKr7.460,38 ± 2.25%) as the krone remains in the ERM.

Post-2004 EU members [edit]

In 2004 the 10 new EU member states had a currency other than the euro; however, those countries are required by their Accession Treaties to join the euro. Some of the following countries have already joined the European Exchange Rate Mechanism, ERM II. They and the others have set themselves the goal to join the euro (EMU III) as follows:

- 1 January 2007 for Slovenia. Date confirmed and finalised by the European Commission, conversion rate 239.640. The final exchange rate was finalised on 2006-07-11. However, the finalised rate will not be effective until 2007-01-01.
- 1 January 2008 for Cyprus, Estonia, Latvia and Malta
- 1 January 2009 for Slovakia and Lithuania
- 1 January 2010 for Bulgaria and the Czech Republic.
- 2011 or later for Hungary, Poland and Romania

Currency	Abbr.	Rate	Fixed in
Slovenian tolar	SIT	239.640	2007-01-01
Currency	**Abbr.**	**Rate**	**Conv goal**
Cypriot pound	CYP	0.585274	2008-01-01
Estonian kroon	EEK	15.6466	2008-01-01
Latvian lats	LVL	0.702804	2008-01-01
Maltese lira	MTL	0.429300	2008-01-01
Bulgarian lev	BGN	1.95583	2010-01-01
Slovak koruna	SKK	38.4550	2009-01-01
Lithuanian litas	LTL	3.45280	2009-01-01
Czech koruna	CZK	--	2010-
Hungarian forint	HUF	--	2010-
Polish złoty	PLN	--	2011-
Romanian leu	RON	--	2011-

Too high an inflation rate postponed the entry of Lithuania as planned on 1 January 2007. Some of these currencies do not float against euro, and a subset of those even were unilaterally pegged with euro before joining ERM II. See European Exchange Rate Mechanism, currencies related to the euro, and individual currency articles for more details.

Cyprus, Estonia, Latvia, Lithuania, Malta, Slovakia and Slovenia have already finalised the design for the country's coins' obverse side.

Bulgaria and Romania are not yet members of the EU, but are scheduled to enter on January 1, 2007.

Public opinion on secession from the euro [edit]

Although the failure of the European Constitution to be ratified would have no direct impact on the status of the euro, some debate regarding the euro arose after the negative outcome of the French and Dutch referenda in mid 2005.

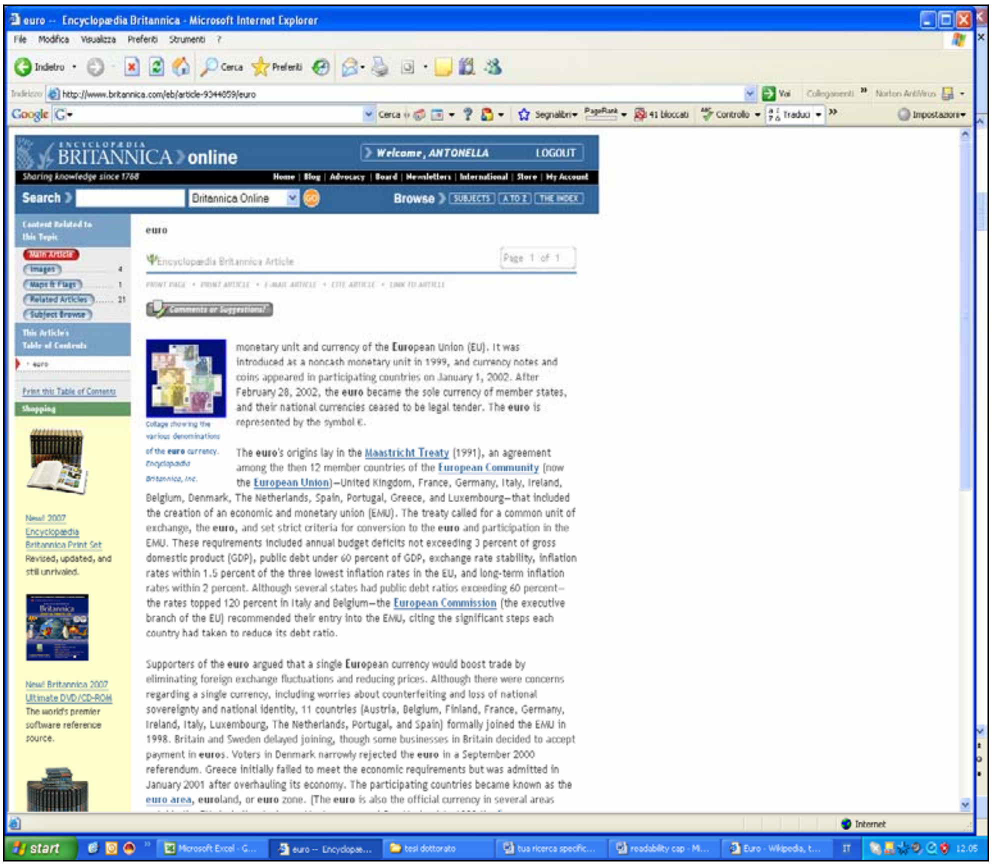
- A poll by Stern magazine released 1 June 2005 found that 56% of Germans would favour a return to the mark.[6]
- Members of the Northern League northern Italian separatist political party have discussed calling a referendum to return Italy to the lira.[7]
- Members of the far-right nationalist Movement for France political party have proposed holding a referendum to return France to the franc.[8]
- In contrast to Germany, a poll in Austria on 7 June 2005 showed the overwhelming support of the euro as 73% of the sample said they preferred to keep the common currency with only 21% in favour of returning to the old currency, the Schilling.[9]
- Soon after these suggestions were made, the European Commission issued a statement denying any possibility of this, stating "the euro is here to stay".

More recently, in April 2006, after the Italian elections, the subject once again came up. Again, the EU strongly refuted this, calling the suggestion "impossible".

The Economics of the Euro [edit]

Optimal Currency Areas [edit]

Main article: Optimal Currency Area, Eurozone


euro -- Encyclopædia Britannica - Microsoft Internet Explorer
File Modifica Visualizza Preferiti Strumenti ?
Indietro Cerca Preferiti
Indirizzo http://www.britannica.com/eb/article-9344059/euro
Vai Collegamenti Norton AntiVirus
Google Cerca Segnalibri PageRank 41 bloccati Controllo Traduci Impostazioni
ENCYCLOPÆDIA BRITANNICA online
Welcome, ANTONELLA LOGOUT
Sharing knowledge since 1768
Home | Blog | Advocacy | Board | Newsletters | International | Store | My Account
Search Britannica Online GO
Browse SUBJECTS A TO Z THE INDEX
Content Related to this Topic
Main Article
Images 4
Maps & Flags 1
Related Articles 21
Subject Browse
This Article's Table of Contents
euro
Print this Table of Contents
Shopping
New! 2007 Encyclopædia Britannica Print Set
Revised, updated, and still unrivaled.
New! Britannica 2007 Ultimate DVD/CD-ROM
The world's premier software reference source.
Encyclopædia Britannica Article
Page 1 of 1
PRINT PAGE • PRINT ARTICLE • E-MAIL ARTICLE • CITE ARTICLE • LINK TO ARTICLE
Comments or Suggestions?
Collage showing the various denominations of the euro currency. Encyclopædia Britannica, Inc.
Internet
start Microsoft Excel - G... euro -- Encyclopæ... tesi dottorato tua ricerca specific... readability cap - Mi... Euro - Wikipedia, t... IT 12.05

euro

monetary unit and currency of the **Euro**pean Union (EU). It was introduced as a noncash monetary unit in 1999, and currency notes and coins appeared in participating countries on January 1, 2002. After February 28, 2002, the **euro** became the sole currency of member states, and their national currencies ceased to be legal tender. The **euro** is represented by the symbol €.

The **euro**'s origins lay in the Maastricht Treaty (1991), an agreement among the then 12 member countries of the European Community (now the European Union)—United Kingdom, France, Germany, Italy, Ireland, Belgium, Denmark, The Netherlands, Spain, Portugal, Greece, and Luxembourg—that included the creation of an economic and monetary union (EMU). The treaty called for a common unit of exchange, the **euro**, and set strict criteria for conversion to the **euro** and participation in the EMU. These requirements included annual budget deficits not exceeding 3 percent of gross domestic product (GDP), public debt under 60 percent of GDP, exchange rate stability, inflation rates within 1.5 percent of the three lowest inflation rates in the EU, and long-term inflation rates within 2 percent. Although several states had public debt ratios exceeding 60 percent—the rates topped 120 percent in Italy and Belgium—the European Commission (the executive branch of the EU) recommended their entry into the EMU, citing the significant steps each country had taken to reduce its debt ratio.

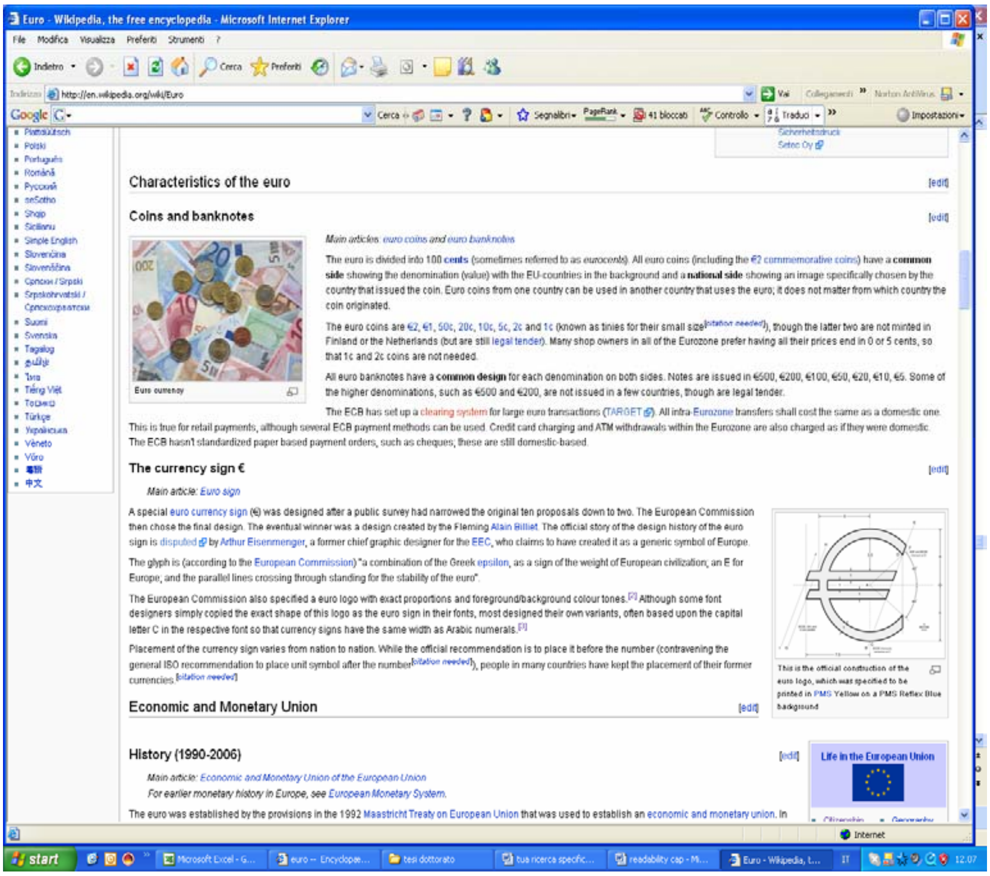
Supporters of the **euro** argued that a single **Euro**pean currency would boost trade by eliminating foreign exchange fluctuations and reducing prices. Although there were concerns regarding a single currency, including worries about counterfeiting and loss of national sovereignty and national identity, 11 countries (Austria, Belgium, Finland, France, Germany, Ireland, Italy, Luxembourg, The Netherlands, Portugal, and Spain) formally joined the EMU in 1998. Britain and Sweden delayed joining, though some businesses in Britain decided to accept payment in **euros**. Voters in Denmark narrowly rejected the **euro** in a September 2000 referendum. Greece initially failed to meet the economic requirements but was admitted in January 2001 after overhauling its economy. The participating countries became known as the euro area, **euro**land, or **euro** zone. (The **euro** is also the official currency in several areas

Fig. 68 a/b Sections and subsections in Wikipedia and Britannica

5.2.4 Paragraphing

In a printed or digital page the basic block of reading is always made up of paragraphs. A paragraph is a self-contained unit of discourse developing a particular point or idea. Generally speaking, a new paragraph marks a change of focus or time, place or speaker in a passage. A new paragraph begins on a new line and it is usually indented or with a one-line gap above it.

In our specific encyclopaedic case, if just one idea is assigned to each paragraph, visitors can easily scan through them, have the general gist of what the paragraph is about, then move on to the next, without overlooking important information because what the paragraph is about is already roughly known.



Characteristics of the euro [edit]

Coins and banknotes [edit]

Main articles: euro coins and euro banknotes

The euro is divided into 100 **cents** (sometimes referred to as *eurocents*). All euro coins (including the €2 commemorative coins) have a **common side** showing the denomination (value) with the EU-countries in the background and a **national side** showing an image specifically chosen by the country that issued the coin. Euro coins from one country can be used in another country that uses the euro; it does not matter from which country the coin originated.

The euro coins are €2, €1, 50c, 20c, 10c, 5c, 2c and 1c (known as tinies for their small size[citation needed]), though the latter two are not minted in Finland or the Netherlands (but are still legal tender). Many shop owners in all of the Eurozone prefer having all their prices end in 0 or 5 cents, so that 1c and 2c coins are not needed.

All euro banknotes have a **common design** for each denomination on both sides. Notes are issued in €500, €200, €100, €50, €20, €10, €5. Some of the higher denominations, such as €500 and €200, are not issued in a few countries, though are legal tender.

The ECB has set up a clearing system for large euro transactions (TARGET). All intra-Eurozone transfers shall cost the same as a domestic one. This is true for retail payments, although several ECB payment methods can be used. Credit card charging and ATM withdrawals within the Eurozone are also charged as if they were domestic. The ECB hasn't standardized paper based payment orders, such as cheques; these are still domestic-based.

The currency sign € [edit]

Main article: Euro sign

A special euro currency sign (€) was designed after a public survey had narrowed the original ten proposals down to two. The European Commission then chose the final design. The eventual winner was a design created by the Fleming Alain Billiet. The official story of the design history of the euro sign is disputed by Arthur Eisenmenger, a former chief graphic designer for the EEC, who claims to have created it as a generic symbol of Europe.

The glyph is (according to the European Commission) "a combination of the Greek epsilon, as a sign of the weight of European civilization; an E for Europe; and the parallel lines crossing through standing for the stability of the euro".

The European Commission also specified a euro logo with exact proportions and foreground/background colour tones.[2] Although some font designers simply copied the exact shape of this logo as the euro sign in their fonts, most designed their own variants, often based upon the capital letter C in the respective font so that currency signs have the same width as Arabic numerals.[3]

Placement of the currency sign varies from nation to nation. While the official recommendation is to place it before the number (contravening the general ISO recommendation to place unit symbol after the number[citation needed]), people in many countries have kept the placement of their former currencies.[citation needed]

This is the official construction of the euro logo, which was specified to be printed in PMS Yellow on a PMS Reflex Blue background

Economic and Monetary Union [edit]

History (1990-2006) [edit]

Main article: Economic and Monetary Union of the European Union

For earlier monetary history in Europe, see European Monetary System.

The euro was established by the provisions in the 1992 Maastricht Treaty on European Union that was used to establish an economic and monetary union. In

Fig. 69 Paragraphing in Wikipedia

Each paragraph should be limited to just one idea; this technique is very effective when combined with front-loading paragraph content. The functions of paragraphing are clearly summarized in Wikipedia *Manual of Style* [45] which emphasizes the concept that its main purpose is to *define, describe, detail and direct*. Each paragraph should contain at least one of these aims. If it performs more than two, it means that it is too complex and unfocused. The *Manual of Style* warns Wikipedians that a

[45] Wikipedia Manual of Style http://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style

paragraph must show a clear organization and that simply grouping together sentences in a block is not enough to create a coherent paragraph.

Analyzing the structure of Britannica and Wikipedia's encyclopaedic articles, it is evident that both take care in properly using paragraphing techniques. Double line spacing isolates paragraphs in both encyclopaedias. Nevertheless, Britannica considers paragraphs, especially in the shortest articles, as its main textual blocks (fig. 68b), whereas in Wikipedia most of the times they are restructured in more general sections and subsections (fig. 69).

5.2.5 Lists and Bullets

Nielsen (2006) recommends an extensive use of lists and bullets when writing for the web. Their use is preferred to long paragraphs because they allow users to read the information vertically rather than horizontally. In addition, the use of lists and bullets makes the text easier to scan, less intimidating and usually make the information provided more concise and simple to remember. While Britannica totally disregards this technique, Wikipedia contributors seem to be aware of the advantages of this practice. Fig. 68a shows, that this technique is extensively used as a strategy to summarize and quickly spot the searched information. Nevertheless, Wikipedia *Manual of Style* suggests a correct and functional use of it by stating:

> *Bulleted lists*
>
> Do not use bullets if the passage reads easily using plain paragraphs or indented paragraphs. If every paragraph in a section is bulleted, it is likely that none should be bulleted.
>
> Do not mix grammatical styles in a list—either use all complete sentences or use all sentence fragments. Begin each item with a capital letter, even if it is a sentence fragment.
>
> When using complete sentences, provide a period at the end of each. When using sentence fragments, do not provide a period at the end [46].

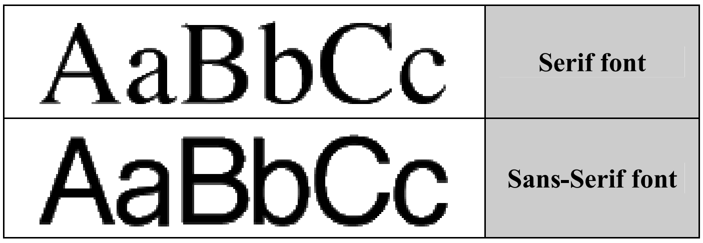
5.2.6 Text Alignment

Left aligned text is easier to read than justified text, which in turn is easier to read than centre or right aligned text. When reading through justified text the spacing between each word is different so that eyes have to search for the next word; this slows down the reading speed. Right and centre aligned paragraphs slow down reading speed even more because each time the reading of one line finishes, the eye has to search for the beginning of the next line (Nielsen, 2006). Both Britannica and Wikipedia seem to be aware of the advantages produced by this writing technique as they both use left-aligned texts (fig. 68a/b).

[46] *Wikipedia Manual of Style* http://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style

5.2.7 Web Font

Typefaces have an important impact on readability. They can be divided into two main classes: Serif and Sans-Serif[47] (fig. 70). Traditionally, typefaces with serifs have been considered easier to read in long passages. As a general rule printed texts, such as newspapers and books, mostly use serif typefaces. Studies on this matter are ambiguous, suggesting that most of this effect is due to the readers' greater familiarity with the *Serif* typefaces. By contrast, websites extensively use modern Sans-Serif fonts, since it is commonly believed that they are more easily readable than serif fonts on low-resolution computer screens. This revelation runs contrary to the long-held understanding that Serif fonts speed reading time.



AaBbCc	**Serif font**
AaBbCc	**Sans-Serif font**

Fig. 70 *Serif* and *Sans-serif* font

The three classic fonts belonging to the Sans-Serif font family used all over the Web are **Verdana**, **Trebuchet MS** and **Arial**. Both Britannica and Wikipedia pay attention to the use of the correct typefaces in order to improve concentration and reading speed. In fact, both encyclopaedias use fonts belonging to the Sans-Serifs family. In particular, Wikipedia has adopted the typeface **Arial Narrow**, while Britannica **Trebuchet MS.**

5.2.8 Font Size

Font size is an important device for giving hierarchy to the content, and the relative sizes of headings, body text and footers have a big influence on the overall feel of a page (Nielsen, 2006). Font size is also very closely linked to other characteristics of the page, such as column width, line-height, and so on.

In the sixteenth century, typographers began to use a common scale for type size, and their approach has been replicated also on the Web, particularly if a traditional and highly legible result is looked for. Most of the websites use as standard font, size 12 and Britannica and Wikipedia employ it as well. While Britannica uses always the same font size, also in the heading and subheadings, Wikipedia prefers to use a bigger one in order to make titles easily recognizable and identifiable (fig. 68a/b).

[47] Serifs comprise the small features at the end of strokes within letters.

5.2.9 Boldface

Lynch *et al.* (2002) give the following description of the concept of typographic emphasis in the *Web Style Guide*:

> A Web page of solid body text is hard to scan for content structure and will not engage the eye. Adding display type to a document will provide landmarks to direct the reader through your content. Display type establishes an information structure and adds visual variety to draw the reader into your material. The key to effective display type is the careful and economic use of typographic emphasis. There are time-honored typographical devices for adding emphasis to a block of text, but be sure to use them sparingly. If you make everything bold, then nothing will stand out and it will seem as if you are shouting at your readers. A good rule of thumb when working with type is to add emphasis using one parameter at a time. If you want to draw attention to the section heads in your document, don't set them large, bold, and all caps. If you want them to be larger, increase their size by one measure. If you prefer bold, leave the heads the same size as your body text and make them bold. You will soon discover that only a small variation is required to establish visual contrast.

According to Nielsen (1999) emphasizing text is a relatively simple way to bring words to life on the Web. Different techniques can be used, the most common and effective method being the use of a bold face from the current font family. Web usability points out that a way to help users to quickly and easily spot information is to use bold font to draw attention to important words in the text. Nevertheless, just two or three words describing the main point of the paragraph should be put in bold. When site visitors scan through the screen, this aspect of the text stands out. It is essential that the bold text makes sense also out of the context. Through the bold words visitors can instantly gain an understanding of what the article is about and decide whether or not to read it. The *Web Style Guide* site (Lynch, 2005) states:

> Boldface text gives emphasis because it contrasts in color from the body text. Section subheads work well set in bold. Boldface text is readable on-screen, though large blocks of text set in bold lack contrast and therefore lose their effectiveness.

The bolding technique is used for this purpose also in the analyzed encyclopaedias. Section and subsection's headings are in bold both in Wikipedia and Britannica. In addition, when Britannica finds the searched encyclopaedic article through the internal search engine, it automatically bolds all the words which have been typed in the search box.

5.2.10 *Italics* and Underlining

Text in *Italics* attracts the eye because it contrasts in shape from body text. Italics is conventionally used when book or periodical titles are listed, or to stress foreign words or phrases within the text. Lynch *et al.* (2002) suggest to avoid setting large blocks of text in Italics because the readability of such a text is much lower than in comparably sized Roman text.

Italics is never used in Britannica. Wikipedia sometimes uses it but not very extensively. This happens for different reasons. Text in *Italics* can suffer on low-resolution monitors for the slanted and more curved shapes of the letters. Nevertheless, this does not stop Wikipedia from using *Italics*, when standard convention approves its use as for example, when foreign words and phrases are quoted, or when books and periodical titles are listed in the references.

Underlined text is a carryover from the days of the typewriter, when options such as *Italics* and **boldface** were unavailable. In addition to its aesthetic shortcomings (too heavy, interferes with letter shapes), underlining has a special functional meaning in web documents as it typically indicates hyperlinked text. This default convention ensures that people colour-blind or with monochromatic monitors can identify links within text blocks (Lynch, 2005). If underlined texts were included on web pages they would certainly be confused with hypertext links. Neither in Wikipedia nor in Britannica underlining is used to highlight parts of a text, since this would confuse the reader in the interpretation of the main function of this technique.

5.2.11 Font and Link Colours

The Web Style Guide site (Lynch, 2002) states:

> Although the use of color is another option for differentiating type, colored text, like underlining, it has a special functional meaning in Web documents. You should avoid putting colored text within text blocks because readers will assume that the colored text is a hypertext link and click on it. Colored text does work well as a subtle means to distinguish section heads, however. Choose dark shades of color that contrast with the page background, and avoid using colors close to the default Web link colors of blue and violet.

Thus, black is the standard colour used for writing on the web. Both Britannica and Wikipedia use this basic font colour in the writing of encyclopaedic articles. Using colour for emphasis can be a rather tricky business. In the past it was common to use distinct colours to give emphasis to a passage of text. On the web, coloured words, just as underlined words, could be mistaken for a link within body text. Nowadays, coloured fonts have become the standard convention used for identifying links. Some colour combinations can frustrate users or make texts virtually unreadable for colour-blind users (Nielsen, 2006) and for many of them, some colours look the same. As colour-blinds cannot distinguish between a large spectrum of colours, it is suggested to strongly contrast the link colours, as most of the times there is also a luminosity loss in their spectrum.

Common standards in web design specify say that the standard hyperlink colours should be: blue for non-visited hyperlinks, purple for visited hyperlinks, and red for active hyperlinks. Britannica and Wikipedia do not respect this standard as both use light blue, for non-visited links and just one colour for both visited and active links. In particular, Wikipedia uses orange for visited and active links, while Britannica prefers grey, a more serious colour.

Thus, both Wikipedia and Britannnica do not respect standard web usability link colours. Britannica makes the worst chromatic choice, as the colour grey does not contrast the black well, the basic textual font colour. Moreover, many colour-blinds do not perceive at all the difference between black and grey for the loss of luminosity in their colour spectrum.

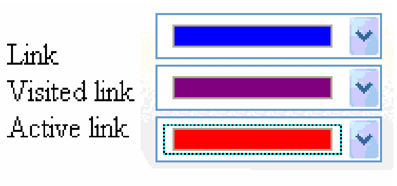
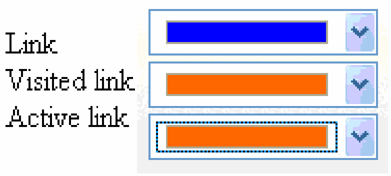
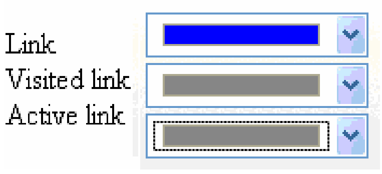
Standard link colours	Wikipedia link colours	Britannica link colours
 Link Visited link Active link	 Link Visited link Active link	 Link Visited link Active link

Fig. 71 Link colours

5.2.12 Capital Letters

Lynch (2002) states in *The Web Style Guide* site:

> Capitalized text is one of the most common and least effective methods for adding typographical emphasis. We recognize words in two ways, by parsing letter groups and by recognizing word shapes. Words or headlines set in all capital letters form rectangles with no distinctive shape. To read a block of text set in all capital letters we must parse the letter groups — read the text letter by letter — which is uncomfortable and significantly slows reading.

Neither in Wikipedia nor in Britannica passages of text fully written in capital letters are found. Even heading and subheadings are written in small letters, with only the first letter in capital. Not only the use of capital letters is considered rude and inelegant, but typographically, and stylistically it is a very poor choice as THE READABILITY OF SENTENCES IN CAPITAL LETTERS is severely inhibited.

5.2.13 Line Spacing and Indentation

Lynch (2002) states in the *Web Style Guide* site:

> One of the most effective and subtle ways to vary the visual contrast and relative importance of a piece of text is simply to isolate it or treat it differently from the surrounding text. If you want your major headers to stand out more without making them larger, add space before the header to separate it from any previous copy. Indentation is another effective means of distinguishing bulleted lists, quotations, or example text.

The use of line spacing is a crucial issue on the Web. The vertical distance between lines of body text can make a huge difference to the legibility and overall style of the text (Nielsen, 1999). The default line-height for most browsers is around 1.2 as 1 is not sufficient for text on screen as in this case the top of one row of characters touches the base of the row above. Britannica and Wikipedia conform their style to the practice dominant on the web, furthermore, they isolate headers' article sections adding space before and after to separate them from the rest of the text. By contrast, neither Britannica nor Wikipedia use indentation, as all the text is uniformly aligned on the left and a blank space has never been found at the beginning of lines.

5.2.14 Text Line Length

The optimal text line length depends upon several factors. It is commonly recommended that shorter line lengths (about 11 words) should be used in place of longer, full-screen lengths. This is because longer lines require greater lateral eye movements, which make it more likely to lose one's position within the text (Horton, 1989; Mills & Weldon, 1987). It has also been pointed out by Horton (1989) that longer line lengths are more tiring to read as shown by the reading pattern eye tracking in fig. 72.

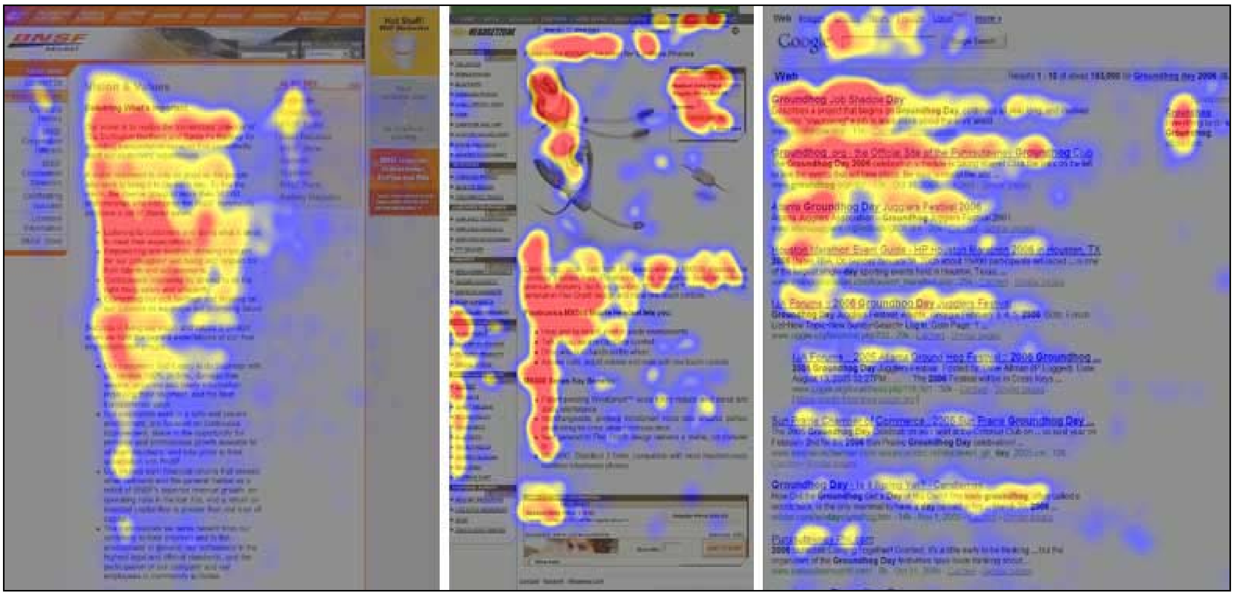


Fig. 72 Reading pattern eye tracking

A recent study (Bernard *et al.,* 2002) on the comparison of three line lengths (24.5, 14.5, and 85 cm, respectively) supports the finding that shorter line lengths are preferred to full-screen line lengths. Horton recommends that lines should be limited to lengths of around 40 to 60 characters, which is approximately 11 words per line.

The web usability analysis has demonstrated that Britannica has followed this fundamental standard convention as its encyclopaedic articles are presented in lines composed of 8-10 words, whereas Wikipedia violates the suggested length, adopting an average line of about 18-20 words (fig. 68 a/b).

5.2.15 Background texture

Most studies have shown that it is usually not a good idea to use complex backgrounds (or images) on a webpage, as they tend to slow down page loading, and can interfere with reading the foregrounded text. Dark characters on a light background have proved to have a higher Index of Readability than light characters on a dark background. Bauer *et al.* (1980) find that participants are 26% more accurate in reading a text when dark characters on a light background are used. Moreover, a survey by Scharff *et al.* (1996) reveals that the colour combination perceived as being the most readable is the traditional black on white background. Being aware of how much background and font colours affect the readability of a text, both Britannica and Wikipedia use white background and always black fonts as standard colours.

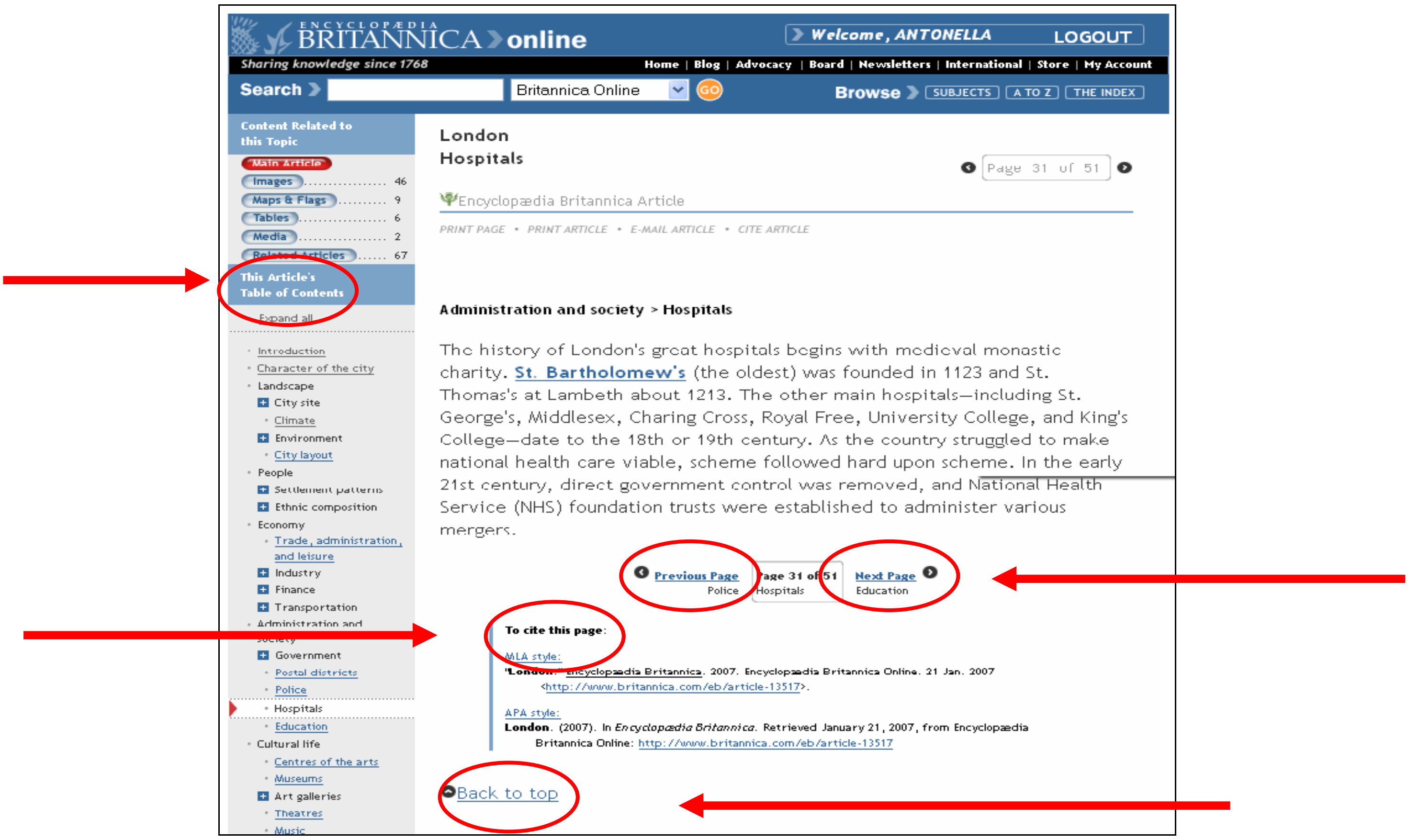
5.2.16 Page Length

Page length is a further important aspect which determines the degree of usability of a webpage. Britannica replicates the structure of a traditional encyclopaedic printed page. It prefers to distribute article content in narrow columns and numerous pages. Interlinking is not widely used in Britannica encyclopaedic articles and content distribution is not homogeneous across webpages. For example, the article *London* is contained in 51 webpages whose length is variable: the 1st introductory page counts 263 words, the 2nd page 930, while the 31th page only 88 words (fig. …), etc.

In Britannica, navigation buttons allow a linear navigation to the *previous* or the *next page,* and internal bookmarks allow to go *back to the top* of a specific webpage. In order to jump from one page to another the *Table of Contents* on the left frame can be used. As the content is parcelled out in many different webpages, each article provides a link to a separate file that contains the full-length text designed as a single page so that the reader can print or save all the article content in one step.

Furthermore, recommendations on how to correctly cite the source according to *MLA* (Modern Language Association) and *APA* (American Psychological Association) are given at the bottom of every article page (fig. 73a).

By contrast, Wikipedia prefers a completely different webpage layout. Each encyclopaedic article is concentrated in a single and long scrollable page, which is easily and directly printable. The content is highly interlinked and, at top of each page, under the opening paragraph, a clickable *Table of Contents* is provided (fig. 73b).


ENCYCLOPÆDIA BRITANNICA online
Welcome, ANTONELLA
LOGOUT
Sharing knowledge since 1768
Home | Blog | Advocacy | Board | Newsletters | International | Store | My Account
Search
Britannica Online
GO
Browse
SUBJECTS
A TO Z
THE INDEX
Content Related to this Topic
Main Article
Images 46
Maps & Flags 9
Tables 6
Media 2
Related Articles 67
This Article's Table of Contents
Expand all
Page 31 of 51
PRINT PAGE • PRINT ARTICLE • E-MAIL ARTICLE • CITE ARTICLE

- Introduction
- Character of the city
- Landscape
  - City site
  - Climate
  - Environment
  - City layout
- People
  - Settlement patterns
  - Ethnic composition
- Economy
  - Trade, administration, and leisure
  - Industry
  - Finance
  - Transportation
- Administration and society
  - Government
  - Postal districts
  - Police
  - Hospitals
  - Education
- Cultural life
  - Centres of the arts
  - Museums
  - Art galleries
  - Theatres
  - Music

London
Hospitals

Encyclopædia Britannica Article

Administration and society > Hospitals

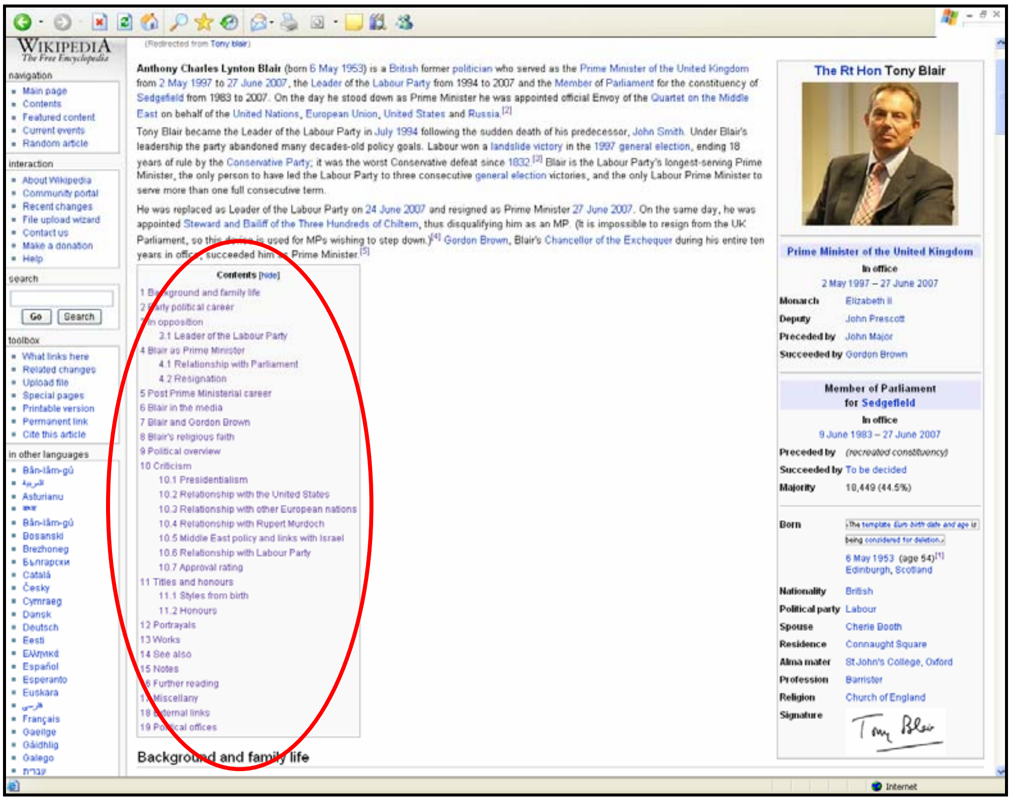
The history of London's great hospitals begins with medieval monastic charity. St. Bartholomew's (the oldest) was founded in 1123 and St. Thomas's at Lambeth about 1213. The other main hospitals—including St. George's, Middlesex, Charing Cross, Royal Free, University College, and King's College—date to the 18th or 19th century. As the country struggled to make national health care viable, scheme followed hard upon scheme. In the early 21st century, direct government control was removed, and National Health Service (NHS) foundation trusts were established to administer various mergers.

To cite this page:

MLA style:
"London." Encyclopædia Britannica. 2007. Encyclopædia Britannica Online. 21 Jan. 2007 <http://www.britannica.com/eb/article-13517>.

APA style:
London. (2007). In *Encyclopædia Britannica*. Retrieved January 21, 2007, from Encyclopædia Britannica Online: http://www.britannica.com/eb/article-13517

Fig. 73a Navigation buttons in Britannica



(Redirected from Tony blair)

Anthony Charles Lynton Blair (born 6 May 1953) is a British former politician who served as the Prime Minister of the United Kingdom from 2 May 1997 to 27 June 2007, the Leader of the Labour Party from 1994 to 2007 and the Member of Parliament for the constituency of Sedgefield from 1983 to 2007. On the day he stood down as Prime Minister he was appointed official Envoy of the Quartet on the Middle East on behalf of the United Nations, European Union, United States and Russia.[2]

Tony Blair became the Leader of the Labour Party in July 1994 following the sudden death of his predecessor, John Smith. Under Blair's leadership the party abandoned many decades-old policy goals. Labour won a landslide victory in the 1997 general election, ending 18 years of rule by the Conservative Party; it was the worst Conservative defeat since 1832.[3] Blair is the Labour Party's longest-serving Prime Minister, the only person to have led the Labour Party to three consecutive general election victories, and the only Labour Prime Minister to serve more than one full consecutive term.

He was replaced as Leader of the Labour Party on 24 June 2007 and resigned as Prime Minister 27 June 2007. On the same day, he was appointed Steward and Bailiff of the Three Hundreds of Chiltern, thus disqualifying him as an MP. (It is impossible to resign from the UK Parliament, so this device is used for MPs wishing to step down.)[4] Gordon Brown, Blair's Chancellor of the Exchequer during his entire ten years in office, succeeded him as Prime Minister.[5]

Contents [hide]

Background and family life

The Rt Hon Tony Blair

Prime Minister of the United Kingdom
In office
2 May 1997 – 27 June 2007

Monarch	Elizabeth II
Deputy	John Prescott
Preceded by	John Major
Succeeded by	Gordon Brown

Member of Parliament for Sedgefield
In office
9 June 1983 – 27 June 2007

Preceded by	*(recreated constituency)*
Succeeded by	To be decided
Majority	18,449 (44.5%)
Born	«The template Euro birth date and age is being considered for deletion.» 6 May 1953 (age 54)[1] Edinburgh, Scotland
Nationality	British
Political party	Labour
Spouse	Cherie Booth
Residence	Connaught Square
Alma mater	St John's College, Oxford
Profession	Barrister
Religion	Church of England
Signature	Tony Blair

Fig. 73b *Table of Contents* in Wikipedia

As Nielsen (1999) points out, long web pages have their advantages. They are often easier for creators to organize and for users to download. Web site managers do not have to maintain as many links and pages with longer documents, and users do not need to download multiple files to collect information on a topic. Long pages are particularly useful for providing information that is not expected to be read online (realistically, that means any document longer than two printed pages). It makes sense to keep closely related information within the boundaries of a single web page, particularly when it is expected that the user prints or saves the text. Thus, keeping the content in one webpage makes downloading, printing or saving easier. In general, text contained in a single long document is easier to maintain (as content is in one piece, not in linked chunks), more like the structure of their paper counterparts. The reading of articles contained in a single page is preferred by netsurfers, as it seems to improve concentration, information retrieval and reading speed (Nielsen, 1999).

5.2.17 Encyclopaedic Article Length and Hypertextual Structure

Although the analysis of Britannica and Wikipedia in this research has revealed similar lexical density as well as same word and sentence length[48], the investigation of the encyclopaedic article lengths has given very divergent results. Articles are much longer in Wikipedia than in Britannica, in spite of the higher number of cross references of the former.

Hypertext is a user interface paradigm for displaying documents which, according to an early definition by Nelson (1970), *branch or perform on request.* Wikipedia's hypertextual structure organizes material attempting to overcome the inherent limitations of traditional printed encyclopaedias. The prefix *-hyper* (Greek term for *over* or *beyond*) means the overcoming of such constraints. Wikipedia articles contain a high number of cross-references to other articles. This is due to the easiness of the *Wiki Markup Language* which allows the user to link pages in a very simple way. Many wikis, especially the earlier ones, used *CamelCase* technique[49] to create links. In most of the recent wikis (such as Wikipedia and other MediaWiki-based wikis), this convention has been abandoned in favour of an explicit link markup, which puts the linking word between double square brackets [[…]].

Wikipedia makes use of different kinds of hyperlinks such as *wikilinks,* which are internal to the encyclopaedia, *interwiki links,* which connect different wiki projects (such as *Wikibooks, Wiktionary, Wikinews,* etc.) and *external links,* which join wiki pages to other net documents.

[48] see chapter 4. sections 1.1.1-1.1.3

[49] *CamelCase* is the practice of writing compound words or phrases where the words are joined without spaces, and each word is capitalized within the compound. The name comes from the uppercase "bumps" in the middle of the compound word, suggesting the humps of a camel.

Surprisingly, Wikipedia's articles are much longer than Britannica's in spite of their dense hypertextual structure which usually should reduce textual length as more in depth information is found in the interlinked pages. This unexpected effect is due to the articulateness of a wider information provided and not to a redundant and repetitive style, as average sentence has practically the same length (22.05 words BAs vs. 22.09 WAs). Article average length has shown to be of 2.472 words in Britannica *vs.* 3.988 in Wikipedia.

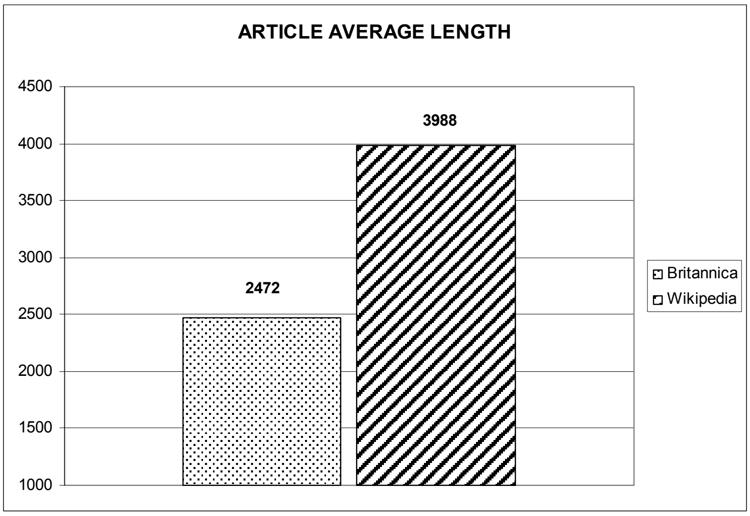


Fig. 74 Article average length in BAs vs. WAs

Fig. 74 shows the difference between average article lengths in the two encyclopaedic corpora, while fig. 75 the length variation in each couple of articles.

The actual difference is higher but some oversized articles in the Britannica corpus [e.g.*Cryptography* (15427 tokens), *London* (17138 tokens) and *Epistemology* (24996 tokens)] have reduced the average length variation.

As fig. 76 shows, article length ranges from 116 words in *Pizza* article to 24996 words in *Epistemology* in the Britannica, whereas the shortest article is of 372 words (*Wind rose*) and the longest is of 11967 words (*Tea*) in Wikipedia.

In conclusion, except for some sporadic cases, Wikipedia articles are longer. The micro analysis has shown that 60% of Britannica articles consist of less than 1000 words, while only 6% contain less than 1000 words in Wikipedia (fig. 76).

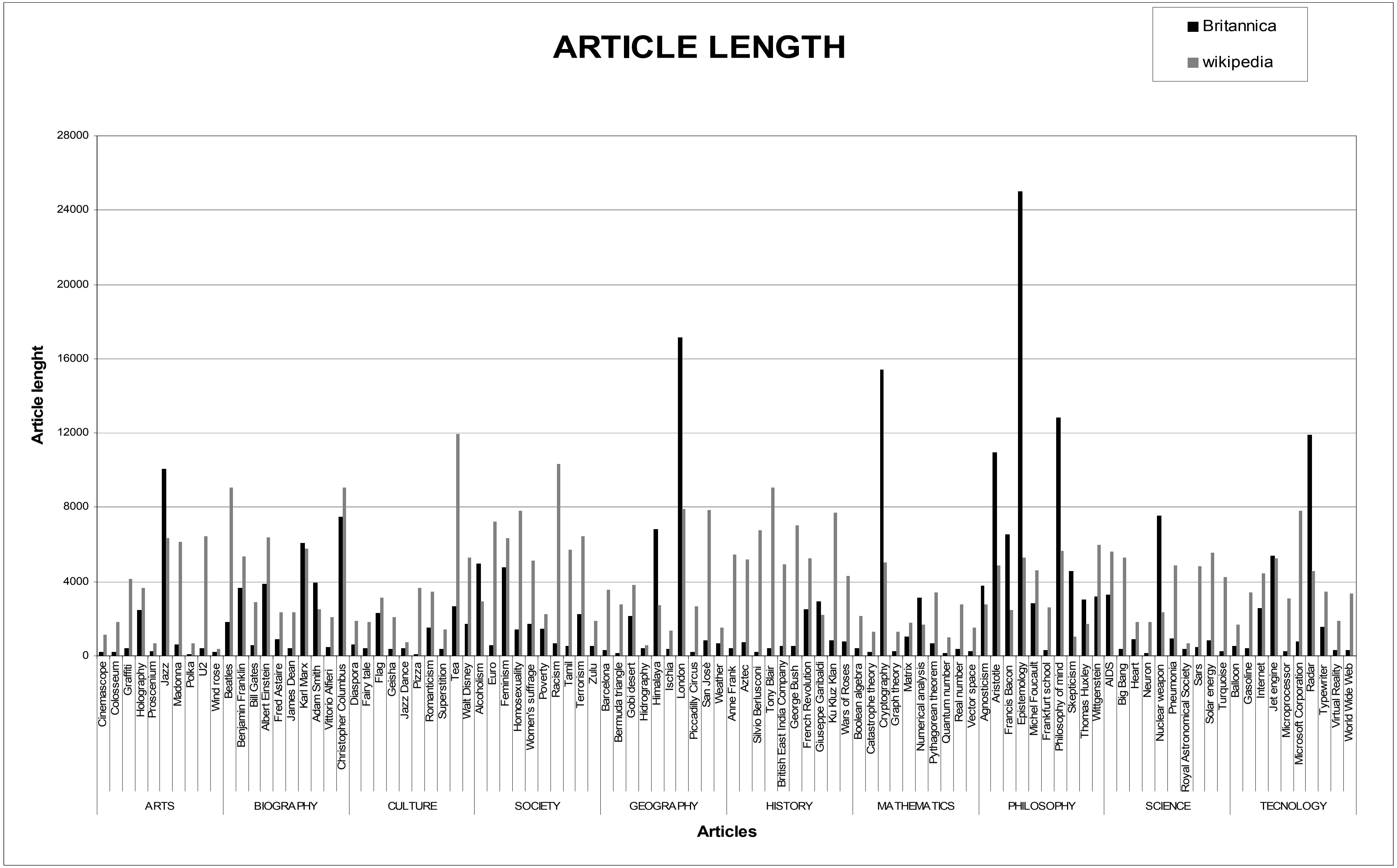


Fig. 75 Article lengths in BAs vs. WAs

60 % Britannica Articles < 1000 Words		6 % Wikipedia Articles < 1000 Words	
Britannica	***Tokens***	**Britannica**	***Tokens***
Pizza	116	U2	438
Polka	127	James Dean	442
Quantum number	135	Vittorio Alfieri	447
Bermuda triangle	154	Sars	456
Neuron	167	Zulu	535
Cinemascope	199	Tamil	544
Colosseum	202	George Bush	546
Catastrophe theory	203	Euro	565
Wind rose	205	Bill Gates	601
Piccadilly Circus	214	Madonna	615
Silvio Berlusconi	227	Diaspora	620
Vector space	256	Pythagorean theorem	673
Graph theory	272	Racism	689
Proscenium	274	Weather	689
Turquoise	277	Aztec	736
Microprocessor	277	Microsoft Corporation	775
Barcelona	294	Wars of Roses	796
Virtual Reality	325	Solar energy	849
Frankfurt school	327	Ku Kluz Klan	859
World Wide Web	338	San Josè	861
Ischia	341	Fred Astaire	905
Real number	341	Heart	917
Superstition	355	Pneumonia	950
Geisha	371	Matrix	1034
Royal Astronomical Society	372		
Big Bang	389	**Wikipedia**	***Tokens***
Graffiti	406		
Jazz Dance	411	Wind rose	372
Hidrography	414	Hidrography	592
Fairy tale	420	Royal Astronomical Society	674
Balloon	505	Polka	697
British East India Company	508	Proscenium	706
Anne Frank	424	Jazz Dance	720
Tony Blair	427		
Boolean algebra	430		
Gasoline	433		

Fig. 76 Specific article length variation in BAs vs. WAs

The complex hypertextual structure is clearly visible in Wikipedia by comparatively glancing at the articles of the two encyclopaedias. Evidence of this has been found by this research counting the number of links in some sample articles taken from the two corpora. This operation has automatically been carried out through the Web Site *Link Analyzer tool*[50] which evaluates a given webpage and returns a table of data containing columns of links subdivided in *external links* (going outside the website) and *internal links* (inside the current website).

To give an example, the article *Agnosticism* has been chosen. Fig. 77 shows that this article has 213 internal links in Wikipedia, whereas Britannica has only 45.

Article	Link typology	Britannica	Wikipedia
Agnosticism	**Internal links**	**45**	**213**
	External links	**29**	**68**
Graffiti	**Internal links**	**64**	**585**
	External links	**27**	**72**

Fig. 77 Links in WAs vs. BAs

More specifically, most of the internal links crosslink content in Wikipedia (a very reduced number of them have been found in Britannica) while navigation buttons are the most frequent typology of internal links found in Britannica; they link each encyclopaedic article to the homepage, the index, the internal search engine, the printing options and the online store. Only one content link has been detected in this specific Britannica article. In addition, external links are visibly more numerous in Wikipedia than in Britannica (68 vs. 29 in *Agnosticism*)[51].

To give a further example in the article *Graffiti* [52], 585 internal links and 72 external links have been found in Wikipedia as opposed to only 64 internal links (of which only 8 content-related) and 27 external links in Britannica (fig. 77).

In brief, the two sample articles have shown that information in Wikipedia is more interlinked. This is probably due to the easiness of the *Wiki Markup Language*, to the power of collaborative authoring, and last but not least, to the network philosophy embraced by the web 2.0. By contrast, hypertextual advantages have not been exploited to the full for content interlinking in Britannica. Articles are here organized in single columns and in short sequential pages with back and forward buttons which only allow linear navigation. Page layout of Britannica's articles seems to be very close to the reproduction of a typical printed page.

In conclusion, this analysis has shown longer articles in Wikipedia despite the dense content interlinking, which should reduce article length as information is expected to be distributed in the connected subpages. Longer Wikipedia articles are not due to a prolix and redundant style. This data has been confirmed by the previously analysis which has shown a similarity in the average sentence

[50] *Link Analyzer tool* http://www.improve-ranking.com/link_analyzer_seo_tool.html
[51] Data refers to 4 February 2006 version of the *Agnostocism* Wikipedia's article
[52] Data refers to 2 February 2006 version of the *Graffiti* Wikipedia's article

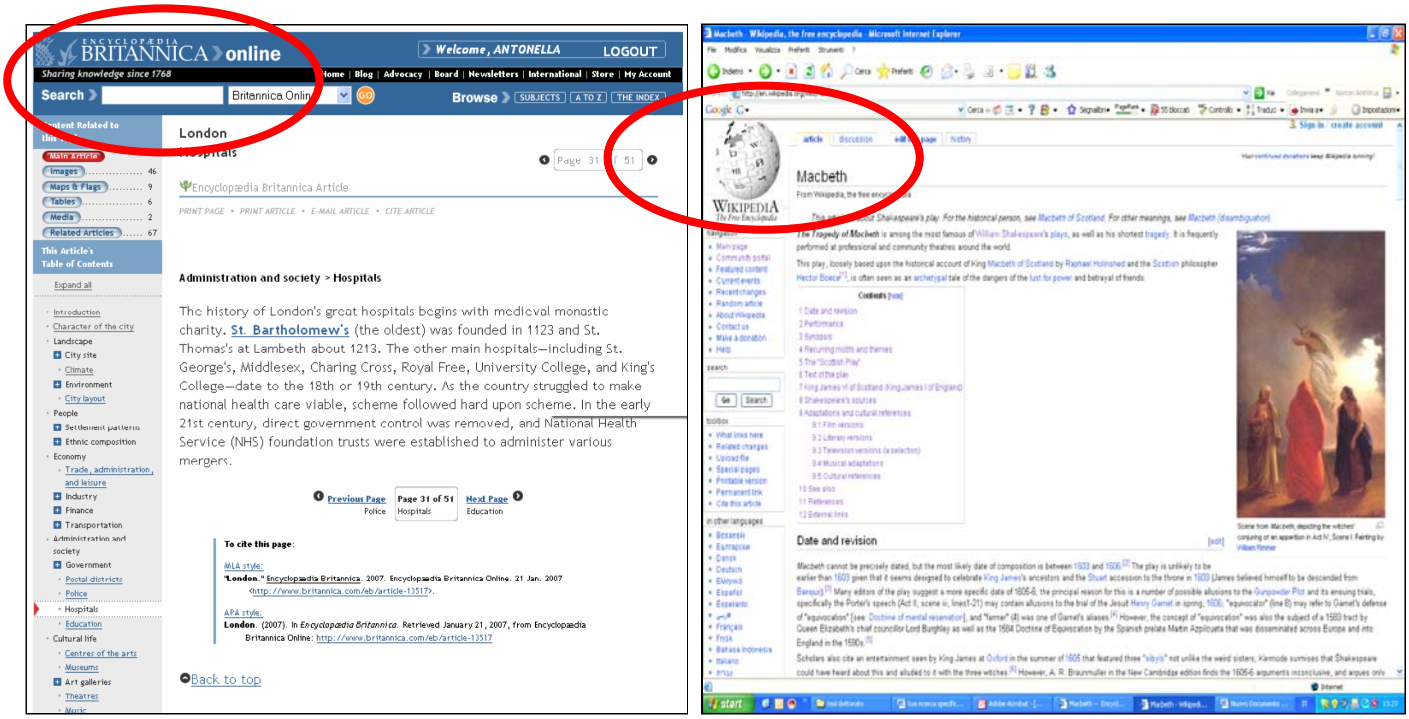
length of Britannica (22.05 words per sentence) and Wikipedia (22.09 per sentence). Thus, Wikipedia articles are longer as they provide additional information.

5.2.18 Multimedia

As most of websites, also Britannica and Wikipedia use graphics and media. If used appropriately, images, video clips, and audio add a remarkable value to the website and facilitate learning as they enrich human understanding. Superiority of hypermedia over simple hypertexts appears to be stronger especially in a learning context, as they are more easily remembered than a monotonous textual explanation.

According to the basic Web Usability principles, multimedia elements have to be used only when they help to convey or support the textual message (Nielsen, 2006). Since they can easily capture the web reader's attention, it is important to have clear and useful reasons for using them so as to avoid unnecessary distractions. Furthermore, some multimedia elements may take a long time to be downloaded, so it is important that the waiting is worthwhile.

The first extra textual elemet which should always be shown in a website is its logo (fig. 78 a/b) which has to be placed in a consistent place on every page.



London
Hospitals

Encyclopædia Britannica Article

Administration and society > Hospitals

The history of London's great hospitals begins with medieval monastic charity. St. Bartholomew's (the oldest) was founded in 1123 and St. Thomas's at Lambeth about 1213. The other main hospitals—including St. George's, Middlesex, Charing Cross, Royal Free, University College, and King's College—date to the 18th or 19th century. As the country struggled to make national health care viable, scheme followed hard upon scheme. In the early 21st century, direct government control was removed, and National Health Service (NHS) foundation trusts were established to administer various mergers.

To cite this page:

MLA style:
"London." Encyclopædia Britannica. 2007. Encyclopædia Britannica Online. 21 Jan. 2007 <http://www.britannica.com/eb/article-13517>.

APA style:
London. (2007). In *Encyclopædia Britannica*. Retrieved January 21, 2007, from Encyclopædia Britannica Online: http://www.britannica.com/eb/article-13517

Macbeth

From Wikipedia, the free encyclopedia

This article is about Shakespeare's play. For the historical person, see Macbeth of Scotland. For other meanings, see Macbeth (disambiguation).

The Tragedy of Macbeth is among the most famous of William Shakespeare's plays, as well as his shortest tragedy. It is frequently performed at professional and community theatres around the world.

This play, loosely based upon the historical account of King Macbeth of Scotland by Raphael Holinshed and the Scottish philosopher Hector Boece[1], is often seen as an archetypal tale of the dangers of the lust for power and betrayal of friends.

Contents [hide]
1 Date and revision
2 Performance
3 Synopsis
4 Recurring motifs and themes
5 The "Scottish Play"
6 Text of the play
7 King James VI of Scotland (King James I of England)
8 Shakespeare's sources
9 Adaptations and cultural references
9.1 Film versions
9.2 Literary versions
9.3 Television versions (a selection)
9.4 Musical adaptations
9.5 Cultural references
10 See also
11 References
12 External links

Scene from Macbeth, depicting the witches' conjuring of an apparition in Act IV, Scene I. Painting by William Rimmer

Date and revision [edit]

Macbeth cannot be precisely dated, but the most likely date of composition is between 1603 and 1606.[2] The play is unlikely to be earlier than 1603 given that it seems designed to celebrate King James's ancestors and the Stuart accession to the throne in 1603 (James believed himself to be descended from Banquo).[3] Many editors of the play suggest a more specific date of 1605-6, the principal reason for this is a number of possible allusions to the Gunpowder Plot and its ensuing trials, specifically the Porter's speech (Act II, scene iii, lines1-21) may contain allusions to the trial of the Jesuit Henry Garnet in spring, 1606; "equivocator" (line 8) may refer to Garnet's defense of "equivocation" [see: Doctrine of mental reservation], and "farmer" (4) was one of Garnet's aliases.[4] However, the concept of "equivocation" was also the subject of a 1583 tract by Queen Elizabeth's chief councillor Lord Burghley as well as the 1584 Doctrine of Equivocation by the Spanish prelate Martin Azpilcueta that was disseminated across Europe and into England in the 1590s.[5]

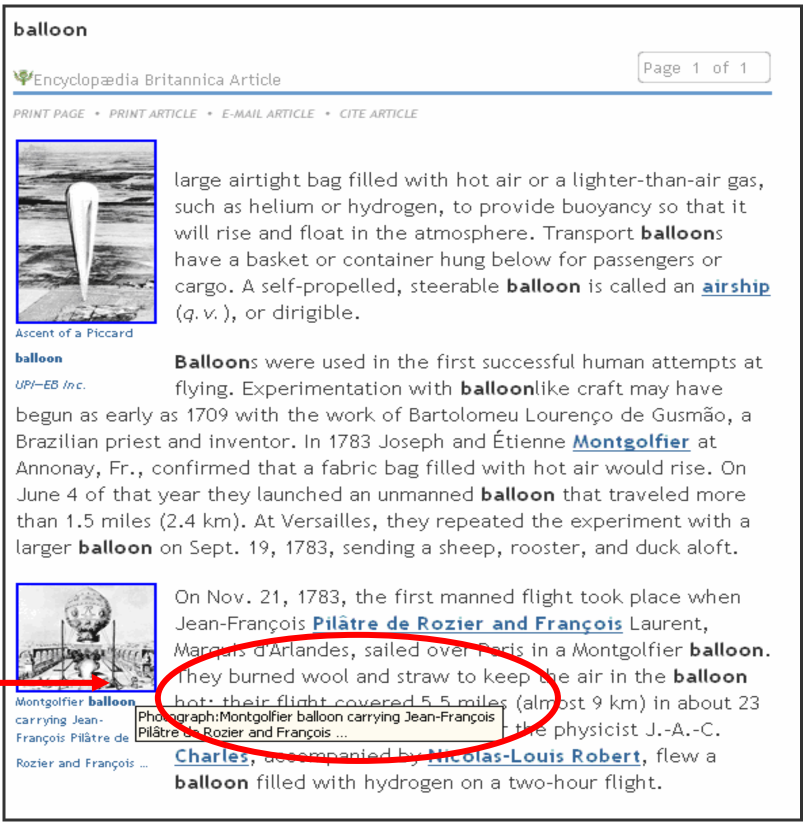
Scholars also cite an entertainment seen by King James at Oxford in the summer of 1605 that featured three "sibyls" not unlike the weird sisters; Kermode surmises that Shakespeare could have heard about this and alluded to it with the three witches.[6] However, A. R. Braunmuller in the New Cambridge edition finds the 1605-6 arguments inconclusive, and argues only

Fig. 78 a/b Encyclopaedia's logo in Britannica and Wikipedia

Users are frequently unaware when they click through to a different website. Having a logo on each page provides a frame of reference throughout a website so that users can easily confirm that they have not left it (Nielsen, 1999). The logo should always be in the same position on each webpage.

Following the standard format, Britannica and Wikipedia's web designers place it at the top in the left corner (fig. 78 a/b). Text and associated images have to be close together in a page so that users can integrate and effectively use them. Users tend to be frustrated if they wait several seconds to download images and then find that they do not add any value to the text. In order to speed downloading time, both Britannica and Wikipedia insert thumbnail versions of larger images in their pages. By using this technique, those who are not interested in the full image are not slowed down by large image downloads. Of course, thumbnail images are linked to the full-size copy. In addition, Wikipedia offers also a download high-resolution version (506 x 800, 55 KB) for all the images which appear on the website, which are free to use as they are released under *Creative Commons Licence* [53], while every content in Britannica is covered by *copyright*.

According to Web Usability principles, Nielsen (1999) prescribes the necessity to **label** images to help users to understand them and their messages. Moreover, *alt text* [54] should accompany every clickable image. Britannica and Wikipedia have respected these basic web usability's rules and therefore they have labelled their images by putting images' pop-up alt text (fig. 79 a/b)



balloon

Encyclopædia Britannica Article

Page 1 of 1

PRINT PAGE • PRINT ARTICLE • E-MAIL ARTICLE • CITE ARTICLE

Ascent of a Piccard **balloon**

UPI—EB Inc.

large airtight bag filled with hot air or a lighter-than-air gas, such as helium or hydrogen, to provide buoyancy so that it will rise and float in the atmosphere. Transport **balloons** have a basket or container hung below for passengers or cargo. A self-propelled, steerable **balloon** is called an airship (*q.v.*), or dirigible.

Balloons were used in the first successful human attempts at flying. Experimentation with **balloon**like craft may have begun as early as 1709 with the work of Bartolomeu Lourenço de Gusmão, a Brazilian priest and inventor. In 1783 Joseph and Étienne Montgolfier at Annonay, Fr., confirmed that a fabric bag filled with hot air would rise. On June 4 of that year they launched an unmanned **balloon** that traveled more than 1.5 miles (2.4 km). At Versailles, they repeated the experiment with a larger **balloon** on Sept. 19, 1783, sending a sheep, rooster, and duck aloft.

Montgolfier **balloon** carrying Jean-François Pilâtre de Rozier and François ...

On Nov. 21, 1783, the first manned flight took place when Jean-François Pilâtre de Rozier and François Laurent, Marquis d'Arlandes, sailed over Paris in a Montgolfier **balloon**. They burned wool and straw to keep the air in the **balloon** hot; their flight covered 5.5 miles (almost 9 km) in about 23 the physicist J.-A.-C. Charles, accompanied by Nicolas-Louis Robert, flew a **balloon** filled with hydrogen on a two-hour flight.

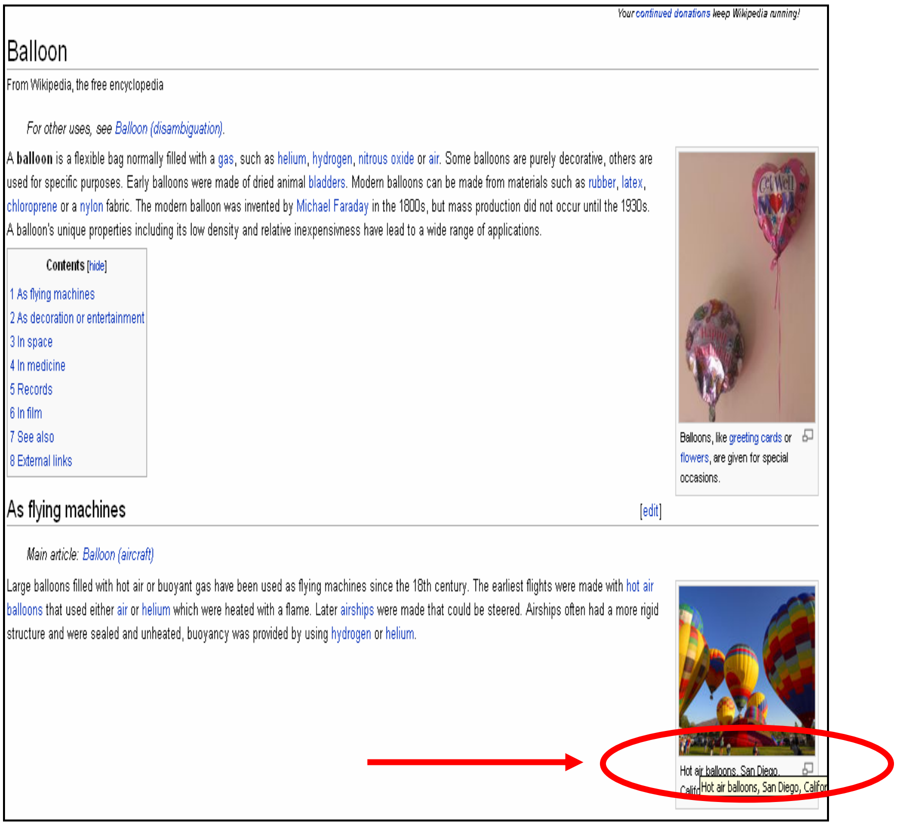
Photograph:Montgolfier balloon carrying Jean-François Pilâtre de Rozier and François ...

Fig. 79a *Alt Text* in Britannica

[53] *Creative Commons licenses* are several copyright licenses released in 2002 by *Creative Commons*, a U.S. non-profit corporation founded in 2001. Many of the licences grant certain "baseline rights" such as the right to distribute the copyrighted work without changes, at no charge.

[54] *Alt (Alternative) text* is an attribute in the HTML image tag. The text contained in the *alt text* attribute will appear in the web page if the image is broken or if the reader has turned off graphics. If the image is not displayed, the *Alt text* can be presented instead. It is especially useful to people using talking web browser. The text should be a brief representation of the purpose of the image, not a description of the image. *Alt text* is often displayed in tool tips when the mouse is moved over images.

As fig. 80 shows, an analysis carried out on a sample of ten encyclopaedic articles randomly chosen from Britannica and Wikipedia (the first one listed in each category has been selected) has proved that the Wikipedia makes a more generous use of images and sound files than Britannica.



Your continued donations keep Wikipedia running!

Balloon

From Wikipedia, the free encyclopedia

For other uses, see Balloon (disambiguation).

A **balloon** is a flexible bag normally filled with a gas, such as helium, hydrogen, nitrous oxide or air. Some balloons are purely decorative, others are used for specific purposes. Early balloons were made of dried animal bladders. Modern balloons can be made from materials such as rubber, latex, chloroprene or a nylon fabric. The modern balloon was invented by Michael Faraday in the 1800s, but mass production did not occur until the 1930s. A balloon's unique properties including its low density and relative inexpensivness have lead to a wide range of applications.

Balloons, like greeting cards or flowers, are given for special occasions.

Contents [hide]

1 As flying machines
2 As decoration or entertainment
3 In space
4 In medicine
5 Records
6 In film
7 See also
8 External links

As flying machines [edit]

Main article: Balloon (aircraft)

Large balloons filled with hot air or buoyant gas have been used as flying machines since the 18th century. The earliest flights were made with hot air balloons that used either air or helium which were heated with a flame. Later airships were made that could be steered. Airships often had a more rigid structure and were sealed and unheated, buoyancy was provided by using hydrogen or helium.

Hot air balloons, San Diego, Calif

Hot air balloons, San Diego, Califor

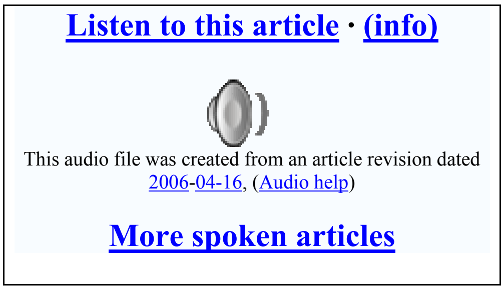
Fig. 79 b *Alt Text* in Wikipedia

Fig. 80 shows that 89 images (pictures, photos, graphs and maps) and 26 audio files have been found in Wikipedia sample articles whereas only 9 images and no audio files in Britannica's. By contrast, Britannica uses, although sporadically, videoclips which are rare in Wikipedia (and not found in the sample analyzed).

	Images		Audios		Videos	
	Brit	**Wiki**	**Brit**	**Wiki**	**Brit**	**Wiki**
Cinemascope	0	5	0	0	0	0
Beatles	2	13	0	25	4	0
Diaspora	0	0	0	0	0	0
Alcoholism	0	2	0	0	0	0
Barcelona	0	33	0	0	0	0
Anne Frank	1	9	0	0	0	0
Boolean algebra	0	5	0	0	0	0
Agnosticism	0	0	0	0	0	0
Aids	3	18	0	1	1	0
Balloon	3	4	0	0	1	0
TOTAL	**9**	**89**	**0**	**26**	**6**	**0**

Fig. 80 Multimedia in BAs vs. WAs

In addition to audio files (speeches, songs, etc.) there are about 800 complete *spoken articles* (up to 13th September 2007)[55] in Wikipedia; their function is to help the understanding of disabled people. When the icon below is found at the end of the page, it means that the article can be listened to.



Listen to this article · (info)

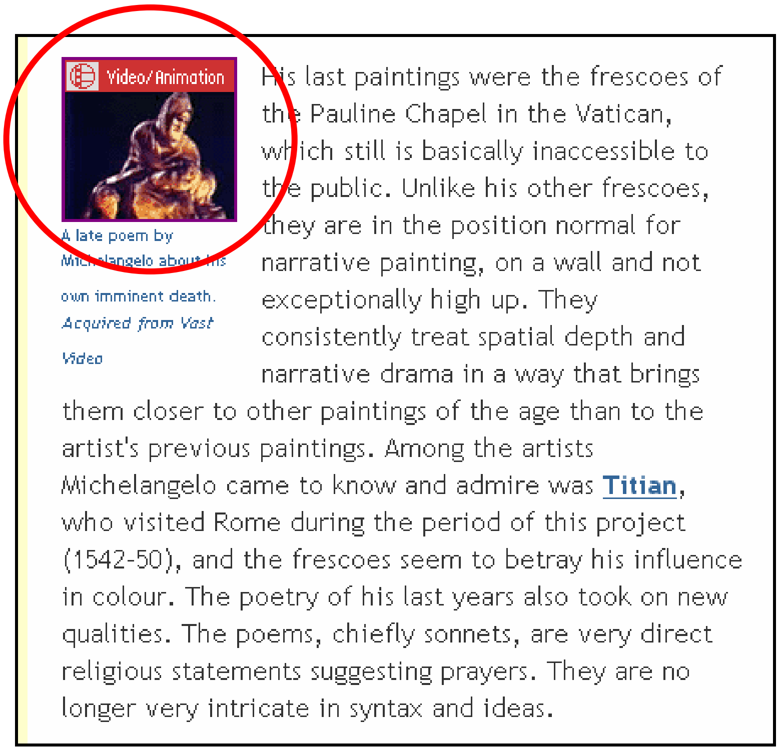
This audio file was created from an article revision dated 2006-04-16, (Audio help)

More spoken articles

Fig. 81 Spoken articles in Wikipedia

Wikipedia media files can be played on almost all personal computers. The software can be freely downloaded from the Internet. Wikipedian sound files generally use *Vorbis* audio format, while video files use *Theora* format. These are roughly similar to other formats used to play digital audio and video such as MP3 and MPEG. The difference is that they are completely free, open, and unpatented.

Music files occasionally use the MIDI format (.MID or .MIDI extension). On the other hand, videos are very rare in Wikipedia, while a collection of more than 2.000 video clips, taken from its archives, has been found in Britannica Online (fig. 82).



A late poem by Michelangelo about his own imminent death. *Acquired from Vast Video*

His last paintings were the frescoes of the Pauline Chapel in the Vatican, which still is basically inaccessible to the public. Unlike his other frescoes, they are in the position normal for narrative painting, on a wall and not exceptionally high up. They consistently treat spatial depth and narrative drama in a way that brings them closer to other paintings of the age than to the artist's previous paintings. Among the artists Michelangelo came to know and admire was **Titian**, who visited Rome during the period of this project (1542-50), and the frescoes seem to betray his influence in colour. The poetry of his last years also took on new qualities. The poems, chiefly sonnets, are very direct religious statements suggesting prayers. They are no longer very intricate in syntax and ideas.

Fig. 82 Video from Britannica's *Michelangelo* article

[55] *Spoken articles* http://en.wikipedia.org/wiki/Category:Spoken_articles

6. Comments and Remarks

A synopsis of the positive and negative features presented in this chapter, is shown in fig. 82 According to the Web Usability principles, the data objectively highlights that the number of positive elements is higher in Wikipedia (22 ☺ - 5 ☹ elements out of 27) than in Britannica (16 ☺ - 11 ☹ elements out of 27). If the Web Usability of the two encyclopaedic corpora could be conveyed just in an easily identifiable score, it could be quantified as 5 for Britannica and 17 for Wikipedia. This means that, although the linguistic analysis has proved that Britannica conveys a more formal expository encyclopaedic style, according to the basic web usability principles, both information transfer and web content fruition is more effective in Wikipedia than in Britannica.

Britannica and Wikipedia's linguistic features with a positive and a negative loading on formal expository style have been measured and compared in order to map and quantitatively define, the encyclopaedic informational production of the two encyclopaedias according to the selected linguistic classes. As fig. 60 shows [56] the variation between the two corpora, from a frequency perspective, is mainly due to a higher frequency of the linguistic classes with a positive loading (79.80% BAs vs. 76.10 WAs), being the total amount of the negative features the same in the two encyclopaedic corpora (7.50%).

Compared to Britannica, the less formal style of Wikipedia is surely due to the more massive and less highly educated mass of contributors, but mainly to a more informal style which is stylistically peculiar of the Web 2.0. In order to define Britannica vs. Wikipedia on the whole, in addition to the purely linguistic perspective, it has been considered equally important to take into account further categories more specific of webgenres, such as, i.e *Index of Readability*, *Web Usability* and *Multimediality* which have a primary weight in defining the total perception of online encyclopaedias as a new webgenre.

Fig.82 highlights the positive and the negative elements found in the two encyclopaedic corpora, according to the basic principles of Web usability. They are definitely higher in Wikipedia than in Britannica (17 ☺ in Wikipedia vs. 5 ☹ in Britannica). Furthermore, technological advantages offered by collaborative wiki software, reinforce the variety and the high informativeness, allow easiest browsing mechanisms, the social editing and tagging (folksomy) and finally the quick updating and interlinking of the information provided by the international multitude of Wikipedian contributors.

[56] *see* chapter 4, section 3

Britannica vs. Wikipedia - Web Usability				
	Britannica		Wikipedia	
Explicit language	☺		☺	
Index of Readability	☺		☺	
Front load content	☺		☺	
Sections and subheadings		☹	☺	
Descriptive subheadings		☹	☺	
Paragraphing	☺		☺	
Lists and bullets		☹	☺	
Text alignment	☺		☺	
Web font	☺		☺	
Font size	☺		☺	
Typographic emphasis (bold)		☹		☹
Italics	☺			☹
Correct use of underlining	☺		☺	
Font colour	☺		☺	
Link colours		☹	☺	
Correct use of capital letter	☺		☺	
Line spacing	☺		☺	
Indentation		☹		☹
Text line length	☺			☹
Page length		☹	☺	
Article length		☹	☺	
Content interlinking		☹	☺	
Search engine	☺		☺	
Background texture	☺		☺	
Images		☹	☺	
Audio		☹	☺	
Video	☺			☹
Total	**16**	**11**	**22**	**5**
Final Score	**5**		**17**	

Fig. 83 Web Usability in Britannica vs. Wikipedia

5. WIKIPEDIA AS VIRTUAL COMMUNITY: FINDINGS

1. Wikipedia: a Community of Practice

Wikipedia is not merely an online encyclopaedia; while its website is useful, popular, and allows anyone to contribute, the site is only the most visible artifact of an active community. Unlike previous reference works which stand on library shelves far from the institutions, people, and discussions from which they arose, Wikipedia is a community and the encyclopaedia is a snapshot of its open ended contribuiting interactions. These interactions reflect and, of course, shape the Wikipedia culture. Thus, the term Wikipedia can be applied to three things: an encyclopaedia (the actual body of work), a project (the effort to make that encyclopaedia) and a community (the group of people working on the project). Wikipedia being based on wiki software, provides an excellent collaborative environment and it represents an efficient model of *Community of Practice* (henceforth CoP). The concept of CoP refers to the process of social networking that occurs when people with a common interest in some subject or problem, collaborate to share ideas, find solutions, and build innovations. More recently, CoPs have become associated with knowledge management as people have begun to see them as ways of developing social capital, promoting new knowledge, stimulating innovation, or sharing existing tacit knowledge within an organization. Nowadays, CoPs are officially accepted as new organizational development prototypes.

Wenger (1998), in defining the idea of practice, which is the basic issue of his theory on community, uses three fundamental concepts: *negotiation*, *participation* and *reification*.

Negotiation refers to the process of dynamic construction of meaning which does not exist autonomously, but is the result of a continuous interaction with the world. P*articipation* means the belonging to communities and the active involvement in social projects while *reification* refers to the process which shapes our experiences producing objects which concretise each experience in what he defines as *thinkness.* Wenger (1998:10) in describing participation argues that:

> If we believe that people in organisations contribute to organisational goals by participating inventively in practices that can never be fully captured by institutionalised processes [...] we will have to value the work of community building and make sure that participants have access to the resources necessary to learn what they need to learn in order to take actions and make decisions that fully engage their own knowledgeability.

The third concept, *reification,* is the central process in every practice. It involves taking what is abstract and turning it into a "congealed" form, represented for example in documents and symbols (in this specific case study, in encyclopaedic articles). Crucially, Wenger describes the relationship between reification and participation as dialogical: no element can be considered in isolation if the process is to be fully understood. He claims (1998:67):

> Explicit knowledge is not freed from the tacit. Formal processes are not freed from the informal. In fact, in terms of meaningfulness, the opposite is more likely [...] In general, viewed as reification, a more abstract formulation will require more intense and specific participation to remain meaningful, not less.

Wenger defines the successful interaction between reification and participation as the *alignment* of individuals with the communal task. Alignment, he claims, *requires the ability to co-ordinate perspectives and actions in order to direct energies to a common purpose*. The challenge of alignment, Wenger suggests, is to connect individual efforts to broader styles and discourses in ways that allow participants to invest their energy in them.

Virtual Community of Practice (henceforth VCoP) is sometimes considered a misnomer as the original concept of CoP is based around a co-located setting. However, the increasing globalization and the exponential growth of the Internet has now led to the acceptance of virtual CoPs. For example, a wiki environment (such as Wikipedia) can be definitely considered as a virtual CoP, or more precisely as a *Community of Purpose* since *Wikipedians*[57] go through the same process, trying to achieve a similar objective. Members of the community assist each other by sharing experiences, suggesting strategies and exchanging information on the process in hand.

A “real” discourse community denotes a group of people with certain things in common: a public goal, a body of specialized knowledge, the use of a specialized lexicon, and a set of beliefs about how knowledge is generated. Members also share an understanding of how to communicate with each other and with the larger community.

To become a member of a discourse community, one must master its theoretical concepts, as well as its language and conventions. This usually means accepting also its beliefs and values. Wikipedia, as a VCoP, shares all its distinctive references with offline CoP. Swales (1990:21-29) defines the concept of discourse community through the six following features:

1. it has a broadly agreed set of common public goals;
2. it has mechanisms of intercommunication among its members;
3. it uses its participatory mechanisms primarily to provide information and feedback;
4. it utilizes and hence possesses one or more genres in the communicative furtherance of its aims;
5. it has acquired, in addition to owning genres, some specific lexicon;
6. it has a threshold level of members with a suitable degree of relevant content and discoursal expertise.

The above mentioned prompts are totally shared by the Wikipedia community. First of all, it has a very definite and pragmatic goal, which is to build a multilanguage encyclopaedia (see point 1), furthermore, its several synchronous and asynchronous intercommunication channels (talk pages, mailing lists, chat, etc.) ensure a high and distributed participation to the community events (see points 2-3). Two different genres have been identified in Wikipedia: the *informational* (encyclopaedic articles) and the *involved production* (talk pages) (see point 4). The community has also developed its specific lexicon, which has been defined in this study as *WikiSpeak Jargon* (see point 5). Furthermore,

[57] *Wikipedians* are the people who write and edit articles for Wikipedia. It has been suggested that *Wikipedist* would be a more appropriate name, as an encyclopedist is someone who contributes to an encyclopedia. *Wikipedian*, though, suggests being part of a group or community.

Wikipedia contributors are shown to have a good level of discourse expertise since they properly use the acquired linguistic competences in the different writing spaces. They use the conversational and unconventional WikiSpeak when they interact in talk pages, while code switching towards a more formal encyclopaedic expository style has been recorded when encyclopaedic articles are written. Here, the official prescriptive rules of the official Manual of Style (see point 6) are strictly observed.

According to Patricia Bizzell (1992), producing text within a discourse community cannot take place unless writers can define their goals in terms of the community's interpretive conventions. In other words, texts cannot be simply produced. They must fit the standards of the discourse community to which they are appealing. For this reason, being a member of a specific discourse community (such as Wikipedia) requires more than just learning its lingo. It requires understanding concepts and expectations within that specific community and acting by precise behavioral norms.

The language used in discourse communities has been defined by Gregory (1967) as *diatype* (Wikipedia, 2007). This term describes a type of language variation which is determined by its social use and purpose. According to Halliday (1985: 12), diatype is usually analyzed in terms of *field* (the subject matter or setting), *tenor* (the participants and their relationships), and *mode* (the channel of communication: spoken, written, or mediated).

Online discourse communities are virtual spaces where people interact with one another mainly by means of written discourse which can take place in synchronous and asynchronous CMC channels such as emails, mailing lists, forums, chats, multi-user virtual games, MSN Messenger, or in the more recent web 2.0 environments such as blogs, wikis, or virtual worlds (e.g. *Second life*[58]*)*. These virtual environments are primarily text-based, but can also be multimodal, since elements such as images, sounds, animation, or emoticons can be co-conveyed.

Wikipedia as VCoP has developed its personal *Computer Mediated Discourse* (CMD) with its peculiar *wired style*. It is thus possible to define the Wikipedia community as unique being a free open content and an encyclopaedia project, since no other community out of the Wiki world combines the above attributes.

The community role, as a sort of Science Fiction super-entity, is to organize and edit individual pages, to structure navigation between pages, to resolve conflicts among individual members and to create rules and patterns of behaviour.

[58] *Second Life* http//secondlife.com is an Internet-based virtual world launched in 2003, developed by Linden Research, Inc. It came to international attention in late 2006 and early 2007. *Residents* interact with each other through motional avatars, providing an advanced level of a social network service combined with general aspects of a metaverse. Residents can explore, meet other *residents*, socialize, participate in individual and group activities, create and trade virtual properties and services from one another.

Part of an interview with Jimbo Wales, the founder of Wikipedia, is reported to explain the main goals of the Wikipedia community[59].

> Wikipedia is first and foremost an effort to create and distribute a free encyclopedia of the highest possible quality to every single person on the planet in their own language. Asking whether the community comes before or after this goal is really asking the wrong question: *the entire purpose of the community is precisely this goal.* I don't know of any real case where there is a genuine strong tension between these two things, either. That is to say, the central core of the community, the people who are really doing the work, are virtually all quite passionate on this point: that we're creating something of extremely high quality, not just goofing around with a game of online community with no purpose.
> *The community does not come before our task, the community is organized *around* our task.* The difference is simply that decisions ought to always be made not on the grounds of social expediency or popular majority, but in light of the requirements of the job we have set for ourselves. I do not endorse the view, a view held as far as I know only by a very tiny minority, that Wikipedia is anti-elitist or anti-expert in any way.
> If anything, we are *extremely* elitist but anti-credentialist. That is, we seek thoughtful intelligent people willing to do the very hard work of getting it right, and we don't accept anything less than that. PhDs are valuable evidence of that, and attracting and retaining academic specialists is a valid goal. There may be some cases of PhDs who think that no one should edit their expert articles, but there are many many more cases of completely unqualified people who think the same thing. It doesn't matter: if someone can't work in a friendly helpful way in a social context, that's a problem for them and for us, and we'll always have to make some very complex judgments about what to do about it. I'm 100% committed to a goal of "Britannica or better" quality for Wikipedia, and all of our social rules should revolve around that. Openness is indispensable for us, but it is our *radical* means to our radical *ends*.
>
> --- Jimbo

Thus, it is evident that members and community are strictly intertwined entities working in tandem. Without members, there would be no community, and no material for the encyclopaedia, but without the community the individual contributions would be meaningless and without context.

Democracy is another important phenomenon which characterizes the Wikipedia community. It is strongly required of contributors not to pay attention to their degree of education, economic status and level of experience, when dealing with human knowledge. This produces a unique egalitarian situation.

The Wikipedia community is engaged in a serious collaborative task. Nevertheless, as most workplaces, it also has its relaxed moments. Like in real life, some people choose to extend the relationships from workplace to an outside context. Thus, also *Wikipedians* metaphorically stop at the *bar* with their workmates and have a few beers. Here, they may joke about situations on the job and talk about their personal lives. Their meeting points are online spaces such as talk pages, mailing lists, edit summaries, user talk pages, person-to-person meetups, private email, IRC chat rooms, etc.

As already outlined, Wikipedia and all the wikis around the world, organize their communities around a written project. Obviously, the specific topic of interest linguistically shapes the participants' style around a common community discourse. Hence, how topics are named and clustered on the wiki indicate what its culture is about, what its values are. For instance, on the entry page of the *Meatball*

[59] *Jimmy Wales* (8 March 2005) *Wikipedia is an encyclopedia* http://lists.wikimedia.org/pipermail/wikipedia-l/2005-March/020469.html

Wiki (an interwiki community), it is specified that it deals with online cultures, especially with how people online come together naturally in groups. The *Portland Pattern Repository* 'WikiWikiPage' explains that it is a web site written by its users, where anyone can change any page or create new pages. By contrast, the *Wikipedia homepage* welcomes its visitors with the following slogan: *Welcome to Wikipedia, the free encyclopedia that anyone can edit,* and the following message has also been recently added:

You can help Wikipedia change the world!

Moreover, inside its website, the *Community Portal* has a high impact on the reader in terms of content, form, and function. It can be considered a supreme synthesis of its free and collaborative background philosophy. It is not coincidental that the highest frequency keywords recorded in this page are: *help* (22 occurrences), *you* (22 occurrences), *article* (18 occurrences), *collaboration* (8 occurrences) and *free* (7 occurrences)"[60].

2. Who are Wikipedians?

The Wikipedia community embraces all editors, ideological supporters, current and even potential readers of all the different Wikipedia's editions, while a narrower definition includes only Wikipedia's contributors. Differently from other online communities the Wikipedia community is multicultural as contributors come from all over the world. Specifically, with reference to the English edition of Wikipedia, Wikipedians are English speaking contributors, mainly from English speaking countries and from those nations where English is the most commonly spoken foreign language.

Wikipedians attempt to understand each other, despite differences in languages, backgrounds, traditions, ethnicities, different cultural approaches and interests. Thus, an intense cross-cultural communication takes place in this world-wide virtual community, which is also heterogeneous since its members, all rigorously volunteers are philosophers, historians, scientists, artists, religious people, specialists, scholars, experts, and also ordinary students and anonymous contributors.

Nowadays, the number of Wikipedians has grown to over 5 million in addition to an unknown large number of unregistered contributors (Wikipedia, 2006). The diversity of Wikipedians renders it nearly impossible to make categorical statements about Wikipedians as a whole. For istance, some of them upload images, some work on humanistic or scientific articles, some clean up grammar, others work on reverting vandalism. Some create new pages or *refactor* old pages, add or correct

[60] This data refers to April 2006 version of the English Wikipedia Community Portal
httalk page://en.wikipedia.org/wiki/Wikipedia:Community_Portal

information, and discuss the nature of the content with other users. Many take on all of these tasks. What Wikipedians definitely have in common is an active commitment in the project's promotion and a strong feeling of belonging to the community. Information on registered Wikipedians can be found on *user pages*. However, it is not compulsory for Wikipedians to have a page of their own, many of them prefer to remain anonymous.

Although the community's goal is to create encyclopaedic articles which have to be objective and possibly without personal biases, the openness of Wikipedia allows total self-expression, as Wikipedians define themselves within the context of the project through their personal interests and cultural goals. Wiki community is knowledgeable and, at the same time, fragile. Its success depends to a large extent on the presence of open-minded and well-informed contributors. If these human qualities are not to be found in its members, the project loses much of its appeal.

The individual commitment of Wikipedians involves two main tasks: writing articles and participating in the community life. A considerable amount of communication and collaboration is needed inside the community, since the purpose of the community is very specific and pragmatic: to create and distribute a free encyclopaedia of the highest possible quality to every single person on the planet in their own language. Since Wikipedia is not a community in the "real world" sense, Wikipedians are bound together mostly by electronic interactions. The community is defined by what exists on the Wikipedia website, and in particular through what is conveyed in the written exchanges carried out in the different community channels (Mailing Lists, IRC, channels, User Pages, etc.) and mainly in talk pages, where commentaries and negotiations are meant to improve the quality of encyclopaedic articles. Nevertheless, regular international face-to-face meetings of Wikipedians (Wikimania conferences [61]), as well as different local meetings (more spontaneous and informal) take place in cities around the world every year.

3. What is the Wikiquette?

Netiquette is a blended word (*network* + *etiquette)* which defines the conventions of politeness and respect recognized in virtual communities. In other words, it is the term which outlines a dynamic set of guidelines for conduct which encourages a pleasant, efficient and agreeable interaction within online communities. Netiquette rules are slightly different in the plethora of existing virtual communities. In this specific case study, the rules and patterns of behavior are outlined in the *WikiQuette* (*wiki* + *etiquette*) (Wikipedia, 2006), which also spells out the *guidelines* of how to deal and work with other Wikipedians. *Wikiquette* rules are more explicit, meticulous and compulsory than precepts of general forums. This happens for different reasons. First of all because a ubiquitous

[61] *Wikimania* http://en.wikipedia.org/wiki/Wikimania is a conference for users of the wiki projects operated by the *Wikimedia Foundation*. The first conference was held in Frankfurt, Germany (August 4–8, 2005); the second ran in Cambridge, Massachusetts, USA (August 4–6, 2006) and the third conference was in Taipei, Taiwan (August 3–8, 2007). Here speakers present studies and experiments on Wikipedia and other projects operated by the Wikimedia Foundation, on wiki culture and technology.

moderator in talk pages is not contemplated, and secondly because *Wikiquette* has to manage a complex process: the writing of a co-authored encyclopaedia. Wikipedia contributors come from many different countries and cultures. They have different points of view and backgrounds. Treating others with sensitivity and respect is the key for avoiding intercultural misunderstanding and for collaborating effectively in building an encyclopaedia. The basic conventions, in force, are included below (Wikipedia, 2007).

WIKIQUETTE RULES

- ***Assume good faith.*** Wikipedia has worked remarkably to assure free editing. People come to WIkipedia to collaborate and write good articles.
- Remember the ***Golden Rule***: Treat others as you would have them treat you – even if they are new.
- ***Be polite***. Keep in mind that raw text is ambiguous and often seems ruder than the same words coming from a person standing in front of you. Irony isn't always obvious, text comes without facial expressions, vocal inflection or body language. Be careful of the words you choose – what you intended might not be what others perceive, and what you read might not be what the author intended.
- ***Sign and date*** your posts to talk pages (not articles!)
- ***Register yourself,*** do not construct a signature
- ***Work toward agreement.***
- **Argue facts, not personalities.**
- **Don't ignore questions.** If another disagrees with your edit, provide good reasons why you think it's appropriate. Concede a point when you have no response to it, or admit when you disagree based on intuition or taste.
- ***Be civil and be prepared to apologize***. In animated discussions, we often say things we later wish we hadn't. Say so.
- ***Forgive and forget***.
- ***Recognize your own biases and keep them in check.***
- ***Give praise when due.*** Everybody likes to feel appreciated, especially in an environment that often requires compromise. Drop a friendly note on users' talk pages.
- ***Remove or summarize resolved disputes that you initiated.***
- ***Help mediate disagreements between others.***
- ***If you're arguing, take a break.*** If you're mediating, recommend a break. If you're angry, take time out instead of posting or editing. Come back in a day or a week. You might find that someone else has made the desired change or comment for you. If no one is mediating, and you think mediation is needed, enlist someone. Walk away or find another Wikipedia article to distract yourself – there are 1,954,451 articles on Wikipedia!
- ***Remember what Wikipedia is not.***
- ***Avoid reverts and deletions whenever possible***. and stay within the three-revert rule except in cases of clear vandalism. Explain reversions in the edit summary box. Amend, edit, discuss.
- ***Remind yourself that these are people you're dealing with.*** They are individuals with feelings and probably have other people in the world who love them. Try to treat others with dignity. The world is a big place, with different cultures and conventions. Do not use jargon that others might not understand. Use acronyms carefully and clarify if there is the possibility of any doubt.

4. Cyberlanguage: from Web 1.0 to Web 2.0

Scholars from different research areas try to give an answer to the following questions: Who speaks online, and how? Is online language *only* text, or is it a *discourse*? To what extent culture affects the language of cyberspace? Thus, approaching these questions from different disciplinary perspectives, cyberlanguage can be variously defined as text, semiotic system, sociocultural discourse, etc. According to Mardziah Hayati (1998) computer networks are changing the way people think and interact. They are redefining the spatial and temporal parameters of the interaction they mediate and online discourse is taking new directions, particularly in the way people write. He focuses on the differences in style and tone between electronic discourse and traditional academic prose.

One important observation made by a number of scholars is that new conventions are evolving and *blurring the past distinctions between writing and talking* (Tornow, 1997:1). Tornow describes the written interaction that occurs in electronic mail and on-line courses as a kind of *written talk*, while Boyd and Brewer (1997:2) use the term *electronic discourse* to refer to written talk *writing that stands in place of voices*. Most scholars generally conclude that online communication is an intermediate stage between oral and written modalities.

Electronic discourse is a relatively new form of discourse with its own peculiar features. On the one hand, it is like conversation in that it presents a number of performance features generally characteristic of in process communicative events and behaviors, such as repetition, direct address and markers of personal involvement, including syntactic and lexical items (Boyd and Brewer, 1997).

On the other hand, since CMC is primarily a written form of communication several authors have focused on the features of digital text. A particular area of interest has been the development of hypertext, whose non-linear, non-sequential, non-hierarchical and multimodal nature (employing images, sound and symbols as well as text) seemed to be in contrast with traditional printed texts.

Many papers discuss evolving conventions in CMC and in particular they analyse its linguistic, pragmatic features, as well as its grammatical, lexical and syntactical aspects. Crystal (2001) also explores s the language of the Internet in depth.

Since the very beginning, each communication technology has reshaped the process of socialization and acculturation, simultaneously changing its discursive horizons. Nowadays, this process has become more complex as the Internet is collapsing in a multitude of new technologies which are emerging very rapidly, creating several distinctive Computer Mediated Discourse Communities. Like forums and blogs, also Wikipedia has developed its peculiar *wired style*, which is a direct consequence of the evolution of a ‘specialized online discourse’ connected to the use of specific webtools. Nowadays Web 2.0 and its philosophy of *Reading/Writing culture* is challenging the more traditional first web generation with its *Read Only* hypertexts and its synchronous and asynchronous CMC tools. Most readers of Internet content have almost no opportunity to create or modify online text, since only a limited number of authors or producers control both content selection and presentation. Recently, new forms of hypertext, such as *wikis,* have blurred the net distinction between author and reader, producer and consumer of online text (Graddol, 2004). While it has already been understood that reading involves the production of meaning, new open-access technologies allow multiple reader-authors to register different interpretations and analysis directly within text, and participate in a new and dynamic collaborative process of co-construction of meaning. As Braga and Busnardo (2004) suggest, these latest developments offer an entire new challenge to communicators, greater than the simple navigation of non-linear texts. Readers will increasingly face multiple-authored texts that exist in a condition of constant change, a situation that is radically challenging existing notions of how knowledge is produced, accessed and divulged.

5. Wikipedia languages: Register Variations and Wikispeak

What happens in the Wikipedia community? What linguistic peculiarities can be identified in the Wikispeak? Although Web 1.0 CMC channels have been extensively investigated in several studies e.g. in Crystal (2001), Herring and Paolillo (1999-2007) studies, etc., linguistic properties of web 2.0 environments, such as blogs and especially wikis, have not yet been systematically investigated.

Similarly to other environments, Wikipedia's contributors also interact in their backstage community by using a hybrid of spoken and written language which has been defined in this study as the *WikiSpeak*. Thus, Wikipedians have developed their own community discourse with its original lexicon by making fresh linguistic adaptations to suit new online circumstances. In this way, a new variety of *NetSpeak Jargon* has come to life. *WikiSpeak* can be considered an unofficial and high-context digital jargon used in the different synchronous and asynchronous CMC channels of the Wikipedia's community.

The aim of the sections which follow is to identify linguistic properties and code switching between *WikiLanguage* an *WikiSpeak* through a comparative analysis. Specifically, in order to identify register variations, talk pages have been analyzed.

As shown in the previous chapter, according to the data and theoretical model proposed by Heylighen and Dewaele (1999) and Biber (1998), formal communication conveys information explicitly, through the linguistic expression. Following a frequency perspective, the empirical measurement of encyclopaedic expository style has demonstrated that Britannica and Wikipedia have in common similar linguistic features, although those ones having a positive incidence on formal expository style are most of the times higher in the Britannica corpus [62]. This data confirms a slightly superior formality of the expository style of the latter. The *Index of Readability* proved to be similar in both corpora[63], by contrast, *Web Usability* has shown to be, decisively, in favour of Wikipedia.

According to Hymes (1972), people should not only know the language, but they should also possess the knowledge derived from the acquired social and cultural experience that may determine, for instance, when to talk, and when not to talk, what to say, to whom, h*ow and in what way.* A community can be defined according to the concept of linguistic competence. A mere group becomes a discourse community when all its members share the same linguistic and communicative competence. Wikipedia, as already claimed, can be considered not only an encyclopaedia, but also a

[62] *see* chapter 4, section 1-2-3
[63] *see* chapter 4, section 5.1

discourse community since all its members share knowledge of the language and specific communicative competences. It will be shown in the following sections, how Wikipedians carry out different roles and show different linguistic competence, inferred from their language performance and conveyed in the different writing space in which they are involved. As Noblia (1998) claims:

> Online communities take shape, generate norms of interaction (for examples rules of network etiquette or netiquette) and conflict resolution procedures. Virtual communities, like communities in real life, protect the interests of its members, and ethical dilemmas result when individual and groups needs come into conflict, as well as certain groups dominate in defining the terms of the discourse.

According to the concept of linguistic competence, Wikipedians can be formal encyclopaedic contributors and informal community's participants. In the first case, when they are involved in the process of collaborative writing, they write, change, edit and improve the articles in the document mode pages; by contrast, when they are informal partakers they speak/write with other Wikipedians in the associated talk pages. The different roles, directly affect their linguistic utterances.

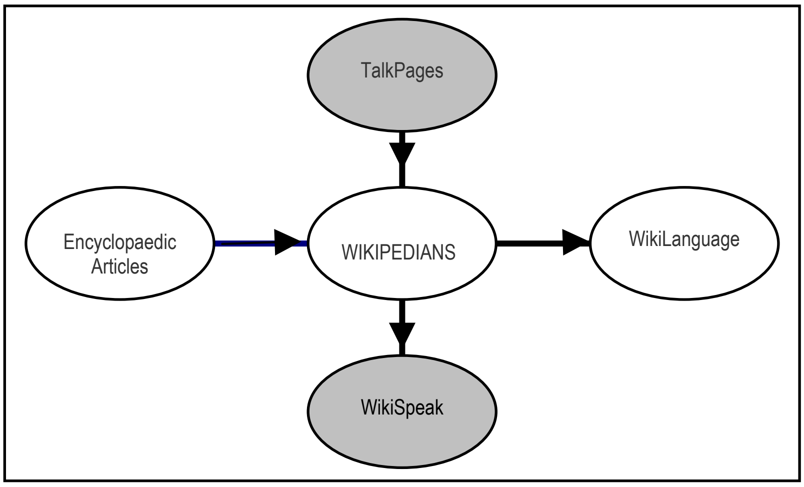


Fig. 1 WAs and Wikilanguage vs. TPs and Wikispeak

With reference to the linguistic style, the opposite of *formality* is defined by Heylighen and Dewaele (1999) as *contextuality*. Their theoretical model suggests that *contextuality* decreases when unambiguous understanding becomes more important and when the separation in space, time or background between actors (writers and readers) increases. The application of this theoretical model to this study has demonstrated how Heylighen and Dewaele's assumptions work well when specifically applied to the different writing spaces of the Wikipedian community: the "front office writing space", that is to say *encyclopaedic articles*, and the back office interactive space represented by *talk pages.*

In talk pages, communication is contextual and often conveys information implicitly, through specific terminology or in apparently encrypted expressions. Every linguistic act refers to the context to some degree (Heylighen and Dewaele, 1999) but in some situations context obviously plays a much central role than in others.

From the anthropological point of view, Hall (1976) has distinguished *high context* and *low context* situations. Communication is explicit and overt in low context situations, stating the facts exactly and in detail (as it happens in the encyclopaedic articles); by contrast, communication is implicit in high context situations, and information is conveyed more by the context than by the verbal expression. Although Hall introduces this concept to distinguish different types of cultures[64], nevertheless, the same distinction can be applied to different communicative contexts also in the case of CMC.

Discourse used in specific situations, both in offline and virtual communities, will appear less ambiguous and comprehensible if beliefs, philosophy, values and the specific lexicon and jargon in use are shared by the members of the community (Duranti & Goodwin, 1992). When compared to the formal style of the *WikiLanguage* conveyed in encyclopaedic articles, *Wikispeak* contextual speech style is more interactive and involving, as it is the result of an immediate reaction to the interlocutors' statements, events or other elements of the context, rather than being a description of things through a detached, impersonal and objective style.

The Community's background assumptions and the essential role of context is fundamental in resolving semantic ambiguity and in understanding the language in use. In formal language things must be expressed explicitly in order to avoid ambiguity, whereas, according to Grice (1975) in natural languages they will be conveyed by *implicatures*. He coined this term to refer to a shared framework and its implications. He points out that if one takes into account the shared context, expressions which appear ambiguous or nonsensical become clear and logical.

The application of Hall's theoretical model to the different Wikipedia writing spaces (document mode encyclopaedic articles vs. thread mode talk pages), reveals empirical variations between what has been defined as *WikiLanguage* and *WikiSpeak*. This allows the definition of different registers and variations between the low context pole (the formal encyclopaedic expository style, the *Wikilanguage*) and the high context pole (the contextual and informal *WikiSpeak*). In order to measure the degree of *contextuality* and *informality* of talk pages, the same methodology based on frequency criteria which was used to analyze the formality of the encyclopaedic expository style will be applied.

While the previous analysis has been based on an intra-genre comparison (*Wikipedia* vs. *Britannica*) this investigation will involve an inter-genre contrastive analysis whose purpose is to compare register variations inside different writing spaces of the Wikipedia community. Specifically,

[64] Hall considers American and Northern European cultures typically as low context, while Mediterranean and Eastern cultures high-context. The terms "high context" and "low context" were coined by Edward Hall (1976) to describe cultural differences between groups, societies and communication systems. In particular, the term "high context" refers to groups where people have strong interpersonal connections and share much common knowledge and presuppositions, thus, many aspects of cultural behaviour are not made explicit since community members have an internalized understanding of what is communicated. "A high-context communication or message is one in which most of the information is either in the physical context or internalised in the person, while very little is in the coded, explicit, transmitted part of the message. A low-context communication is just the opposite, i.e. the mass of information is vested in the explicit code" (Ibid, 91). In high-context cultures, knowledge is situational and relational and decisions and activities focused around personal interactions.

the analysis will compare the Wikipedia encyclopaedic corpus (which in the previous chapter was compared to Britannica corpus to measure the stylisitc formality of the encyclopaedic expository style) with the associated talk pages.

The Wikipedia encyclopaedic corpus is made up of 391,637 tokens, whereas the talk pages corpus counts 601,745 tokens. Just comparing the length of the two corpora, the first interesting data which emerges is that discussions involve an extensive "verbal" commitment of Wikipedian contributors as they, in total use 1.5 more words to discuss, than to write the associated encyclopaedic articles. In the previous section and according to a frequency perspective, the encyclopaedic expository style was shown to be formal, static and rigid and very similar in both Britannica and Wikipedia corpora. By contrast, WikiSpeak contextual speech, as will be shown in the next section, is more flexible, personal and uses a more informal register.

6. Forums vs. Talk pages: Differences and Similarities

A comparison between forums and talk pages will be outlined in this section.

In following discussions in talk pages it is possible to understand to what extent the process of collaborative writing is different from the traditional individual writing: here the importance of the negotiation visibly emerges as the essential element of writing literacy in hypertext environments.

The core of democracy, represented in Wikipedia by *talk pages*, can be associated, to a certain extent, to online forums, even if talk pages usually tend to be meta-discussions on the topic or on the validity of the content, rather than on the content itself.

The process of linguistic accommodation is common to both Internet Forums and Wikipedia talk pages. David Crystal (2001:147) in observing the dynamics in WELL Community[65], notices a linguistic adjustment in forums and claims:

> The members accommodate to each other. Although they come from many different backgrounds, and write in many different styles, their contributions progressively develop a shared linguistic character. Everyone comes to use certain types of grammatical construction, jargon, or abbreviations.

This phenomenon has also been observed in talk pages where contributors communicate, exchange opinions and discuss technical and editing operations. In forums, users may post chronological messages, individually answering a discussion thread. Text, images or audio files may be provided, in most cases, as attachments. The meaning of the forum speech act is aggregative, since it accumulates ideas and phrases. Quoting is a method used to create meaning through aggregation. Ong (1982) claimed that speech acts are redundant in oral discourse; it is necessary to repeat previous

[65] *WELL* http://www.well.com stands for 'Whole Earth Electronic Link'. It is a Community which was founded in 1985. It had more than 250 groups by mid 2000 referred to as conferences.

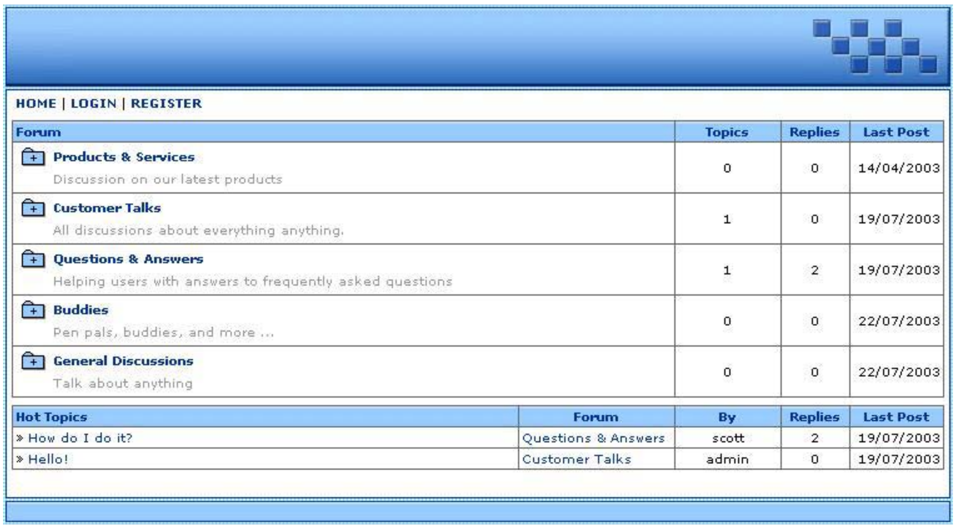
messages in most of the forums both to remind the reader and to inform new readers of what has been said in the discussion. Thus, as December pointed out (1993:10-13): *the online quoting process acts as a repetition in oral speech.* This phenomenon has not been detected in talk pages where the different architectural structure allows contributors to immediately add their post in a new paragraph, avoiding to quote the contribution they are replying to. The last post will be the first one. In the case of long discussions, a dynamic *Table of Contents* is added by contributors at the beginning of each page. Its aim is to allow a quick retrieval of the desired topical discussion. Unlike forums, files may never be attached in talk pages. In the same way as for talk pages, the range of topics discussed in forums is very extensive, as a website running forums may have more than one *room*, each dedicated to a different topic. Talk pages and forums usually provide anyone with the authority to start a new discussion (known as a *thread*), or to reply to an existing thread. But, whereas talk pages allow every visitor to post comments in reply, the number of contributors who can post in forums is usually limited and, consequently, so is the range of viewpoints and beliefs expressed.

Both in forums and talk pages, participant posts are public for many years, especially when an archive is provided. For example, *Google Groups* includes Usenet articles dating back to 1981. Nevertheless, while posts are permanently stored in talk pages, sooner or later forum messages will be deleted.

Furthermore, the architecture of the two environments is different. In most cases forum interventions are organized in separate messages posted in topical rooms (fig. 2); they are chronologically ordered and can be clicked on to expand the content. By contrast, posts by Wikipedia contributors are published in a unique and long scrollable talk page (fig. 3). It is a sort of ‘backstage agora’, a public and interactive written open space whose topic of discussion is associated to the main encyclopaedic article. Talk pages are accessible by everyone and they allow people to edit other contributors’ messages, whereas in forums only moderators may be authorized to do so in order to monitor and control content. In this way moderators can control spams, add details and comments. The presence of moderators and their behaviour can shape tendencies, address issues, and set the tone of the discussions according to a more formal or informal style. David Crystal (2001:132), referring to forums, claims:

> the more specialized is the topic the more likely the content will be focused and several groups use moderators to ensure that the conversation doesn’t diverge from the subject too much (go off-topic).

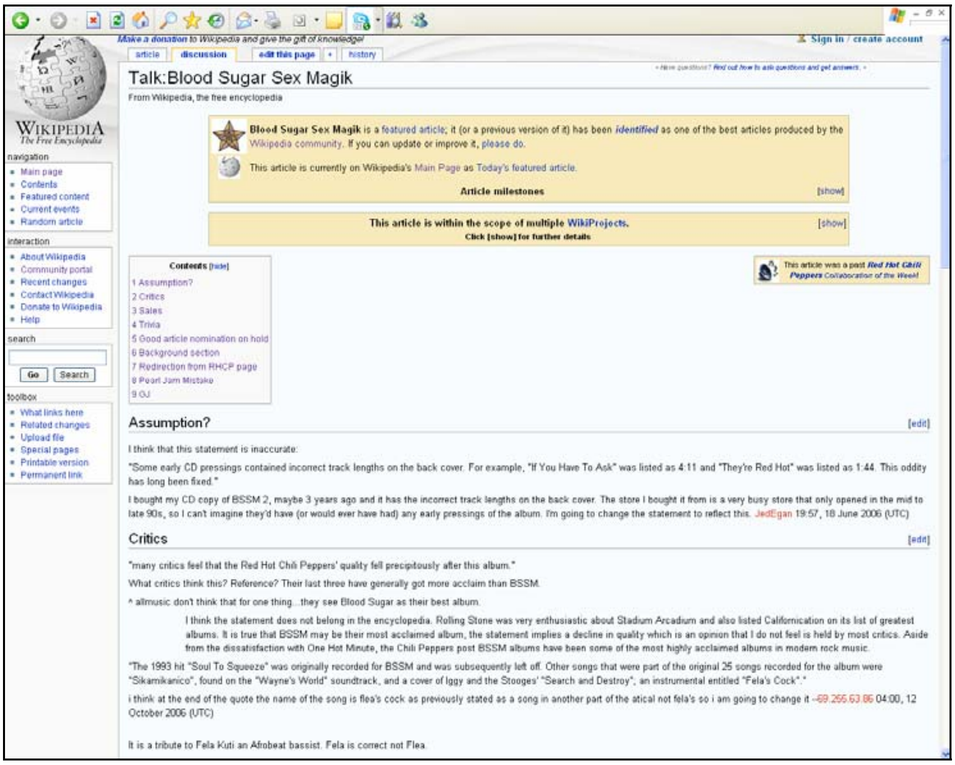
By contrast, official moderators are not provided on talk pages. This role, as already mentioned, has been replaced by a democratic mechanism of consensus, mutual control and peer to peer review. Nevertheless, one of the Wikipedia community's role, as a super-parties entity, is to resolve conflicts among members, if and when they happen, through the *ArbCom* (Arbitration Committee). The *ArbCom* only deals with the most serious disputes and cases of rule-breaking and it imposes binding solutions to Wikipedian disputes. It is the last step in the dispute resolution process, the final resort to be turned to when all else has failed.



Forum	Topics	Replies	Last Post
Products & Services Discussion on our latest products	0	0	14/04/2003
Customer Talks All discussions about everything anything.	1	0	19/07/2003
Questions & Answers Helping users with answers to frequently asked questions	1	2	19/07/2003
Buddies Pen pals, buddies, and more ...	0	0	22/07/2003
General Discussions Talk about anything	0	0	22/07/2003

Hot Topics	Forum	By	Replies	Last Post
» How do I do it?	Questions & Answers	scott	2	19/07/2003
» Hello!	Customer Talks	admin	0	19/07/2003

Fig. 2 An example of forum



Talk:Blood Sugar Sex Magik

From Wikipedia, the free encyclopedia

Blood Sugar Sex Magik is a featured article; it (or a previous version of it) has been *identified* as one of the best articles produced by the Wikipedia community. If you can update or improve it, please do.

This article is currently on Wikipedia's Main Page as Today's featured article.

Article milestones [show]

This article is within the scope of multiple WikiProjects. [show]
Click [show] for further details

This article was a past ***Red Hot Chili Peppers*** *Collaboration of the Week!*

Contents [hide]
1 Assumption?
2 Critics
3 Sales
4 Trivia
5 Good article nomination on hold
6 Background section
7 Redirection from RHCP page
8 Pearl Jam Mistake
9 OJ

Assumption? [edit]

I think that this statement is inaccurate:

"Some early CD pressings contained incorrect track lengths on the back cover. For example, "If You Have To Ask" was listed as 4:11 and "They're Red Hot" was listed as 1:44. This oddity has long been fixed."

I bought my CD copy of BSSM 2, maybe 3 years ago and it has the incorrect track lengths on the back cover. The store I bought it from is a very busy store that only opened in the mid to late 90s, so I can't imagine they'd have (or would ever have had) any early pressings of the album. I'm going to change the statement to reflect this. JedEgan 19:57, 18 June 2006 (UTC)

Critics [edit]

"many critics feel that the Red Hot Chili Peppers' quality fell precipitously after this album."

What critics think this? Reference? Their last three have generally got more acclaim than BSSM.

^ allmusic don't think that for one thing...they see Blood Sugar as their best album.

> I think the statement does not belong in the encyclopedia. Rolling Stone was very enthusiastic about Stadium Arcadium and also listed Californication on its list of greatest albums. It is true that BSSM may be their most acclaimed album, the statement implies a decline in quality which is an opinion that I do not feel is held by most critics. Aside from the dissatisfaction with One Hot Minute, the Chili Peppers post BSSM albums have been some of the most highly acclaimed albums in modern rock music.

"The 1993 hit "Soul To Squeeze" was originally recorded for BSSM and was subsequently left off. Other songs that were part of the original 25 songs recorded for the album were "Sikamikanico", found on the "Wayne's World" soundtrack, and a cover of Iggy and the Stooges' "Search and Destroy", an instrumental entitled "Fela's Cock"."

i think at the end of the quote the name of the song is flea's cock as previously stated as a song in another part of the atical not fela's so i am going to change it --69.255.63.86 04:00, 12 October 2006 (UTC)

It is a tribute to Fela Kuti an Afrobeat bassist. Fela is correct not Flea.

Fig. 3 An example of talk page

7. Talkpages Analysis

The aim of this section is to outline peculiar linguistic features of WikiSpeak through a contrastive analysis of *talk pages* (henceforth TPs) vs. *Wikipedia articles* (WAs).

WikiSpeak's informal and contextual register undoubtedly shares some peculiarities both with face-to-face spoken language and with spoken-written language used in Internet forums. In the analysis of TPs, references have been made to the theoretical model of *contextuality* expressed by

Heilighen and Dewaele (1999) and to Biber's multidimensional approach (1988) which defines the same concept using the term *involved production.*

Since linguistic classes which will be considered in this section have already been exhaustively treated from the theoretical and methodological point of view in the previous chapter, in this section only findings and data will be reported and discussed. The WikiSpeak which emerges in the TPs has been analyzed through the same methodology which has been applied to compute the style of document mode encyclopaedic pages[66]. The findings, have then been normalized (to obtain the *relative frequency)* and the final score has outlined the WikiSpeak linguistic features, according to a frequency perspective. At this stage, the purpose of the contrastive analysis is to investigate whether there is the existence, or not, of a code switching in the two Wikipedian writing spaces.

The previous intra-genre analysis (*Britannica* vs.*Wikipedia*) had given similar findings. This result has highlighted the use of a similar expository register in the two encyclopaedias, although slightly less formal in Wikipedia. In line with Biber's theoretical assumptions of *involved* vs. *informational production*, the inter-genre analysis (*talk pages* vs. *encyclopaedic articles*) is expected to detect frequency variations in the linguistic classes analyzed, resulting from the more personal, involved and interactive style used by Wikipedians in TPs. Specific linguistic analyses follow in the next sections.

7.1. Findings: (+) linguistic features

7.1.1 Lexical Specificity

In the U.K. Yates (1996) compares three corpora made up, respectively, of *CMC, Spoken* and *Written* texts. He analyses several aspects of language use. According to Yates (1996: 39) CMC users *package information in text in ways that are more written than speech-like* because they may be exhibiting what he calls as *textualization of sociality*, where they bring their *literate production practices to an interactive, social and orally-oriented interaction.*

The specific investigation carried out on Wikipedian TPs vs. WAs has shown that the average word length in TPs is slightly lower than in the encyclopaedic corpus, as it is of 4.1 letters vs. 5.2 characters in the encyclopaedic corpus. Sentences have also proved to be more concise with an average length of 13.5 words vs. 22.09 in Wikipedia corpus. According to what has been prevously claimed [67] this means that shorter words and sentences produce a more informal and involved style.

Chafe and Danielewicz (1987:88) claim that speakers, in contrast to writers, produce language *on the fly* and therefore tend to use the first words that occur to them, consequently the vocabulary of spoken language is more limited in variety. On the other hand, Yates (1996) finds that in terms of *vocabulary use* based on type/token ratios, CMC is more similar to written than spoken language. Thus, CMC is more like written than spoken language in terms of lexical density. In accordance with

[66] *see* chapter 3, section 2
[67] *see* chapter 4, section 1.1.2

Yates' findings, the data of this research has also shown that lexical density (*type/token ratio*) of TPs is lower, but not significantly, if compared to WAs (40% TPs vs. 43.6% WAs). Thus, only a slightly less variety in the language in use has been noticed in TPs.

7.1.2 Nominalizations, Gerunds and Present Participial Forms

The same linguistic classes already taken into account in the analysis of encyclopaedic corpora (WAs vs. BAs) have been considered in the intergenre analysis (TPs vs. WAs).

As shown below (fig. 4) the frequency of nominalizations (singular + plural forms) is practically the same in the two corpora although slightly lower in TPs (4.59% TPs vs. 4.62% WAs), while the frequency of gerunds and present participial forms is identical (2.41 % in TPs and WAs) (fig. 4).

Nominalizations				
	Talk Pages	**%**	**Articles**	**%**
- tion	12515	2.08	7371	1.88
- ity	2297	0.38	2469	0.63
- ment	2496	0.41	2001	0.51
- ence	2369	0.39	1457	0.37
- age	3032	0.50	1328	0.34
- ism	2095	0.35	1135	0.29
- ance	868	0.14	1031	0.26
- sion	1651	0.27	1016	0.26
- ness	300	0.05	302	0.08
Total	**27623**	**4.59**	**18110**	**4.62**

Gerunds and Present Participial Forms				
	Talk Pages	**%**	**Articles**	**%**
Total	**14.527**	**2.41**	**9456**	**2.41**

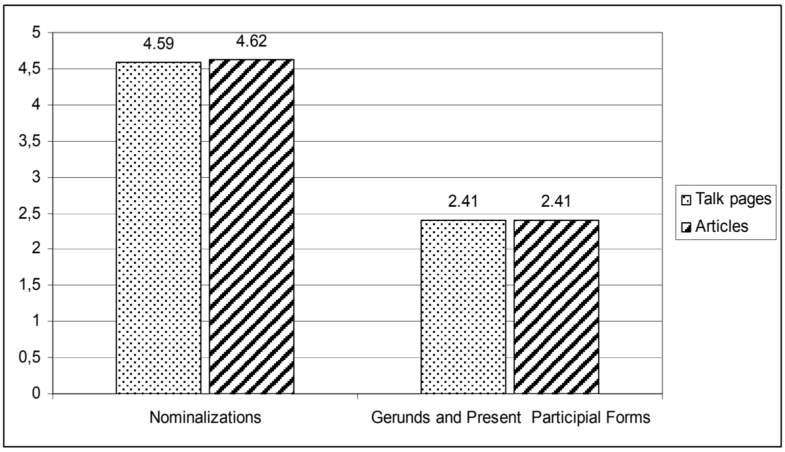
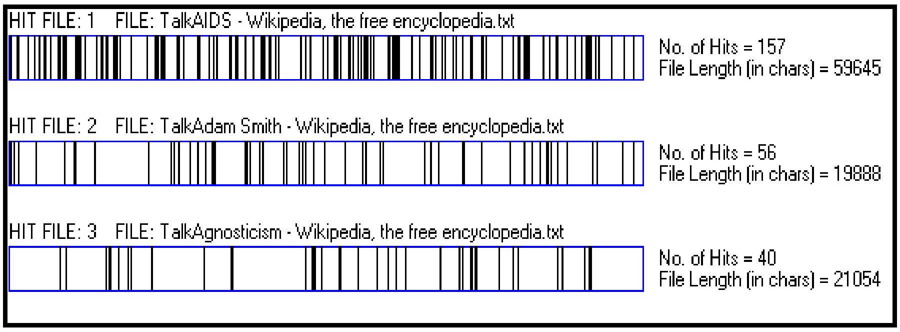


Fig. 4 Nominalizations, Gerunds, Present Participial Forms
in TPs vs. WAs

Although the total frequency of the above mentioned linguistic features is very similar (nominalizations) or identical (gerunds and present participial forms) in the two corpora, their distribution in each specific TP is not homogenous.

This data clearly emerges also observing the graphical distribution of nominalizations, gerunds and present participials in the samples shown (fig. 5a,b,c,d).

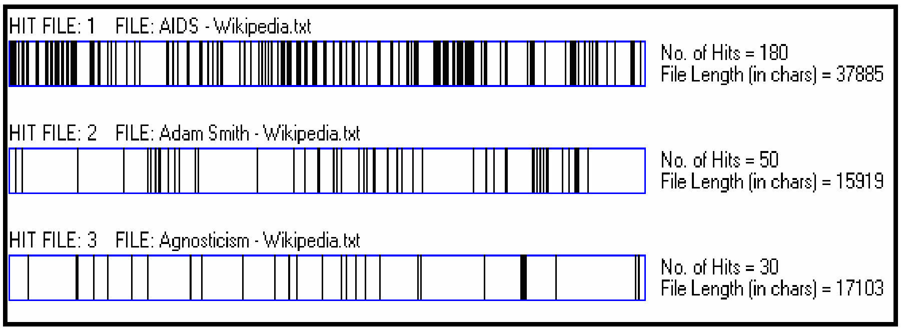


HIT FILE: 1 FILE: TalkAIDS - Wikipedia, the free encyclopedia.txt
No. of Hits = 157
File Length (in chars) = 59645

HIT FILE: 2 FILE: TalkAdam Smith - Wikipedia, the free encyclopedia.txt
No. of Hits = 56
File Length (in chars) = 19888

HIT FILE: 3 FILE: TalkAgnosticism - Wikipedia, the free encyclopedia.txt
No. of Hits = 40
File Length (in chars) = 21054

Fig. 5a Distribution of nominalizations in three samples of TPs

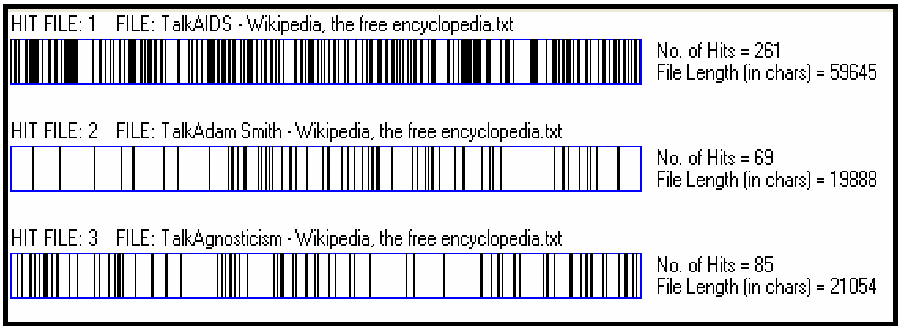


HIT FILE: 1 FILE: AIDS - Wikipedia.txt
No. of Hits = 180
File Length (in chars) = 37885

HIT FILE: 2 FILE: Adam Smith - Wikipedia.txt
No. of Hits = 50
File Length (in chars) = 15919

HIT FILE: 3 FILE: Agnosticism - Wikipedia.txt
No. of Hits = 30
File Length (in chars) = 17103

Fig. 5b Distribution of nominalizations in three samples of WAs

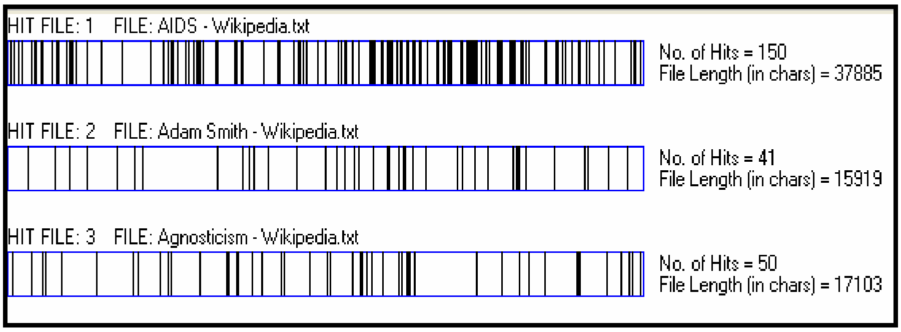


HIT FILE: 1 FILE: TalkAIDS - Wikipedia, the free encyclopedia.txt
No. of Hits = 261
File Length (in chars) = 59645

HIT FILE: 2 FILE: TalkAdam Smith - Wikipedia, the free encyclopedia.txt
No. of Hits = 69
File Length (in chars) = 19888

HIT FILE: 3 FILE: TalkAgnosticism - Wikipedia, the free encyclopedia.txt
No. of Hits = 85
File Length (in chars) = 21054

Fig. 5c Distribution of gerunds and participial forms in three samples of TPs



HIT FILE: 1 FILE: AIDS - Wikipedia.txt
No. of Hits = 150
File Length (in chars) = 37885

HIT FILE: 2 FILE: Adam Smith - Wikipedia.txt
No. of Hits = 41
File Length (in chars) = 15919

HIT FILE: 3 FILE: Agnosticism - Wikipedia.txt
No. of Hits = 50
File Length (in chars) = 17103

Fig. 5 d Distribution of gerunds and participial forms in three samples of WAs

The works of some of the most important sociolinguists such as Hymes, Labov and Gumperz, have described systematic linguistic variations across a wide range of social and situational parameters, including the social class and ethnic group of participants, the social and situational relationship between the participants, the setting, and the purpose of communication. Normally it is expected that the higher the academic level of a person, the richer the vocabulary the person uses and the wider the outlook. With regard to this subject, Heylighen and Dewaele claim (1999:23):

> Academically educated persons express their thoughts in a more precise and less subjective way, that is to say with more formality also in informal context, as cognitively more skilled individuals are less inclined to avoid formality. Thus we might hypothesize that formality would correlate positively with the general factor of intellect.

Probably this is the reason why a deep discrepancy has not been detected in the comparative analysis of lexical density and nominalization frequency in TPs and WAs. The findings show that Wikipedians unconsciously convey their cultured background and their attitude towards the use of a more elaborated language also when they discuss in less official and more relaxed writing spaces as talkpages are.

7.1.3 Articles, Nouns, Adjectives, Prepositions

Significant frequency variations have been noticed with regard to definite and indefinite articles (7.98% TPs vs. 9.68% WAs), nouns (24.03% TPs vs. 29.28% WAs), adjectives (6.43% TPs vs. 10.06% WAs) and prepositions (10.55% TPs vs. 13.42 % WAs) (fig. 6).

As can be observed, occurrences of these four linguistic classes, are definitely lower in TPs corpus. These variations identify the most significant differences between the expository style of WAs and the contextual style of TPs. As already mentioned, total frequency of nouns and adjectives has been deducted through a sample of 10.000 tokens tagged by Claws POS (tagset 5).

Graph in fig. 6 highlights the differences detected in the two corpora. The bars show that the frequency in TPs is constantly lower with regard to the four linguistic classes analyzed.

Definite/indefinite Articles

	Talk Pages	%	Articles	%
the	32.549	5.41	27.780	7.09
a	13.076	2.17	8.573	2.19
an	2.394	0.40	1.576	0.40
Total	**48.019**	**7.98**	**37.929**	**9.68**

Nouns

	Talk Pages	%	Articles	%
Total	**144598**	**24.03**	**114671**	**29.28**

Adjectives

	Talk	%	Articles	%
Total	**38692**	**6.43**	**39398**	**10.06**

Prepositions

	Talk Pages	%	Articles	%
Of	15.205	2.53	15084	3.85
To	15.053	2.50	8915	2.28
Within	13972	2.32	164	0.04
In	9.922	1.65	10026	2.56
For	4441	0.74	3148	0.80
With	3148	0.52	2666	0.68
On	2562	0.43	4037	1.03
By	2.481	0.41	2906	0.74
From	2173	0.36	1924	0.49
At	2065	0.34	1466	0.37
Than	1033	0.17	560	0.14
Out	820	0.14	280	0.07
Into	632	0.11	539	0.14
Without	386	0.06	169	0.04
Between	376	0.06	415	0.11
Against	283	0.05	260	0.07
Through	247	0.04	307	0.08
Off	220	0.04	87	0.02
Per	130	0.02	77	0.02
During	112	0.02	442	0.11
Versus	99	0.02	10	0.00
Among	92	0.02	173	0.04
Toward/s	92	0.02	126	0.03
Except	80	0.01	34	0.01
Upon	76	0.01	86	0.02
Througho	62	0.01	74	0.02
Plus	34	0.01	13	0.00
Opposite	29	0.00	20	0.01
Onto	25	0.00	30	0.01
Total	**63489**	**10.55**	**52566**	**13.42**

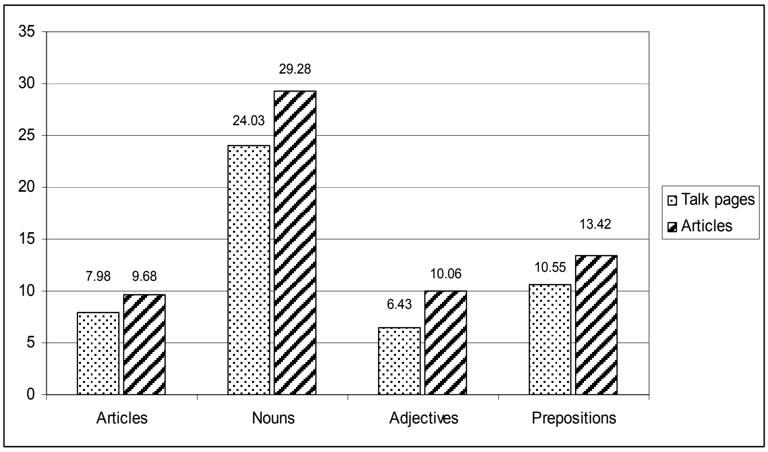


Fig. 6 Articles, Nouns, Adjectives, Prepositions in TPs vs. WAs

7.1.4 Passives

According to Yates (1996) there are more active than passive verbs in spoken language. The analysis made on Britannica and Wikipedia's encyclopaedia has shown a clear preference in both, for the use of active rather than passive verbs[68]. The frequency of passive verbs, as was expected, is even lower in talk pages (0.68% TPs vs. 0.96% WAs) as fig. 7 shows.

Passives				
	Talk Pages	**%**	**Articles**	**%**
has been	150	0.02	117	0.03
have been	114	0.02	93	0.02
had been	23	0.00	58	0.01
is (adv)	961	0.16	**939**	0.24
are (adv)	419	0.07	409	0.10
was (adv)	751	0.12	**1034**	0.26
were (adv)	197	0.03	388	0.10
be (adv)	**1501**	0.25	**730**	0.19
Total	**4116**	**0.68**	**3768**	**0.96**

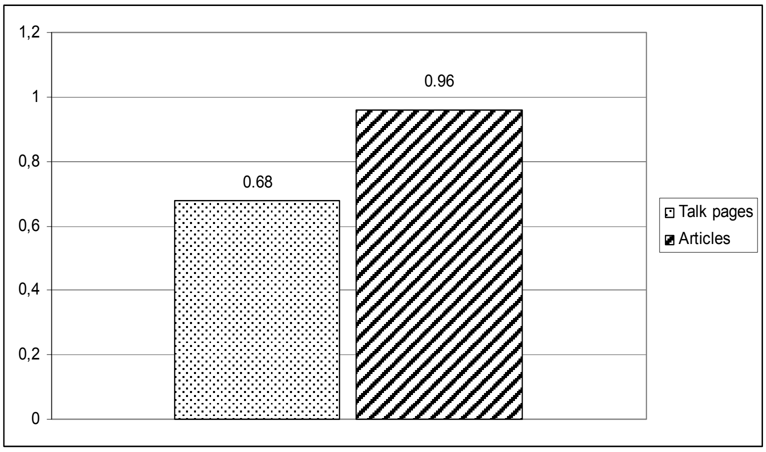


Fig. . 7 Passives in TPs vs. WAs

Some concordances of passives detected in TPs are shown below:

```
   Portal:Dance  Portal:Dance has been started. Please have a look. -
d Hunger      Little progress has been made in tackling world hunger
een   1908 and 1925, and much has been discovered since then.    circ
dian priority, but this claim has been widely criticized. The chronol
cross if you believe the job  has been done:    Rewrite the "History"
Schutz     The first of these has been done, but it is not clear to m
to have 1 error; this   error has been fixed."  Out of curiosity, wha
though it continues to exist, has been at least widely addressed, the
acism against Asian Americans has been totally ignored, even Asian
easel wording   (example; "It has been argued that ...") --Wiley 12:5
d makes is read like       it has been established that Chinese offic
r 2005   (UTC)    The article has been moved to Severe acute respirat
 acute respiratory syndrome   has been reported this year or in late
nce June 2003 means that   it has been eradicated. It's like saying,
```

[68] *see* chapter 4, section 1.5

7.1.5 Subordination and Coordination features

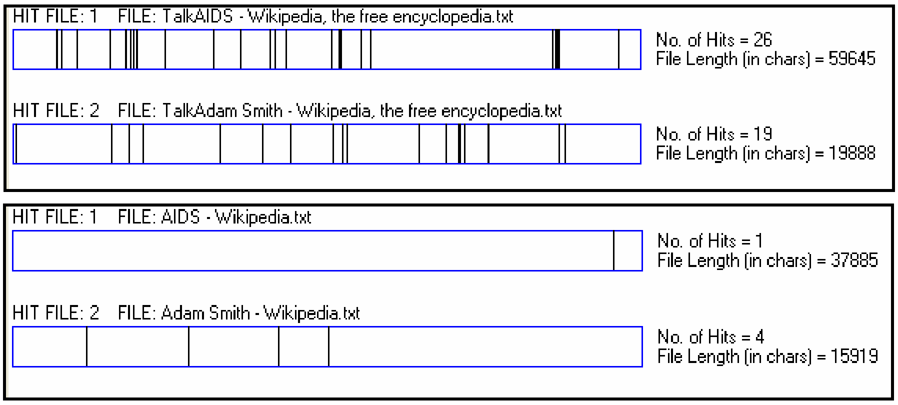
Yates (1996) claims that the subordination structure detected in CMC are very similar to those typically used in spoken discourse. Moreover, Walters (in Farr, 1993:15) claims that their occurrence is less significant if compared to the plentiful hypotactic structures detectable in academic writing:

> there is a strong tendency to structure short chunks of speech so that only one predicate is attached to a referent at a time, whereas in formal written language, information related to a particular referent can be concentrated in heavily modified noun phrases.

Findings of this research show the theoretical inadequacy of any proposal that attempts to characterize subordination as a functional unified construct. In the analysis of TPs, four subordination features have been shown to have a higher incidence when compared to the encyclopaedic corpus, that is to say clauses and sentences introduced by *wh-words* (0.81% TPs vs. 0.69% WAs), *that-clauses* (1,01 % TPs vs. 0.59% WAs) and *conditional subordinators* (0.44 % TPs v.s 0.10% WAs).

By contrast, the frequency of *concessive* (0.10% TPs vs. 0.11% WAs), and *causative adverbial subordinators* (0.32% TPs vs. 0.26 % WAs), and what has been grouped in the miscellaneous category as *other adverbial subordinators* has proved to be very similar in the two corpora (0.09% TPs vs. 0,11% WAs) (fig. 9).

As the concordance plot shows (fig. 8), the allocation of subordination features (in this specific case of *what* in two TPs and WAs random samples*)* proves, once again, that their specific distribution is not homogeneous.



HIT FILE: 1 FILE: TalkAIDS - Wikipedia, the free encyclopedia.txt
No. of Hits = 26
File Length (in chars) = 59645

HIT FILE: 2 FILE: TalkAdam Smith - Wikipedia, the free encyclopedia.txt
No. of Hits = 19
File Length (in chars) = 19888

HIT FILE: 1 FILE: AIDS - Wikipedia.txt
No. of Hits = 1
File Length (in chars) = 37885

HIT FILE: 2 FILE: Adam Smith - Wikipedia.txt
No. of Hits = 4
File Length (in chars) = 15919

fig. 8 Distribution of *What* in two TP and WA samples

Wh-Words				
	Talk Pages	%	Articles	%
What	1888	0.31	297	0.08
Which	1747	0.29	1552	0.40
Who	1119	0.19	699	0.18
Whom	40	0.01	61	0.02
Whose	76	0.01	85	0.02
Total	4870	0.81	2694	0.69

That Clauses				
	Talk Pages	%	Articles	%
That	6119	1.01	2322	0.59

Conditional Adverbial Subordinators				
	Talk Pages	%	Articles	%
If	2552	0.42	383	0.097
Unless	153	0.02	15	0.003
Total	2705	0.44	398	0.10

Concessive Adverbial Subordinators				
	Talk Pages	%	Articles	%
Although	223	0.04	267	0.07
Though	385	0.06	158	0.04
Total	608	0.10	425	0.11

Causative Adverbial Subordinators				
	Talk Pages	%	Articles	%
Since	557	0.09	331	0.08
As	483	0.08	374	0.10
Because	923	0.15	315	0.08
Total	1963	0.32	1020	0.26

Other Adverbial Subordinators				
	Talk Pages	%	Articles	%
While	427	0.07	366	0.093
Whereas	26	0.00	28	0.007
Whereby	10	0.00	9	0.002
As soon as	13	0.00	2	0.0005
As long as	50	0.01	10	0.002
Total	526	0.09	415	0.11

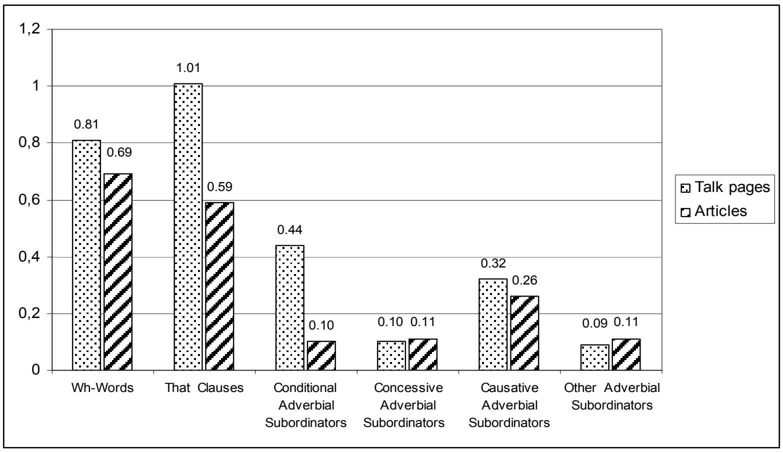


Fig. 9 Subordinations in in TPs vs. WAs

Surprisingly, if the overall occurrence of subordination features is taken into account, the total frequency proves to be higher in TPs than in WAs (2.79% TPs vs. 1.86% WAs) as fig. 10 clearly shows.

Total Subordination Features				
	Talk Pages	**%**	**Wikipedia**	**%**
Wh-	4870	0.81	2690	0.69
That	6119	1.02	2322	0.59
Condition	2705	0.45	398	0.10
Concessive	608	0.10	425	0.11
Causative	1963	0.33	1020	0.26
Other	526	0.09	415	0.11
Total	**16791**	**2.79**	**7270**	**1.86**

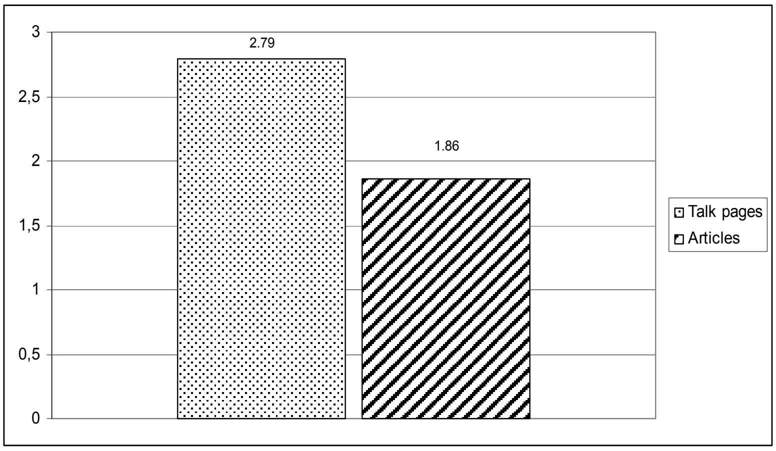


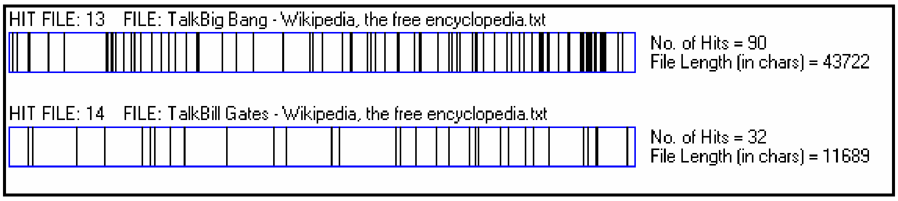
Fig. 10 Total Subordination Features in TPs vs. WAs

In conclusion, a higher number of subordination features in TPs than in WAs has been detected. Thus, the intergenre analysis has shown that their overall frequency is higher in the involved production.

This data runs counter to the general expectations of many previous studies, which have claimed that all dependent clauses are syntactically complex and therefore occur more frequently in informational production since they reflect textual elaboration.

Findings of this analysis are supported by previous suggestions of Halliday (1976) and Biber (1988) who claimed that certain subordination features are typically more frequent in involved rather than in informational production. Although the frequency and the use of some specific structures is strictly connected to the personal writing style, as the concordance plot highlights, certain hypotactic structures are represented to a variable level in specific texts and sometimes an overlapping in their use can be detected. Nevertheless, the frequency of some subordination structures can provide interesting information on the nature of a text, e.g. whether it is written rather than spoken or mediated, formal rather than informal, etc.

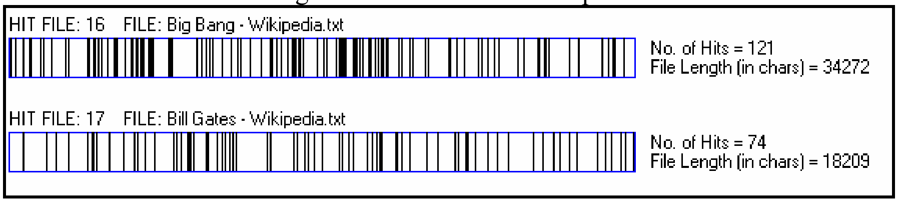
Differently from subordination, the analysis of coordinating conjunctions (fig. 12) has proved that their frequency is lower in TPs than in WAs (3.03% TPs vs. 3.64% WAs) as the visual distribution of the most frequent conjunction, *and* shows in two random samples of TPs and WAs (fig. 11a/b). Specific and total occurrences of coordinating conjunctions are shown in fig. 12.



HIT FILE: 13 FILE: TalkBig Bang - Wikipedia, the free encyclopedia.txt
No. of Hits = 90
File Length (in chars) = 43722

HIT FILE: 14 FILE: TalkBill Gates - Wikipedia, the free encyclopedia.txt
No. of Hits = 32
File Length (in chars) = 11689

Fig. 11a *And* in two TPs samples



HIT FILE: 16 FILE: Big Bang - Wikipedia.txt
No. of Hits = 121
File Length (in chars) = 34272

HIT FILE: 17 FILE: Bill Gates - Wikipedia.txt
No. of Hits = 74
File Length (in chars) = 18209

Fig. 11b *And* in two WAs samples

Coordination Features				
	Talk Pages	**%**	**Articles**	**%**
And	11417	1.90	11255	2.87
But	3426	0.57	928	0.24
Or	3096	0.51	1945	0.50
Nor	127	0.02	49	0.01
Yet	182	0.03	62	0.02
Total	**18248**	**3.03**	**14239**	**3.64**

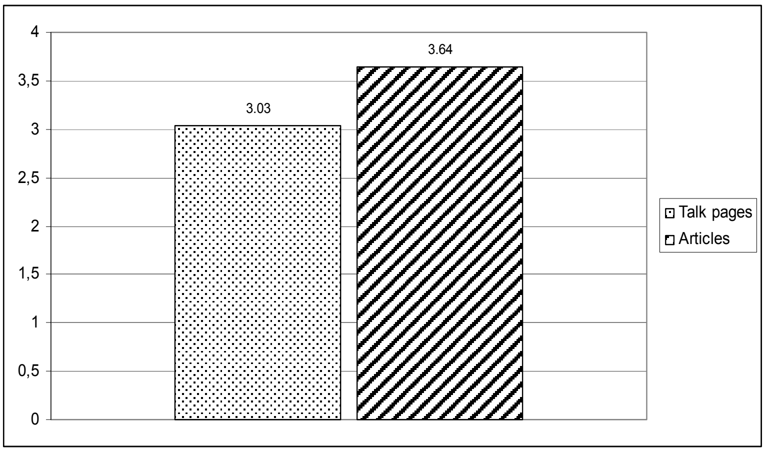


Fig. 12 Coordination Features in TPs vs. WAs

In conclusion, the comparison of the total frequency of coordination and subordination features in the two corpora has shown that Wikipedians use more subordinating structures when they discuss technical and editing operations in TPs, than when writing official WAs (2.79 TPs% vs. 1.86 % WAs).

By contrast, more coordination conjunctions have been detected in encyclopaedic pages (3.03 % TPs vs. 3.64% WAs). Fig. 13 provides the overall visual representation of subordination and coordination features in the two corpora.

Overall Subordination and Coordination Features				
	Talk Pages	%	Articles	%
Subordina	16791	2.79	7270	1.86
Coordinati	18248	3.03	14239	3.64

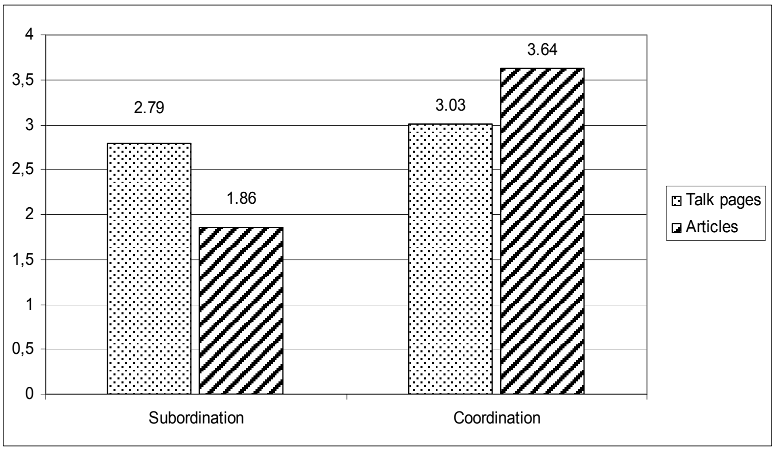


Fig. 13 Subordinations and Coordinations in TPs vs. WAs

7.1.6 Conjuncts

Conjuncts add information to sentences and connect them with previous parts of discourse [69]. The investigation into conjuncts has not shown a significant variation in their use in the two different writing spaces. In particular, Wikipedians use more extensively the conjuncts *therefore, rather, that is*, *otherwise* and *never* in TPs than in WAs. On the other hand, *however* which is in absolute the most recurrent conjunct, has a lower frequency in TPs than in WAs (0.08% TPs vs 0.11% WAs). Just to give an example, an excerpt of *however*, is shown below:

```
poverty caused by the war.     However, at the time of publication, n
al rationales for capitalism. However, Smith never used the term
lot to be    desired. I am     however surprised that you seem to be
ally during the 19th century. However, Smith criticised a number of
ion is not the correct one;    however, neither is this John Rae (edu
o hand. Whilst on my travels, however, I  discovered another John R
For most of his career he was however a  Deist, and recognised as s
 some victims of the disease. However, having just watched the famou
e reassurance and background, however, I was very dissappointed in t
 then maybe you have a point. However...ÔÇöBob 18:25, 15 June 2006 (
t get worked up. WeÔÇÖre not, however, going to assume most people a
aph to include the zero line. However, IÔÇÖd think that a consensus
fe styles. It should be clear however that these beliefs are held by
nsertive party. Keep in mind, however small this may be, it is worth
```

The overall conjunct frequency is very similar, although slightly higher in TPs than in WAs, as the graph comparing the overall findings clearly shows (fig. 14).

[69] *see* chapter 4. sections 1.11

Conjuncts				
	Talk Pages	**%**	**Articles**	**%**
However	465	0.08	445	0.11
For	201	0.03	166	0.04
Rather	389	0.06	139	0.04
Thus	120	0.02	126	0.03
Instead	165	0.03	98	0.03
Therefore	154	0.03	97	0.02
That is	391	0.06	91	0.02
In	51	0.01	51	0.01
As a result	20	0.00	44	0.01
Hence	45	0.01	37	0.01
Otherwise	99	0.02	30	0.01
Never	308	0.05	22	0.01
Similarly	24	0.00	22	0.01
On the	33	0.01	22	0.01
Nonethele	16	0.00	20	0.01
Conseque	5	0.00	20	0.01
Furtherm	35	0.01	18	0.00
Moreover	32	0.01	16	0.00
Total	**2553**	**0.42**	**1464**	**0.37**

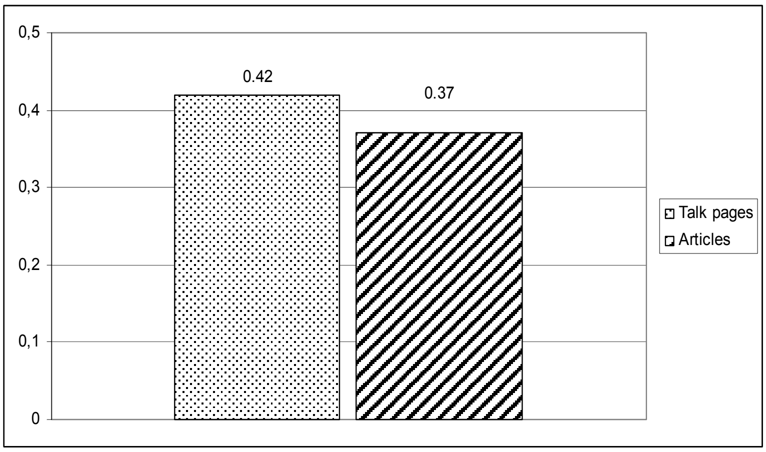


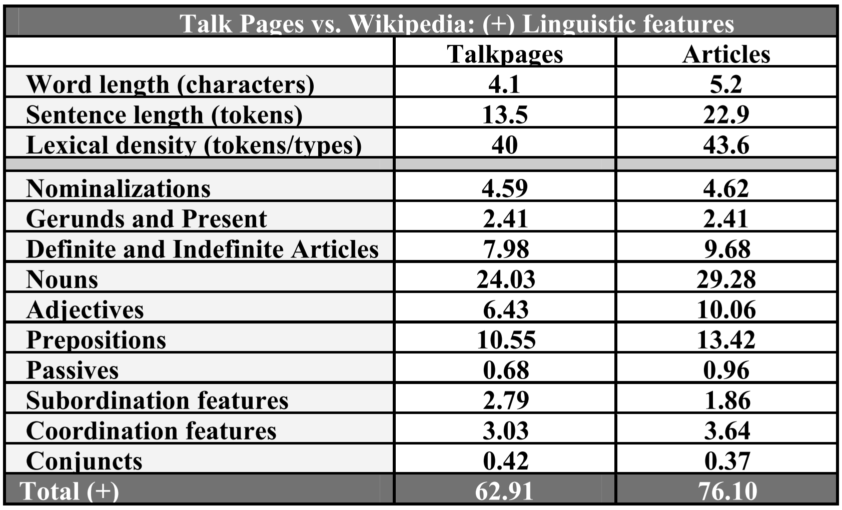
Fig. 14 Conjuncts in TPs vs. WAs

7.2 Comments and Remarks

The overall findings, in most cases, show a constant lower frequency in TPs, for all the linguistic classes which are typical of a formal register. To summarize, average word and sentence length is shorter and lexical density is lower in TPs than in WAs (fig. 15). Furthermore, nouns, definite and indefinite articles, adjectives, prepositions and passive forms are less frequent in TPs than in WAs. It can be noticed (fig. 16) that only the frequency of gerunds and present participial forms have the same value in the two corpora. The frequency of conjuncts is similar, although slightly higher, in TPs.

Surprisingly, and unlike most of the theoretical approaches on informational vs. involved production, the use of subordination, instead of coordination structures, is higher in TPs.

To conclude, the totals of the linguistic features analyzed in this section have been shown to have an overall frequency lower in TPs than in WAs (62.91% TPs vs. 76.10% WAs).



Talk Pages vs. Wikipedia: (+) Linguistic features		
	Talkpages	Articles
Word length (characters)	4.1	5.2
Sentence length (tokens)	13.5	22.9
Lexical density (tokens/types)	40	43.6
Nominalizations	4.59	4.62
Gerunds and Present	2.41	2.41
Definite and Indefinite Articles	7.98	9.68
Nouns	24.03	29.28
Adjectives	6.43	10.06
Prepositions	10.55	13.42
Passives	0.68	0.96
Subordination features	2.79	1.86
Coordination features	3.03	3.64
Conjuncts	0.42	0.37
Total (+)	62.91	76.10

Fig. 15 (+) Linguistic features in TPs vs. WAs

This data demonstrates that Wikipedia (according to the selected linguistic criteria), uses a less formal register in TPs. The value of the first three classes (*word length, sentence length* and *lexical density*) have not been included in the final computation as they are not homogeneous to the other classes listed.

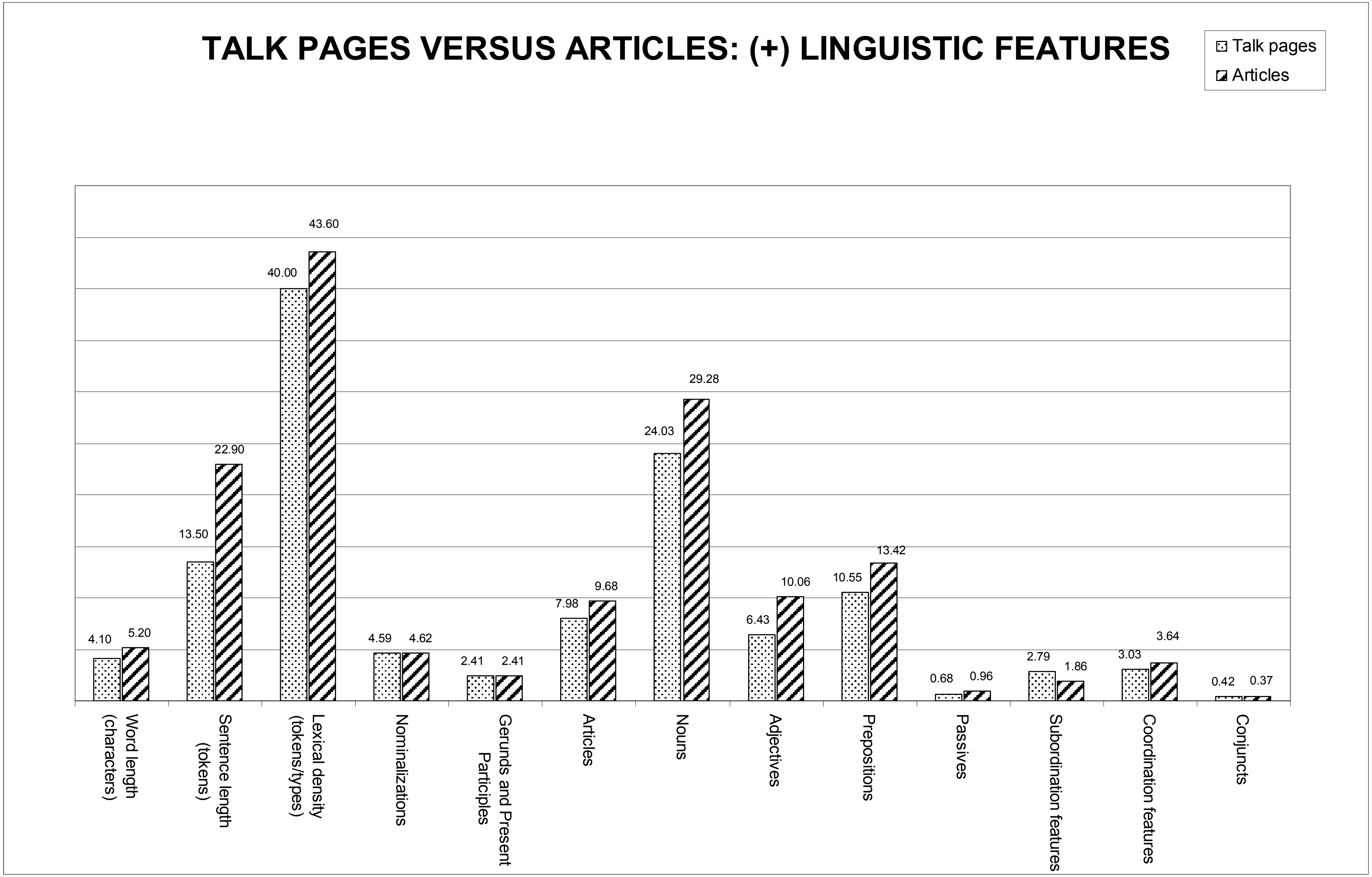


Fig. 16 (+) Linguistic features in BAs vs. WAs

7.3 Findings: (-) linguistic features

Findings of the linguistic classes which convey involvement of the writer will be presented in this section. It is expected that TPs belonging to CMC spoken written genre, share many linguistic features with spoken discourse or with the more general *involved production.* Thus*,* a higher frequency of these linguistic features is expected if compared to the more formal written register of WAs.

Generally speaking, in*volvement* refers to those linguistic classes which reflect the fact that actors involved in the verbal exchange typically interact with one another while writer and reader typically do not. Due to this interaction, speakers use a lot of words with a deictic function which make reference to the specific spatio-temporal or communicative context (Heilighen and Dewaele, 1999). Levelt (1989:45) distinguishes four types of deixis: referring to person (e.g. *we, him my,)* place (e.g. *here, those*), time (e.g. *now, later, yesterday*) and discourse (e.g. *therefore, however*). Further examples of discourse deixis are exclamations or interjections (e.g. *ooh, ok, well*) which are typically concerned with the expressions of personal thoughts and feelings (e.g. marked by use of *first person pronouns, affective forms* such as emphatics and amplifiers, and *private verbs* such as *think* and *feel*). As a result of this concern, involved production often has a distinctly non-informational and fuzzy character (marked also by *hedges*, and by other forms of *reduced* or *generalized content*).

With reference to the above assumptions, the linguistic classes mentioned have been investigated and the relative findings interpreted in the sections which follow.

7.3.1 Place, Time Adverbials, Demonstratives

As already mentioned, Chafe and Danielwicz (1986) include *place and time adverbials* as markers of involvement and Biber (1986) interprets their distribution as marking situated instead of abstract textual content. What happens in CMC? The analysis has shown that total frequency of place adverbials is higher in TPs than in WAs (0.40% TPs vs. 0.28% WAs) as fig. 17 shows.

Here is the most frequently used place adverbial in TPs. Its frequency variation, when compared to WAs is remarkable since it occurs 18 times more often than in WAs (0.18% TPs vs 0,01% WAs). This data proves that speaker/writer makes reference to the immediate online spatial context where discussions take place. Some concordances are provided below.

```
e dont need the quote section here on Wikipedia now, or at leas
to the very first paragraph.  Here's the bigger part of the chang
of ideas that I will present  here undermines Smith's theory and
ch as a model for the editors here. Regards, Durova 17:55, 18 Oct
es the state of peoples minds here, to wake-up would be from a
s encyclopeda format. Wiki is here to be a represention of re
 I wrote out  my explanation  here: Agnosticism wording Origi
6:42, 19 May 2006 (UTC)  See  here that worldwide, about 50% is a
6:51, 19 May 2006 (UTC)  See  here then, that in the USA, in the
6, 15 June 2006 (UTC) I came  here wondering about this as well.
ving oral sex to a woman. But here I found none. I would be reall
```

The use of *here* is followed by *above*, which has, as the concordances below show, a deictic function related to the immediate textual space of reference of the electronic page.

```
ds  to a refernce to the       above mentioned site. Which is .nl,
y off-putting.  Sort out the   above points and I reckon the articl
about a week since I made the  above suggestions, so in true X
pyright restrictions", so the  above image is linked, does  linkin
e Wrote   Someone added the    above section. Don't know whether it
d States, but I   agree the    above phrasing should be more precis
hink are     important --      above and beyond the criteria that d
vise folk that, following the  above comments, I wrote a
ly, I notice that the comment  above this one is also about transit
be in  order, so created the   above titled new section in the corr
 office. Its  been discussed   above about Blair perhaps outlasting
  Impeachment As discussed     above, at the moment this paragraph
ustify  given the discussion   above. At the moment this para begin
n V. Then, if V satisfies the  above eight axioms, it is a   vect
sfies the eight axioms listed  above." However,  there are 10 axio
owever,  there are 10 axioms   above. I guess this is just a mistak
```

Place Adverbials				
	Talk Pages	**%**	**Articles**	**%**
Here	1076	0.18	57	0.01
Above	308	0.05	108	0.03
Under	300	0.05	267	0.07
Around	217	0.04	177	0.05
Below	96	0.02	65	0.02
Toward(s)	92	0.02	126	0.03
Outside	83	0.01	84	0.02
Near	64	0.01	71	0.02
Ahead	52	0.01	9	0.00
Behind	42	0.01	55	0.01
Inside	41	0.01	35	0.01
Nowhere	21	0.00	3	0.00
Next to	21	0.00	10	0.00
Nearby	7	0.00	24	0.01
On top of	6	0.00	2	0.00
Total	**2426**	**0.40**	**1193**	**0.28**

Fig. 17 Place adverbials in TPs vs. WAs

During the analysis a recurrent use of *http://* [70] has also been noted. This acronym and the *url* [71] which follows, specifically indicate the spatial collocation of the information provided in the external virtual space. Since this recurrence has been considered highly meaningful, its frequency has been investigated. Plenty of these direct references (456 occurrences) have been found in TPs while none of them in WA. Some random examples in context are provided.

[70] *http://* means *Hyper Text Tansfer Protocol*. It indicates the communication protocol which enables Web browsing . It is used to transfer data over the World Wide Web.

[71] *Url* means *Uniform Resource locator.* It indicates the exact address where a web document is.

A. Hello. Why does the graphic "City of San Jose Capital of Silicon Valley 10th Largest U.S. City" on your web page (the one in the upper right corner) not have an accent mark in "San Jose"? I am a big fan of consistency. If we are to write San Jose with an accent mark, why don't *you* do it consistently? (Here is a link to at least one page with the graphic I am taking about: httalk page://www.sanjoseca.gov/feedback.html) Yours,-- Marek Lugowski
(from San José talk page)

B. Rewrote section on China. IMHO the CNN article was a very bad summary of what was actually said at the press conference. If yo go to www.xinhua.org, you see this page on SARS httalk page://news.xinhuanet.com/ziliao/2003-04/15/content_832545.htm and even if you can't read Chinese there is enough there to make it clear that the official media is no longer an official blackout on the story. *(from SARS talk page)*

C. The intro para should provide a very short synopsis of the article, so any history there should be 5-6 lines at most. We could have a full section on history, and here are a few more references;
httalkpage://www.southface.org/solar/solar-roadmap.htm
httalk page://www.vidyaonline.net/arvindgupta/assolarpower.htm
(from SARS talk page)

The overall occurrence of demonstratives, which are markers of generalized pronominal reference, have been shown to be higher in TPs than in WAs (1.76% TPs vs. 0.98% WAs) as fig. 18 shows. Their frequency is almost double in the former corpus.

Demonstratives				
	Talk Pages	**%**	**Articles**	**%**
This	5810	0.97	1754	0.45
That	3574	0.59	1210	0.31
These	688	0.11	639	0.16
Those	534	0.09	227	0.06
Total	**10606**	**1.76**	**3830**	**0.98**

Fig. 18 Demonstratives in TPs vs. WAs

This is the demonstrative which more often occurs. Its high frequency (0.97% TPs vs. 0.45% WAs) indicates a spatial deixis which proves a greater reference to the immediate context in TPs once compared to WAs. Some examples are provided below.

```
    April 2006 (UTC)   Why     this article is a mess I haven't b
ess I haven't been following this article closely and I didn't r
iff. I haven't been following this article much either but I
ate a major rewrite proposal. This article needs some love: com
e article is about. I tracked this down to two edits, 16th Jan by
eply motivated to   improve   this article and make it both more
oposal for a major rewrite of this article. We should be   able
p talking and strive to bring this fundamental article back to
o..." (06:59, 18 June 2006 on this article). Maybe it's time
ne 2006 (UTC)   Piotr Blass   This guy in his vanity article clai
ter "w" in Hebrew. Where does this idea that www =  666 come fro
```

Frequency of time adverbials has also been investigated and compared in TPs and WAs (fig. 19). Examples of some concordances for *before* and *yesterday* follow.

```
source page Added this page   yesterday, due to the amount of referenc
```

```
lance will have to be struck. Yesterday I   removed "a set of object
are correct    (well as of    yesterday morning @0705 UTC (0805 BST)).
 Liberty Leading The People   Yesterday I removed the Delacroix painti
 BBC2 documentary broadcasted yesterday noted two additional facts abo
front page  featured article  yesterday, it or course attracted a floo
trying to edit Minelli's text yesterday, I find it to be almost worthl
tain the figures of today and yesterday. Therefore, you  cannot compa
```

```
hat a consensus would be nice before removing it (I know that I wou
er people wanted to delete it before I fix it. Ideogram 18:42, 15 J
dition I have a question but  before I get into it I want to make i
e rate of disease progression before diagnosis of HIV infection or
sentation of the truth.       Before I hop off my high horse, let m
futation of    these ideas    before making such changes. Perhaps w
ugh and see how long it takes before they are reverted in order
r arguments       there       before making a change. As it is, the
identity must to be mentioned before any other identity  even if w
```

Unlike spoken language, which relies on spatial and time references more often than the formal written language, a slightly higher occurrence of time adverbials has been surprisingly detected in WAs (0.68% TPs vs 0.88% WAs).

Time Adverbials				
	Talk Pages	**%**	**Articles**	**%**
When	871	0.14	531	0.14
Now	625	0.10	230	0.06
While	427	0.07	366	0.09
Before	404	0.07	214	0.05
After	377	0.06	537	0.14
Again	269	0.04	79	0.02
Later	187	0.03	344	0.09
Until/til	172	0.03	166	0.04
Early	141	0.02	323	0.08
Today	126	0.02	155	0.04
Once	116	0.02	105	0.03
Earlier	83	0.01	59	0.02
Recently	73	0.01	60	0.02
Late	53	0.01	118	0.03
Immediate	39	0.01	34	0.01
Whenever	19	0.00	2	0.00
Yesterday	19	0.00	6	0.00
The first	17	0.00	31	0.01
Next/last	17	0.00	-	-
Tomorrow	17	0.00	5	0.00
Initially	14	0.00	40	0.01
As soon as	13	0.00	2	0.00
Afterward	10	0.00	10	0.00
Tonight	10	0.00	2	0.00
Formerly	6	0.00	16	0.00
Lately	4	0.00	1	0.00
By the	3	0.00	9	0.00
Everytime	2		-	-
Total	**4114**	**0.68**	**3445**	**0.88**

Fig. 19 Time adverbials in TPs vs. WAs

This data can be differently interpreted. My personal point of view is that the lower frequency of time adverbials in TPs can be probably due to the fact that each TP is a micro independent cosmos, a self-contained space where most of the verbal acts are temporally encapsulated in the artificial time conveyed in that specific TP. Most of the actions which take place here are self-referential. By contrast, encyclopaedic articles report episodes, biographies, describe facts, historical events, which need accurate temporal collocation to be correctly identified and understood by the reader.

Probably this need is the main reason for the higher frequency of time adverbials such as *after, later, early* which are respectively used two, three and four times respectively more often in WAs than in TPs.

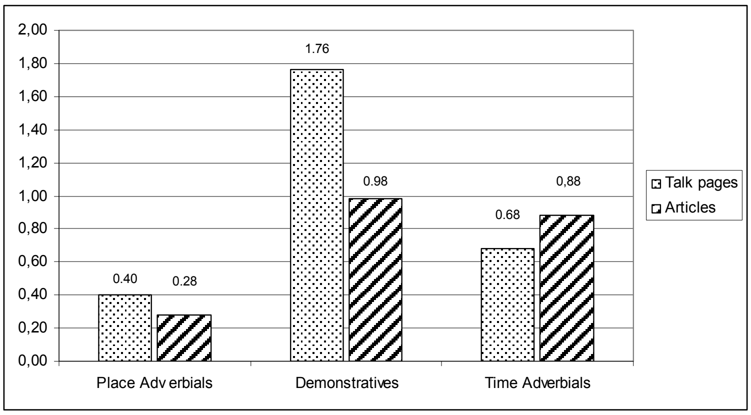


Fig. 20 Place adverbials, demonstratives, time adverbials in TPs vs. WAs

In addition to the most common time adverbials (fig. 19), *UCT* [72] acronym has been searched. The acronym *UCT* is always preceded by the exact time, day, month and year of every post, as well as by the personal nickname of Wikipedian contributor, as the examples below show. 6821 occurrences of this acronym have been found in TPs.

Kjkolb 18:36, 6 April 2006 (UTC)
Ethan Mitchell 18:57, 10 May 2006 (UTC)
Tamino 08:59, 11 May 2006 (UTC)
Nmcmurdo 20:57, 31 October 2006 (UTC)

Moreover, if further time references are searched in TPs, looking for the cluster *this page was last modified* it is possible to exactly identify time and date of the last change, or addition, made on the original TP.

```
This page was last modified 02:52, 20 February 2007
This page was last modified 19:50, 13 February 2007
This page was last modified 02:13, 2 November 2006
This page was last modified 00:10, 22 February 2007
This page was last modified 05:18, 1 November 2006
This page was last modified 11:47, 21 October 2006
This page was last modified 04:28, 18 February 2007
This page was last modified 17:29, 21 February 2007
This page was last modified 07:54, 18 October 2006
This page was last modified 19:18, 21 October 2006
```

[72] *UTC (Universal Time Coordinated)* is based on the Greenwich Meridian used by the military and in aviation. *GMT* (*Greenwich Mean Time)* approximately equivalent to *UTC* has now been considered obsolete and replaced by *UTC*. Using this time zone, standard errors and problems associated with different time zones and summer times operational in different countries are avoided.

In conclusion, data shown demonstrates that Wikipedians use time adverbials differently in the two writing spaces and that their overall frequency is lower in TPs than in WAs (fig. 19). Neverthless, the new functionalities offered by wiki software allow to track the exact date and time of page modification and also provide personal references of involved contributors.

7.3.2 Personal Pronouns and Indefinite Pronouns

As already shown[73] all pronominal forms mark interpersonal focus, a more informal style, a lower informational load and less accuracy in referential identification. The distribution of each pronoun is quite different in different kinds of text. In his analysis, Yates (1996) explores personal pronoun use in spoken, written and CMC corpora**.** In terms of overall frequency of pronoun use Yates observes a higher occurrence of personal pronouns in spoken than in written and CMC texts. In particular, his comparison of CMC and spoken texts shows great similarities in first and second person pronoun use.

The present investigation has highlighted a wide variation in the overall frequency of personal pronouns (and associated object and reflexive personal pronouns and possessives) in TPs vs. WAs. In particular, the highest occurrence has been recorded for first person pronoun use. The high occurrence of *I* (1,85 % TPs vs. 0,06% WAs), which specifically occurs 31 times more often in TPs, indicates an explicit ego-involvement of Wikipedian contributors in the textual production. Furthermore, the high occurrence of the inclusive *we*, whose frequency is about six times higher in TPs (0.28% TPs vs. 0.05% WAs) conveys the identification of single contributors with the Wikipedia community and its collaborative project.

I/WE (+ object/reflexive p. p. and possessives)				
	Talk Pages	**%**	**Articles**	**%**
I	11131	1.85	238	0.06
We	1711	0.28	183	0.05
My	1168	0.19	74	0.02
Me	888	0.15	65	0.02
Mine	44	0.01	8	0.00
Us	316	0.05	18	0.00
Our	229	0.04	63	0.02
Ours	6	0.00	1	0.00
Myself	125	0.02	5	0.00
Ourselves	12	0.00	1	0.00
Total	**15630**	**2.60**	**656**	**0.17**

Fig. 21 First person pronouns in TPs and WAs
(+object/reflexive p.p. and possessives)

Some concordances of *I* and *we* in their original context of use follow.

[73] *see* chapter 4, section 2.2

```
I  am not restoring it, but   I strongly suggest that when rem
ch a substantative statement. I'm not   sure what references
en adding (and so few do, and I've    actually had people c
ally had people complain when I do!), but doesn't it concern y
even a comment? In this case, I have no idea    of the fact
 idea    of the facts, but    I've run across a lot of cases o
 2004 (UTC)   Sorry folks -   I removed it. The Bantu, as a la
ves from Congo   is absurd.   I will be working on the Bantu p
e working on the Bantu pages. I ripped a lot of stuff from
eresting, but quite German .. I was going to do   the same f
e Talk page. Wizzy  Colors    I think that colors of infobox s
infobox should be changed and i said this when Jmabel  asked
ox for ethnic groups? Thought I'd check before  reverting. --
abel 17:19, 6 Apr 2004 (UTC)  I wanted to make every part of t
s could be other than the one I chose but it is important that
```

```
, we don't use 'right' names, we use common names. Its common n
  East India Company. Since   we have to disambiguate that, Eng
a   large measure of blame)   we can only speculate what would
rime with "unoffical", as in "We did not break the    law, w
 did not break the    law,    we only imported the stuff 'unoff
torical sources (not legends) we know that: a) yes gladiators f
 emperors after Nero's death  We know that one chamber in Nero'
meones photo album. I  think  we need to decide which images ad
 changed, too: it sounds like we have the death penaly in Italy
 death penaly in Italy, while we  obviously don't Alessio Dama
what about the Colosseum? Do  we actually know what people thou
isgracefully continue even as we stand on our homeland. What i
s were denied entry. However, we know that for the next several
sking them to come to talk so we can get a better understanding
respect. Language is one tool we have. This is a diaspora. Peo
nt. If man were not involved, we would have a higher death toll
```

From the quantitative point of view, the frequency of first person pronouns *I/we* is followed by the third person pronoun *it.* The higher frequency of *it* (and of the associated object and reflexive personal pronouns and possessives) in TPs (1.77%TPs vs. 0.63% WAs) indicates that its lexical content is not so explicit as it is in encyclopaedic pages. This data confirms the higher fuzziness and semantic vagueness conveyed in TPs.

IT (+ associated object/reflexive p. p. and possessives)				
	Talk Pages	**%**	**Articles**	**%**
It	9644	1.60	1695	0.43
Itself	221	0.04	70	0.02
Its	791	0.13	690	0.18
Total	**10656**	**1.77**	**2455**	**0.63**

Fig. 22 Third Person Pronoun *It* in TPs vs. WAs
(+ associated object/reflexive p. p. and possessives)

Finally, direct reference to the interlocutor is conveyed through the highest frequency of the second personal pronoun *you* which occurs about 25 times more often in TPs (0.70% TPs vs. 0.03% WAs). This data indicates direct involvement of contributors with their addressees. Some examples of *you* in their original context of use follows (fig. 23).

YOU (+ associated object/reflexive p. p. and possessives)				
	Talk Pages	**%**	**Articles**	**%**
You	4184	0.70	113	0.03
Yourself	62	0.01	2	0.00
Your	50	0.01	44	0.01
Yours	27	0.00	2	0.00
Yourselves	6	0.00	1	0.00
Total	**4329**	**0.72**	**162**	**0.04**

Fig. 23 Second Person Pronouns in TPs vs. WAs
(+ associated object/reflexive p. p. and possessives)

```
 to have to directly confront you this way, but are you ac
onfront you this way, but are you actually physically capa
 the middle of it? I mean, do you see it? I can see it qui
. That is amateur quality! If you truly care about making sure t
 is a "high-quality article," you would concede that an amateuri
August 2006 (UTC)  Well, if   you're going to rewrite the intro,
e going to rewrite the intro, you need to make sure that your
  comprehensible does it do   you intend to emulate that common
not   sure what references    you are asking for that
 do!), but doesn't it concern you when people delete subst
d a fact uncongenial, haven't you? -- Jmabel 05:38, 15 Apr
ut it: Any particular reason  you edited the infobox on Zulu int
```

In conclusion, as fig. 24 shows, a wide variation in personal pronoun use has been detected in the two corpora. This data (5.09% TPs vs. 0.84% WAs), in agreement with the findings of Yates (1996) and Biber (1998) proves that the TP channel has a personal pronoun distribution very dissimilar to WA. This is due to the fact that the TPs have a strong interpersonal focus, determining a more interactional and involved form of communication.

Total Personal Pronouns				
	Talk Pages	**%**	**Articles**	**%**
1st	**15630**	**2.60**	**656**	**0.17**
2nd	**4329**	**0.72**	**162**	**0.04**
3rd	**10656**	**1.77**	**2455**	**0.63**
Total	**30615**	**5.09**	**3273**	**0.84**

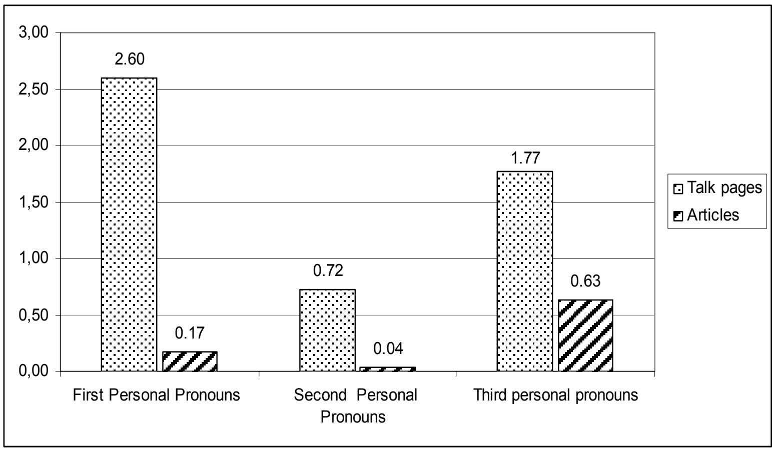


Fig. 24 Total Personal Pronouns in TPs vs. WAs
(+ associated object/reflexive p. p. and possessives)

As well as personal pronouns, TPs are also full of explicit nominal references related to the specific contributor's identity. If we search for the word *user*, the names of the participants involved in the discussion appear, as the examples below show. Most of these references are linked to personal *user pages* [74], where more detailed information on contributors can be obtained.

> A. I've put in the relevant quotes from Pais (the definitive biography) which cite the Nature Paper, and a ref to the Nature Paper. As user:Sparkhead pointed out the mention of the Nature paper should not be in the quotes section, but the previous one. Since the other "blind/lame" quote is slightly different and cited to a secondary source I have not deleted it, because he may have said it twice, but the primary source Nature and definitive bio should take priority IMHO. NBeale 23:52, 19 December 2006 (UTC)
> *(from Albert Einstein talk page)*

> B. User:NoraBG has charged rightly that Human sacrifice in Aztec culture lacks primary sources. This is true because of the reasons explained above. We now need people that have familiarity with the sources to review all of the above-mentioned articles to make sure that they are adequately sourced. Thanks. Richard 16:38, 13 September 2006 (UTC)
> *(from Aztec talk page)*

Thus, only in TPs it is possible to track the identity of contributors while the formal expository style of WA imposes total objectivity in writing. The author's identity is strictly anonymous in WAs since all personal contributions are merged in encyclopaedic articles and author's individuality is suppressed in favour of the neutral and collaborative project.

In addition to personal pronouns, also indefinite pronouns have proved to be more frequent in TPs than in WAs (2.30% TPs vs. 1.72% WAs) (fig. 25). Although the most frequent indefinite pronoun is *all* in both corpora, the contrastive analysis has revealed that the frequency of *any* is more significant, since it is used about twice more often than in WAs (0.19% TPs vs 0.08% WAs). As the examples show, indefinite pronouns add fuzziness and vagueness to the text. This is the reason why their frequency is decisively lower in WA as the prescriptive rules of precision and accuracy are attained in the writing of encyclopaedic articles. The contrastive frequency of indefinite pronouns in the two corpora is shown in fig. 25. The excerpt below reports a few concordances for the indefinite pronoun *any*.

```
pretty  much everybody with   any claim on the throne was a desc
ll recieve less sunlight.     Any precession in a planet's orbit
  ed towards the sun, so at   any given NH lattitude more sunlig
n colonists of Australia were any better than their opposite  n
rried by the largest majority any referendum in  history. The o
th important, I can't   see   any justification for going into t
dent country, and has been by any definition since   the passi
ome up with. Does anyone have any ideas on how it could come off
 since 1964? I  don't recall  any restrictions being on there fo
such as Gopher, or Archie, or any others that  may have allowed
```

[74] A *user page* is a web-based display of information relating to its author. User pages are usually associated with social networking web sites thus, they are a way for internet users to communicate with each other. In addition to the author's username, a user page might include further details such as occupation, interests, website url, and can also provide extra features (such as photos, videos and music).

Indefinite Pronouns				
	Talk Pages	%	Articles	%
All	1682	0.28	712	0.18
More	1643	0.27	806	0.21
Some	1543	0.26	805	0.21
Other	1292	0.21	902	0.23
Any	1161	0.19	302	0.08
Most	763	0.13	707	0.18
Much	713	0.12	287	0.07
Many	710	0.12	780	0.20
Something	553	0.09	72	0.02
Someone	528	0.09	36	0.01
Anyone	401	0.07	32	0.01
Another	292	0.05	235	0.06
Little	288	0.05	103	0.03
a/few	277	0.05	125	0.03
Anything	267	0.04	22	0.01
Nothing	250	0.04	35	0.01
Either	247	0.04	100	0.03
Others	244	0.04	175	0.04
Each	218	0.04	201	0.05
Several	206	0.03	235	0.06
Everythin	104	0.02	19	0.00
Everyone	99	0.02	12	0.00
Somebody	82	0.01	0	0.00
No one	75	0.01	12	0.00
None	63	0.01	15	0.00
Nobody	45	0.01	3	0.00
Anybody	44	0.01	2	0.00
Everybody	36	0.01	6	0.00
Total	13826	2.30	6741	1.72

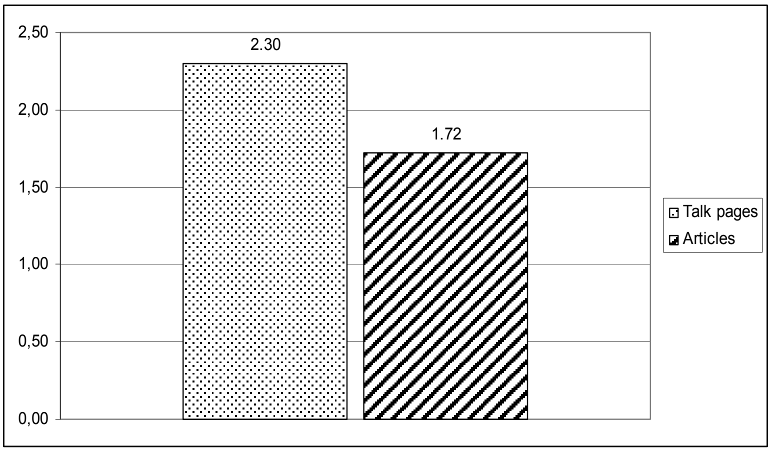


Fig. 25 Indefinite Pronouns in TPs vs. WAs

7.3.3 Mitigating and Boostering Devices

As already seen [75] mitigation devices are a basic interactive dimension of spoken language (Stubbs, 1983) since they serve the purpose of facilitating cooperation between the partners. They mark politeness or deference towards the addressee and avoid threatening the hearer (Holmes in

[75] *see* chapter 4, section 2.5

Calude, 2005). In the specific case of downtoners, when they occur in the encyclopaedic corpus, most of the times they represent the reliability of the information provided, marking uncertainty toward a preposition (Chafe and Danielwics, 1986). Two excerpts of the queries made for the downtoners *slightly* and *somewhat* are reported below.

```
 title could be broadened     slightly. What about slide rules? ;-)
verge. This is akin to taking slightly wrong initial data for the
el, but the arguments   are   slightly different. Not much too it. Yo
 Oven, Wood Oven, Thin Crust (slightly different from Wood and Brick
 glossary. I'm neutral about (slightly opposed   to) this. Anybody?
le who are smart tend to make slightly better  decisions. Renalcat
s; they just keep it  hidden  slightly better...not because they are
like to say that this term is slightly  different from ÔÇÿWhiteÔÇÖ o
be used in a rigorous (though slightly painful) definition, I have am
ary 2006 (UTC)   I disagree   slightly. Truzzi considered himself a Ô
d  therefore tend to produce  slightly more power.  That's highly m
ite cute. But then again I am slightly      insane.- Amorwikiped
go at re-wording this section slightly in a few days, but on this  t
 with it; I've made it sounds slightly less like    a glorious pos
```

```
 audience, still leaves him   somewhat obscure. A  publisher who wou
TM3270 media processor. It's  somewhat similar to a DSP/GPU, but is a
 2007 (UTC)  I'm neutral to   somewhat in favor of mentioning it. If
sm section could be shortened somewhat because the seperate article i
 the study in depth of a few, somewhat arbitrarily    selected, to
dealistic monism is currently somewhat uncommon    in the West.
y at    all. I just have a    somewhat provocative and confrontationa
! What I saw in the intro was somewhat disturbing    though (I don
crust of a St. Louis pizza is somewhat  crisp and cannot be folded e
t section of this article is, somewhat amusingly, "Notable     pe
arlier, the definition is     somewhat of a claim about reality; it s
m   Webster and Oxford seem   somewhat like an exaggerated version of
n is right or wrong. You seem somewhat   confused as to what the pr
are transcendental.  This is  somewhat puzzling, because the link to
"San  Jose, California"? I'm  somewhat picky about naming articles pr
```

As the concordances show, although with a different function, they occur in both corpora, even if more frequently in TPs than in WAs. (0.34 %TPs vs. 0.23% WAs) (fig. 26).

Downtoners				
	Talk Pages	**%**	**Articles**	**%**
Only	1049	0.17	524	0.13
Rather	389	0.06	139	0.04
Pretty	198	0.03	1	0.00
Merely	69	0.01	23	0.01
Fairly	68	0.01	19	0.00
Slightly	56	0.01	30	0.01
Somewhat	55	0.01	28	0.01
Hardly	53	0.01	5	0.00
Relatively	40	0.01	54	0.01
Nearly	38	0.01	53	0.01
partly	16	0.00	22	0.01
Partially	14	0.00	16	0.00
Practically	11	0.00	5	0.00
Barely	9	0.00	1	0.00
Mildly	5	0.00	0	0.00
Scarcely	1	0.00	2	0.00
Total	**2071**	**0.34**	**922**	**0.23**

Fig. 26 Downtoners in TPs vs. WAs

Biber (1989), Chafe and Danielwicz (1986) have explored the use of *hedges* and *amplifiers* in spoken discourse. They agree on the fact that these linguistic classes mark fuzziness in the discourse and co-occur frequently with other interactive features such as first, second person pronouns and questions. The specific analysis has also shown a higher occurrence of *hedges* in TPs than in WAs (0.12% TPs vs. 0.03% WAs). Their frequency is exactly four times higher in the first corpus (fig. 27).

Hedges				
	Talk Pages	**%**	**Articles**	**%**
Maybe	294	0.05	4	0.00
Almost	157	0.03	91	0.02
Kind of	129	0.02	23	0.01
Sort of	89	0.01	15	0.00
Something	66	0.01	1	0.00
More or	23	0.00	10	0.00
At about	3	0.00	12	0.00
Total	**761**	**0.12**	**156**	**0.03**

Fig. 27 Hedges in TPs vs. WAs

Some concordances of *maybe*, the most frequently hedge used in TPs, are provided below.

```
t just doesn't have  belief?  maybe you should include strong athe
 offending quotation, because maybe it should be restated or by so
article, and of this website. Maybe one day, people will learn to
 it is just pure coincidence, maybe or maybe not because of the vi
st pure coincidence, maybe or maybe not because of the viruses all
ersion; meanwhile, admins (or maybe  even long-term registered us
orrectly read the graph, then maybe you have a point. However...ÔÇ
7, 15 June 2006 (UTC)  Hmmm,  maybe I should just remove the graph
hy of mention in the article. Maybe under ÔÇ£StigmaÔÇØ or ÔÇ£Alter
```

Equally higher is the occurrence of *amplifiers* as shown in fig. 28. They occur twice more frequently in TPs than in WAs (0.33% TPs vs. 0.15% WAs).

Amplifiers				
	Talk Pages	**%**	**Articles**	**%**
Very	804	0.13	275	0.07
Of course	185	0.03	13	0.00
Clearly	173	0.03	32	0.01
Completel	130	0.02	34	0.01
Obviously	112	0.02	6	0.00
Entirely	80	0.01	31	0.01
Absolutely	69	0.01	7	0.00
Extremely	64	0.01	29	0.01
Highly	62	.	60	0.02
Intensely	62	0.01	2	0.00
Totally	62	0.01	10	0.00
Strongly	50	0.01	21	0.01
Fully	40	0.01	23	0.01
Perfectly	40	0.01	4	0.00
Greatly	17	0.00	38	0.01
Altogether	12	0.00	8	0.00
Thoroughl	11	0.00	6	0.00
Utterly	7	0.00	2	0.00
Enormous	3	0.00	4	0.00
Total	**1983**	**0.33**	**605**	**0.15**

Fig. 28 Amplifiers in TPs vs. WAs

Emphatics also indicate emotional expression marking involvement with the topic. They frequently occur in conversational genres (Chafe 1985) and, as data in fig. 29 shows, they have also been detected in TPs where their frequency proves to be about twice higher than in the encyclopaedic corpus (0.55% TPs vs 0.28% WAs).

Emphatics				
	Talk Pages	**%**	**Articles**	**%**
Just	1429	0.24	106	0.03
Really	591	0.10	37	0.01
Real + adj	387	0.06	147	0.04
Such a +	128	0.02	49	0.01
A lot + adj	25	0.00	20	0.01
For sure +	16	0.00	1	0.00
Most	763	0.13	707	0.18
Total	**3339**	**0.55**	**1067**	**0.28**

Fig. 29 Emphatics in TPs vs. WAs

Some concordances of the emphatic *really* follow.

```
  ted to Lanier  in 1989. This  really should be mentioned earlier in
  r things in the  article, or  really attempting to learn and furthe
on, because the source is not  really clear. It would be interestin
   later became), but I don't  really know. I'm not really sure abou
 I don't really know. I'm not  really sure about connections  betwe
   2006  (UTC)  Relevance??    really not quite sure why this is in
 know its in brackets and its  really interesting but my journalisti
ense is  tingalling, is this  really relevant and it does clutter t
)  Yeah I agree completly, i  really think this article could do wi
suffrage in all states cannot  really be claimed before 1965 with th
```

All the mitigating and boostering devices explored in this section (downtoners and hedges, amplifiers and emphatics) reduce, although in a different way, the discourse neutrality. As fig. 30 shows, they occur approximately twice more frequently in TPs than in WAs (1.34% TPs vs. 0.69% WAs). In particular emphatics are the more recurrent (0.55%) but comparing the two corpora, the use of hedges is shown to be the more significant, as they occur 4 times more often than in TPs (0.12% TPs vs 0.03% WAs).

Thus, with reference to this linguistic aspect, it is possible to associate a similar use of these devices in TPs with spoken language rather than with written texts. Their overall higher frequency, while softening or amplyfing the force of the utterance, reduce the objectivity and neutrality of the text which is, by contrast, more preserved in encyclopaedic presentation of facts.

Total Mitigating and Boostering Devices				
	Talk Pages	%	Articles	%
Downtone	2071	0.34	922	0.23
Hedges	761	0.12	156	0.03
Total	**2832**	**0.46**	**1078**	**0.26**
Amplifiers	1983	0.33	605	0.15
Emphatics	3339	0.55	1067	0.28
Total	**5322**	**0.88**	**1672**	**0.43**
Total	**8154**	**1.34**	**2750**	**0.69**

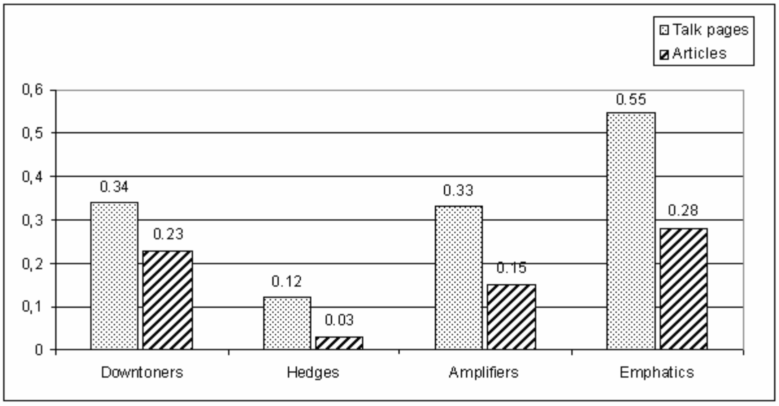


Fig. 30 Mitigating and Boostering Devices in TPs vs WAs

7.3.4 Modal verbs

According to Biber (1998) *modal verbs* indicate uncertainty or lack of precision in the presentation of information[76], and they are especially common in conversation. Yates (1996) observes the different use of modal verbs in his three spoken, written and CMC corpora. His findings show that the usage of modals in CMC is significantly higher than in speech and writing, with writing having the lowest usage of the three. Yates claims that CMC differs significantly from both spoken and written discourse in all cases of modals except possibility modals (e.g. *may, might*). Furthermore, he finds a similarity in the modal usage between oral and CMC communication. Some concordances of *can* and *should* in their original context of use are provided below.

```
tence"  - social subsistence  can be considerably different than
rocessor market. Corporations can be    assumed to be sophist
sleading and should be fixed? Can someone comment    on this?
t ingenious men of his time.  Can we please have 1) a cite and
ms that Adam Smith was gay. I can find no corroborating  source
roborating  source for this.  Can anyone check? If this is not t
omething in   discussion. I   can't argue that the original vers
s nothing about belief so one can be theistic or spiritual and
UTC) Also, agnostic atheists  can be strong atheists as well. Th
eistic   agnosticism (if it   can be called that) is possible, I
, 6 November 2006 (UTC)  You  can't believe, or disbelieve, that
```

[76] See chapter 4, section 2.6

```
  for example, this article   should do a general overview of women
le, in my personal openion. I should here refer again to the
penning" of the action, if it should continue      to be, majo
ent).      Other editors      should be encouraged to edit rather t
he planet. Both histories     should be reduced to a couple of para
/states to timeline Maybe we  should add the dates when women were
USSR and of Yugoslavia. Also, should  countries which have had the
TC) The one listed under [3]  should be editted to say for the U.A.
on here: [4] Same thing below should have  Qatar included for 1997
was a leader in  this. There  should also be sections on more count
 the Wide World ;-). The text should explain elementary notions in
"Java and Javascript" section should rather be called "Dynamic cont
```

A comparison of modals' frequency in the present research has revealed a higher overall occurrence in TPs than in WAs (1.60% TPs vs. 0.72% WAs). They occur more than twice in the first corpus. In particular, a more frequent use of possibility modals has been recorded (*can, may, etc.*), followed by predictive (*would, will, shall)* and necessity modals *(should, must, etc.)* as fig. 31 shows.

Nevertheless, comparing the specific frequency of different modals in TPs and WAs, a higher frequency of *should* has been detected as it occurs approximately seven times more often in TPs than in WAs (0.29% TPs vs. 0.04 WAs%).

Modals				
	Talk Pages	**%**	**Articles**	**%**
Possibility				
Can	1906	0.32	753	0.19
May	1140	0.19	547	0.14
Could	878	0.15	208	0.05
Might	442	0.07	79	0.02
Total	***4366***	***0.73***	***1587***	***0.40***
Predictive				
Would	1932	0,32	561	0.14
Will	971	0,16	328	0.08
Shall	32	0,01	16	0.00
Total	***2935***	***0.49***	***905***	***0.23***
Necessity				
Should	1719	0.29	144	0.04
Have to	303	0.05	24	0.01
Must	270	0.04	141	0.04
Ought	34	0.01	5	0.00
Total	***2326***	***0.39***	***314***	***0.08***
Total	**9627**	**1.60**	**2806**	**0.72**

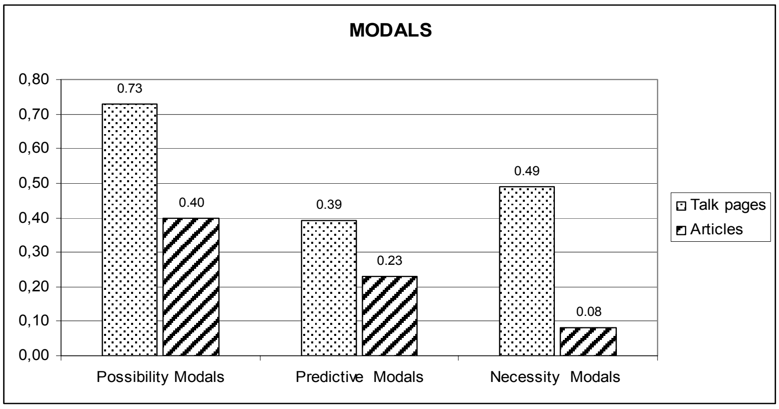


Fig. 31 Modal verbs in TPs vs WAs

7.3.5 Lexical, Public, Private, Suasive and Perception Verbs

Biber (1998) in comparing involved vs. informational production has distinguished five restricted classes of verbs which have specific functions*: lexical, public, private, suasive* and *perception***.** Their frequency has been investigated in this research and compared in TPs and WAs in order to detect similarities or differences in their use.

Lexical Verbs [77]				
	Talk Pages	**%**	**Articles**	**%**
Think	1951	0.32	185	0.05
See	1404	0.23	577	0.15
Say	1360	0.23	235	0.06
Make	1308	0.22	503	0.13
Know	1135	0.19	508	0.13
Get	770	0.13	251	0.06
Mean	702	0.12	210	0.05
Go	588	0.10	183	0.05
Give	585	0.10	326	0.08
Take	581	0.10	575	0.15
Want	511	0.08	46	0.01
Come	438	0.07	192	0.05
Total	**11333**	**1.88**	**3791**	**0.97**

Fig. 32 Lexical verbs in TPs vs. WAs

According to Biber (2006) the most common lexical verbs are very frequently used in conversational genres [78] and, I add, in TPs. Findings of this analysis (fig. 32) have shown almost a double occurrence of the selected lexical verbs in TPs than in WAs (1.88% TPs vs. 0.97% WA). The most recurrent lexical verbs in TPs are *think*, *see* and *say*, whereas in encyclopaedic pages are *see*, *take* and *know*. The most recurrent verb is to *think* which occurs eight times more often in TPs (0.32 % TPs vs. 0.04 % WAs). In particular, it is very frequently associated with the first personal pronoun *I* (*I think* occurs 947 times) as the concordances reported show.

```
Follow". For The Joshua Tree, I think "Where The Streets Have No Nam
there have been many changes. I think it's a good idea to review
he page to change this, since I think it's an error. If you   don'
2006 (UTC)   Commutativity    I think it should be interesting to no
re too   complicated: While   I think that "abelian group" is OK for
le to do with set theory, and I think connecting     linear alge
" is inherently   abstract;   I think it's only logical to give a pr
VS is to the first paragraph. I think the main   point is to relat
ually in the latter camp, but I think there is wide agreement   th
e 2006 (UTC)     Alright,     I think this is better. Put it into th
```

[77] Each figure represents the sum of the occurrences found for each verb at the simple present, past tense and perfect tense.

[78] *see* chapter 4, section 2.9

The overall occurrence of *public* verbs, was also shown to be higher in TPs than in WAs (fig. 33). They are used almost twice more often in the first corpus (0.50% TPs vs. 0.24 % WAs). The high frequency of the verb *write* (fig. 33) conveys a further important information. It indicates how Wikipedia is mainly an active written project which involves a continuous editing process; in addition, the high occurrence of verbs such as *agree, disagree, suggest, claim* and *explain,* conveys the collaborative atmosphere which is typical of this community. Some concordances of the verb *write* are provided below.

```
might have stemmed from" You  write: "Einstein's refusal might ha
lativity, but not enough  to  write anything about "geometrization
nglish teacher who told us to write a 8-10 sentence piece about an
th unless the middle got a re-write. I called it ÔÇÿillogicalÔÇÖ a
ons for a more  prety way to  write it? +MATIA ÔÿÄ 18:46, 24 Octob
ods. I am      trying to      write something about it, but i thin
ld gods. etc. I have tried to write about this, but i still can fi
his sentence? If it helps to  write an equivalent sentence in Span
e because: 1. Tlacaelel didnt write anything,    at least not a
least twelve books, he didn't write "La  mujer dormida debe dar a
ess someone takes the time to write about it,    the article sh
t. If you know how to do  it  write e-mail to by this address (mgl
```

Specific frequencies related to public verbs are reported below (fig.33).

Public Verbs				
	Talk	%	Articles	%
Write	595	0.10	250	0.06
Claim	510	0.08	214	0.05
Agree	500	0.08	40	0.01
Mention	353	0.06	34	0.01
Suggest	290	0.05	110	0.03
Explain	230	0.04	53	0.01
Report	160	0.03	82	0.02
Disagree	105	0.02	15	0.00
Deny	53	0.01	29	0.01
Admit	47	0.01	25	0.01
Reply	40	0.01	10	0.00
Assert	39	0.01	21	0.01
Remark	30	0.00	5	0.00
Insist	26	0.00	13	0.00
Declare	20	0.00	31	0.01
Complain	17	0.00	7	0.00
Promise	12	0.00	9	0.00
Protest	8	0.00	2	0.00
Swear	6	0.00	2	0.00
Total	**3041**	**0.50**	**952**	**0.24**

Fig. 33 Public verbs in TPs vs. WAs

According to Biber, *private verbs* (defined as verbs of cognition in other studies) express intellectual states (e.g. *believe*) and clearly convey the cognitive position of contributors (*I assume, I find, I hope*, etc.). Some concordances of the cluster *I believe* are provided.

```
mell in natural gas is added; I believe the smell in gasoline/petrol
respect are they inaccurate?  I believe that "misinterpreted" is the
Andover, Yale Cheerleader     I believe that there needs to be some me
ts related to the article and I believe relevant to show hateful   g
however keeping the full text I believe is not  appropriate. --zero f
 I've removed the "unsolved". I believe that the proof is generally ac
icle before they are defined. I believe the  link to the non-specific
8 November 2006 (UTC)         I believe that graph theory is a complex
d  endorsed by authority.    (I believe it is also good Sanscrit.) Whe
ve phase. With this in mind,  I believe that the wavefield exiting the
```

Private Verbs				
	Talk	**%**	**Articles**	**%**
Think	1951	0.32	179	0.05
See	1404	0.23	577	0.15
Know	1135	0.19	508	0.13
Find	764	0.13	260	0.07
Believe	402	0.07	227	0.06
Show	300	0.05	248	0.06
Understand	286	0.05	65	0.02
Feel	263	0.04	57	0.01
Hope	158	0.03	10	0.00
Notice	137	0.02	8	0.00
Suppose	135	0.02	20	0.01
Assume	117	0.02	24	0.01
Guess	116	0.02	0	0.00
Prove	105	0.02	50	0.01
Imply	90	0.01	3	0.00
Determine	59	0.01	69	0.02
Realize	59	0.01	18	0.00
Imagine	57	0.01	10	0.00
Hear	56	0.01	26	0.01
Demonstrate	55	0.01	34	0.01
Forget	53	0.01	0	0.00
Indicate	49	0.01	57	0.01
Discover	43	0.01	45	0.01
Estimate	34	0.01	57	0.01
Recogonize	34	0.01	25	0.01
Reveal	14	0.00	31	0.01
Infer	9	0.00	2	0.00
Total	**7885**	**1.31**	**2610**	**0.67**

Suasive Verbs				
	Talk Pages	**%**	**Articles**	**%**
Suggest	290	0.05	110	0.03
Ask	156	0.03	54	0.01
Request	102	0.02	14	0.00
Propose	90	0.01	51	0.01
Recommend	28	0.00	22	0.01
Grant	26	0.00	35	0.01
Insist	26	0.00	13	0.00
Demand	21	0.00	8	0.00
Arrange	9	0.00	22	0.01
Urge	7	0.00	8	0.00
Beg	4	0.00	0	0.00
Command	2	0.00	5	0.00
stipulate	2	0.00	1	0.00
Total	**763**	**0.12**	**343**	**0.08**

Fig. 34 Private and Suasive Verbs in TPs vs. WAs

The frequency of the private verbs [79] shown in fig. 34, proves that their occurrence is about twice higher in TPs than in WAs (1.31% TPs vs. 0.67% WAs). This data is very interesting as it reveals how private and personal attitudes, thoughts, and emotions are clearly conveyed in TPs.

S*uasive verbs* imply the intention to bring changes in the future. *Suggest, ask, request* are the most recurrent suasive verbs detected in TPs (fig. 34). Following the general trend, their overall occurrence proves to be higher in TPs than in WAs (TPs 0.12 % vs. WAs 0.08%). They undoubtedly convey the collaborative atmosphere and the negotiating activity typical of this peculiar encyclopaedic working community. Some examples in context of the verb *suggest* are provided below.

> A. Since cultural references sometimes get deleted without discussion, I'd like to *suggest* this as a model for the editors here. Regards, Durova 15:53, 17 October 2006 (UTC)
>
> B. To the casual reader, it *suggests* that women were treated differently to men until 1962. If Egil wants to continue to split hairs, he'd better do it in a clearer, more accurate way. If Egil doesn't do it I will, when I have time. I *suggest* finding out the nature of the exception in each of the five cases, and writing something like [...]
>
> C. What does Swiss women being given the right to study have to do with women voting in Britain? I *suggest* that this be either removed completely or severely pruned to only include things which are actually relevant.

S*eem* and *appear,* defined as *perception* verbs, can be used to mark evidentiality with respect to the reasoning process. The present analysis has proved they are used about three times more often in TPs than in WAs (0.18% TPs vs. 0.06 % WAs) (fig. 35).

	Perception verbs			
	Talk Pages	**%**	**Articles**	**%**
Seem	876	0.15	78	0.02
Appear	218	0.04	160	0.04
Total	**1094**	**0.18**	**238**	**0.06**

Fig. 35 Perception Verbs in TPs vs. WAs

Chafe (1985) considers its use a strategy for hedging in academic writing. In most of the cases, their use proves to accomplish a similar function in TPs, adding uncertainty to the utterances, as the examples which follow show.

> What were the cons of women's rights? I've researched it many times, but there doesn't *seem* to be anything on what was tragic about it. Getting in trouble with the law from illegal strikes and such, losing time, and hunger from the hunger strikes were bad about it. That's all I could come up with.
>
> The overall structure is poor; the order and choice of sections *seems* arbitrary. For example, the "Java and Javascript" section should rather be called "Dynamic content", or something similar, and cover more than these two particular technologies.

[79] Private verbs have been selected from Quirk (1985:1180-1183)

The bars in fig. 36 clearly indicate a higher frequency of lexical, public, private, suasive and perception verbs in TPs.

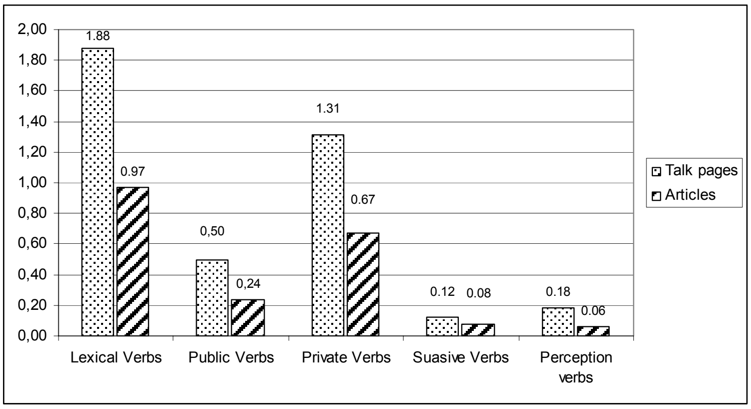


Fig. 36 Lexical, Public, Private, Suasive and Perception verbs in TPs vs. WAs

7.3.6 Interrogative Sentences, Reduced and Negative Forms and Discourse Particles

Interrogative sentences, as already claimed [80] indicate a concern with interpersonal functions and involvement with the addressee (Biber, 1998) especially when they occur with second personal pronoun. This is the reason why, similarly to what happens in spoken language, a very high frequency of interrogative sentences has been detected in TPs and not in WAs. They occur 63 times more often in TPs than in WAs (fig. 37).

Interrogative sentences				
	Talk Pages	**%**	**Articles**	**%**
Total	**3802**	**0.63**	**17**	**0.00**

Fig. 37 Interrogative sentences in TPs vs. WAs

Some examples in context follow.

```
 mean to "stand for election"?? Georgia guy 22:53, 1 June 2006
   Other definitions,  anyone ? WBardwin 01:22, 2 June 2006 (U
ia as mentioned in the article? Subdivisions  of other countr
ir current name in parentheses? For example,  Myanmar, a numb
ormation really that important? Futhermore, I am 99% sure  th
sm or valid edit on December 9? And again on December 14? I d
er 9? And again on December 14? I don't know if this edit, ma
ere the cons of women's rights? I've researched it many times
ould come off as a bad  thing ? ÔÇöThe preceding unsigned comm
ments against women's suffrage? If there was a debate  going
```

All the reduced forms[81] are very recurrent in conversation being a consequence of fast and easy production; by contrast, their use is forbidden in formal academic production. Thus, as was expected,

[80] *see* chapter 4, section 2.8
[81] *see* chapter 4, section 2.10

they are very frequently used in TPs, while they are completely absent in WAs, except in cases of direct quotations (fig. 38). Some examples of reduced forms are shown below.

Reduced forms				
	Talk Pages	**%**	**Articles**	**%**
Total	**10077**	**1.67**	**0**	**0**

Fig. 38 Reduced forms in TPs vs. WAs

```
 marginal interest --  though I'd appreciate alternativ
 his article is a mess I haven't been following this article c
  article   closely and I didn't realize it has become  such
 rect  definition with the W3C's politically correct one (see
2, 18 April 2006  (UTC)     It's hard from your links to see w
rom your links to see what you're referring to - only the midd
   one shows a diff.   I haven't been following this article m
lesale revert of over 7 months' work - there must have   bee
in   the discussion above.   I'm glad to see that some people
 that some people care, so let's go ahead   and clear the me
mers as well as the experts. I'd love to read your comments
tely been improved, although I'm sure we will   be able to m
even better  over  time.   Let's keep talking and strive to br
n the agression, Coolcaesar. I'm referring to your edit   co
on this   article). Maybe   it's time for you to re-read some
   June 2006  (UTC)  Fine,   I'll concede that my comment was
August 2006 (UTC)  666   I don't believe there is a letter "w"
```

Many linguists have theorized a higher occurrence of *negative forms*[82] in speech than in writing. Coherently to what happens in spoken discourse, TPs have also proved to have a higher occurrence of negative forms, (three times higher) in TPs than in WAs (1.11% vs. 0.42%) (fig.39).

Negative forms				
	Talk Pages	**%**	**Articles**	**%**
No	1415	0.24	356	0.09
Neither	75	0.01	22	0.01
Nor	127	0.02	49	0.01
Not	5106	0.85	1209	0.31
Total	**6723**	**1.11**	**1636**	**0.42**

Fig. 39 Negative forms in TPs vs. WAs

Some concordances of *negative forms* are shown below.

```
lien splotch in the middle is not high    quality. That is am
hat goal.    Finally, I am    not going to remove the flare beca
d to retouching, which I have not dabbled    in since I was i
, Wikipedia:What Wikipedia is not, Wikipedia:Neutral   point o
to cover the WWW, but decided not to claim rights to  the WWW.
against SBC and BT decided to not appeal. --Coolcaesar 02:15, 22
standard, but it is certainly not common usage. Take    a loo
   the educated. That does    not necessarily suggest those who
essarily suggest those who do not     write   it   as  so   a
  it stops with a full-stop,  not mid-sentence. However, this is
very day. A URI and a URL are not the same thing, but the  conc
g Wikipedia:What Wikipedia is not, Wikipedia:Verifiability and
```

[82] *see* chapter 4, section 2.7

In association with the interactional style of spoken language, *discourse particles* are normally used to maintain coherence in conversation and are rare outside conversational genres.

Chafe (1985) describes discourse particles as devices which monitor the information flow in involved discourse. For this reason, their frequency has been searched in TPs. As can be noticed (fig. 40), the selected *discourse particles* very rarely occur in encyclopaedic articles, by contrast, they are 50 times more frequent in TPs (0.05% vs. 0 %).

Discourse Particles				
	Talk Pages	**%**	**Articles**	**%**
Well	141	0.02	2	0.00
Anyway	140	0.02	4	0.00
Anyways	10	0.00	0	0.00
Anyhow	8	0.00	0	0.00
Total	**299**	**0.05**	**6**	**0.00**

Fig. 40 Discourse Particles in TPs vs. Was

Two excerpts of of *anyway* (fig. …) and *well* in their original context of use are shown below.

```
t did. It didn't add anything anyway . -- Someone else 22:31 Nov
ike they originally were, but anyway that's neither here nor there)
ch of what is in the article. Anyway, if you think my explanation i
, but they would have done so anyway, so the        decision
sense the definition is wrong anyway, because numerical      a
o Aristotle seems superfluous anyway. Mel Thompson in  'Teach Your
n't had in a good    while     anyway. --Francesco Franco aka Lacato
'80s - definitely after 1970, anyway; I have a photo taken in 1970
e classical   Italian pizza    anyway), I'm not sure how you can say
ticle, it isnt THAT popular.. anyway ive never seen one with mushro
```

```
osexuality or bisexualityÔÇØ. Well, in short, there is an associati
st, it actually sounds quite, well, prejudiced is the word which sp
heMat (talk ÔÇó contribs) .   Well, there is actually a reference o
) Wow.  That was a whole lot. Well, wikipedia is not a soap box (se
17:20, 14  August  2006 (UTC) Well, what did you mean exactly when
4   December   2006 (UTC)     Well, I have come across Einstein's q
 should reflect this doubt as well, and it shouldnÔÇÖt be seen as a
srs  14:40,  3 May 2006 (UTC) well, maybe the van Daan name should
i 20:15, 11 Aug    2004 (UTC) Well, now I see a sense in which the
in).  18:12, 7th  September   Well, this can of worms isn't really
ge—it  is very informative as well, but a little long  for
better to  rename   those as  well, and make it easier to split the
```

Fig. 41 shows a comparison of the four elements analyzed in this section. They are ordered in descendent frequency order . They never, or very rarely, occur in the associated WAs.

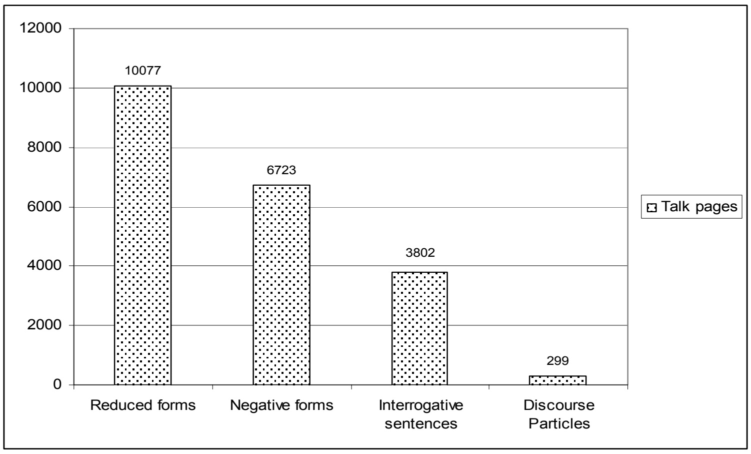


Fig. 41 Reduced and Negative Forms, Interrogative Sentences, Discourse Particles in TPs and WAs

7.3.7 Punctuation Marks, Interjections and Mispellings

A different attitude towards the use of punctuation marks is another interesting aspect which emerges from the analysis of Wikipedian pages. A more generous use of traditional punctuation marks has been detected in TPs (fig. 42). As total figure show, they occur approximately three times more often in TPs than in WAs. In particular, the use of *commas* (11.93% TP vs 5.12% WA), *full stops* (6.96% TPs vs 2.48% WAs), and *semicolons* (0.33% TPs vs 0.17% WAs), is about twice higher in TPs than in WAs, while the use of dots which simulate a pause or indecision in the discourse, are very frequent in TPs while, as expected, they are completely absent in WAs.

My personal point of view is that their higher occurrence in TPs is due to the use of shorter sentences, to the higher frequency of hypotactical structures and to the conversational and interactive style of TPs which uses punctuation marks as a functional device to reproduce the natural pause, hesitation and intonation of oral language.

Punctuation Marks				
	Talk Pages	%	Articles	%
Commas	71773	11.93	20066	5.12
Full stops	41872	6.96	9721	2.48
Semicolon	2004	0.33	667	0.17
Exclamati	1091	0.18	2	0.00
Dots (…)	725	0.12	0	0.00
Total	117465	19.52	30456	7.77

Fig. 42 Punctuation Marks in TPs vs. WAs

The use of *exclamation marks* in TPs conveys the emotional attitudes of interlocutors. *Interjections* carry out a similar function, although in a more complete and expressive way. Usually invariable in form, they do not have a precise grammatical function but they typically express emotions, such as surprise, or sharply call attention to something.

As expected, *interjections* have been detected only in TPs. Thus, in this regard, they can also be associated with the spoken conversational genre. By contrast, this linguistic feature is constantly absent in the objective and neutral expository production of WA (fig. 43).

Interjections				
	Talk Pages	**%**	**Articles**	**%**
Yes	230	0.04	0	0
Oh	62	0.01	0	0
Yeah	48	0.01	0	0
Hey	30	0.00	0	0
Hello	23	0.00	0	0
Ah	18	0.00	0	0
Dear	12	0.00	0	0
Eh	11	0.00	0	0
Ha	10	0.00	0	0
Bye	5	0.00	0	0
Yep	4	0.00	0	0
Aye	1	0.00	0	0
Ooh	1	0.00	0	0
Aha	0	0.00	0	0
Goodbye	0	0.00	0	0
Total	**455**	**0.07**	**0**	**0**

Fig. 43 Interjections in TPs vs. WAs

Some concordances for the interjections *yes*, *ah*, and *dear* are shown below.

```
 Word  count  Over 6,400.     Yes I know it's boring, but one ha
imeline of Tony Blair's life  Yes?No?87.113.24.112 20:58, 18 Aug
 approval rating early on and yes that should be mentioned
 is no Thatcher, he is a pure yes man with no iron in his
word's literal meaning today, yes. The article is relating the
ed Sholes Glidden typewriters yes, but there were many  other t
:04, 14 Jan 2004   (UTC)      Yes, good article, but... Yes, th
  Yes, good article, but...   Yes, the article is exemplary, but
xperience, without checking. (Yes, I know...) After reading the
```

```
0:58, 18 October 2006 (UTC)   Ah, right. That was part of the i
8:15, 7 June 2006 (UTC)       Ah, one more thing: the name. I b
vember 2005 (UTC)             Ah, no, not in the intro, I suppo
4:51, 24 April 2006 (UTC)     Ah, the Company of Scotland of Da
:38, 5 September 2005 (UTC)   ah, yes, it happened when I was a
ording to this encyclopaedia. Ah well. --Kiand    20:00, 29
04:10, 9 October 2006 (UTC)   Ah, I left for Ottawa, at the end
brown 19:09 6 Oct 02 (UTC)    Ah, I just moved the graph down w
9, 12 September 2006 (UTC)    Ah, yes. Much like the Ande, the
  15:46, 1 May  2006 (UTC)    Ah. Well McLellan says that Demut
```

```
ot agree with my choices, but oh well. Brutannica   01:39, 10
 00:37, Sep 15, 2004 (UTC)    Oh. O.K.... Brutannica 05:08, 15
but we must not suddenly say "Oh, even though you think  Aztec
:29, 17 November 2006  (UTC)  oh, and it appears the next comme
 a ton to learn about Wiki... oh well.    Totnesmartin 22:06
    3   January 2007 (UTC)    Oh yes! And there's always new in
  31 October   2005  (UTC)    Oh, or maybe you're saying that t
November   2005  (UTC)        Oh, I see, yes. This is probably
hopefully) clarify the issue. Oh, I had a   closer look at th
 you see on the main page....(oh...well the second..the f
n 07:10, 29 July  2006 (UTC)  Dear Baaa, I am really astonished a
what's true.   You   forgot,  dear Anonymouse, to mention that hi
 go from here ...? Maybe you, dear fellow  Wikipedian, can help
1, 31 August    2006   (UTC)  Dear Zerofaults, I considered it be
12:32, 6 October 2006 (UTC)   dear slurbenstein! as a staunch Mar
3, 7 October 2006 (UTC)  yes, dear friend and thats why i will tr
20       October 2006 (UTC)   dear Gronky, I raised the same ques
, 23   September 2006 (UTC)   Dear Bejnar. I do agree in what you
:24, 5 October 2006 (UTC)     Dear Bejnar what do you mean as a t
 makes him unreliable?   ok   dear, thanks for the True and fair
 I think you may find, my     dear Cock-er-nee David, that Lady G
```

8. New Writing Conventions in Wikipedia Community

Electronic discourse has brought new linguistic conventions. As already shown, the functions performed by voice quality, intonation and pauses in oral discourse, have been traditionally performed in written language by capitalization, punctuation, italicization, and paragraphing (Brown and Yule, 1983:10-11). The use of new original and unconventional practices which have emerged in CMC and TPs have been affected by their use.

Wikipedians, as well as bloggers, chatters, forum participants and texting writers, make use of non standard spellings which reflect pronunciation (e.g. *yep, nope, yay, sokay*) or convey personal emotions, by using a varying number of vowels and consonants (*noooooooo, yayyyyyyy*). The use of repeated interjections (*ah, ah, ah*) or punctuation marks (*Yes!!!!, WHAT????),* dots (…….) and commas („„„„) is also very common in TPs.

Original use of Punctuation Marks				
	Talk Pages	**%**	**Articles**	**%**
-----	2342	0.39	0	0
…..	1337	0.22	0	0
!!!!	230	0.04	0	0
*****	221	0.04	0	0
???	177	0.03	0	0
#	107	0.02	0	0
@	47	0.01	0	0
Total	**4461**	**0.74**	**0**	**0**

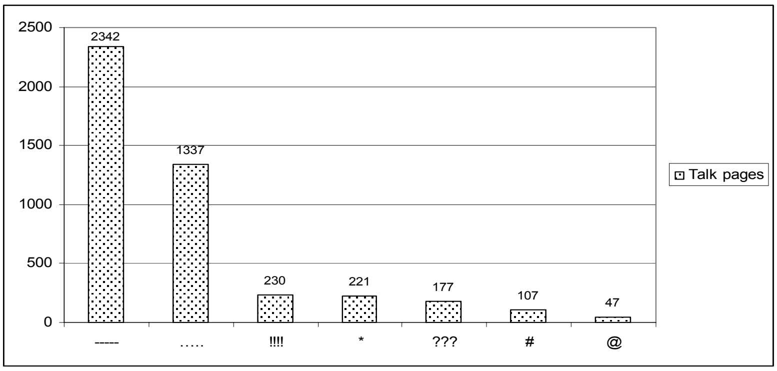


Fig. 44 Original Use of Punctuation Marks in TPs

Nevertheless, *punctuation marks* sometimes tends to be minimalist or completely absent in TPs. Much depends on the user's personality: some Wikipedians make an excessive use of them, some are scrupulous about maintaining traditional punctuation while others do not use them at all.

An increased use of symbols, not normally part of the traditional punctuation system, such as the dash (#) , hyphens (--) or asterisks (***) has also been observed in TPs. In brief, an original and unconventional use of punctuation marks, strictly banned in WAs, has been detected in TPs. Fig. 44 shows the distribution of the above mentioned elements in TPs corpus.

Some concordances for the unconventional use of repeated question and exclamation marks, dots and repeated dashes are shown below.

```
6 (UTC)  It says Talk section ??? Fact idiot 19:24, 30 September
ven if the have to be repeated??? ÔÇöThe preceding  unsigned co
f the empire in sq km or sq ft??? Thanks.  Mmace91 04:11, 13 De
 the idea of  America"... Wha ??? -Eisnel 06:13, 18 January 2006
y sought after by the Chinese ??? <comic book guy voice>: In t
hy satisfactory... HMMMMMMMMM ???  Actually, no, it wouldn't b
 flags comes form russian flag??? what is this??? who wrote thi
 russian flag??? what is this ??? who wrote this??? those flags
what is this??? who wrote this??? those flags have nothing to d
s which are  exact squares.   ???? Oh my God, who wrote this???
```

```
EST SELLING ARTIST OF ALL TIME!!!!!!!!! CHECK YOUR FACTS!
e do not remove my    tag    !!!!--HalaTruth(ßêÉßêïßëâßêà) 11:35
 00:03, 27 May 2006 (UTC)     !!!!What about the Berlusconi's fal
on of terms.  Giuseppe Italy !!!!! This article will be neutr
GAL IS A MAJOR PRODUCER OF TEA!!!!!!! Did you even bother to re
st, general, community feeling!!!! ÔÇöThe preceding unsigned comm
r not is none of YOUR business!!!! Mtoussieh 07:29, 17 November 2
m so tired of reaching nowhere!!!!!!!!! Here's something to ma
 so tired of reaching nowhere!!!!!!!! Here's something to mak
people are spelling this word !!!! In standard American English
```

```
lity made Smith  ingenious?   .......I have removed this recentl
r username for you? Hmmmmm......... Tess Tickle 01:26, 16 August
h   needs to know about that  .....-JLSWiki 15:28, 11 February 200
aid "because Im in  the KKK." .....they black guy was talkin about
 the KKK like it was no  deal .....he had the white robe w/ black
hen wearing it), had the books.......it REALLY needs to  be added
ck people who  are in the KKK ......hmm. not that I come to think
Y AWARDS..............NOT FIVE.....AND IS THE 3RD     BIGGEST
)   While the rooster's away  ..... Can any of you folks with pro
```

```
. ´++Talk:Poverty  /Archive1  ---  "Whatever you do will be in
nblocking their accounts :) ? ----212.199.22.211  22:00, 30 Apr
blocking their accounts :) ?  ---212.199.22.211  22:00, 30 Apri
quations!! What's going on?   ---Â-®+Ö+¦+¦ÔÖÑ+ñ-»+¦+®+ÿ 09:52, 2
st planet and its rapid heat, ---69.255.16.162 20:28, 5 August 2
T REMOVE THIS SECTION  AGAIN! ---Halaqah 11:48, 19 November 2006
A History of Western  Society ---McKay, Hill, Buckler), and the
 to be extremely uninformed.  --- User:Roadrunner  I haven't k
t will be fruitful. --¦-í-¦   --- I've got graphs on world numbe
lobe would be very helpful.   --- Hackeru  Solar land area Ima
```

The replacement of plural –s by –z (e.g. downloadz, filez, gamez) has been found in Wikipedia TPs. This original spelling practice has been probably inherited from the crackers'subculture (which systematically puts “z” for “s” at the end of words to denote an illegal or cracking connection; e.g. *codez, passwordz, MP3z, sitez, FTPz*) or from Afro-American youth culture which extensively replaces the final “s” with “z” in Hip-hop lyrics. Nowadays, this practice has been exported in the advertisement (e.g. *eXperienz,* a Belgian brand of clothes for women, *Allianz*, one of the largest insurers in UK., etc.).

The spontaneity of TPs discourse sometimes leads to misspellings. Spelling and grammatical errors have been frequenty detected in TPs. By contrast, they have been very rarely found in WA as one of the main tasks of Wikipedian contributors is to accurately revise and improve form, grammar and cohesion of encyclopaedic articles. On the other hand, spelling and grammar mistakes in TPs do not appear to reflect, as in mobile texting, a lack of education of Wikipedians, but are simply regarded as typing inaccuracy, the result of a hurried communication in this unconventional writing space. Non-standard and original spellings, as already mentioned, are used without sanction in TPs. By contrast, they are heavily eschewed in traditional and academic writing and, consequently, their use is absolutely forbidden also in WAs.

8.1 Emoticons

Every form of electronic discourse uses punctuation and all-capital letters to signal humor, irony, or intimacy. *Emoticons* have been specifically created to better convey personal feelings (Wilkins, 1991; Boyd & Brewer, 1997) in CMC.

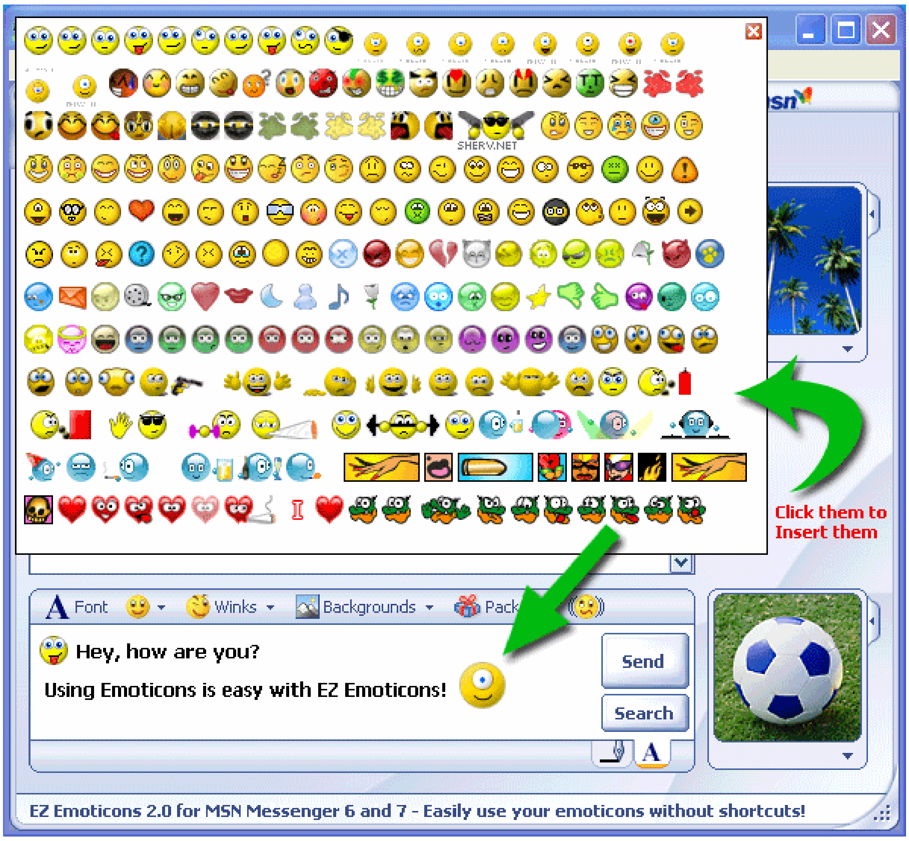
Differently from face to face communication, the lack of paralinguistic features (such as facial expressions, gestures, body posture, and distance conventions) and physical context in CMC, has often led to the misinterpretation of even the simplest utterance. This deficiency has been the main cause of the alternative introduction of *emoticons*, also defined as *smileys* (Rezabeck *et al.,*1995) which represent an extended interpunctuation symbolic system used in most CMC channels.

An emoticon is a small piece of specialized ASCII art used in text messages as informal markup to indicate emotions and attitudes. They are intended to be relatively simple to type, easy to recognize, and most commonly represent stylized facial expressions. Traditionally, the emoticon in Western style is written from left to right, the way one reads and writes in most Western cultures.

David Crystal (2001) also suggests that emoticons are used to fill a void in online communication. He claims that they help to accentuate or emphasize the tone or meaning during message creation and they can be considered a creative and visually-salient way to add expression to an otherwise completely textual form, Constantin *et al.* (2002) claim that emoticons help to establish a current mood or impression of the author. A smile is often represented with a basic smiley :-). The colon represents the eyes, the hyphen is for the nose, and the parenthesis for the mouth. Many variants exist with different symbols substituted for the basic ones. Midget smileys, for example, omit the symbol for the nose, e.g. :) or ;) . A list of some of the most common emoticons follows.

:-)	or :)	Smile or Happy
:-(	or :(	Frown or Sad
:-D	or :D	Open-mouthed smile
:-p	or :p	Smile with tongue out
:-S	or :S	Confused Smile
:-/	or :/	Blank Smile

With the advent of the latest CMC software, *textual emoticons* have been replaced by *graphic emoticons* ☺☺☺☺☺☺☺☺☺☺☺☺☺☺☺) usually based on the generic smiley. Nevertheless other evocative imagery such as hearts, ❤ lips 💋, hands with different communicative functions 👎👍✋ are nowadays very frequently used in the different CMC channels.



Hey, how are you?

Using Emoticons is easy with EZ Emoticons!

EZ Emoticons 2.0 for MSN Messenger 6 and 7 - Easily use your emoticons without shortcuts!

Fig. 45 Examples of graphic emoticons (from MSN Messanger)

Originally, emoticons were fairly simple but over time they have become so complex that are often input using a menu, or popup windows, which sometimes list hundreds of items some also with embedded sounds to bring emoticons to full life (fig. 45). Some correspondences of text based and graphic emoticons are shown in fig. 46.

Type of emoticon	Text-based	Graphical
Happy	:) or :-)	🙂
Sad	: (or :-(	☹
Angry	>:(or :O	😠
Flirty	;) or :P	😘

Fig. 46 Examples of text-based and graphical emoticons

A very recent phenomenon is the emergence of very short video clips, now referred to as *EmotiClips*. They are video snippets containing an expression of emotion. They can be shared on websites, in emails, and through mobile phone messaging to express feelings not unlike video greeting cards. This new form of communication has been used recently by *MTV* and *Paramount Home Entertainment.*

The descriptive analysis of Huffaker *et al.* (2005) proves that more than half (63%) of the total population of bloggers use emoticons, whether in the form of a graphic or a text-based smiley, while the majority of emoticons are *happy* or *sad*, bloggers sometimes use *angry, flirty*, or *tired* emoticons. The following figure portrays the percentage of emoticon types in the study by Huffaker e*t al.* (2005).

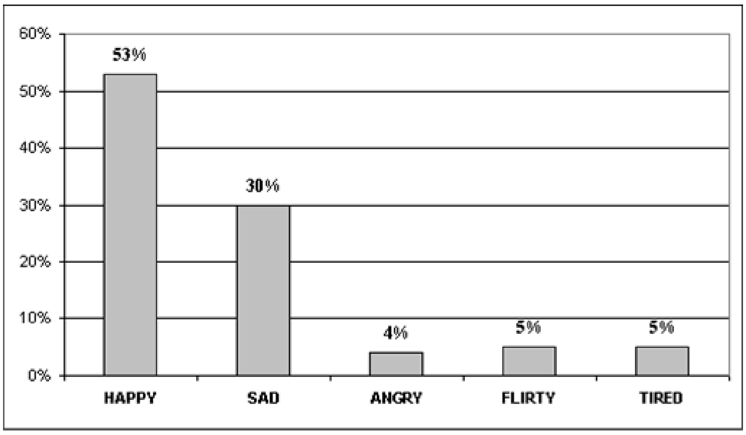


Fig. 47 Overall emoticon use in a blog study (Huffaker *et al.* 2005)

Emoticons are very generously used in weblogs. Their frequency is overflowing in online interactions, partly because they are now often built into CMC applications (such as instant messaging, chat rooms, forums, and blogs).

Wikipedians also use textual symbols to express their feelings and emotions. The most widespread *textual emoticons* (e.g. **:-)** pleasure, humor; **:-(** sadness, dissatisfaction, **; -)** winking, **; -(** crying, etc.) have been detected in TPs. However, textual *midget smileys* are also very common (e.g. **:)** **:(**). By contrast, *joke emoticons* (e.g. **:*)** user is drunk; **:-@** user is screaming; **:-[** , user is a vampire, etc.) are

rarely found while *graphic emoticons* (e.g.: 😀 😆 ☺ 🙂 , etc.) widespread in forums, chats and blogs, are completely absent in TPs (Elia, in press).

Textual emoticons have been found disseminated mainly in TPs and also in other synchronous and asynchronous channels used by Wikipedians to communicate within their community (fig. 48).

Nevertheless, their use is not so widespread as it is in blogs. The total absence of graphic emoticons is certainly due to the wiki syntax which does not allow an immediate and friendly inclusion of image files in TPs. Furthermore, the moderate use of textual emoticons is probably also due to the adult average age of contributors who, differently from teenagers are not very fond of them. Most Wikipedians seem to prefer a more explicit and traditional writing style (already proved by the high lexical density, high nominalization frequency detected in TPs [83]).

However, TPs being a free writing space, no prescriptive rules have to be followed, thus Wikipedians can freely express themselves in their posts, revealing in this way their age, personal writing style and cultural level.

Differently from TPs, the compulsory rules of *Wikipedia Manual of style* strictly forbid the use of emoticons. Thus, as was expected, none of them has never been detected in encyclopaedic pages. The frequency of the most popular emoticons is shown in fig. 48. Some examples of the most popular emoticons follow.

Emoticons				
	Talk Pages	**%**	**Articles**	**%**
:)	138	0.02	0	0
:-)	108	0.02	0	0
;-)	48	0.01	0	0
;)	36	0.01	0	0
:-(	35	0.01	0	0
:-/	21	0.00	0	0
:-1	18	0.00	0	0
:(	15	0.00	0	0
Total	**419**	**0.07**	**0**	**0**

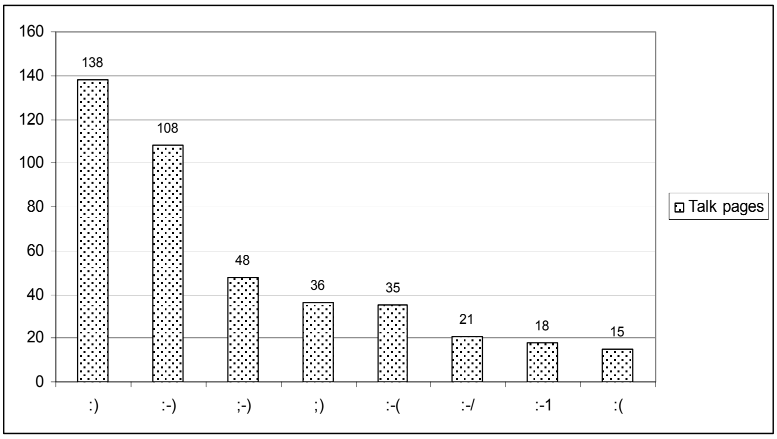


Fig. 48 Emoticons in Wikipedia's TPs

[83] *see* chapter 5, section 7.1

```
y/may not/did/will exist etc. :-) ChrisRed  09:02, 1 December 2
that might provoke a reaction :-). I donÔÇÖt care enough about t
 this article in its entirety :-) Rossrs 14:40, 3 May 2006 (UTC)
Never mind, found it in 10.9  :-) WhiteC 16:40, 28 October 2005
scussion (and summer) I guess :-). I think I had in mind a quite
is article not beeing neutral :-) --Jurgensen 11:20, 4 August 20
 guide writing style, I guess :-) --Jurgensen 12:05, 5 August 20
enue for   being exclusive.   :-) --KSmrqT 21:02, 19 May 2006 (U
 relation to English Grammar? :-) To my knowledge that s   fin
 all have Wikipedia articles. :-)  The others are just waiting
di Air Force Fighter Aircraft :-)(ChrisR) 28/7/05 A flag repers
```

```
m working on it as we  speak  :)), but others seem te be mentio
ce to the readership. Thanks. :) Jesset77 16:33, 29 June 2006 (
e this helps clear things up. :) JoeSmack Talk 21:23, 15 June 2
yourselves. Good job to all!  :) JoeSmack Talk 19:23, 16 June 2
irst place - thus the images! :)  JoeSmack Talk 02:16, 22 Augu
 Thanks for the invitation... :) First off,    let's establi
     ew" or   something?      :) In his native tongue, of cours
  (UTC)  Hoorah! thanks       :) ÔÇöQuiddity 18:50, 17 January
ore tomorrow, here ya go [3]! :) JoeSmack Talk(p-review!) 00:49
 danged word. it gets tiring. :) JoeSmack Talk(p-review!) 16:41
 the content.  Glad to help.  :)  BTW, thereÔÇÖs a perfectly g
e bold and see where it goes. :) JoeSmack Talk(p-review!) 17:05
```

8.2 Wikispeak Jargon

As McLuhan (1964) pointed out in the sixties: “the medium is the message”. This means that every form of textual expression cannot be decontextualized as the medium has not only a functional role, but it intrinsically shapes the nature of the message. As a consequence, the electronic channel has also definitely affected digital writing since the use of the keyboard implies a mediated and slower form of communication when compared to face to face interaction, as well as a restricted available space delimited by the monitor’s width.

The language spoken/written on the Internet, the *NetSpeak Jargon* is one of the most creative domains of contemporary English. It is also known as *Netlingo* or *Weblish*. What makes it so interesting, as a branch of ICT language and as a new form of online discourse, is the way it relies on characteristics belonging to both speech and writing. Tracking its development is an interesting way of linguistically documenting the progression of the ICT language which is evolving on a national and international level. Like most jargons, internet slang boosts authors and readers, making them appear to share their specialized knowledge of a complex medium. *Netspeak Jargon* currently spoken by net surfers on the web, has its origins in the technological vocabulary once used only by computer programmers and hackers. Thus, the matrix of *Netspeak Jargon* is the hackers’ *Jargon file*[84] which is a collection of slang terms used by various subcultures of computer hackers for fun, social communication and technical debate.

[84] The *Jargon File* is a glossary of hacker slang. The original Jargon File was a collection of slang words from technical cultures including the *MIT AI Lab*, the *Stanford AI Lab* (SAIL), and others of the old *ARPANET* communities.

Hackers, as a rule, love wordplay and are very conscious and inventive in their use of language. Linguistic invention in most subcultures of Western countries is largely an unconscious process. Hackers, by contrast, regard slang formation and use as a game to be played for conscious pleasure. Their inventions display an almost unique combination of the enjoyment of language-play with an educated and powerful intelligence. Electronic media have well adapted to the spreading of this new slang. The results of this process give perhaps a uniquely intense and accelerated view of linguistic evolution in action. In a page of *The Jargon File*[85] website one can read:

> It is usually claimed that low-context communication (characterized by exactness, clarity, and completeness of self-contained utterances) is typical in cultures which value logic, objectivity, individualism, and competition; by contrast, high-context communication (emotive, elliptical, heavily coded, nuance-filled, multi-modal) is associated with cultures which value subjectivity, consensus, cooperation, and tradition. ICT linguistic domain is themed around extremely low-context interaction with computers and exhibits primarily "low-context" values, but on the other hand, it cultivates an almost absurdly high-context slang style.

Thus, *Jargon file* challenges the traditional linguistic and anthropological assumptions, since it is a miscellanea of *low-context* and *high-context* languages and cultures (Hall, 1976).

The dynamicity of Netspeak Jargon reflects the very rapid development of new concepts and the need to communicate them. Netsurfers have coined new words for their new world, to say new things for which they do not yet have adequate references. Neologisms are normally strictly connected to the terminological paradigm actually dominating a specific field of knowledge. Teenagers, the most frequent users of instant messaging and other forms of simultaneous online communication, have been the chief proponents and users of this emerging style.

Netspeak variety, talked by Wikipedians inside their community, has been here defined as *Wikispeak Jargon*. It is the jargon which Wikipedians use when they talk about technical operations and activities connected to their authorial and collaborative writing work in the different community channels and especially in TPs. One of the main peculiarities of *WikiSpeak Jargon* lies in the new lexicon invented. Discourse communities have the freedom to break the conventional linguistic code, creating and altering the language in use. A new discourse community is barely conceivable without the use of neologisms, or new interpretations of old words to describe and explain reality in new ways. Thus, a large number of new words, defined as wikilogisms[86], have been coined inside the Wikipedia community.

8.3 Wikispeak word formation process

During the last century, various linguists have developed taxonomies for classifying the different types of word formation. According to David Crystal (1995:429) there are a number of

[85] *Jargon File*, version 4.4.7 httalk page://catb.org/esr/jargon

[86] A *wikilogism* is a form of neologism originating on a wiki project page

common processes for word formation in the new ICT domain (affixation, backformation, compounding, conversion, acronym, initialism, blending, clipping). Both Eble (1996) and Gotti (2002), working on different slangs, college-age and criminal slang respectively, found that the same processes in the formation of slang resemble those commonly adopted in the coining of neologisms in the standard language. In addition, Gotti (2002) found examples of onomatopoeia, borrowing, clipping, graphic iteration, metonymy, synecdoche, metaphor, personification, specialization, generalization, semantic shift, and ellipsis.

Other processes identified in slang formation also include acronyms and ironic semantic shift (Munro 1989:6). Algeo (1999) in the *Cambridge History of English language*, develops a classification system for word-formation based on four factors depending on whether:

- the word has an etymon based on earlier words;
- the word omits any part of an etymon;
- the word combines two etyma;
- any of the etyma is from another language.

He uses six groupings: *composites* (prefixes, suffixes, compounding) *shortenings* (acronyms, initialisms, clipping, backformation), *blending*, *shifting* (functional or semantic shifts), *loans*, and *new creations*.

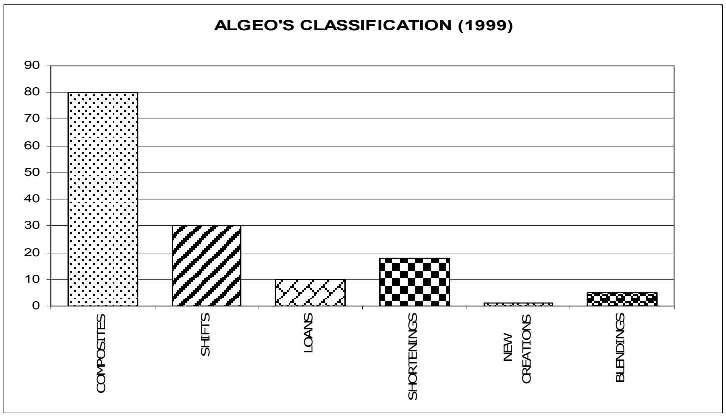


Fig. 49 Algeo's classification (1999)

Algeo shows the different percentages of word-formation types in a set of samples taken from different linguistic domains (fig. 49). He demonstrates the high frequency of composites in his analysis (especially compounds), over other types of word-formation.

Frequency of shifts and shortenings is also high, but less significant, occurrences of blends and loan words are low and the presence of new creations is irrelevant (Algeo, in Shortis 2001: 53-57).

The distinctiveness of WikiSpeak Jargon certainly lies in its lexicon where many word formation processes take place, including several ludic innovations. The amount of new words coming into WikiSpeak provides an opportunity to see Algeo's classification in action. Since new linguistic phenomena have been detected, to explore the occurrences of the most common new word formation processes in WikiSpeak Jargon, the *Wiki glossary* [87] available for the newbies in the Wikipedia community pages, has been annotated using Algeo's word-formation taxonomy with the addition of the following three new categories more appropriate to the specific case study: *double clipping (<2cl>)*, *soft semantic shifts* (<sss>) and *loans from ICT Language* (<ictll>) (fig. 50).

D*ouble clipping*, e.g. *dicdef (*dictionary definition*), medcab* (mediation committee*), medcom* (mediation committee*), permcat (*permanent category) means a word in which two joined clipped words have been found. The term *soft semantic shift* has been used to define a light change in the original meaning of the word which has been semantically recontextualized in the new Wikipedia community (e.g. *article, mediation, shortcut, vandalism).* Thus, the new specific connotation of the word can be easily understood, as the example below from the Wikipedia glossary shows.

> *In the outside world,* vandalism means a willful or malicious destruction or defacement of public or private property. *On Wikipedia,* it means deliberate defacement of Wikipedia pages. This can be by deleting text or writing nonsense, bad language, etc. The term is sometimes improperly used to discredit the views of an opponent in edit wars.

Finally, the tag <ictl> (*loans from ICT language)* has been attached to the terms which properly belong to the ICT language domain. ICT terms have been included in Wikipedia Glossary by its compilers, as their meaning can be unintelligible to a novice (e.g., *boilerplate, bot, cruft, template, tag, etc.*). The *Wikipedia Glossary Corpus* is made up of 242 words (14 September 2007). Fig. 50 shows the annotation system which has been used to analyze the word formation process in action in WikiSpeak Jargon with the specific and the total occurrences for each category:

SHORTENINGS	Initialisms	<in>	53	85
	Clippings	<cl>	14	
	Double	<2cl>	10	
	Acronyms	<ac>	8	
COMPOSITES	Compoundings	<co>	47	67
	Prefixes	<pr>	19	
	Suffixes	<su>	1	
SHIFTS	Semantic	<ss>	33	62
	Soft semantic	<sss>	23	
	Functional	<fs>	6	
LOANS	***ICT language***	<ictll>	14	14
NEW		<nc>	9	9
BLENDINGS		<bl>	5	5
		TOTAL 242		

Fig. 50 Wikipedia Glossary Annotation

[87] *Wikipedia Glossary* httalk page://en.wikipedia.org/wiki/Wikipedia:Glossary

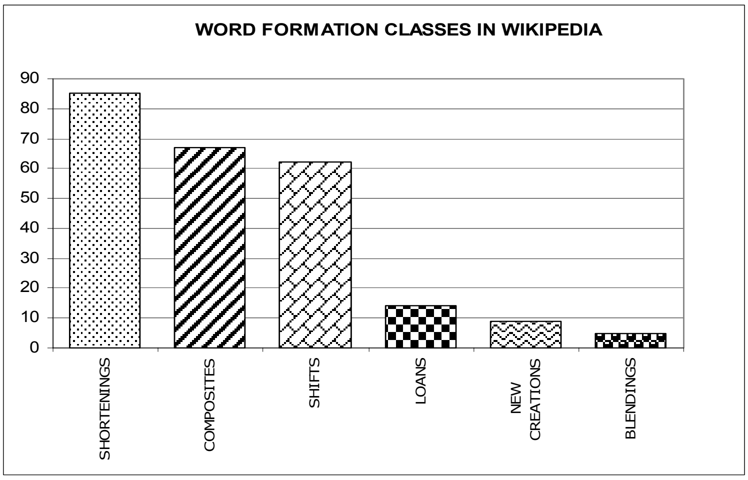


Fig. 51 Word formation classes in Wikipedia

Some concordances of the queries made on the glossary corpus are shown below.

```
Wikipedia:convenience links  <co>  Copyedit  A change to an ar
ght to a subordinate topic.  <co>  Editcountitis  A humorous t
ernal link  See free link.   <co>  Interwiki  A link to a sist
tion?; Wikipedia:Mediation.  <co>  MediaWiki  The software beh
binations.  See also Meta.   <co>  Metapage  Page that provid
ki use, "not applicable".    <co>  Namespace  A way to classif
ee also Wikipedia:Redirect.  <co>  Redlink  A wikilink to an a
ve. See also "rouge admin".  <co>  Rollback  To change a page
g via [edit] links" option.  <co>  Self-link  A Wikilink conta

Wikipedia:Mediation Cabal.   <cl>  Cat  "Category" or "Categoriz
r topic to work on or read.  <cl>  Contribs  Short for contribut
erative editing for a week.  <cl>  'Crat  Short for Bureaucrat,
omputer and video games.     <cl>  dab  See Disambiguation.  <i
ipedia is not a dictionary.  <cl>  Diff  The difference between
ge.  See also m:Help:Diff.   <cl>  Disambig  See Disambiguation.
ough for a Wikipedia entry.  <cl>  nom  Short for "nomination,"

:Collaboration of the week.  <2cl>  ArbCom  Abbreviation for Wik
so known as "sea of blue").  <2cl>  Dicdef  Also used: Dictdef.
 to it.  See also Repoint.   <2cl>  Dupe  Short for a duplicate
ween them largely academic.  <2cl>  medcab  The Mediation Cabal
 Wikipedia:Mediation Cabal.  <2cl>  medcom  The Mediation Commi
 See Wikipedia:Peer Review.  <2cl>  Permcat  A permanent catego
 include CFD, RFA, and AFD.  <2cl>  Prod  Proposed deletion.
re for the original usage.)  <2cl>  Sysop  See Admin.

Cross-namespace redirects.   <ac>  COI  Acronym for Wikipedia:
pedia:Conflict of interest.  <ac>  COIN  Acronym for Wikipedia
ts consensus for promotion.  <ac>  FAC  Featured article candi
 certain points of view.  I  <ac>  IANAL  An abbreviation for
in nomination for deletion.  <ac>  NOR  The Wikipedia policy t
well as on many user pages.  <ac>  OR In Wikipedia, original r
on. Used in edit summaries.  <ac>  SPA  Short for Single Purpo

aggato 15 tags 245 terms     <in> 1RR  See three-revert rule  <
1RR  See three-revert rule   <in> 3RR  See three-revert rule
 users.  Also used: Sysop.   <in> AfD  The Wikipedia:Articles f
 of some terms used on AfD.  <in> AGF  Abbreviation for "assume
a page#Links,_URLs,_images.  <in> AOTW  Abbreviation of Wikiped
umber of edits required.     <in> BFN Bad faith nomination  A
:Be bold in updating pages.  <in> BJAODN  Abbreviation for Wiki
```

```
r inclusion of biographies. <bl>  Ghits  "Google hits" - the
kin, Nostalgia, and Simple. <bl>  Smerge  A contraction of "s
 See also Wikipedia:Portal. <bl>  Wikipediholic  Also used: W
also Wikipedia:WikiProject. <bl>  Wikiquette  The Wikipedia e
ikistress Meter, Wikistress <bl>  Wiktionary  A Wikipedia sis

of Style, should be plain). <fs> Deletionist  Someone who act
ges and media for Deletion. <fs> Inclusionist  A user who is
to have undergone link rot. <fs> Listify  To delete a categor
lp:Merging and moving pages <fs> Mergist  A user who adheres
e also Wikipedia:Userboxes. <fs> Userfy  To turn a page in th
e also Wikipedia:WikiFairy. <fs> Wikify  To format using Wiki

if you have not logged in.  <ss> Anchor An HTML term for code
rats.  Also used: Crat.     <ss> Cabal  Sometimes assumed to
ed.  Fancruft  See Cruft.   <ss> Forest fire  A flame war whi
ipedia:Mediation Committee. <ss> Meat puppet  An account crea
e also Wikipedia:Meta page. <ss> Mirror  A website other than
s are repeatedly recreated. <ss> Sandbox  A sandbox is a page
rtcuts for a complete list. <ss> Skin  The appearance theme i
Wikipedia:Snowball clause.  <ss> Sock puppet  <ss> Sock  An
ut only one's own comments. <ss> Stub  An article usually con
11.  See also m:Transwiki.  <ss> Troll  A user who incites or
User:AmiDaniel/VandalProof. <ss> Village pump  The main commu

pedia:Arbitration Committee <sss> Arbitration The final step in
 usually given less weight. <sss> Article An encyclopedia entry
bans a troublesome editor.  <sss> Barnstar Barnstars are a li
are European or American.   <sss> Cut and paste move  Moving a
pedians.  See also Wikify.  <sss> Dead-end page  Page that has
re and MeatBall:ForestFire. <sss> Fork  A splitting of an enti
 See also Wikipedia:Revert  <sss> Revert war  See Edit war.
te is called a "wolf vote". <sss> Shortcut  A redirect used wi
es.  See also m:Vandalbot.  <sss> Vandalism  Deliberate deface

articles on living people.  <ictl> Boilerplate text  A standard
Wikipedia:Boilerplate text. <ictl> Bot A program that automatica
imit comments in HTML code. <ictl> Community Portal One of Wikip
dia:WikiProject Laundromat. <ictl> Link rot  Because websites ch
ld) indicates a minor edit. <ictl> Main Page  The page to which
e also Wikipedia:Namespace. <ictl> Newbie test  Also used: newb
lso Wikipedia:PokÚmon test. <ictl> Portal    Portal <in>   POTD
Countering systemic bias    <ictl> Tag In addition to its usual
See also Wikipedia:Taxobox. <ictl> Template A way of automatica

Wikipedia:Cleanup process.  <nc> Climbing the Reichstag  A hu
G <in> GA  Good article.    <nc> Gdanzig  An edit war over wh
 cleanup or stub sorting.   <nc> Per, Per Nom, Per X  A comm
dia:This page is protected. <nc> Protologism  A word that is
kiProject_Red_Link_Recovery <nc> Refactor  To restructure a d
ols to do this more easily. <nc> Rouge admin  A misspelling o
ee also Wikipedia:Subpages. <nc> Suitly emphazi  A phrase wit
r.  See also polarization.  <nc> Tyop  A cute misspelling of
tton with a high frequency. <nc> Wikipe-tan  Also used: Wiki-
```

The findings (fig. 51) have shown a high frequency of *shortenings* in TPs (85 occurrences). In particular *initialisms* have recorded the highest frequency (53 occurrences) followed by *clippings* (14 occurrences) and *double clippings* (10 occurrences). Fewer acronyms have been detected (8 occurrences) and the frequency of composites is slightly lower. Most of them are compounds (47 occurrences) and are followed by prefixed words (19 occurrences).

Only one suffixed word has been found in the Wikipedia Glossary. *Shifts* seem to be the third most important category. *Soft Semantic Shifts* are the most numerous ones (34 occurrences). As in Algeo's analysis, also in Wikipedia community pages a low rate of *blends*, *loans*, and *new creations*, has been noticed.

A comparison of the macro categories in Algeo's classification and in Wikipedia is provided in fig. 52 a,b. In the two graphs, the main discordant data is related to the higher number of *shortenings* in WikiSpeak (*initialisms* and *clippings*) while the number of *composites* seems more or less to be similar. The higher number of wiki *shortenings* is probably due to the need for quick typing and to the use of specific lexicon related to daily technical and editing operations already familiar to the community members. With reference to Algeo's classification, the corpus analysis has revealed that a recurrent word-formation process in Wikispeak is the use of *affixation* (especially prefixes).

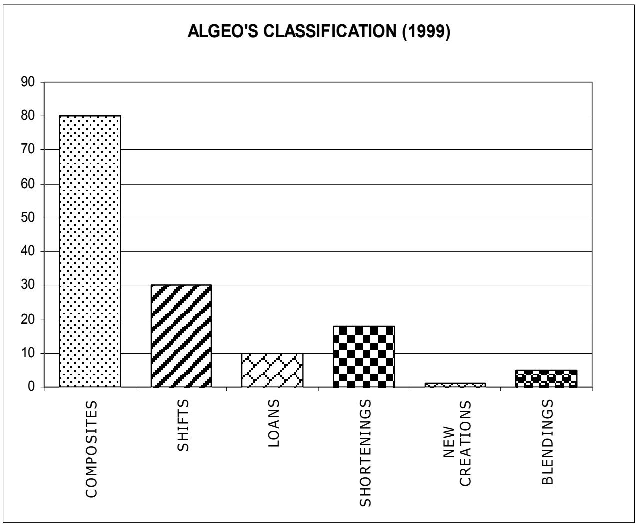


Fig. 52a Algeo's word formation classification

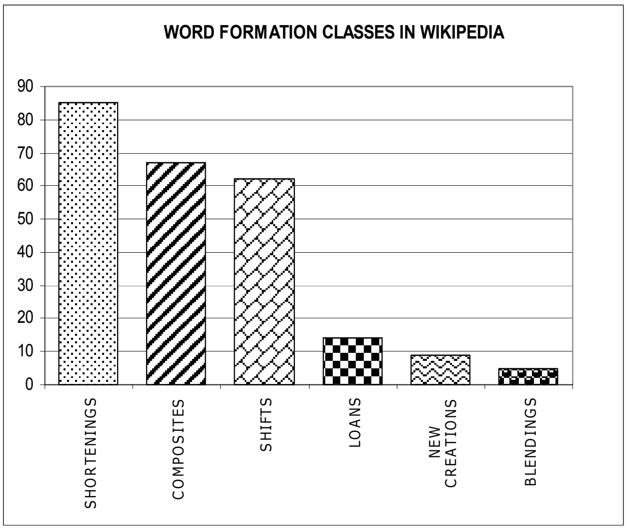


Fig. 52b Word formation classes inWikipedia

Prefixes, as a group of letters at the beginning of a word, change its basic meaning. They can make words negative or make words with opposite meaning. Typical Wikipedian prefixes include *de-* (*desysop, dewikify*), *un-*(*unencyclopedic, unwiki*), *sub-*(*subpage*), *trans-* (*transwiki, transclusion*).

Suffixes, on the other hand, as a group of letters at the end of a word change the word's meaning and often its grammatical function. As already claimed, they are rarely used in Wikipedia Community Pages, with the only example found in the *glossary* being the suffix *-bot* (*vandalbot*).

A popular method of creating new words is *compounding*, that is the combination of two existing words (e.g. noun + noun, adjective + noun) to make new words (e.g. *namespace, mediawiki, rollback, infobox*). In Wikipedia some compoundings are written as single words (*Editcountitis*), some as series of words (*Meta page*) and some with hyphens (*Sock-puppet*). Some have an obvious meaning, while others are more complicated. The element which repeatedly turns up is obviously the word *wiki* (e.g. *wikibooks, wikilink, wikispam*, *wikislap,* etc.). Veterans avoid an excessive use of *wiki-* compounding as it is considered "unencyclopaedic" and cliché. However, it is tolerated when it refers to technical terms (e.g. *wikilink*), to an existing subject (such as *Wikimedia Foundation*), or when it is useful in communicating wiki-specific ideas (e.g. *Wikifairy, WikiGnome*).

Within ICT domain, *shortenings* are so commonly used that their full forms are rarely found. Text based facilities like email, chat or forums, blogs and wikis are riddled with short forms, and it is not only the question of new words, but also the way of combining the elements of written language which need to be taken into account.

As Crystal (1995) claims *to use an abbreviated form is to be in the know' part of the social group to which the abbreviation belongs*. Shortenings are made up of acronyms, initialisms and clippings. Since WikiSpeak is a written form of CMC, it makes an extensive use of initialisms; by contrast, a similar broad use of acronyms has not been detected in the Wikipedia community[88].

Some of the most common and recurrent initialisms met in TPs are: *NPOV* (Neutral point of view), *POV* (Point of view), *COTW* (Collaboration of the week), *IFD* (Images for Deletion), *RC* (Recent changes), *RfM* (Request for mediation), *VP* (Village Pump), *AOTW* (Article of the week). Initialisms found in TPs are not restricted to words, but can also imply sentences e.g. *IMHO* (In my humble opinion), *IMO* (In my opinion), *WDYS* (what did you say?), *CIO* (check it out), *CID* (consider it done), *RTM* (read the manual) etc. Some of them are like puzzles as the sound value of the letter, or numeral, acts as a syllable of a word, e.g. *B4N* (bye for now), *CYL* (see you later), *U R* (you are), *2L8* (too late), *2g4u* (too good for you), etc. A query made for some of the most widespread acronyms has given the results shown in fig. 53.

Acronyms (an axample)		
	Talk Pages	**Articles**
POV	531	0
NPOV	137	0
IMHO	49	0
IMO	53	0

Fig. 53 An example of acroyms' frequency in TPs

Some concordances of *NPOV*, *POV*, and *IMHO* are reported below:

```
     2006  eh TMS was trash   NPOV ... what today is considered c
0.214 doesn't seem to respect NPOV. I'm not sure  whether everyo
 page does not show up in the NPOV  disputes category. -- Kjko
ok then, i guess that is more NPOV. --Tsinoyboi 06:32, 2 October
n the debate and departs from NPOV. Well, they are mentioned in
y opinion undue weight per WP:NPOV. Oops, now IÔÇÖm on the soapbo
behind it, etc, maintaining a NPOV; or removed. As it stands it i
view held by some, it isnÔÇÖt NPOV. Thryduulf 16:32, 14 Jan 2005
profound writer? That isnÔÇÖt NPOV and this paragraph needs to be
 0:35, 11  April 2006 (UTC)   NPOV This article seems to have a
icle needs to be reviewed for NPOV violations (and general   ac
```

[88] *Acronyms and initialisms* are abbreviated words. There are different interpretations of the two terms. Acronym is a pronounceable word (scuba, Nato radar), usually written in lower case letters and governed by phonological rules, they tend to have a vowel in the middle of consonant clusters. On the other hand, initialisms are free from constraints, are usually written in capital letters, sometimes with a full stop between letters and each individual letter is pronounced (XML,TLA, BBC)

```
 .") Whether or not Pinker's  POV is correct is   not for us to
 countries section and  other  POV issues are old news to you guys
icle which can deal  with its  POV problems on its own. It's absol
absolutely riddled with wildly POV and  unsourced claims and it d
ry of "wiki" to  protect from  POV problems. Since I am just a mer
UTC) Most honest editor put a  POV flag, they cite original resear
ng   organized mob pushing a   POV unfortunately. Rcnet 18:57, 13
urce is of little value due to POV bias, hence under the circumsta
nk you for that self-righteous POV stream. You are not original, a
erits. This section is grossly POV and deliberately includes

tive bio should take priority IMHO.  NBeale 23:52, 19 December 2
tants of the region, at least imho, fit the definition of ÔÇ£any
ed will return.  In summary,  imho it has the potential to become
gument, so forget about that. IMHO,    the value of a currency
y sacrifice, I mean to hurt). IMHO, this is hardly going to happe
nism for Good Article review. IMHO any problems this page has  a
he subject of french history (IMHO). Anyone that shows up with
links (usually the best part, IMHO, as domain specific webpages
ical   for this discussion.   IMHO  Probably the simplest correc
9 November 2006 (UTC)         IMHO, the Internet is the largest i
e just back to the original,  IMHO) unhelpful opening. ND 06:03,
 extreme, the "pun"  theory.  IMHO, both distort Marx' argument,
 16:54, 17 July 2006 (UTC)    Imho the first bombastic sentence s
osition in this market space. IMHO, the statement  should stay.
age against another. that is, imho,  wikipedia not a software co
```

Basically, it is recommended that a moderate use of *TLAs* (Three Letter Acronyms), *initialisms* and *wikilogisms* be made in TPs, otherwise the message could be incomprehensible especially to newbies. Furthermore, when encyclopaedic articles are titled with wikilogisms they are immediately elected for deletion as no type of jargon expressions can be used in encyclopaedic articles. This rule has been explicitely expressed at the beginning of the *Wikipedia Glossary* page [89] which declares:

> While the definitions below may be useful for understanding and communicating on project and talk pages, and with edit summaries, remember to explain jargon in encyclopedic articles, and write them in language which is readily understandable without specific knowledge of the Wikipedia project. This is an encyclopedia, not text messaging! Don't overdo the use of Wikipedia jargon such as shortcuts on talk pages and edit summaries, either, at least not without providing explanatory links to the appropriate pages.

Then, to demonstrate its incomprehensibility, the following example is reported:

> **WTF? OMG! TMD TLA. ARG!**
> Basically, when WP:EDIANS CITE pages IN the PRJ NS, they often refer TO them using CUTS like "BEANS", "BALLS", and "NFCC". While these ABB are GREAT for RDRing to a particular page you USE often, it's probably a BAD idea to make A POINT of using these TLAs in daily TALK, lest your discussion end UP as NONSENSE like the TITLE of this page.

It means:

> ***W**hat **T**he **F**uck? **O**h **M**y **G**od! **T**oo **M**any **D**amn **T**hree **L**etter **A**cronyms. **ARRRGGGHHH!***
> When Wikipedians refer to pages in the Wikipedia namespace, they often use shortcuts like "WP:BEANS", "WP:BALLS", and "WP:NFCC". While these are quick jargon, and get you quickly to a particular page you use often, it's probably a bad idea to habitually use these three letter acronyms in daily conversation, lest your discussion end up as nonsensical as the title of this page.

[89] Wikipedia Glossary httalk page://en.wikipedia.org/wiki/Wikipedia:Glossary

Typical of WikiSpeak is the use of clippings (abbreviating or cutting off a word, at the beginning, at the end or at both ends of a word), i.e *admin* (administrator), *disambig* (disambiguation), c*ontribs* (contributions), *diff* (difference), *nom* (nomination), *dupe (*duplicate article), etc. An original association of the above-mentioned word formation processes can be noticed in the matching of two clipped words to make new compound words, e.g. *ArbCom (*Arbitration Committee), *CopyViol (*Copyright Violation), *Medcom* (Mediation Committee*)*, *DicDef* (Dictionary Definition), *SysOp* (System Operator), etc.

Although it is the less frequent category, blended words have also been detected in WikiSpeak Jargon. An older term to define this technique is *portmanteau.* This linguistic phenomenon is not recent as fashion for such formation began in the 1890s. Through the blending process part of one word is joined to part of another, and enough of each word is retained so that the elements are recognizable e.g. *wikipediholic, wikiquette*, *wiktionary* and of course *Wikipedia.*

Semantic shift is a linguistic phenomenon which has been very frequently detected in Netspeak. Many words are taken from standard and colloquial English and applied to new ideas or protocols. On a page of the *Study World*[90] website it is written:

> A gopher is not a furry rodent on the Internet. A *gopher* is a software program designed to gopher through the vast amount of information so that the user can find what she's looking for.A *server* is not a waitress or waiter; a server is another computer that tells your machine what it needs to know to communicate on the net. A *handle* is not a part of a coffee cup; a handle is a nickname. A *shell* isn't the thing a clam lives in; it's the command system that allows you to enter commands to communicate with the machine on the other end.

Numerous semantic shifts have been detected in Wikipedia Glossary (e.g. *forest fire, sock-puppet, pokemon test, village pump troll, etc.)* nevertheless, *Soft Semantic Shifts* are also very frequent in Wikipedia community pages (e.g. *orphan, stub, reincarnation,* etc.) [91].

8.4 Wiki graphology

Distinctive graphology is also an important feature of WikiSpeak Jargon. All orthographic features have been affected. For example, the status of capitalization varies greatly. As in forums, chats and blogs, a strong tendency to use lowercase (e.g. *i want*) has been noticed also in TPs. The "lower-case default mentality" has a long tradition in cyberspace and implies a different use of capitalization.

Within the Internet the capitals are, therefore, a specially marked form of communication. Messages wholly in capitals are considered to be "shouting" and usually avoided. Although asterisk and spacing have the same function (e.g. it's VERY important, it's *very* important, it's _very_

[90] *Netspeak*: An Analysis of Internet Jargon in *Study world* http//www.studyworld.com/

[91] A list of the most frequent *Wikispeak Jargon Terms*, according to the word formation criteria, is provided in *Appendix*.

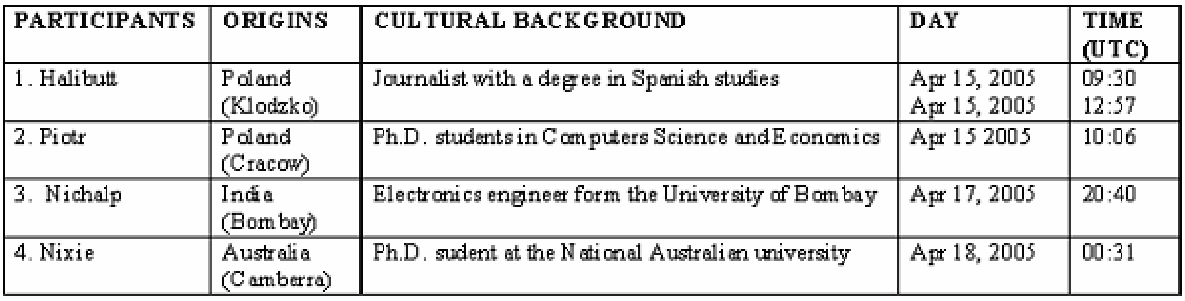
important, it's v e r y important) words in CAPITALS add extra emphasis because of their intertextual relation to the comic culture which is very influential on all artefacts of cyberculture (Wyss, 2000).

A distinctive feature of wiki graphology lies in the way two capitals are used: one initial, one medial. This phenomenon is called *BiCaps* (bicapitalization) or *CamelCase*, and is widespread in TPs (e.g. *MediaWiki, WikiProject,* etc). It is the practice of writing compound words or phrases where the words are joined without spaces, and each word is capitalized within the compound. The name comes from the uppercase "bumps" in the middle of the compound word, suggesting the humps of a camel.

CamelCase is a very interesting example of how a programming language influences the wired style. It was originally used in hackers' communities as a word joiner alternative to the underscore based style and later in the original wiki markup language to create hypertextual links before the invention of [[_]] double square brackets. Nowadays it has become fashionable in marketing to identify names of products and companies. Outside these contexts, however, BiCaps are rarely used in formal written English, and most style guides recommend against its use.

8.5 An analysis of a Talk Page: Klodzko

In order to give an empirical demonstration, an analysis of a TP (18, April 2006) related to the article Klodzko[92] is shown. The investigation of both the Klodzko TP and the contributors' user pages have revealed linguistic habits, identity, nationality, and cultural background of the four Wikipedians involved, as well as date and time of the contributions posted (fig. 54). The nationality of the first two Wikipedians (Halibuttis and Piotr) is Polish, while the other two (Nichalp and Nixie) are Indian and Australian respectively. Their different origins and high level cultural background [93] (journalist, doctoral students, engineer) demonstrate how heterogeneous and culturally rich the global collaborative writing process is.



PARTICIPANTS	ORIGINS	CULTURAL BACKGROUND	DAY	TIME (UTC)
1. Halibutt	Poland (Klodzko)	Journalist with a degree in Spanish studies	Apr 15, 2005 Apr 15, 2005	09:30 12:57
2. Piotr	Poland (Cracow)	Ph.D. students in Computers Science and Economics	Apr 15 2005	10:06
3. Nichalp	India (Bombay)	Electronics engineer form the University of Bombay	Apr 17, 2005	20:40
4. Nixie	Australia (Camberra)	Ph.D. sudent at the National Australian university	Apr 18, 2005	00:31

Fig. 54 Participants in Klodzko TP

[92] *Klodzko* httalk page://en.wikipedia.org/wiki/Klodzko

[93] Personal information on Wikipedians have been inferred by personal *user pages*

The TP contains 496 words organized in 37 lines through 6 different turn takings. The length of the interaction was 4 days (from the 15th to the 18th April 2006). The atmosphere and the tone of the discussion was highly positive, revealing the open-mindedness of the contributors, for instance: *input from others is particularly important* (line 1), *I appreciate any comment and/or corrections* (line 2).

Politeness has been manifested through expressions which communicate agreement and through the use of the conditional tense which convey consideration and kindness, for instance: *good history, but…* (line 4), *Agreed!* (line 11), *yes, my mistake* (line 13), *I'd expect* (line 20), *The history section should be shortened* (line 24);

The Wikipedians'collaborative and helping attitude has been expressed through comments such as: *See wikipedia references …*(line 7), *try to avoid…*(line 7), *I recommend using …*(line 8).

Occurrences of the new lexicon used in WikiSpeak are found in the use of initialisms, acronyms, clippings and wikilogisms, e.g.: *FAs* (Featured articles - lines 8/17), (also FArticles - line 16), *AFAICT* (As Far As I Can Tell -line 17), *FAC* (Featured Article Candidates - line 26), *Pics* (Pictures - line 17), *Dividers club* (line 17), *UTC* (Coordinated Universal Time), *30k* (thousand = kilobyte →K - line 20).

It is interesting to notice that the plural form has been preserved, even though the nouns have been shortened, as in the case of the acronym 'FAs' and the clipped word 'Pics'. To convey emotions, a *basic smiley* (line 13) and two *midget smileys* have been used in this TP: e.g. :-((dissatisfaction - line 13), :> (sarcastic remark - line 20), :) (joking statement -line 22).

486 *tokens* and 257 *types* have been found in Klodzko TP, thus the *lexical density* of the linguistic interaction is very high (52.9 %) although lower than the associated WA (60.9 %). Sentences have an *average length* of 9.92 words in Klodzko TP, thus also in this case, they are decisively shorter than in the WA where 19.51 average words per sentence have been detected.

Examples of the informal style in use in Klodzko TP, have been reported in the concordances below. In particular, four occurrences of *reduced forms* and three occurrences of *synthetic negations,* typical of spoken language (Biber, 1988), have been found. Furthermore, the 14 occurrences of the deictic personal pronoun *I* and the 5 occurrences of *you* testify to the high subjectivity and addressivity of the analyzed TP.

```
  ey have  that  Klodzko doesn't: Johannesburg, Marshall, Texa
2005   (UTC)    Agreed.      I'm starting right away    You
ht? Wikipedia:References doesn't say    much.    Yes, my
odzko :>). With economy etc. I'd expect it to double in
 As  pointed above, there  isn't much to say on the city. Sinc
rtant commerce centre but that's about   all;   Transport
ation chart is a good start, I'd consider   turning it into
```

```
ipedia:Peer review/Klodzko    I wrote virtually all the text m
 is particularly  important.  I think it is factually accurate
accurate and fairly complete. I think it  would make a great
ake a great featured article. I appreciate any comments and/
 in cases of sources and such I recommend using Wikipedia:Foot
 2005   (UTC)    Agreed.      I'm starting right away    Yo
      s, my mistake. Will do (I guess you mean the source for
rticles, the basic difference I see is the length (which is a
lodzko :>). With economy etc. I'd expect it to double in
 to double in    size, but    I doubt it would became a target
e the better. Say, Halibutt - I    wonder - why Klodzko? :)
ll town  of 30k inhabitants,  I would put you on to Kalimpong,
) These are some things that  I think could be included to mak
lation chart is a good start, I'd consider   turning it into
```

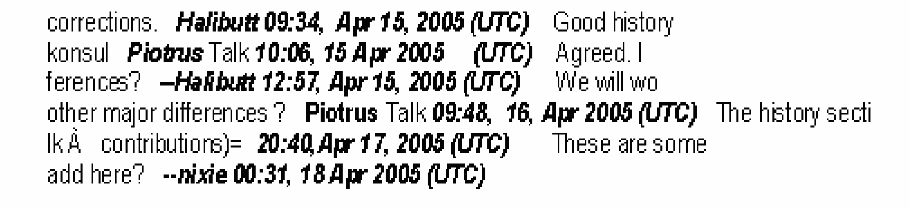
```
  I'm starting  right away   You mean Wikipedia:Cite sources,
  stake.   Will do (I guess you mean the source for the pop
 statistical information that you can   get your hands on
dd some more pictures here if you have some, and are  there
here any local festivals that you could add here?  --nixie
```

Conversational, interactive and direct style of TPs is confirmed in the five occurrences of interrogative sentences. The spatial deictic *here* has also been detected three times. The above mentioned occurrences are reported below.

```
Wikipedia:Cite sources, right ? Wikipedia:References doesn't s
ource for the pop table, right?)    Yes.    As to other
s. Any other major differences?  --Halibutt 12:57, Apr 15, 20
- I    wonder  - why Klodzko  ? :) --Piotr Konieczny aka Proko
tivals that you could add here?  --nixie 00:31, 18 Apr 2005 (
```

```
and railways passing through  here), perhaps a more detailed
tions, add some more pictures here if you have some, and are
festivals that you could add  here? --nixie 00:31, 18 Apr 2005
```

As can be inferred from the examples below, the explicit mention of authors (e.g. Halibutm Piotrus, Nixie), in addition to date and time of the published posts, represent a clear contextual reference.



corrections. **Halibutt 09:34, Apr 15, 2005 (UTC)** Good history
konsul **Piotrus** Talk **10:06, 15 Apr 2005 (UTC)** Agreed. I
ferences? **–Halibutt 12:57, Apr 15, 2005 (UTC)** We will wo
other major differences ? **Piotrus** Talk **09:48, 16, Apr 2005 (UTC)** The history secti
lk Å contributions)= **20:40,Apr 17, 2005 (UTC)** These are some
add here? **--nixie 00:31, 18 Apr 2005 (UTC)**

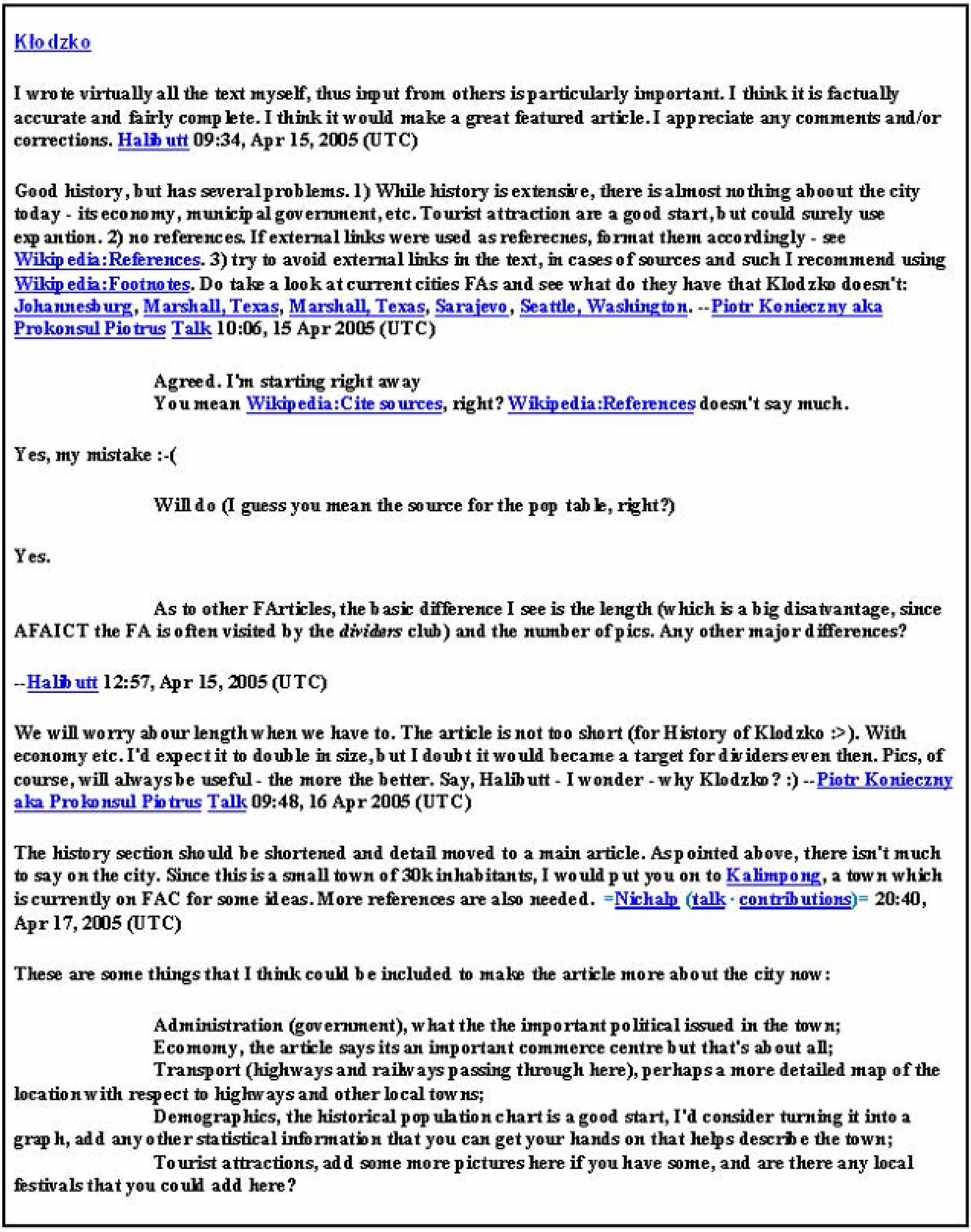
By contrast, in the Klodzko associated encyclopaedic article (1571 tokens/562 types), no occurrence of first and second personal pronouns *I/You* has been detected, but only the third personal pronoun ‘it’ (14 occurrences).

The occurrences of *modals* are also very frequent as they occur 12 times (*can 1, could 3, would 3 will 3, should 1, have to 1).* In addition the occurrence of the verb *think* in the first person (3 times) and *write* (only once) *and say* (3 times) clearly convey the cognitive position of the contributors.

```
Wikipedia:References doesn't  say    much.    Yes, my mist
useful - the more the better. Say, Halibutt - I    wonder - w
ed above, there isn't much to say on the city. Since this is a s
```

```
s particularly  important. I  think it is factually accurate and f
curate and fairly complete. I think it  would make a great featur
 These are some things that I think could be included to make the
```

Grammar and spelling errors have been found in the Klodzko talk page. There are five typing errors: *aboout* (line 4), *disatvantage* (line 16), *referecnes* (line 6), *the the* (line 19), *ecomomy* (line20). A grammar error has also been detected e.g.: *would became* (line 20). Mistakes, as already mentioned, are not taken seriously in this discussion writing space as here, unlike in WA, content over form and the communicative function are considered more relevant.



Klodzko

I wrote virtually all the text myself, thus input from others is particularly important. I think it is factually accurate and fairly complete. I think it would make a great featured article. I appreciate any comments and/or corrections. Halibutt 09:34, Apr 15, 2005 (UTC)

Good history, but has several problems. 1) While history is extensive, there is almost nothing aboout the city today - its economy, municipal government, etc. Tourist attraction are a good start, but could surely use expantion. 2) no references. If external links were used as referecnes, format them accordingly - see Wikipedia:References. 3) try to avoid external links in the text, in cases of sources and such I recommend using Wikipedia:Footnotes. Do take a look at current cities FAs and see what do they have that Klodzko doesn't: Johannesburg, Marshall, Texas, Marshall, Texas, Sarajevo, Seattle, Washington. --Piotr Konieczny aka Prokonsul Piotrus Talk 10:06, 15 Apr 2005 (UTC)

Agreed. I'm starting right away
You mean Wikipedia:Cite sources, right? Wikipedia:References doesn't say much.

Yes, my mistake :-(

Will do (I guess you mean the source for the pop table, right?)

Yes.

As to other FArticles, the basic difference I see is the length (which is a big disatvantage, since AFAICT the FA is often visited by the *dividers* club) and the number of pics. Any other major differences?

--Halibutt 12:57, Apr 15, 2005 (UTC)

We will worry abour length when we have to. The article is not too short (for History of Klodzko :>). With economy etc. I'd expect it to double in size, but I doubt it would became a target for dividers even then. Pics, of course, will always be useful - the more the better. Say, Halibutt - I wonder - why Klodzko? :) --Piotr Konieczny aka Prokonsul Piotrus Talk 09:48, 16 Apr 2005 (UTC)

The history section should be shortened and detail moved to a main article. As pointed above, there isn't much to say on the city. Since this is a small town of 30k inhabitants, I would put you on to Kalimpong, a town which is currently on FAC for some ideas. More references are also needed. =Nichalp (talk · contributions)= 20:40, Apr 17, 2005 (UTC)

These are some things that I think could be included to make the article more about the city now:

Administration (government), what the the important political issued in the town;
Ecomomy, the article says its an important commerce centre but that's about all;
Transport (highways and railways passing through here), perhaps a more detailed map of the location with respect to highways and other local towns;
Demographics, the historical population chart is a good start, I'd consider turning it into a graph, add any other statistical information that you can get your hands on that helps describe the town;
Tourist attractions, add some more pictures here if you have some, and are there any local festivals that you could add here?

Fig. 55 An excerpt from klodzko TP

9. Comments and Remarks

The analysis of Klodzko TP has highlighted the spoken-written style of TPs which, for their involved nature, are stylistically different form the informational WAs. Nonetheless, Wikispeak register is not very dissimilar from that conveyed in Internet forums and in spoken conversation. A peculiar difference resides in the specific lexicon used in WikiSpeak Jargon, which is associated with the new editing process typical of this new We b2.0 collaborative writing space.

As shown (fig. 57a) the analysis has highlighted a less formal and objective style in TPs as compared to WAs. The different frequency of some linguistic devices has produced the stylistic variation recorded. First of all both shorter average words and more concise sentences, have contributed to the tone variation. In addition, the lexical density has also proved to be lower in TPs together with the frequency of nominalizations, passives, prepositions, definite and indefinite articles, nouns, adjectives, and finally coordinating conjunctions. Unexpectedly, the frequency of both subordination features and of conjuncts has been higher. By contrast, a high occurrence of linguistic features typical of conversational speech (Biber 1988) has been recorded in Wikipedian TPs. Specifically, a high frequency of place adverbials, personal pronouns, demonstratives, indefinite pronouns, mitigating and boostering devices, modals, lexical verbs, negative forms, reduced forms and interrogative sentences has been recorded (fig. 56) [94].

To conclude, the lower frequency of the linguistic features typical of the encyclopaedic formal expository style in TPs, and by contrast, the higher occurrence of those linguistic classes typical of spoken and conversational discourse, is further evidence of the more informal and involved modality conveyed in the WikiSpeak spoken/written by Wikipedians in TPs.

Fig. 56 which follows,visualizes the specific loading of the linguist classes typical of the involved production in TPs compared to WAs, while fig. 57 contrastively highlights positive and negative linguistic features in TPs vs. WAs, and the total scores recorded (according to the selected criteria) in the two corpora (fig. 57b).

In 18 out of 20 linguistic classes analyzed, by comparing the two corpora, the Chi-Square test shows that the data has a 99.99 % of reliability, with a P-value of 0.0001. Only in the case of gerunds and present participles, with the relative frequency identical (2.41%) in both corpora, the Chi Square test has shown that there is 100% probability that the coincidence may be due to pure chance.

Relative frequency of nominalizations is also very close in the two corpora (4.62% WAs vs. 4.59% TPs), and this means that this similarity has a 56.52% probability that this result is due to random variation (fig. 58).

[94] Specific findings on the frequency of positive and negative linguistic classes analysed are shown in chapter 6, section 1, fig. 1.

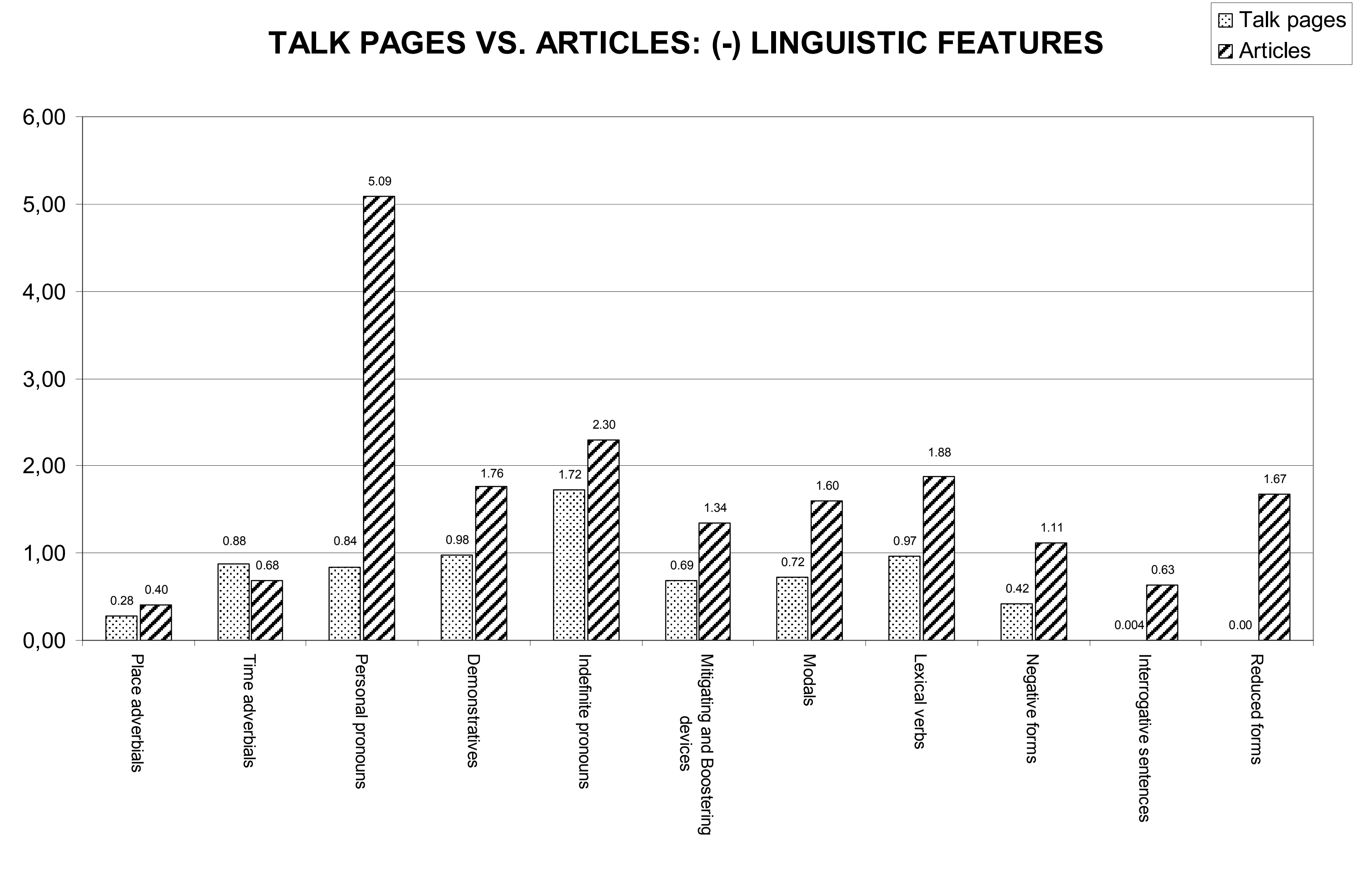


Fig. 56 Overall (-) linguisitc features in TPs vs. WAs

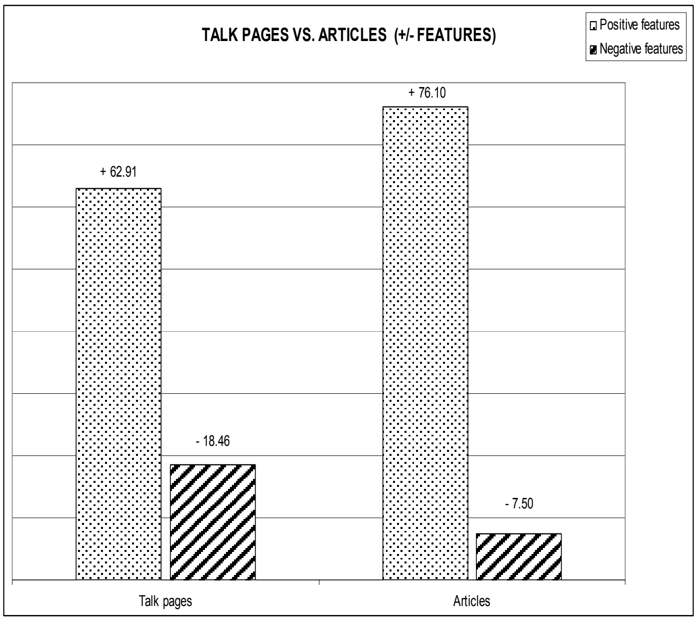


Fig. 57a (+/- features in TPs vs. WAs)

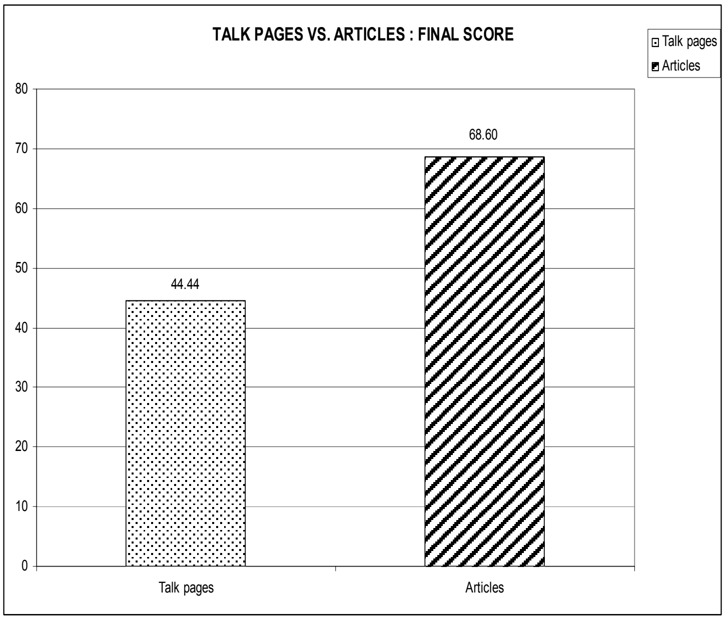
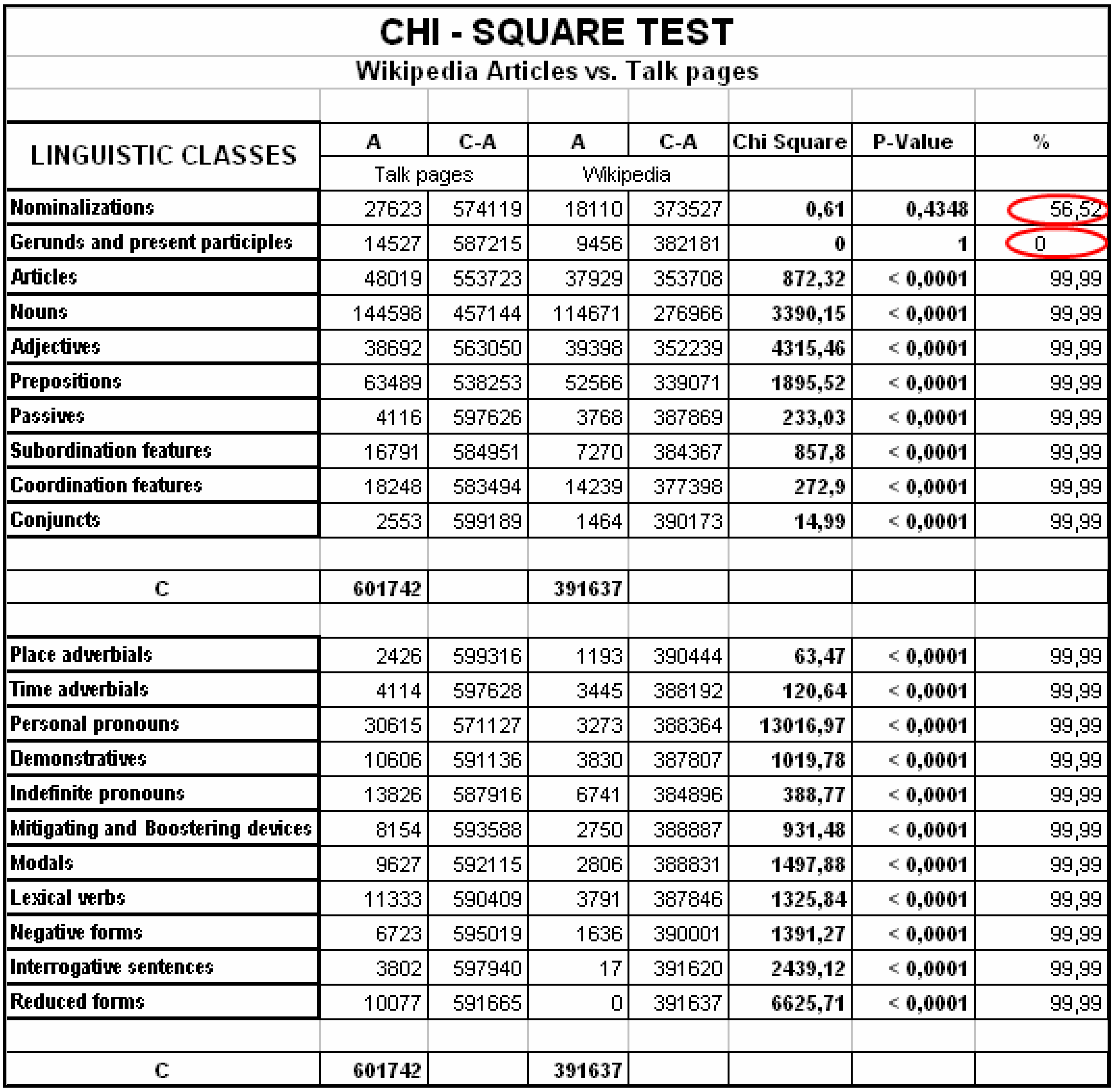


Fig. 57b Final Score TPs vs. WAs



CHI - SQUARE TEST

Wikipedia Articles vs. Talk pages

LINGUISTIC CLASSES	A	C-A	A	C-A	Chi Square	P-Value	%
	Talk pages		Wikipedia				
Nominalizations	27623	574119	18110	373527	**0,61**	**0,4348**	56,52
Gerunds and present participles	14527	587215	9456	382181	**0**	**1**	0
Articles	48019	553723	37929	353708	**872,32**	**< 0,0001**	99,99
Nouns	144598	457144	114671	276966	**3390,15**	**< 0,0001**	99,99
Adjectives	38692	563050	39398	352239	**4315,46**	**< 0,0001**	99,99
Prepositions	63489	538253	52566	339071	**1895,52**	**< 0,0001**	99,99
Passives	4116	597626	3768	387869	**233,03**	**< 0,0001**	99,99
Subordination features	16791	584951	7270	384367	**857,8**	**< 0,0001**	99,99
Coordination features	18248	583494	14239	377398	**272,9**	**< 0,0001**	99,99
Conjuncts	2553	599189	1464	390173	**14,99**	**< 0,0001**	99,99
C	**601742**		**391637**				
Place adverbials	2426	599316	1193	390444	**63,47**	**< 0,0001**	99,99
Time adverbials	4114	597628	3445	388192	**120,64**	**< 0,0001**	99,99
Personal pronouns	30615	571127	3273	388364	**13016,97**	**< 0,0001**	99,99
Demonstratives	10606	591136	3830	387807	**1019,78**	**< 0,0001**	99,99
Indefinite pronouns	13826	587916	6741	384896	**388,77**	**< 0,0001**	99,99
Mitigating and Boostering devices	8154	593588	2750	388887	**931,48**	**< 0,0001**	99,99
Modals	9627	592115	2806	388831	**1497,88**	**< 0,0001**	99,99
Lexical verbs	11333	590409	3791	387846	**1325,84**	**< 0,0001**	99,99
Negative forms	6723	595019	1636	390001	**1391,27**	**< 0,0001**	99,99
Interrogative sentences	3802	597940	17	391620	**2439,12**	**< 0,0001**	99,99
Reduced forms	10077	591665	0	391637	**6625,71**	**< 0,0001**	99,99
C	**601742**		**391637**				

Fig. 58 Chi-Square Test

6. CONCLUSIONS

1. Implications

This research has been methodologically inspired by Biber's multidimensional approach which has identified the underlying linguistic parameters of variation in a range of several different registers, from conversation to academic writing.

In the first part of this research intra-genre register variations in Britannica vs. Wikipedia encyclopaedic expository style has been mapped in accordance with Biber's dimension: *Informational vs. Involved Production*. In addition, Wikipedia Index of Readability and its consistency with Web Usability principles has also been explored. In the second part of this research, inter-genre variations between Wikipedian encyclopaedic articles and interactive talk pages has been analyzed.

To summarize, thirteen linguistic classes which, according to Biber, have a positive loading in defining the informational production, have been investigated. A micro/macroscopic contrastive analysis has been carried out to define the positive incidence of the selected linguistic classes on the formal register of the encyclopaedic expository style and to map and highlight similarities and differences in the two corpora through selected examples and excerpts. The overall data outlined in fig. 1 clearly shows that all the positive findings are not very dissimilar, although constantly slightly higher (in most of the cases) in Britannica. As pointed out[95] the linguistic features with a positive loading on formal encyclopaedic expository texts are also very frequent in academic writing, considered by Biber as the most typical and extreme formal expression of the informational production. None of them is either difficult to use or stylistically demanding. However, combining them in large numbers evidently requires the writer's attention, as well as extra care and time. The final result is a more accurate style and a more sophisticated and complex modality of expression.

As shown [96] one of the main peculiarities of informational production is associated with a high lexical density and with the use of longer words and sentences. Sentences are expanded through a variety of devices, some of the most frequent ones, beings nominalizations, gerunds and present participial forms, prepositions, definite and indefinite articles, nouns, adjectives, and an extensive use of subordination and coordination devices. Furthermore unlike in academic writing, a reduced number of passive constructions and conjuncts has been detected in both encyclopaedic corpora. Fig. 1 outlines the linguistic classes with a positive and a negative loading on the formality of the encyclopaedic expository style. Of course, word length, sentence length and lexical density have not been computed in the totals, as they are not categories grammatically homogeneous and consistent with the other linguistic classes listed. The aim of subtotals and final scores is to provide unique and indicative figures in order to easily identify differences and similarities in the two encyclopaedic corpora. Subtotals show a higher formality of the Britannica expository style (79.80%). Thus, the

[95] *see* chapter 3, sections 1.1, 1.2
[96] *see* chapter 4, section 1

Wikipedia's less formal style (76.10%) demonstrates a lower conformity of Wikipedians with the prescriptive linguistic and stylistic norms which have a determinant incidence in defining the formal register of its expository style[97]. Ten linguistic classes with a negative loading on informational production have been investigated. Fig. 1 shows that the overall occurrences of these linguistic classes are practically the same in the two encyclopaedic corpora (7.50%).

	LINGUISTIC CLASSES	**Britannica**	**Wikipedia**	**Talkpages**
	Word length (characters)	*5.3*	*5.2*	*4.1*
	Sentence length (words)	*22.05*	*22.09*	*13.5*
	Lexical density (tokens/types)	*45.5*	*43.6*	*40*
⊕	**Nominalizations**	5.26	4.62	4.59
	Gerunds and present participles	2.38	2.41	2.41
	Definite and Indefinite Articles	10.02	9.68	7.98
	Nouns	29.90	29.28	24.03
	Adjectives	10.54	10.06	6.43
	Prepositions	14.23	13.42	10.55
	Passives %	0.96	0.96	0.68
	Subordination features	2.31	1.86	2.79
	Coordination features	4.11	3.64	3.03
	Conjuncts	0.47	0.37	0.42
	(+) SUBTOTALS	***+79.80***	***+76.10***	***62.91***
⊖	**Place adverbials**	0.23	0.28	0.40
	Time adverbials	0.77	0.88	0.68
	Personal pronouns	1.05	0.84	5.09
	Demonstratives	0.75	0.98	1.76
	Indefinite pronouns	1.74	1.72	2.30
	Mitigating and Boostering devices	0.75	0.69	1.34
	Modals	0.82	0.72	1.60
	Lexical verbs	0.84	0.97	1.88
	Negative forms	0.52	0.42	1.11
	Interrogative sentences	0.035	0.004	0.63
	Reduced forms	0	0	1.67
	(-) SUBTOTALS	***- 7.50***	***-7.50***	***18.46***
	FINAL SCORE	***72.30***	***68.60***	***44.44***

Fig. 1 Frequency of analysed linguistic classes in Wikipedia, Britannica and Talk Pages

In particular, it has been found that three classes (place and time adverbials, demonstratives and lexical verbs) have a slightly higher frequency in Wikipedia, whereas personal pronouns, indefinite pronouns, mitigating and boostering devices, modal verbs, negative and interrogative forms are more extensively used in Britannica. The same final result (7.50%) proves to what extent both Britannica authors and Wikipedia contributors are very careful in avoiding the use of those linguistic features which can invalidate the objectivity, neutrality and formality of the encyclopaedic expository style.

[97] See chapter 4, section 2

Nevertheless, the final score is highly respectable and encouraging for Wikipedia (72.30% BA vs. 68.60% WA), the overall stylistic frequency variation being only 3.70%.

In fig. 2 the totals of positive and negative features detected in Britannica, Wikipedia articles and Talk pages are compared. In fig. 3 final scores representing *Informational Production* of Britannica and Wikipedia vs. *Involved Production* of Talk pages are presented. The significance of the data has been tested by the Chi-Square (tab. 1 in *Appendix*). The overall data related to the comparison of BAs, WAs and TPs shows a very high degree of reliability, being the statistics and the resulting P-value a symptom of a true underlying significant difference (Baroni, 2006:4) and thus not explainable as a result of a random variation.

The higher formal expository style detected in Britannica is certainly due to encyclopaedic articles written individually by paid scholars, professionals and experts and further revised by an editorial board which ensures the stylistic consistency. By contrast, for Wikipedia articles, which are always in progress and subject to never ending improvements and changes, the mechanism of collaborative writing, the volunteer work of contributors (sometimes also anonymous), the open nature of Wikipedia writing space (and risks associated to Vandalism and content falsification), the lack of an official editorial committee which supervises style and content (delegated to mutual control and reciprocal consensus) are all features which can justify the inferior "orthodoxy" of Wikipedia expository style.

As has been shown, it is easier to avoid the use of the forbidden informal linguistic features, by contrast, it is more difficult to guarantee the best stylistic practices. Thus, it is not a coincidence that the best articles, corresponding to the *featured* articles, have undergone continuous revisions, refactoring and improvements thus, attaining in this way a better textual cohesion, coherence and a higher degree of stylistic formality. On the other hand, the final result of what Surowiecki (2004)[98] defines as the *Wisdom of Crowds*, Deleuze (1980) *Ryzhome* and Pierre Levy the *Collective Intelligence* [99] is the variety of the information provided guaranteed by the free and democratic participation of many to the *Massively Distributed Collaboration*.

[98] *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations* (2004) is a book written by James Surowiecki about the aggregation of information in groups, resulting in decisions that, he argues, are often better than those made by any single member of the group.

[99] The concept of *Rhizome* elaborated by Gilles Deleuze and Felix Guattari, and the idea of *Collective Intelligence* by Pierre Levy have been considered within the philosophical frame of Wikipedia. Deleuze and Guattari in *Mille Plateaux. Capitalisme et schizophrenie* (1980) claimed that a rhizome is any structure in which each point is necessarily connected to each other point, where no location may become a beginning or an end. Deleuze labels the rhizome as a "multiplicity," resistant to structures of domination. He claims that "...many people have trees growing in their heads, but the brain is more like grass than a tree. We're taught to act like trees and forced to think like trees, but he believes that we more naturally think like a rhizome" (p. 17). Pierre Lévy in *Iintelligence Collective*. *Pour une antropologie du cyberspace* (1994), affirmed that it is a form of universally distributed intelligence, constantly enhanced, coordinated in real time and resulting in the effective mobilization of skills and knowledge, where no one knows everything but everyone knows something. Thanks to it the sharing of ideas in cyberspace has the potential to liberate us from the social and political hierarchies and to develop a real distributed knowledge.

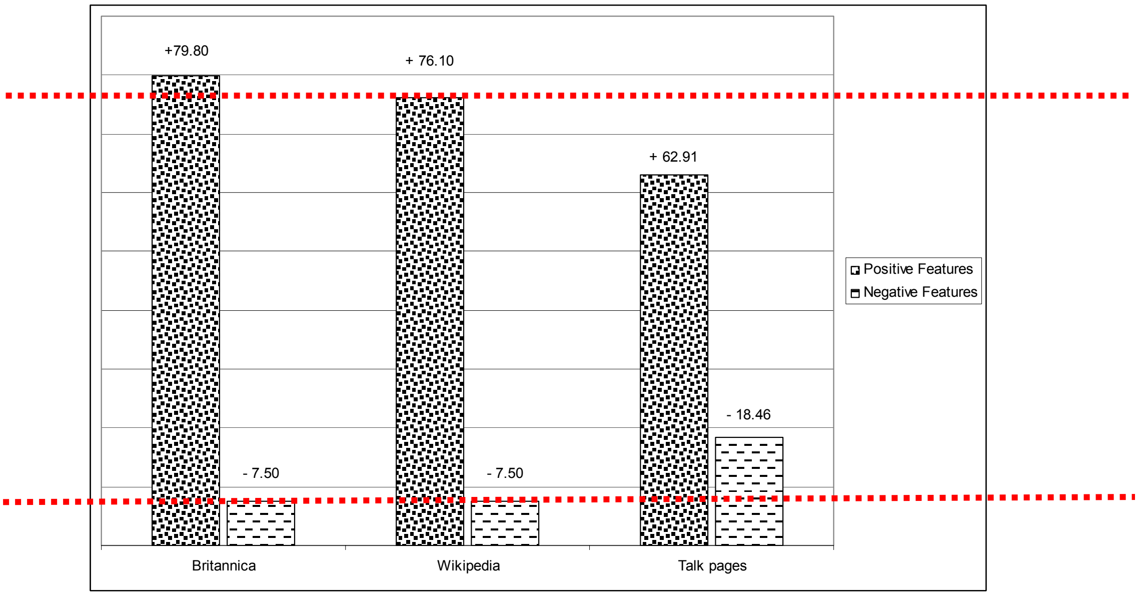


Fig. 2 (+ /-) Linguisitc features in Britannica, Wikipedia, Talk Pages

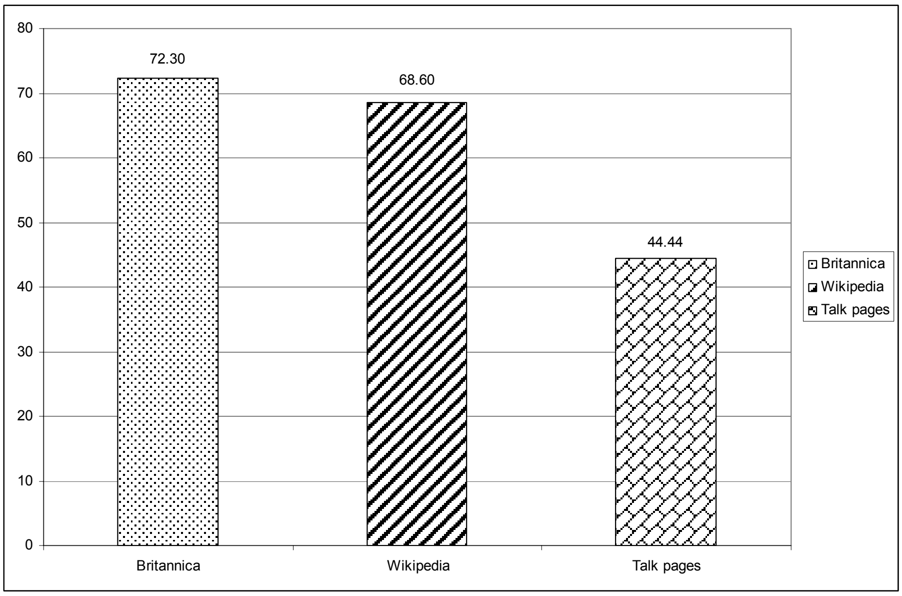


Fig. 3 Final Score: Britannica, Wikipedia, Talk Pages

As fig. 2 shows, the variation between the two encyclopaedic corpora, is mainly due to a higher frequency of the linguistic classes with a positive loading, while the total amount of the negative features corresponds across the two corpora (7.50). The interpretation of the collected data, seems to suggest that despite the collective editorial control, the language used in Wikipedia co-authored articles has a formal and standardized expository style not very far from Britannica.

Even though Britannica's production is more formal than that of Wikipedia (fig. 2,3), the stylistic difference is not so marked as expected. If Britannica is considered the best encyclopaedia in the English speaking world, the different individual vs. collaborative authorial production, the copyright vs. copyleft licence and, finally, the different authorship (professional paid writers *vs.* volunteer and anonymous amateurs), are controlled variables which do not deeply invalidate Wikipedia "fair" and correct formal production.

In addition, the empirical data suggests a number of clear correlations between formal expository style and different situational variables. The formality of what has been defined as *WikiLanguage* is mainly due to the need to avoid misinterpretation of the message. It is used in case of a wide reading audience in order to avoid intercultural misunderstanding as it could happen in the specific case of Web delivered contents. Furthermore, formality of the encyclopaedic expository style is higher, as it is in the specific case of Britannica, when writing and reading settings are not shared by sender (writer) and receiver (reader), since the latter is not allowed to freely and directly participate in the writing process. Finally, formality seems to be higher when the temporal interval between textual production and its reception is longer.

It is my personal point of view that Wikipedia can be taken as an example of the evolution of an *extant* traditional genre (encyclopaedias) into a *variant* one (co-authoring web 2.0 encyclopaedias). Nevertheless, since formal style and traditional conventions of the genre have been officially preserved in the encyclopaedic superficial form, this suggests that when collaborative users have to conform to stylistic established norms (*Wikipedia Manual of Style)* and shared social working ethics (*Wikiquette*), diversity and controversy tends to be successfully erased and the official requested style is observed also within an open co-authoring system.

As shown in the first part of this study, contributors' voices are merged and homogenized in the encyclopaedic formal expository style (Emigh, Herring, 2005). It seems that Wikipedians, belonging to the second web generation, have metaphorically sacrificed the "I" in Internet, to exploit the "We" in Web. Compared to Britannica, the less formal style of Wikipedia is surely due to the more massive and less "highly educated" number of contributors distributed all over the world. My personal point of view is that it is also due to a more informal mode of communication which is stylistically dominant in the Web 2.0.

In order to define Britannica vs. Wikipedia on the whole, in addition to the purely linguistic perspective, further categories, more specific of webgenres, have been considered equally important. *Index of Readability* and *Web Usability* have a crucial role in defining the total perception of online encyclopaedias as a webgenre. Specifically, the Index of Readability proves to be quite similar (11.5 BAs vs. 11 WAs) in the two encyclopaedic corpora. In particular, the results from Gunning Fog analysis have shown that Wikipedia articles are slightly simpler to be read and understood by a learner audience.

Furthermore, fig. 82 [100] points out the positive ☺ and the negative ☹ features detected in the two encyclopaedic corpora, according to the Web Usability principles defined by Nielsen (1999, 2006).

[100] *see* chapter 4, section 5.1.2

With reference to his standards the final score is definitely higher in Wikipedia than in Britannica (17 ☺ in Wikipedia vs. 5 ☺ in Britannica).

In brief, with reference to the selected criteria, a lower linguistic formality has been detected in Wikipedia than in Britannica. By contrast, Wikipedia succeeds more than Britannica in reaching a better Index of Readability, Web Usability and variety of the information provided.

Technological advantages offered by collaborative wiki software, reinforce the granularity of the information provided which comes from contributors distributed all around the globe. Consequently, both the number and the length of encyclopaedic articles prove to be definitely higher in Wikipedia than in Britannica[101].

Since the average sentence length is similar in the two corpora[102], it has been shown that longer articles in Wikipedia are not due to a prolix style, as some critics have argued, but to a higher informative content of Wikipedia. If we consider that the content is more interlinked in Wikipedia than in Britannica, this feature should reduce the length of the main entries as information provided should go into more depth in the interconnected pages. Nonetheless, article average length proves to be higher in Wikipedia. Content interlinking and quick updating of the information is guaranteed by the easiness of wiki software which every day attracts greater and greater numbers of contributors.

Furthermore, Wikipedia allows a democratic participation thanks to the collaborative writing of encyclopaedic articles and to the social tagging which, through the Wikipedia *Folksomy*, enables the practice of collaborative categorization made directly by the contributors who spontaneously cooperate to organize articles into categories. Consequently, the body of encyclopaedic information becomes increasingly easier to be searched, discovered, and navigated over time.

As McLuhan (1964) pointed out in the sixties *the medium is the message.* With reference to this principle, Wikipedia observes the traditional linguistic expository style of the encyclopaedic genre but, at the same time, it adds new peculiarities to the extant genre in terms of production, function, and reception. In this way, it ensures the stability of the genre replicating its linguistic superficial form and style and, in the meantime, it reflects the evolution of knowledge and the migration of information into the Web 2.0 networked cyberspace.

As has been highlighted in the second part of this research, Wikipedia is not only an encyclopaedia, but it is the most widespread global Community of Practice on the web. Clicking on the *Community Portal* of the Wikipedia website, we enter a new world based on principles of democracy and collaborativeness. Behind the encyclopaedic façade, there is a working community in the back office. From the linguistic point of view, the analysis of TPs has highlighted the code switching between *Wikilanguage* (in WAs) and *Wikispeak* (in TPs).

TP style has been specifically analyzed in this research. Talk pages represent the writing space where Wikipedians carry out discussions related to the specific encyclopaedic articles written in

[101] The number of articles is 2,034,000 in English Wikipedia vs. 100,000 in Britannica Online, up to 18th October 2007.

[102] *see* chapter 4, section 1.1.3

Document Mode pages. Methodologically, this research has demonstrated the usefulness of corpus-based empirical research in assessing language use in different online writing spaces.

Halliday (in Yates, 1996). claims that an important difference between genres and modes of communication lies in the *field,* in which communication takes place, and in the *topic*, the main discursive object. In the case of CMC discourse, there is no field in the virtual space beyond the text of the interaction. Thus, the text of the CMC interaction coincides with the *field* Such lack of a defined field may explain the high levels of modality within CMC wiki discourse. The text not only carries the social situation, but also the participants' relationship to the situation, their perception of the relationships between the knowledge and topics under discussion.

Halliday (in Yates, 1996) considers also the *tenor* of the communication which is essentially defined by social roles, often made clear by the social situation in which the participants are placed

Wikipedians as tenors are again limited by their virtual existence to those presentations of self which take place within and through the different synchronous and asynchronous channels offered by Wikipedia community, especially by talk pages and user pages. Wikipedians live, as cyber beings, through their communicative utterances. The necessity to present oneself may be a factor behind the high levels of first and second person pronoun use in TPs. In this way, they try to recover, their personal identity submerged, in document mode by the nature of the collective encyclopaedic project.

Finally, the *mode* of TPs, as a new communicative medium, is neither simply speech-like nor simply written-like. Though, as has been shown[103], TPs bear similarities in their textual aspects (e.g., lexical density, nominalizations, etc.) to written discourse, they differs greatly in others, namely personal pronouns, modal verbs and use of deixis. As a whole, these similarities and differences exemplify the complexity of TPs as a communication mode. Similar to written and spoken discourse, TPs are influenced by numerous social and situational factors defined by the communicative acts.

In particular, the contrastive analysis has highlighted the code-switching between language in use in TPs and WAs. TPs are unconventional and unpredictable and visibly show the different points of view and personalities of contributors. Posts are signed and written in the first person, using a dialogical and informal style which is exploratory, flexible and involving although less informative than WAs.

A comparison to forums has also offered the opportunity to highlight differences and similarities with TPs[104]. Peculiarities of the linguistic style conveyed in Wikipedia TPs have been specifically outlined[105]. *WikiSpeak* has been defined the variety of the NetSpeak jargon used by Wikipedians while they are engaged in their editing and technical operations. The word formation process of the new lexicon coined inside the community has also been specifically analysed[106]. The investigation of a

[103] *see* chapter 5, sections 7.1-7.2
[104] *see* chapter 5, section 6
[105] *see* chapter 5, sections 7.1-7.2
[106] *see* chapter 8, section 8

specific TP (related to Klodzko article) has brought to light the collaborative atmosphere, the informal register in use, different graphology and spelling practice, and last but not least, the new words in use.

A high occurrence of linguistic features typical of conversational speech (Biber, 1988) has been recorded in Wikipedian TPs. Specifically, a high frequency of reduced forms, interjections, negative forms (*don't, no, not*) exclamative and interrogative sentences, modal verbs have been detected. Different types of deixis, such as personal pronouns (*I, you*), locative and temporal adverbials (e.g. here, now) as well as a high frequency of proximal and distal deictics (demonstratives: *this, that*) have also been found.

Furthermore, a more involved and informal style has been conveyed by a shorter average word and sentence length. In addition, the lexical density has also proved to be lower in TPs and the frequency of nominalizations, passives, prepositions, definite and indefinite articles, nouns, adjectives and, finally, coordinating conjunctions. Unexpectedly, the frequency of subordination features and conjuncts has been higher. The lower frequency of the linguistic features mentioned above, typical of the encyclopaedic formal expository style, is further evidence of the more informal and involved modality conveyed in the WikiSpeak spoken/written by Wikipedians in TPs. By contrast, in most of the cases, the measurement of those linguistic features typical of spoken and conversational discourse, has proved to have, a higher frequency in TPs once compared to WAs [107].

In conclusion, if the linguistic production has to be considered a *continuum* from spoken to written language (Biber, 1988) the new genre of CMC conveyed in Wikipedian TPs is clear evidence of this continuity.

TPs use a written conversational style which is educated, and similar to a forum, but at the same time not extremely informal as can be a *Face to Face* conversation or a written exchange in chat. On the other hand, WikiSpeak, with its jargon expressions and wikilogisms, the original use of punctuation marks, interjections, emoticons and unconventional expressions, contributes in defining TPs and WikiSpeak, as an intergenre between the spoken and written discourse.

Talk pages share many linguistic peculiarities with spoken language, while encyclopaedic pages show to have linguistic features more similar to formal academic writing. The Internet as a new medium has transformed what has traditionally been oral communication into a new mediated form of written communication. For this reason the definition of CMC as a spoken written genre has been very appropriate, as it is a combination of both forms of communication.

Moreover, the migration and blending of genres has also joined together the figures of the two actors of communication: reader and writer. In this process of innovation they have been remodelled and reshaped. The prototypical reader and writer have acquired interchangeable functions in Wikipedia since they have become *writing readers* and *reading writers* thanks to the more interactive, social networking and collaborative values conveyed in these new writing spaces which increasingly characterize the new millennium. In the Web 2.0, Wikipedia blurring the classical distinction between

[107] *see* fig. 1 chapter 6

author and reader, has allowed multiple reader-authors to participate in a dynamic and collaborative process of construction of meaning. Thus, Wikipedia can be defined as a 'new dialogic digital genre' as it has recognized the importance of both collaborative and constructivist philosophy. It supports the ideal of an open textuality free from the control of the single author in favour of a collective reconfigured author.

The concept of knowledge and information seems to be often confused. Computophiles tend to see information processing as practically the same as knowledge. Knowledge rather consists of those human cognitive structures which give data their meaning and value as information. Knowledge, therefore, is socially mediated through language and individually and collectively constructed. It cannot be transmitted but must be recreated by individual minds. Thus, my personal point of view is that, it is essential to always keep in mind that knowledge is a matter of participating in a relatively well-defined discourse sustained and enriched by debate, as Wikipedia specific case clearly testifies. New constructivist scenarios and online collaborative environments such as wikis have contributed to the establishment and refinement of the concept of knowledge as the result of a collaborative construction. Furthermore, the concept of knowledge is intimately related to the medium which supports the discourse, and definitely it contributes to the establishment and refinement of the knowledge paradigm and philosophical framework which gives coherence to the overall reference work.

Furthermore, Wikipedia cannot be decontextualized from its main philosophical and political goals which are to pursue freedom of content and information. Encyclopaedia Britannica is a reference work without any political meaning hosted by a commercial website (*.com*), while the original French Encyclopédie from Diderot and D'Alembert was mainly a political project designed to propagate the ideas of the Enlightenment and to establish the reign of reason as the basis of modern public debate. It was not simply a knowledge catalogue, but a "reasoned" dictionary of arts, science and crafts, as its final title states. The adjective "reasoned" is not to be understood as "organized", but as being part of a wider political project to bring out reason as the basis of public and political debate in the 18th century in Europe (Soufron, 2005). Similarly, in the current age of Information and Technology, Wikipedia can be considered a post-modern Encyclopaedia, a copyleft reference work with a non-profit cultural goal (.org) affording a political project rather than merely a scientific one. In fact it is aimed at changing the society of the 21st century by giving control over content to everyone and thus enhancing freedom of expression and recovering the original aim of the World Wide Web inventor Sir Tim Berners Lee who wanted the web to be a boundless library of Babel and not a global supermarket as it has become in the *dot.com* era.

Observing the increasing traffic towards Wikipedia, it is reasonable to suppose that in the not so distant future, proprietary encyclopaedias will probably be limited[108], small, out of date and generally irrelevant and obsolete by comparison to Wikipedia and to the many other non-proprietary reference

[108] *Encyclopaedia Britannica* contains about 100.000 articles vs. 2.045.000 in Wikipedia (18th October 2007).

works. It is my personal point of view that good content can be free on line. Wikipedia specific case reveals pure love towards knowledge separated from economic interests and income. If there are costs associated with “first-class” content, users look elsewhere for comparable content and they will find it, simply because the number of Internet content-producers is huge. Academics, hobbyists and journalists want to educate the public because they have a natural desire to communicate and “help to change the world”. Many scholars concentrate their forces in building an open content encyclopaedia. There is considerable value in the collaboration that can be found in a general encyclopaedia project and in the uniformity and high quality of the results. This value cannot be found in the activities of writers posting content independently.

The Wikipedia's status as an encyclopaedia is very controversial. It has often been criticized for a perceived lack of reliability, comprehensiveness and authority. Many librarians, academics and editors of more formally written encyclopaedias have considered it to be of limited utility as a reference work. Nevertheless, Wikipedia's content is generally considered useful, so people link to it. Google and other search engines have already discovered the project and the daily traffic they send to it produces a steady stream of new readers and contributors. The greater the number of Wikipedia articles, the greater the number of links to them, and therefore the higher the rankings and number of listings on Google. Hence, it is conceivable that the articles’quality, reliability, verifiability and formality will actually increase over the coming years. Wikipedia content is getting constantly better as people go back again and again to old articles improving their quality, something which will increasingly come to the notice of experts. In the beginning, Wikipedia had a number of limited participating experts, but it has now attracted a higher number of graduate students, professors and professionals and it will probably attract the attention of many more experts in the next future. As the Wikipedia project improves and becomes better known, it is reasonable to expect that it will obtain wider academic recognition as many American institutions have already done. It is also reasonable to suppose that, in the coming years increasing numbers of academics will take part in the project seeing the increasing value of being associated with it. After all, many online courses, which can be read free of charge, demonstrate a very encouraging enthusiasm on the part of distinguished academics, to associate themselves with imparting free knowledge.

The LINGUIST List started a "Wikipedia Update Project" in mid-June 2007. Recently O'Donnell (2007), reporting on the Wikipedia phenomenon, has suggested that academics need to accept Wikipedia open-based collaborative model and view further contributions to it as a unique form of community service scholarship. He claimed:

> We are in a position to contribute to the construction of individual articles in a uniquely positive way by taking the time to help clean up and provide balance to entries in our professional areas of interest.

The first philosophical statement used by Renée Descartes *Cogito ergo sum* (I am therefore I exist) a foundation element of Western philosophy, makes it impossible to doubt one’s personal existence. The well-known quotation can be recontextualized and pluralized in *Cogitamus ergo sumus* in Wikipedia, since the individual existence is intrinsically intertwined and coincides with the collective existence of the virtual community. It is a further counter proof of the value of the encyclopaedic project which justifies the collaborative writing of its contributors.

In conclusion, although from the linguistic perspective a difference has been detected between Encyclopaedia Britannica and Wikipedia corpora, on the other hand a noteworthy difference has been found in the type of system at work. Wikipedia is an open system while Encyclopaedia Britannica is a closed system, furthermore, quality in Wikipedia is consensually defined by the collective, while Encyclopaedia Britannica defines knowledge in terms of absolutism, thus the statement of facts is how it was, is now and ever shall be. On the other hand, Wikipedia has a more evolving relativist view of knowledge. Each article becomes a photograph of the best view of knowledge at that instant. These different methods could yield more differences between Britannica and Wikipedia, but it goes beyond the scope of this work. The linguistic analysis carried out in this study provides just an initial layer of investigation that may inform other methods and probably aid future research directions.

My wish is that the findings of this research will help to clarify and scientifically demonstrate the positive effects of both the technology and the collaborative authoring on the conventions and on the quality of Web 2.0 online encyclopaedias.

2. Limitations

This study does not want to be the end of a line of research but just a beginning. The linguistic investigation carried out is to be intended just as an initial step in the exploration of variations between different reference works.

A limitation of this study can be certainly found in the corpus size. Although Wikipedia and Britannica offer thousands of articles only 200 have been selected. A larger corpus may give different and probably more reliable results. Furthermore, selected articles have been collected using a random sampling techniques but other techniques (e.g. *systematic, stratified or cluster sampling*) could have been used and, different sampling methods might have produced different results, especially if we take into account the variable length and the articulateness of each specific encyclopaedic entry. Another limitation, which has restricted the investigation, has also been the nature of the corpus. Since it is annotated nor tagged (only small samples), this factor has limited the range of linguistic queries.

3. Future Research

Wikipedia has recently become an important topic of communication studies. Although it is sometimes very difficult to persuade people to agree on a simple decision, millions of people find an agreement everyday on a wide range of topics in Wikipedia community. Thus, it could be interesting for future research to investigate the social dynamics of consensus and to what extent the typology of social roles carried out in the Wikipedia community, e.g *newbie, lurker, flamer, troll, ranter,* etc. (Scott, 2004) can affect the diachronic development of encyclopaedic articles towards better quality standards.

Before concluding, a new project, which has recently come out in the news, deserves to be mentioned. Its name is the *Citizendium Project* [109]. It is *a Citizens Compendium of Everything* launched by Larry Sanger, co-founder of Wikipedia with Jimmy Wales, in March 2007. The project has been initially described as a progressive fork of Wikipedia, a mirror of the Wikipedia site which allows anyone to contribute changes to articles, merging public participation with "gentle expert guidance". The final aim of the *Citizendium* is to improve the Wikipedia model by requiring all contributors to use their real names, by strictly moderating the project for unprofessional behaviors.

What will this new anti-populist project mean? How much and how will the quality and reliability of the information provided progress? To what extent will the formal style of encyclopaedic articles be improved and how will talk pages and WikiSpeak differ in the future? Will a new cultured community discourse come to life as more scholars, professionals, educators and experts contribute and edit topics? It would indeed be interesting to observe these phenomena in future research.

[109] *Citizendium Project* http://citizendium.org

REFERENCES

Ahrens, F. (9 July 2006). Death by Wikipedia: The Kenneth Lay Chronicles. In *Washington Post.com* http://www.washingtonpost.com/wp-dyn/content/article/2006/07/08/AR2006070800135.html

Algeo, J. (1999). *Cambridge History of English language.* Cambridge: Cambridge University Press.

Allen, N.J., Atkinson, D., Morgan, M., Moore, T., & Snow, C. (1987). What experienced collaborators say about collaborative writing. In *Journal of Business and Technical Communication*, Vol. 1. No. 2, 70-90.

Alteberg, B. (1986). Contrastive linking in spoken and written English. In Tottie G., Backlund I. (Eds.) *English and Speech in writing: a symposium*, 13-40. Studia Anglistica Upsaliensia 60. Stockholm: Almqvist and Wiksell.

Angermeier, M. (2005). Web 2.0 Mindmap. In *Kosmar*. Retrieved 15 June 2006, from http://kosmar.de/archives/2005/11/11/the-huge-cloud-lens-bubble-map-web20

Baker, P. (2006). *Using Corpora in Discourse Analysis* NewYork-London: Continuum.

Ball, A.F. (1992). Cultural preference and the expository writing of African-American adolescents. In *Written Communication, 9* (4), 501-532.

Baroni, M., Evert, S. (20 September 2006). *Chapter 38: Statistical methods for corpus exploitation*. Retrieved 2 October 2007 from, http://www.cogsci.uni-osnabrueck.de/~severt/PUB/BaroniEvertHSK38_manuscript.pdf

Bauer, D., Cavonius, C. R. (1980). Improving the legibility of visual display units through contrast reversal. In Grandjean E., Vigliani, E. (Eds.), *Ergonomic aspects of visual display terminals*, 137-142. London, UK: Taylor & Francis.

Beaman, K. (1984). Coordination and subordination revisited. In Tannen D. (Eds.) *Coherence in spoken and written discourse*, Norwood, N.Y. Ablex.

Bernard, M., Fernandez, M., Hull, S. (2002). The Effects of Line Length on Children and Adults' Online Reading Performance. In *Usability News 4.2 2002*. Retrieved 14 June 2006 from, http://psychology.wichita.edu/surl/usabilitynews/42/text_length.htm.

Biber, D. (June 1986). Spoken and Written Textual Dimensions in English: Resolving the Contradictory Findings in *Language*, Vol. 62, No. 2, 384-414. Retrieved 5 December 2007 from, http://links.jstor.org/sici?sici=00978507(198606)62%3A2%3C384%3ASAWTDI%3E2.0.CO%3B2-1#top

Biber, D. (1988). *Variation across speech and writing.* Cambridge: Cambridge University Press.

Biber, D. (1995). *Dimensions of register variation – across linguistic comparison*. Cambridge: Cambridge University Press

Biber, D., Conrad, S., Reppen R. (1998). *Corpus linguistics: investigating language structure and use*. Cambridge: Cambridge University Press.

Biber, D., Conrad S. (2005). Register variation: A corpus approach. In Schiffrin D, Tannen D., Hamilton E.D. (Eds.). *The Handbook of Discourse Analysis*. Oxford: Blackwell Publishing, 175-196.

Bizzell, P. (1992). *Academic Discourse and Critical Consciousness*. Pittsburgh: University of Pittsburgh Press.

Boyd, H. D., Brewer, J. P. (1997). Electronic discourse: linguistic individuals in virtual space. Albany, NY: State University of New York Press.

Boyd, H. D. (4 January 2005). Academia and Wikipedia - Many-to-Many. In *Corante Blog* . Retrieved 26 March 2006, from http://many.corante.com/archives/2005/01/04/academia_and_wikipedia.php

Braga, D. B., Busnardo, J. (2004). Digital Literacy for Autonomous Learning: Designer Problems and Learning Choices. In Snyder I., Beavis C., *Doing Literacy Online: Teaching. Learning and Playing in an Electronic World.* Cresskill, NJ: Hampton Press, 45-68.

Brown, G., Yules G. (1983). *Discourse Analysis.* Cambridge: Cambridge University Press

Calude, A. S. (2005) Subordination in Spoken New Zealand English. Retrieved 27 July 2005 from, http://www.calude.net/andreea/expandedRP.pdf

Capocci, A., Servedio V., Colaiori, F., Buriol, L. S., Donato, D., Leopardi, S., Caldarelli, G. (2006). *Preferential attachment in the growth of social networks: the case of Wikipedia*. In *Physical Review*. Retrieved 30 August 2007 from, http://www.inf.ufrgs.br/~buriol/papers/Physical_Review_E_06.pdf

Chafe, W. (1982). Integration and involvement in speaking, writing, and oral literature. In Tannen D. (Eds), *Spoken and written language: exploring orality and literacy,* 35-53. Norwood, New Jersey: Ablex.

Chafe, W. (1985). Linguistic differences produced by differences between speaking and writing. In Olsonm N. Torrance, Hildyard A. *Literature, language and learning: the nature and consequences of reading and writing*. Cambridge: Cambridge University Press, 105-23

Chafe, W., Danielewicz, J. (1987). Properties of spoken and written language. In R. Horowitz & S. J. Samuels (eds), *Comprehending Oral and Written Language.* New York: Academic Press, 83-113

Chafe, W., Tannen, D. (1987). The relation between written and spoken language. In *Annual Review of Anthropology*. 16, 383-407.

Leuf, B., Cunningham, W. (2001). *The Wiki way: quick collaboration on the Web*. New York: Addison-Wesley.

Collison R.L. (1966). *Encyclopaedias: Their History throughout the Ages*. N.Y.: Hafner Publishing Co.

Coulmas, F. (ed.) 1997. *The handbook of sociolinguistics*. Oxford: Blackwell.

Crawford, H. (2001). Encyclopedias. In Bopp, R., Smith L. C.. *Reference and information services: an introduction*, 433-459. Englewood, CO: Libraries Unlimited.

Crystal, D. (1995). *The Cambridge Encyclopedia of English Language*. Cambridge: Cambridge University Press.

Crystal, D. (2001). *Language and the Internet*. Cambridge: Cambridge University Press.

Crowston, K., Williams, M. (2000). Reproduced and emergent genres of communication on the World- Wide Web. In *The Information Society*, Vol. 16, No. 3, 201-216. Retrieved December 2006 from http://www.cindoc.csic.es/cybermetrics/pdf/41.pdf

December, J. (1993). *Characteristics of Oral Culture in Discourse on the Net*. Proceedings of the twelfth annual Penn State Conference on Rhetoric and Composition, University Park, Pennsylvania. Retrieved 28 April 2007 from, http://www.december.com/john/papers/pscrc93.txt

Deleuze G., Guattari F. (1980). *Mille Plateaux. Capitalisme et schizophrenie* Paris: Minuit, *Eng. tr.* (1987) *A Thousand Plateaus. Capitalism and Schizophrenia*. Minneapolis: University of Minnesota Press.

Diatype (9 April 2006). In *Wikipedia, The Free Encyclopedia*. Retrieved 9 April 2006, from http://en.wikipedia.org/wiki/Diatype

Duranti, A., Goodwin, C. (1992). *Rethinking Context: Language as an Interactive Phenomenon*. Cambridge: Cambridge University Press.

Eble, C. (2000). *Slang and Lexicography* in James Copeland et al. (eds.) *Functional Approaches to Language, Culture, and Cognition.* Amsterdam: John Benjamins, 499-511.

Ede, L., Lunsford, A. (1990). *Singular texts/plural authors: Perspectives on collaborative writing*. Carbondale: Southern Illinois University.

Elia, A. (2006). An analysis of Wikipedia digital writing. In Proceedings of 11th Conference of the European Chapter of the Association for Computational Linguistics, 16-21.

Elia, A. (2007a). Gesso in silicio e Costruttivismo: i wiki e i nuovi ambienti collaborativi dell' e-learning 2.0. In *JE-LKS Journal of e-learning and knowledge Society* . Milano: Giunti No. 1, 99-108.

Elia A., (2007b). Wikis, Wikipedia and collaborative technology: linguistic changes and new challenges. In *New Media and Linguistic Change*. Hong Kong: University of Hong Kong (In press).

Ellis, C. A., Gibbs, S. J., Rein, G. L. (1991). Groupware: Some issues and experiences. In *Communications of the ACM*, Vol. *34*, No. 1, 39-58.

Emigh W. and Herring S. (2005). Collaborative Authoring on the Web: A Genre Analysis of Online Encyclopedias. In *Proceedings of the Thirty-Eighth Hawai'i International Conference on System Sciences, HICSS-38,* Los Alamitos: IEEE Press.

Encyclopaedia (2007). In *Encyclopædia Britannica.* Retrieved 19 November 2006, from http://www.britannica.com/eb/article-32027

Encyclopaedia (18 March 2006). In *Wikipedia, The Free Encyclopedia*. Retrieved 18 March 2007, from http://en.wikipedia.org/wiki/Encyclopedia

Erickson, T. (1995). Social Interaction on the Net: Virtual Community as Participatory Genre. In *Proceedings of the Thirtieth Annual Hawaii International Conference on System Sciences*. Maui, Hawaii, 1997, Vol. 6, 13-21. Digital Libraries. McLean, Virginia, 13-20. Retrieved 12 January 2006 from, http://www.pliant.org/personal/Tom_Erickson/VC_as_Genre.html

Farr, M. (1993). Essayist Literacy and Other Verbal Performances. Written Communication, Vol. 10, No. 1, 4-38. Retrieved 7 April 2006 from, *Sage Journal Online* from, http://wcx.sagepub.com/cgi/content/abstract/10/1/4

Foley, W., Van Valin, R. (1984). *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.

Galegher, J., Kraut, R. E. (1990). Computer-mediated communication for intellectual teamwork.

In *Proceedings of the 1990 ACM conference on Computer-supported cooperative work* **,** Los Angeles, 65-78. Retrieved 15 July 2005 from, http://portal.acm.org/citation.cfm?id=99343&dl=ACM&coll=portal

Gere, A. R. (1987). *Writing groups: History, theory, implications*. Carbondale: Southern Illinois University.

Giles, J. (26 March 2006). Internet encyclopaedias go head to head. In *News@nature.com.* Retrieved 28 August 2006 from, http://www.nature.com/doifinder/10.1038/438900a

Givon, T. (ed. 1983). *Topic continuity in discourse: a quantity cross language study.* Amsterdam: John Benjamins

Gomes Lee*,* (15 August 2007). Forget the Articles*,* Best Wikipedia Read Is Its Discussions. In *the Wall Street Journal online.* Retrieved 28 September 2007 from, http://online.wsj.com/public/article/SB118712061199497533.html

Gotti, M. (2002). The origin of seventeenth century canting terms. In *A Changing World of Words: Studies in English Historical Lexicography, Lexicology, and Semantics.* Amsterdam-New York: Rodopi. 165-196.

Graddol, D. (27 February 2004). The Future of Language. In *Science Magazine*. Vol. 303. No. 5662, 1329-1331. Retrieved 18 April 2005 from http://www.sciencemag.org/cgi/content/abstract/303/5662/1329?etoc

Graham, P. (November 2005). Web 2.0. In *Hackers News*. Retrieved 5 October 2006 from, http://www.paulgraham.com/web20.html

Gregory, M. (1967). Aspects of Varieties Differentiation. In *Journal of Linguistics* 3, 177-197.
Grice, H.P. (1975) Logic and conversation. In *Syntax and semantics 3: Speech acts*, New York: Academic Press, 41-58

Grossman, L. (13 December 2006). Time's Person of the Year: You. In *Time*. Retrieved 28 January 2007 from, http://www.time.com/time/magazine/article/0,9171,1569514,00.html

Kister, K. F. (1994). *Best encyclopedias: A guide to general and specialized encyclopedias*. Phoenix, AZ: Oryx Press.

Haas S., Grams E. (1998). Page and Link Classifications: Connecting Diverse Resources. In *Proceedings of Digital Libraries '98* – Third ACM Conference on Digital Libraries, 99-107.

Hall, E. T. (1976). *Beyond Culture*. New York: Doubleday

Halliday, M.A.K , Hazan, R. (1976).*Cohesion in English.* London: Longman

Halliday, M.A.K. (1979). Differences between spoken and written language: Some implications for literacy teaching. In Page G., Elkine J., O'Connor B. (Eds). *Communication Through Reading: Proceedings of the Fourth Australian Reading Conference.* Vol. 2, 37-52. Adelaide: Australian Reading Association.

Halliday, M.A.K. (1985a). *An Introduction to Functional Grammar*. London, Baltimore: Arnold.

Halliday, M. A. (1985b). *Spoken and Written Language*. Oxford: Oxford University Press.

Herring, S. (1996). *Computer Mediated communication: Linguistic, social and cross-cultural perspectives.* Philadelphia: John Benjamins Publishing Co.

Herring, S. (2001). Computer-mediated discourse. In Tannen D., Schiffrin D., Hamilton H. (Eds.), *Handbook of discourse analysis*. Oxford: Blackwell, 612-634.

Herring, S. (2004). Computer-mediated discourse analysis: An approach to researching online behavior. In Barab, S. A., Kling, R., Gray J. H. (Eds.), *Designing for Virtual Communities in the Service of Learning*. New York: Cambridge University Press, 338-376.

Herring, S. (2007). A faceted classification scheme for computer-mediated discourse. In *Language@Internet*, article 761.

Heylighen, F., Dewaele, J. M. (1999) *Formality of Language: definition, measurement and behavioural determinants,* Leo Apostel University of Brussels. Retrieved 9 February 2006 from, http://pcp.lanl.gov/Papers/Formality.pdf

Hyland, K. (2000). Hedges, Boosters and Lexical Invisibility: Noticing Modifiers in Academic Texts. In *Language Awareness* Vol. 9, No. 4, 179-197. Retrieved 10 September 2005 from, http://www.channelviewpublications.net/la/009/0179/la0090179.pdf

Hymes, D. (1972). Models of the interaction of language and social life. In Gumperz J., Hymes D. (eds.) *Directions in Sociolinguisitcs. The ethnography of communication*. New York: Holt, Rinehart and Witson, INC.

Hodge, R., Kress G. (1988). *Social Semiotics*. Cambridge: Polity

Holloway, T., Bozicevic M., Börner K. (2005). Analyzing and visualizing the Semantic Coverage of Wikipedia and Its Authors. In *Complexity*, Special issue on *Understanding Complex Systems*. Retrieved 25 Febraury 2007 from, http://arxiv.org/ftp/cs/papers/0512/0512085.pdf

Holmes, J. (1984). Hedging your bets and sitting on the fence: some evidence for hedges as support structures In *Te Reo* 27: 47-6

Horton, W. (1989). *Designing and writing online documentation: Help files to hypertext*. New York: John Wiley & Sons

Huffaker, D. A., Calvert, S. L. (2005). Gender, identity, and language use in teenage blogs. In *Journal of Computer-Mediated Communication,* Vol. 10, No. 2. Retrieved 7 March 2007 from, http://jcmc.indiana.edu/vol10/issue2/huffaker.html

Kapor, M. (9 November 2005). *Content Creation by Massively Distributed Collaboration.* In UC Berkley School of Information. Retrieved 15 January 2006 from, http://www.ischool.berkeley.edu/about/events/dls11092005

Kiefer, F. (1994). Modality. In Asher, R. E. (Ed.), *The Encyclopedia of language and linguistics.* Oxford: Pergamon Press, 2515-2520

Labov, W. (1984), Intensity. In Schiffrin D. (Ed.). *Meaning, form and use in context: linguistic applications*, Washington: Georgetown University Press, 43-70.

Lakoff, R. (1982). Some of my favourite writers are: the mingling of language and literate strategies in written communication. In *spoken and written language: exploring orality and literacy* Norwood, N.Y.: Abelex, 239-260.

Lamb, B. (September-October 2004). Wide Open Spaces: Wikis, Ready or Not. In *EDUCAUSE Review.* Vol. 39, No. 5, 36-48. Retrieved 19 Febraury 2007 from, http://www.educause.edu/pub/er/erm04/erm0452.asp

Landow, G. P. (1997). *Hypertext 2.0: The Convergence if Contemporary Critical Theory and Technology*, Baltimore and London: The John Hopkins Press.

Lanier, J. (2006). On Digital Maoism: The Hazards of the New Online Collectivism. In *Edge*, 183. Retrieved 8 March 2007, from http://www.edge.org/documents/archive/edge183.html

Larsen Freeman, D. (1997). *Grammar dimensions: Form, meaning, and use.* Boston: Heinle & Heinle. Retrieved 15 April 2007 from, http://www.cal.org/resources/digest/larsen01.html

Lawler, C. (2005). *Wikipedia as a Learning Community.* Master Thesis. Manchester: University of Manchester.

Levelt, J.M. (1989). Speaking: From Intention to Articulation. Cambridge Masschusetts: MIT Press.

Lèvy, P. (1994). *L'intelligence Collective. Pour une antropologie du cyberspace*, Paris : La Découverte.

Lih, A. (2004). Wikipedia as participatory journalism: reliable sources? Metrics for evaluating collaborative media as a news resource. In *Proceedings of 5th International Symposium on Online Journalism*. Retrieved 17April 2007 from, http://jmsc.hku.hk/faculty/alih/publications/utaustin-2004-wikipedia-rc2.pdf

Lowry, P.B., Curtis, A., Lowry, M.R. (2004). Building a taxonomy and nomenclature of collaborative writing to improve interdisciplinary research and practice. In *The Journal of Business Communication* 41, 66. Retrieved 23 April 2006 from http://www.questia.com/googleScholar.qst?docId=5002069499

Lynch P. J., Horton S. (2002). *Web Style Guide*. Retrieved 26 June 2005 from, http://webstyleguide.com/

MacCormick, K., Pursel, J. (1982). A Comparison of the Readability of the Academic American Encyclopedia, the Encyclopedia Britannica, and World book. In *Journal of Reading.* Vol. 25, No.4, 322-25.

Mardziah Hayati, A. (1998). Electronic Discourse: Evolving Conventions in Online Academic Environments. In *ERIC Digest Clearinghouse on Reading English and Communication*. Retrieved 30 August 2006 from, http://www.ericdigests.org/1999-2/online.htm

McArthur, T. (1986). *Worlds of reference: lexicography, learning and language from the clay tablet to the computer*. Cambridge: Cambridge University Press.

McHenery, T., Wilson A. (2001). *Corpus Linguistics*. Edinburg: Edinburg University Press.

McHenry R., (15 November 2004).The Faith-Based Encyclopedia. In *Tech Central Station* . Retrieved 7 April 2006 from, http://www.techcentralstation.com/111504A.html

McLuhan, M. (1964). The Medium is the Message. In *Understanding Media: The Extensions of Man*, New York: Signet.

Mejias Ulises A. (4 March 2005). Social Literacies: some observations about writing and wikis. In *IDEANT* . Retrieved 28 September 2007 from http://ideant.typepad.com/ideant/2005/03/social_literaci.html

Merriam-Webster's Dictionary Online. Retrieved 9 February 2005, from http://www.m-w.com/dictionary/

Mills, C. B. & Weldon, L., J. (1987). Reading text from computer screens. In *ACM Computing Surveys,* No. *4,* 329-358.

Miller, N. (2005). Wikipedia and the disappearing "author". In *ETC: A Review of General Semantics* Volume**,** No**.** 62, International Society for General Semantics

Morgan, M.C. (2006). BlogsandWikis. In *Bemidjistate State University.* Retrieved 22 January 2006 from, http://ferret.bemidjistate.edu/~morgan/cgi-bin/blogsAndWiki.pl

Munro, P. (1989). *Slang U.* New York: Harmony Books.

Neurath, O. (1938). International Encyclopaedia of Unified Science, Vol. I, 26. Chicago: The University of Chicago Press,

Nielsen, J. (1999). Designing web usability, Indianapolis: New Riders Publishing.

Nielsen, J., Loranger, H. (2006). *Prioritizing Web Usability.* Berkeley:New Riders Press.

Noblia, M. V. (1998). The Computer Mediated Communication. A new way of understanding the language. In *Proceedings IRRIS '98 conference*.

Ochs, E. (1979). Planned and Unplanned discourse. In Givón T., (Ed.). *Syntax and semantics*, Vol. 12, 51-80. New York: Academic Press.

Ong, W. J. (1982). *Orality and Literacy* London: Routledge

Orlikowski, W., Yates J. (1994). Genre repertoire: The structuring of communicative practices in organizations. In *Administrative Science Quarterly*, vol. 39, No. 4, 541-574. Retrieved 7 June 2005, from http://findarticles.com/p/articles/mi_m4035/is_n4_v39/ai_16987482

Panagiota, A. (2002). To wire or not to wire? Encyclopaedia Britannica versus Microsoft Encarta. In *Educational Technology & Society* 5(1). Retrieved 26 December 2006 from, http://www.ifets.info/journals/5_1/alevizou.html

Paolillo, J. (1999). The Virtual Speech Community: Social Network and Language Variation on IRC. In *Journal of Computer-Mediated Communication*, Vol. 4, No. 4. Retrieved 5 June 2007 from, http://www.ascusc.org/jcmc/vol4/issue4/paolillo.html

Pentzold C., Seidenglanz, S. (2006). Foucault@Wiki: first steps towards a conceptual framework for the analysis of Wiki discourse*s*. In *Proceedings of the 2006 international symposium on Wikis***,** Odense, Denmark. Retrieved 3 January 2006 from, http://www.wikisym.org/ws2006/proceedings/p59.pdf

Pombo, O., Guerreiro A., Alexandre, A. F. (2006). *Enciclopédia e Hipertexto*. Lisboa: Editora Duarte Reis.

Poster, M. (2001) *What's the matter with the Internet*? Electronic MNediations.Vol. 3 Minneapolis-London: University of Minnesota Press.

Poole, M., Field, T.W. (1976). A comparison of oral and written code elaboration. In *Language and Speech*, Vol. 19, No.4, 305-31.

Posner, I. R., Baecker, R.M. (10 January 1992). How people write together. In Proceedings of the *Twenty-Fifth Hawaii International Conference on System Sciences*, Kauai. Vol. 4, No 4, 127-138 Retrieved from 4 May 2007 from, http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=183420

Prakash, B. (18 July 2007). Weaving it together: Web 2.0. In *Rediff News*. Retrieved 24 June 2007 from, http://www.rediff.com/news/2007/jul/18bsp.htm

Preacher, K.J. (April 2001). Calculation for the chi-square test: an interactive calculation tool for chi-square tests of goodness of fit and independence [computer Software]. Retrieved 10 September 2007 from, http://www.psych.ku.edu/preacher/chisq/chisq.htm

Quirk, R., Greenbaum S., Leeh, G., Svartvik, J. (1985). *A comprehensive grammar of the English Language.* London: Longman.

Raymond, E. S. (1999). *The Cathedral & the Bazaar.* Cambridge, USA: O'Reilly.

Read, B. (15 July 2005). *Romantic Poetry Meets 21st-Century Technology, With wikis, the new Web tool, everybody's an editor and a critic*. Retrieved 28 March 2006 from, http://chronicle.com/free/v51/i45/45a03501.htm

Reagle, J. (2006). *A Case of Mutual Aid: Wikipedia, Politeness, and Perspective Taking*. Retrieved 28 January 2007 from, http://reagle.org/joseph/2004/agree/wikip-agree.html

Rehm, G. (2007). Hypertext Types and Markup Languages. In: *Linguistic Modelling of Information and Markup Languages*, Dieter Metzing, Andreas Witt (eds.), Springer. In press. Retrieved February 2006 from, http://georg-re.hm/pdf/Rehm-Hypertext-Types.pdf

Resnick P., Hansen D., Riedl J., Terveen L., Ackerman M. (2005). Beyond Threaded Conversation. In *Proceedings of the SIGCHI conference on Human factors in computing systems (HCI'05)* (Portland, OR, April 2-7, 2005) Retrieved 9 October 2006 from, http://portal.acm.org/citation.cfm?doid=1056808.1057126

Rezabeck, L., Cochenoer, J. (1994). Emoticons: Visual cues for Computer –Mediated Communication. In *Imagery and Visual Literacy*: Selected Readings from the Annual Conference of the International Visual literacy Association October 12-16, Arizona.

Richards, J. Platt, J., Platt, H. (1997). *Dictionary of Language Teaching and Applied Linguistics.* London: Longman.

Salas, R. A. (6 August 2007). Wikipedia keeps up with events. In *Star Tribune.* Retrieved 11 June 2007 from, http://www.startribune.com/10204/story/1339608.html

Sanger, L. (15 September 2006). Toward a New Compendium of Knowledge. In *Citizendium.org*. Retrieved 19 November 2006 from, http://www.citizendium.org/essay.html

Santini, M. (2007). Characterizing Genres of Web Pages: Genre Hybridism and Individualization. In *Proceedings of the 40th Hawaii International Conference on System Sciences*.

Schiff, S. (31 July 2006). Know It All. Can Wikipedia conquer expertise? In the *Annals of Information*, The New Yorker. Retrieved 5 January 2007 from, http://www.newyorker.com/archive/2006/07/31/060731fa_fact

Schmidt, K., Simone, C. (1996). Coordination Mechanisms: Towards a Conceptual Foundation of CSCW Systems Design. In *Computer Supported Cooperative Work*, Vol. 5, No.2/3, 155-200.

Selcer, D. (2007). The Uninterrupted Ocean: Leibniz and the Encyclopedic Imagination. In *Caliber* the Journals of University of California press. No. Spring 2007, 25-50. Retrieved 19 September 2007 from, http://caliber.ucpress.net/doi/abs/10.1525/rep.2007.98.1.25?journalCode=rep

Scharff, L. F., Hill, A., Austin, S. F. (1996). *Color test results.* Retrieved 18 February 2005 from, http://www.thecube.com/color/survreslts.html

Scott, A., Donath, J. (2004). Social roles in electronic communities. In *Proceedings Internet research 5.0,* 19-22 September, 2004 Brighton. Retrieved 18 January 2006 from, http://web.media.mit.edu/~golder/projects/roles/golder2004.pdf

Shah, S.(2005). Productive Controversy. In *Proceedings of the Wikimania'05* (Frankfurt am Main, Germany, August 4-7, 2005.

Sharples, M. (1992). Representing writing: External representations and the writing process. In Holt, P., Williams, N. (Eds.), *Computers and writing: State of the art*. Oxford, UK: Intellect; Kluwer Academic.

Sharples, M. (1993). Adding a little structure to collaborative writing. In *Literature Genre & Cybergenre*. London: Springer-Verlag.

Shepherd, M., Watters, C.R. (1998). The evolution of cybergenres. In *Proceedings of the Thirty-First Annual Hawaii International Conference on System Sciences* (HICSS '98). Hawaii, Vol. 2, 97-109.
Schiffrin D, Tannen D., Hamilton E.D. (Eds.). *The Handbook of Discourse Analysis*. Oxford: Blackwell Publishing

Shortis, T. (2001). *The language* of ICT Informatiopn and Communiation Technology. London: Routledge.

Sinclair, J. (1991). *Corpus, concordance, collocation.* Oxford: Oxford University Press.

Sinclair, J. (2003). *Reading Concordances.* London: Pearson- Longman.

Sloane, J. (2007). Wikimedia Foundation Moving To San Francisco. Posted in *Wired News,* 10 October 2007. Retrieved 29 October 2007 http://blog.wired.com/business/2007/10/wikimedia-found.html

Smith, L. C. (1989). Wholly new forms of encyclopedias: electronic knowledge in the form of hypertext. In *Proceedings of the forty-fourth FID Congress*. Helsinki, Finland, 245-250.
Stubbs, M. (1983). *Discourse Analysis: The Sociolinguistic Analysis of Natural Language*. Oxford: Basil Blackwell.

Stvilia, B., Twidale, M.B., Gasser, L., Smith, L.C. (2005).*Information Quality Discussion in Wikipedia.* Technical Report ISRN UIUCLIS-- 2005/2+CSCW. Retrieved 5 November 2006 from, http://www.isrl.uiuc.edu/~stvilia/papers/qualWiki.pdf

Soojung-Kim Pang, A. (2000). The work of the encyclopedia in the age of electronic reproduction . In *First Monday.* Retrieved 23 October 2006 from, http://www.firstmonday.org/issues/issue3_9/pang/
Soufron, J.B. (16 November 2004). The political importance of the Wikipedia Project, the only true Encyclopedia of our days. In *Around Wikipedia.* Retreived 29 December 2006 from, http://soufron.typhon.net/article.php3?id_article=71

Surowiecki, J. (2004). *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. New York: DoubleDay.

Tannen, D. (1989). Talking voices: Repetition, dialogue, and imagery in conversational discourse. In *Studies in Interactional Sociolinguistics,* Vol. 6, Cambridge: Cambridge University Press.

The American Heritage Dictionary of the English Language (2000). Boston: Houghton Mifflin

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam: John Benjamins.

Tornow, J. (1997). Link/age: Composing in the on-line classroom. Logan, UT: Utah State University Press.

Torsten, K. (February, 2005). World of Knowledge. The Wikipedia Project in *Linux Magazine*. Retreived 19 August 2006 from, http://w3.linux-magazine.com/issue/51/Wikipedia_Encyclopedia.pdf

Tottie, G. , Bengt A., Lars H. (1983) *English in Speech and writing* (ETOS Report 1) Lund: Engelska Institutionen.

Sanger L. (25 July 2001). Britannica or Nupedia? The future of free Encyclopedia. In *Kuro5hin* Retrieved 28 May 2007 from, http://www.kuro5hin.org/story/2001/7/25/103136/121

Sinclair, J. (1991). *Corpus, concordance, collocation.* Oxford: Oxford University Press.

Sinclair, J. (2003). *Reading Concordances.* London: Pearson- Longman.

Stvilia B., Twidale M., Gasser L., Smith L. (2005). Information quality discussions in Wikipedia, ICKM05. Retrieved 27 September 2007 from http://mailer.fsu.edu/~bstvilia/papers/qualWiki.pdf

Thomson, S.A. (1982) Subordination in formal and informal discourse. In Shiffrin, D. (Ed). *Meaning form , and use in context: linguistic applications,* GURT 84 Washington, D.C.: GeorgeTown University press, 85-94.

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam: John Benjamins.

Yates, S. J. (1996). Oral and written linguistic aspects of computer conferencing: A corpus-based study. In Susan C. Herring (Ed.), *Computer-mediated communication: Linguistic, social and cross-cultural perspectives.* Philadelphia: John Benjamins Publishing Co.

Yates, S. J., Sumner T. (1997). Digital Genre and the New Burden of Fixity. In *Proceedings of the 'Thirtieth Annual Hawaii International Conference' on System Sciences*, Maui, Hawaii. Vol. 6, 3-12.

Viegas, F., Wattenberg, M., Dave, K. (2004). Studying cooperation and conflict between authors with history flow visualizations. In *Proceedings of CHI 2004.* Vienna, 575-582. Retrieved 2 March 2007 from, http://alumni.media.mit.edu/~fviegas/papers/history_flow.pdf

Vincent J. (2001), Talk-speak: gioco e ideologia nei logonimi inglesi, in Vallini C., ed., *Le Parole per le parole - I logonimi nelle lingue e nei metalinguaggi.* Roma: Il Calamaio, 701-738.

Waldman, S. (26 October 2004). Who knows? In *The Guardian*. Retrieved 12 June 2007 from http://www.guardian.co.uk/guardian/

Wales, J. (8 March 2005). Wikipedia is an encyclopedia. In *Wikimedia* . Retrieved 23 June 2006 from http://lists.wikimedia.org/pipermail/wikipedia-l/2005-March/020469.html

Wei, C., Maust, B., Barrick, J., Cuddihy, E., Spyridakis, J. H. (2005). Wikis for Supporting Distributed Collaborative Writing. In *STC Proceedings.* Retrieved 18 July 2007 from, http://www.uwtc.washington.edu/research/pubs/jspyridakis/STC_Wiki_2005_STC_Attribution.pdf

Wenger, E. (1998). *Communities of Practice: Learning, Meaning, and Identity.* Cambridge: Cambridge University Press.

Wikipedia (16 January 2006). In *Wikipedia, The Free Encyclopedia*. Retrieved 30 January 2006, from http://en.wikipedia.org/wiki/Encyclopedia

Wiki (6 March 2007). In *Wikipedia, The Free Encyclopedia*. Retrieved 7 March 2007, from http://en.wikipedia.org/w/index.php?title=Wiki&oldid=113158225

Wikipedia.org is more popular than… (2007, March 6). In *Wikipedia, The Free Encyclopedia*. Retrieved 7 March 2007, from http://meta.wikimedia.org/wiki/Wikipedia.org_is_more_popular_than…

Wikipedia:Neutral point of view (20 March 2007). In *Wikipedia, The Free Encyclopedia*. Retrieved, March 20, 2007, from http://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view.

Wikipedia:Proposed deletion (14 March 2007). In *Wikipedia, The Free Encyclopedia*. Retrieved, March 14, 2007, from http://en.wikipedia.org/wiki/Wikipedia:Proposed_deletion

Wikipedia:How to edit a page (14 March 2007). In *Wikipedia, The Free Encyclopedia*. Retrieved, March 14, 2007, from http://en.wikipedia.org/wiki/Wikipedia:How_to_edit_a_page.

Wikipedia:Talk page (14 March 2007). In *Wikipedia, The Free Encyclopedia*. Retrieved, March 14, 2007, from http://en.wikipedia.org/wiki/Help:Talk_page.

Wikipedians (7 May 2006). In *Wikipedia, The Free Encyclopedia*. Retrieved 18 January 2006, from http://en.wikipedia.org/wiki/Wikipedia:Wikipedians

Wikiquette (11 July 2006). In *Wikipedia, The Free Encyclopedia*. Retrieved 12 July 2006, from http://en.wikipedia.org/wiki/Wikipedia:Etiquette

Wilkins, H. (1991). Computer Talk: Long-Distance Conversations by Computer in Written. In SAGE Publications, Vol. 8, No. 1, 56-78. Retrieved 19 Apei http://wcx.sagepub.com/cgi/content/abstract/8/1/56

Winter, E. (1982). *Towards a contextual grammar of English: the clause and its place in the definition of sentence*. Londo: George Allen and Unwin.

Wyss, E. L. (2000). *Iconicity in the Digital World,* Amsterdam, Philadelphia: Benjamins.
Zipf, G.K. (1949). *Human Behaviour and the principles of least effort* (Cambridge, Mass: Addison-Wesley)

Zlatić, V., Božičević, M., Štefančić, H., Domazet, M. (2006). *Wikipedias: Collaborative web-based encyclopedias as complex networks.* Physical Review E, Vol. 74, No.1, 6–11.

APPENDIX

GLOSSARY

A list of the most frequent *WikiSpeak Jargon Terms* [110]
organized according to the word-formation strategies

··· SHORTENINGS ···

INITIALISMS

COTW
(*Collaboration Of The Week*) an article which needs improvement. It is selected by vote to be the subject of widespread cooperative editing for a week.

GPL
(GNU General Public Licence) a license created by the Free Software Foundation. The purpose of the GPL is to grant any user the right to copy, modify and redistribute programs and source code from developers that have chosen to license their work under the GPL. Wikipedia's software is released under this license.

IfD
(*Images For Deletion)* a list of images which are unneeded. Images that have been listed here for more than 5 days are eligible for deletion if a consensus has been reached and no objections to deletion have been raised.

NPOV
(*Neutral Point Of View*) to present possibly subjective content in an objective, neutral, and substantiated manner, in order to avoid edit wars between opposing sides. As a verb it means to remove biased statements. As an adjective, it indicates that an article observes Wikipedia's NPOV policy.

POTD
(*Picture Of The Day*) an image which is dynamically updated each day from Wikipedia:Featured pictures.

RC
(*Recent changes*) a page dynamically generated. It lists all edits in descending chronological order. RC are checked regularly by editors doing *RC Patrol*, which means checking all suspicious edits to discover vandalism as early as possible.

RfM
(*Request for mediation*) an action part of the dispute resolution process.

ACRONYMS

COI
(*Conflict Of Interest)* an incompatibility between the purpose of Wikipedia to produce a neutral, verifiable encyclopedia and the motivations of some editors who promote themselves or other individuals, companies, or groups. When an editor disregards the aims of Wikipedia to advance outside interests, they stand in a conflict.

COIN
(*Conflict of Interest Noticeboard)* a noticeboard for reporting and discussing the application of the Wikipedia:Conflict of interest guideline to incidents and situations where editors have close personal or business connections with article topics.

FAC
(Featured Article Candidate) an article that has been proposed to be featured as one of the best in Wikipedia.

[110] *Wikipedia Glossary* http://en.wikipedia.org/wiki/Wikipedia:Glossary#AFD

IANAL
(I Am Not A Lawyer) an editor who gives his opinion on a legal matter as he understands it, although he is not professionally qualified to do so, and may not fully understand the law in question. May be generalized to other occupations, e.g. *IANAA* (administrator), *IANAD* (doctor).
NOR
(No Original Research) Wikipedia policy that does not allow in citing personal and creative works in articles.

OR
(*Original Research)* the material added to articles that has not been already published by a reputable source. As an encyclopedia, Wikipedia is not the appropriate place to publish original research.

POV
(*Point Of View*) originally referred to each of many perspectives on an issue which need to be considered in an encyclopedic article, nowadays it is often used as a synonym for *bias*, *not neutral* in Wikipedia Community.

CLIPPINGS

ADMIN
(*administrator)* already in use in computer tech jargon, refers to a user with extra technical privileges on Wikipedia, e.g. deleting and protecting pages and blocking users.

CONTRIBS
(*contributions)* edits made by a user.

DIFF
the *Difference* between two versions of a page, as displayed using the *Page history* feature, or *Recent Changes Page.*

DISAMBIG
(*disambiguation,* also used *dab*) the process of resolving the conflict that occurs when articles about two or more different topics have the same natural title.

DUPE
(*Duplicate Article*) identification of a duplicate page that needs to be merged with another.

NOM
(*Nomination*) often found as part of the phrase *Delete per nom*, the term indicates a voter's assent to the main nomination for deletion.

DOUBLE CLIPPINGS

ARBCOM
(*Arbitration Committee)* a group of users that exists to impose compulsory solutions to Wikipedia disputes.

COPYVIOL
(copyright violation) also used *copyviol,* and occasionally *CV,* the term is used when are deleted copyrighted material which have been added without complying with Wikipedia copyright verification procedures.

DICDEF
(*Dictionary Definition*, also used *Dictdef)* commonly used on “Wikipedia Articles for deletion” when referring to an article that is more similar to a dictionary article than to an encyclopedia entry. It is usually a good reason for *transwikifying* it to Wiktionary.

MEDCAB
(*Mediation Cabal)* a group of volunteers who provides unofficial and informal mediation for disputes on Wikipedia.

MEDCOM
(Mediation Committee) part of the formal dispute resolution process on Wikipedia, it was set up in January 2004 by Jimmy Wales, along with the Arbitration Committee, to assist in resolving disputes between users.

PERMCAT
(*Permanent Category*) a category into which an article is assigned to aid reader navigation.

SYSOP
(*System Operator)* a user with extra technical privileges on Wikipedia, specifically, deleting and protecting pages and blocking users.

▪▪▪ COMPOSITES ▪▪▪

COMPOUNDINGS

EDITCOUNTITIS
usually applied to one trying to make as many edits as possible, it refers to an unhealthy obsession with the number of edits that a person makes to Wikipedia.

NAMESPACE
a way to classify pages. Wikipedia has namespaces for encyclopedia articles, pages about Wikipedia (project namespace), user pages (User:), special pages (Special:), template pages (Template:), talk pages (Talk:, Wikipedia talk:, and User talk:) etc.

MEDIAWIKI
the software behind Wikipedia and its sister projects.

META PAGE
a page which provides information about Wikipedia.
ROLLBACK
to change a page back to the version before the last edit.

WIKIBOOKS
a Wikipedia sister project that works to develop free textbooks, manuals, and other texts online.

WIKILINK
a link to another Wikipedia page, as opposed to an external link.

WIKIFAIRY
slang term for a wiki editor who beautifies wiki entries by organizing chaotic articles, and adding style, color and graphics. The efforts of *WikiFairies* are normally welcome in the Wikipedia community, though they do not affect the content of the articles they edit.

WIKIGNOME
a Wikipedian who makes minor, helpful edits without requiring for attention or praise for what he does.

WIKISLAP
provides someone with the URL of a Wikipedia article when a lack of knowledge about a particular topic is expressed.

WIKISPAM
articles or sections created to promote a product. Spamming can also include adding extraneous or irrelevant links to promote an outside site, particularly for commercial purposes.

WIKIMEDIA FOUNDATION
a non-profit organization that provides a legal, financial and organizational framework for Wikipedia and its sister projects and provides the necessary hardware.

AFFIXATIONS

DE-SYSOP
(also used: *De-admin*) to take away someone's sysop status. It is used very rarely and only when someone has voluntarily elected to resign such status, or is judged to have misused their sysop powers.

DE-WIKIFY
(*also used Un-Wikify)* to remove (*de-link)* a wikification of an article. This can be done to remove self-references or excessive common-noun Wikification.
UNENCYCLOPEDIC
(also unencyclopaedic) implies that something is not expected to appear in an encyclopedia, and thus not in Wikipedia.

UN-WIKI
To go against the character of a Wiki. Saying that something is *un-wiki* means that it makes editing more difficult or impossible.

SUBPAGE
a page connected to a parent page. *Subpages* do not have to be used in the main article space.

TRANSCLUSION
inclusion of part of a document into another document by reference.

TRANSWIKI
to move a page to another Wikimedia project, in particular *Wiktionary*, *Wikibooks* and *Wikisource*.

VANDALBOT
a kind of *bot* used for vandalism or spamming. It is recognizable by the fact that one or a few IP-addresses make many similar clearly vandalist edits in a short time. In the worst cases vandalbots can vandalize hundreds of pages in different Wikipedia's articles in few minutes.

▪▪▪ SHIFTS ▪▪▪

SEMANTIC SHIFTS

CABAL
Sometimes assumed to be a secretive organization responsible for the development of Wikipedia, the word is usually used as a sarcastic hint to *lighten up* when discussions seem to become too paranoid. Discussions involving the term may have links to POV/NPOV issues or admin problems.

FOREST FIRE
a flame war which uncontrollably spreads beyond the pages where it began into unrelated articles' talk pages. The flame war is normally kept under control thanks to well-established boundaries for user conduct, clear guidelines for article content, and a formal dispute resolution process

MEAT PUPPET
an account created only for the illegitimate strengthening of another user's position in votes or discussions. Unlike a sock puppet, the account is used by another person.

SANDBOX
a page that users may edit whenever they want. It helps users experiment to gain familiarity with Wiki markup.

SOCK-PUPPET
a user account who has been created secretly by an existing Wikipedian, generally to manufacture the illusion of support in a vote or argument.

VILLAGE PUMP

(also *VP*) the main community forum of Wikipedia where proposals, policy changes, technical and internal problems are announced and discussed in front of a wider audience than a topic-specific page would have.
TROLL
a user who incites or engages in disruptive behavior (the verb is *to troll*) and enjoys causing conflict. However, these are few in number and one should *always assume goof faith* in other editors.

SOFT SEMANTIC SHIFTS

ARTICLE
an encyclopedia's entry.

BUREAUCRAT
A Wikipedia Administrator who has been entrusted with promoting users to SySops.

FORK
a splitting of an entity to satisfy different groups of people. In Wikipedia, this can either mean a project-wide split or the split of an article, usually to accommodate different POVs.

MEDIATION
an attempt by a third party to resolve an *edit war* or other conflicts between users.

ORPHAN
a page with no links from other pages.

REINCARNATION
a new user account created by a banned user to evade the block. This action creates a sock puppet.

REVERT
an edit that reverses edits made by someone else, thus restoring the prior version.

SHORTCUT
a redirect used within Wikispace to enable editors to get to a project page more quickly.

STUB
an article usually consisting of one short paragraph or less.

VANDALISM
a deliberate defacement of Wikipedia pages. This can be done by deleting text, writing nonsense, using bad words, etc.

FUNCTIONAL SHIFTS

INCLUSIONIST
(from inclusion) a user who thinks that Wikipedia should contain as much information as possible. There are varying degrees of Inclusionism. *Radical* inclusionists vote "Keep" on every AfD they come across, while more *moderate* ones merely express their desire for a wide variety of topics to be covered.

LISTIFY
the verb (from the noun *list*) deletion of a category whose content is turned into a list, because this is the best way to present the specific content.

MERGIST
(from the verb *to merge*) a user who adheres to the principle of *Mergism*. The term indicates a compromise between the *Inclusionist* and *Deletionist* principles. A *Mergist* is of the opinion that while many topics merit inclusion, not every topic deserves its own article, and tries to combine these topics into longer and less specific articles.

USERFY

(form the noun *user*) the action of turning a page in the article into a user page or subpage. A common case is where an inexperienced user (a newbie) who is not a notable person has created an article about himself/herself. The article would be deleted after userfying — moving its content to a user page.

WIKIFY
(from the noun *wiki*) sometimes shortened to *wfy* the verb means to format using Wiki markup (as opposed to plain text or HTML) and add internal links to material incorporated into Wikipedia.

▪▪▪ LOANS FROM ICT LANGUAGE ▪▪▪

ARCHIVE
subpage of a talk page to which some parts of the discussion are transferred, to reduce the size of the talk page.

BOILERPLATE TEXT
a standard message which can be added to an article using a template.

BOT
a program that automatically, or semi-automatically, adds or edits Wikipedia-pages.

COMMUNITY PORTAL
one of Wikipedia's main pages. It is found on the left sidebar. It is a page that lists the collaboration of the week, outstanding tasks that need to be addressed, and several other useful information and resources.

LINK ROT
When an article's link is outdated and no longer working, the article is said to have undergone *link rot*.

MIRROR
a website other than Wikipedia that uses content original to Wikipedia as a source for at least some of its content.

RENDER
in the World Wide Web, *rendering* is the operation performed by the user's browser of converting the web document (in *HTML*, *XML*, etc. plus image and other included files) into the visible page on the user's screen.

TAG
In addition to its usual HTML meanings, a tag can simply mean a category or a template that will assign an article to a category. "To tag an article" means to either add a category or a stub template.

▪▪▪ NEW CREATIONS ▪▪▪

GDANZIG
an edit war over which of several possible names should be used for a place. The word is a blending of Gdańsk and Danzig, the two names about which a venerable edit war ensued.

CLIMBING THE REICHSTAG
a humorous way of indicating that an editor has over-reacted during an argument such as an edit-war in order to gain some advantage.

ROUGE ADMIN
a misspelling of "rogue admin" occasionally used by vandals and trolls. Now used jokingly by many Wikipedia administrators, usually to describe themselves performing actions which the affected users may not like (such as blocking vandals and deleting pages).

▪▪▪ BLENDINGS ▪▪▪

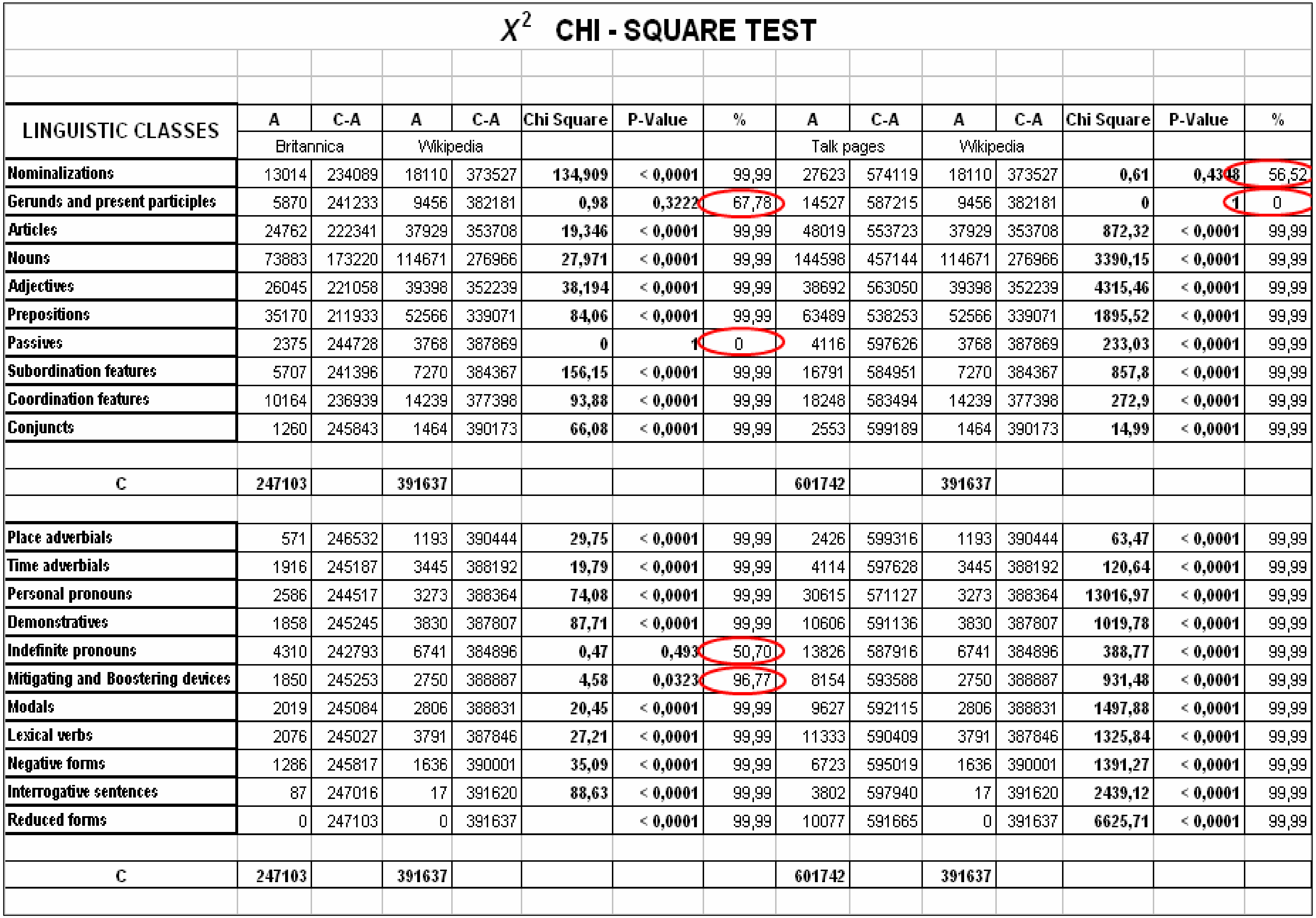
GHITS
short for *Google hits*. The term indicates the number of successful searches for a particular word or phrase using the Google search engine.

SMERGE
(*slight + merge)* sometimes used in *Articles for deletion discussions*, when a topic deserves mention in another article, but not to the extent and detail that is already.

WIKIPEDIHOLIC
(*wikipedia + alcoholic,* also used *Wikiholic)* refers to someone with a serious addiction to Wikipedia. One of the most common characteristics is the victim having a web browser window constantly open to the *Recent Changes* section of Wikipedia (or on the *Watchlist*), and pressing the "Reload" or "Refresh" button with a high frequency.

WIKIQUETTE
(*Wikipedia + netiquette*) the etiquette which defines how working with others on Wikipedia.

WIKTIONARY
(*wiki + dictionary)* a Wikipedia sister project whose aim is to create a free online dictionary of every language.



X^2 CHI - SQUARE TEST

LINGUISTIC CLASSES	A	C-A	A	C-A	Chi Square	P-Value	%	A	C-A	A	C-A	Chi Square	P-Value	%
	Britannica		Wikipedia					Talk pages		Wikipedia				
Nominalizations	13014	234089	18110	373527	134,909	< 0,0001	99,99	27623	574119	18110	373527	0,61	0,4348	56,52
Gerunds and present participles	5870	241233	9456	382181	0,98	0,3222	67,78	14527	587215	9456	382181	0	1	0
Articles	24762	222341	37929	353708	19,346	< 0,0001	99,99	48019	553723	37929	353708	872,32	< 0,0001	99,99
Nouns	73883	173220	114671	276966	27,971	< 0,0001	99,99	144598	457144	114671	276966	3390,15	< 0,0001	99,99
Adjectives	26045	221058	39398	352239	38,194	< 0,0001	99,99	38692	563050	39398	352239	4315,46	< 0,0001	99,99
Prepositions	35170	211933	52566	339071	84,06	< 0,0001	99,99	63489	538253	52566	339071	1895,52	< 0,0001	99,99
Passives	2375	244728	3768	387869	0	1	0	4116	597626	3768	387869	233,03	< 0,0001	99,99
Subordination features	5707	241396	7270	384367	156,15	< 0,0001	99,99	16791	584951	7270	384367	857,8	< 0,0001	99,99
Coordination features	10164	236939	14239	377398	93,88	< 0,0001	99,99	18248	583494	14239	377398	272,9	< 0,0001	99,99
Conjuncts	1260	245843	1464	390173	66,08	< 0,0001	99,99	2553	599189	1464	390173	14,99	< 0,0001	99,99
C	247103		391637					601742		391637				
Place adverbials	571	246532	1193	390444	29,75	< 0,0001	99,99	2426	599316	1193	390444	63,47	< 0,0001	99,99
Time adverbials	1916	245187	3445	388192	19,79	< 0,0001	99,99	4114	597628	3445	388192	120,64	< 0,0001	99,99
Personal pronouns	2586	244517	3273	388364	74,08	< 0,0001	99,99	30615	571127	3273	388364	13016,97	< 0,0001	99,99
Demonstratives	1858	245245	3830	387807	87,71	< 0,0001	99,99	10606	591136	3830	387807	1019,78	< 0,0001	99,99
Indefinite pronouns	4310	242793	6741	384896	0,47	0,493	50,70	13826	587916	6741	384896	388,77	< 0,0001	99,99
Mitigating and Boostering devices	1850	245253	2750	388887	4,58	0,0323	96,77	8154	593588	2750	388887	931,48	< 0,0001	99,99
Modals	2019	245084	2806	388831	20,45	< 0,0001	99,99	9627	592115	2806	388831	1497,88	< 0,0001	99,99
Lexical verbs	2076	245027	3791	387846	27,21	< 0,0001	99,99	11333	590409	3791	387846	1325,84	< 0,0001	99,99
Negative forms	1286	245817	1636	390001	35,09	< 0,0001	99,99	6723	595019	1636	390001	1391,27	< 0,0001	99,99
Interrogative sentences	87	247016	17	391620	88,63	< 0,0001	99,99	3802	597940	17	391620	2439,12	< 0,0001	99,99
Reduced forms	0	247103	0	391637		< 0,0001	99,99	10077	591665	0	391637	6625,71	< 0,0001	99,99
C	247103		391637					601742		391637				

Tab. 1 Chi-Square Test - Britannica vs. Wikipedia

4,7	4,7	19,1	25,2	2290	3170	35,5	34,6	3,1	1,8	12,1	12,1	1,8	2,0	2,5	1,9	3,7	3,2	11,9	10,9	2,8	2,5	9,0	9,7
5,0	5,1	21,8	14,7	371	2085	57,7	50,3	3,0	2,6	10,5	9,2	1,9	2,9	1,3	1,8	3,8	3,8	12,9	9,6	3,0	2,7	13,3	9,7
5,3	4,5	20,6	18,0	411	720	45,0	48,5	10,7	8,6	9,0	8,8	2,2	3,2	1,0	2,2	6,6	2,9	10,9	8,3	3,2	3,3	9,8	8,1
5,3	5,1	23,2	19,4	116	3671	68,1	75,0	6,9	1,6	10,3	8,6	2,6	0,0	2,6	2,0	5,2	3,2	9,5	9,6	0,0	2,6	8,4	8,9
5,7	5,4	26,5	30,1	1511	3465	45,2	44,0	6,8	7,5	10,5	9,8	1,5	1,6	1,0	1,5	6,9	6,1	12,0	15,0	2,1	3,6	12,1	13,9
5,6	5,1	25,4	22,2	355	1418	59,7	56,2	5,9	3,5	5,4	9,9	2,0	2,5	2,5	1,8	2,3	1,8	10,4	9,5	7,0	5,6	10,8	11,3
5,0	4,9	23,1	20,0	2678	11967	32,7	34,2	2,7	2,3	9,6	8,9	4,2	3,1	3,2	2,8	4,6	2,9	12,3	10,7	1,7	2,9	10,4	9,6
5,3	5,2	25,2	25,3	1716	5311	40,9	36,8	4,8	3,6	9,9	9,4	1,6	2,3	0,6	1,5	5,5	3,2	10,8	10,3	3,4	2,7	12,2	11,1
5,8	5,7	29,2	25,6	4996	2919	34,5	35,3	8,2	8,0	8,2	8,0	2,6	2,6	1,8	0,9	3,8	3,0	11,9	10,2	3,1	3,8	12,0	12,8
5,4	5,1	21,7	25,4	565	7253	49,6	45,4	5,1	4,2	11,7	11,1	3,0	2,2	1,2	1,9	5,3	2,2	10,6	11,1	1,9	3,4	10,7	13,3
5,5	5,5	26,3	28,4	4786	6332	34,8	30,2	7,5	7,2	9,0	6,7	2,3	2,2	1,3	1,2	3,4	3,3	11,5	10,5	3,6	4,7	11,7	12,0
5,6	6,0	25,5	22,8	1426	7810	43,1	42,0	6,5	7,1	7,4	7,9	1,6	2,4	1,6	1,7	3,8	3,1	12,8	11,4	3,4	4,1	14,5	13,3
5,2	5,1	27,0	21,9	1756	5152	32,5	34,8	6,2	5,8	9,6	9,4	2,2	2,4	2,3	1,7	3,2	2,1	15,7	12,3	1,7	2,7	10,6	10,4
5,7	5,5	23,5	16,3	1459	2252	41,7	38,8	6,7	6,8	8,1	7,5	2,4	3,8	1,6	1,4	2,9	3,2	13,0	12,5	3,6	3,7	14,5	10,7
5,5	5,4	32,8	23,8	689	10318	49,2	47,2	8,0	5,7	6,1	8,3	1,2	2,5	1,9	1,6	7,1	3,0	11,3	11,7	4,6	4,2	15,8	11,9
5,3	5,3	24,7	16,3	544	5739	48,5	48,8	3,7	3,9	11,9	10,6	1,5	2,1	1,8	1,9	5,3	3,7	11,0	11,3	1,1	2,6	12,7	9,6
5,6	5,5	28,9	23,5	2257	6463	38,1	36,8	6,8	5,8	9,2	9,0	1,7	2,9	1,2	1,4	4,7	2,4	12,3	11,9	3,0	3,3	10,9	11,7
5,0	5,5	23,3	17,2	535	1875	51,8	48,5	3,7	2,2	9,0	8,9	1,3	2,2	2,2	1,7	4,1	3,4	11,6	11,5	0,9	2,2	10,4	8,6
5,5	5,1	18,4	26,2	294	3543	55,1	55,5	4,4	4,2	14,3	11,1	1,7	2,0	2,0	1,1	4,4	2,9	8,5	10,1	2,4	2,0	10,3	11,7
5,4	5,2	19,3	26,0	154	2778	68,2	67,3	3,9	4,5	8,4	10,7	1,3	2,8	4,5	1,6	4,5	2,0	10,4	10,7	2,6	3,6	8,5	11,1
5,3	5,0	22,6	29,8	2172	3850	38,9	33,7	4,1	2,7	11,6	12,6	1,4	1,4	1,2	1,0	5,3	4,1	10,5	11,3	1,5	2,6	9,4	11,0
5,5	6,1	24,4	19,7	414	592	52,9	47,5	2,7	5,7	11,8	7,1	1,9	1,2	1,9	1,5	3,9	4,7	13,5	12,0	1,4	3,5	11,7	15,4
5,2	5,4	24,1	16,9	6808	2710	32,1	33,3	2,7	2,7	12,5	13,2	2,2	1,9	1,4	1,2	3,7	3,4	12,3	10,5	2,4	1,8	10,4	8,9
5,0	4,7	22,7	19.51	341	1370	55,7	60,9	3,2	2,6	10,9	12,5	1,8	2,0	1,8	2,0	4,7	3,1	12,3	14,5	1,8	2,0	7,7	9,5
5,3	5,2	23,0	28,9	17138	7918	29,0	26,1	5,1	4,6	11,2	10,0	2,7	2,2	1,3	1,0	3,7	3,5	12,5	11,0	1,9	2,4	10,7	10,5
5,3	5,2	23,8	23,9	214	2700	55,6	53,7	3,3	3,6	11,7	11,4	3,3	2,9	3,7	2,7	2,8	2,9	13,1	11,0	1,9	2,9	10,2	11,6
5,5	5,2	22,1	25,7	861	7859	47,7	38,8	4,6	4,7	9,2	9,2	2,0	2,2	1,0	0,9	5,1	3,0	8,6	9,6	1,5	2,0	9,8	9,1
5,7	5,2	25,5	22,9	689	1512	50,7	49,7	4,4	2,6	8,6	9,7	2,5	2,0	0,7	0,9	6,5	2,9	11,9	10,5	2,3	2,6	12,1	11,3
6,4	4,9	21,2	22,4	424	5446	54,7	48,1	2,1	2,9	9,4	10,2	2,1	2,4	3,3	1,9	2,4	3,8	12,0	11,3	1,9	3,7	11,0	11,2
5,2	5,2	18,9	17,6	736	5209	49,7	42,5	3,5	3,3	11,4	11,6	1,6	1,9	1,2	1,8	4,2	2,8	13,0	12,0	3,3	3,0	10,5	10,5
5,6	5,3	20,6	30,8	227	6747	58,1	52,6	6,2	5,2	8,8	9,6	2,6	2,7	1,3	1,4	4,8	2,5	11,0	11,7	0,9	2,9	11,5	13,3
5,2	4,9	17,8	27,1	427	9063	53,2	51,1	5,9	4,8	10,5	9,8	2,8	2,5	0,9	1,5	2,1	1,9	13,8	10,6	2,6	3,4	7,5	11,7
5,2	5,1	18,1	20,4	508	4903	42,7	44,9	3,5	3,3	13,2	13,1	2,2	2,6	1,6	1,2	3,1	3,4	13,8	14,3	2,4	2,2	12,6	10,8
5,2	5,4	26,0	23,9	546	7043	51,8	46,2	6,0	6,0	11,5	7,5	1,8	2,6	1,8	0,8	3,1	2,8	11,5	10,4	3,8	1,8	8,9	9,6
5,3	5,3	20,9	19,3	2530	5241	37,1	35,8	4,7	5,2	14,5	14,0	2,2	3,1	1,4	0,6	4,3	3,1	12,5	11,7	2,3	2,7	12,0	11,3
5,0	5,1	23,7	18,8	2916	2203	35,4	38,3	4,2	3,7	9,0	10,8	1,9	2,0	1,0	1,5	2,7	2,9	13,7	12,9	3,2	1,9	12,1	10,7
5,5	5,2	22,6	25,5	859	7703	49,5	47,3	5,4	3,5	11,8	11,2	1,9	2,4	1,3	1,4	3,6	3,4	12,7	11,5	2,3	3,5	11,1	11,2
5,1	5,0	19,4	25,1	796	4298	43,7	42,4	2,6	2,8	10,6	9,7	2,9	2,2	1,0	1,6	3,4	3,2	14,3	13,5	1,8	3,0	8,0	11,2
4,9	4,6	30,7	26,7	430	2166	36,5	42,2	7,0	6,6	9,3	8,4	0,5	2,1	1,9	1,5	3,7	3,8	9,8	8,8	4,2	3,2	14,2	8,0
5,1	5,3	25,4	23,4	203	1312	56,7	57,8	2,5	4,0	13,8	12,4	0,5	2,8	3,4	1,7	1,5	2,4	13,3	11,1	5,9	3,4	15,3	15,0
5,1	5,5	21,1	20,6	15427	5016	26,1	28,2	4,4	4,6	11,7	8,4	2,5	1,7	2,1	1,8	2,4	2,8	12,1	8,6	3,9	3,2	12,5	13,3
5,0	5,2	19,4	23,2	272	1299	54,8	42,4	1,8	2,8	14,7	9,6	3,7	2,4	3,7	2,2	3,3	2,0	9,6	9,8	4,0	3,4	10,7	10,8
3,9	4,7	23,5	20,7	1034	1760	24,4	28,4	4,3	4,0	15,5	10,8	0,8	2,0	2,4	1,4	3,7	3,0	9,1	9,5	4,4	3,7	9,9	9,0
5,6	5,7	21,9	19,4	3136	1703	36,5	33,7	6,9	7,7	8,2	10,0	4,0	2,5	1,8	2,2	3,2	2,6	11,5	10,2	2,0	3,2	13,8	12,1
5,3	4,8	25,9	25,1	673	3418	45,3	38,4	1,6	1,8	12,3	13,8	1,8	1,6	1,2	1,5	1,2	2,6	11,1	10,6	3,4	3,3	11,3	10,0
5,5	5,1	22,5	22,9	135	1009	56,3	50,7	2,2	3,3	14,8	12,1	1,5	1,0	0,0	1,8	3,7	1,9	11,1	9,9	2,2	3,0	11,5	8,5
5,2	5,1	24,4	24,0	341	2765	41,9	45,9	2,9	4,0	11,7	11,2	2,3	1,3	1,8	1,7	2,3	2,2	9,1	9,7	5,3	4,3	12,6	9,3
5,3	4,4	28,4	22,4	256	1499	50,0	42,6	3,9	4,3	10,2	8,7	1,2	1,0	1,6	1,9	3,1	2,4	10,2	7,0	5,5	3,8	15,4	7,3
5,2	4,8	22,3	18,1	3768	2787	36,7	31,3	5,2	4,9	8,6	7,9	1,8	1,8	1,1	1,2	2,7	2,9	11,7	9,5	5,4	6,2	13,8	11,0
5,1	5,4	20,0	22,3	10959	4872	31,3	29,2	4,3	3,9	10,4	8,4	1,7	2,2	1,4	1,6	3,5	2,7	13,2	11,1	3,7	5,3	13,4	11,6
5,0	5,0	22,5	18,6	6539	2488	38,9	40,6	4,6	4,2	8,7	9,1	2,2	2,6	1,2	1,3	3,0	3,2	13,2	12,8	3,8	3,1	12,6	10,2
5,1	5,0	17,4	18,1	24996	5273	21,5	23,8	4,6	4,7	8,7	8,1	2,5	2,7	1,8	1,0	2,4	1,8	12,8	10,6	6,1	6,8	10,7	10,7
5,3	5,3	20,8	19,2	2820	4626	39,9	36,3	5,1	5,3	9,7	7,8	1,8	2,1	0,7	1,1	3,8	2,7	13,0	12,0	2,9	3,8	12,8	11,4
5,8	5,7	15,6	24,3	327	2620	57,5	57,2	4,6	7,1	8,6	11,4	0,6	1,8	1,2	0,3	3,1	1,2	13,1	13,0	3,7	3,6	11,9	15,3
5,1	5,4	30,1	23,2	12832	5679	24,1	24,6	4,7	5,4	9,2	9,2	2,5	1,2	1,7	1,4	2,5	2,7	11,3	11,8	6,5	4,6	13,1	11,8
5,6	5,9	23,1	23,0	4581	1058	41,5	44,2	5,8	8,4	7,5	7,0	2,6	0,5	1,6	1,8	3,5	1,9	11,2	9,8	4,3	5,9	11,3	12,1
5,5	5,2	24,4	20,5	3021	1722	47,5	42,2	5,4	3,3	9,4	9,6	2,8	2,6	0,9	1,5	3,2	2,8	11,3	13,4	3,4	2,8	11,8	10,6
5,3	5,3	26,4	24,5	3218	5963	33,3	34,4	3,4	2,5	7,9	8,2	3,2	2,4	1,1	0,6	2,5	2,9	12,0	12,0	5,1	4,3	11,8	10,2
5,4	5,6	23,1	21,3	3322	5615	33,6	30,2	5,7	7,2	9,6	6,7	2,9	2,7	1,1	1,3	2,5	3,5	11,8	11,4	3,3	2,8	14,7	11,4
5,4	5,4	17,7	23,7	389	5284	55,5	49,2	5,4	6,3	11,8	11,5	0,8	1,9	2,3	1,7	1,8	2,3	11,3	11,4	4,1	4,0	11,6	12,8
5,1	5,1	19,5	19,9	917	1831	38,3	33,7	3,4	2,8	14,9	13,4	2,4	1,7	2,4	1,8	3,8	3,0	10,8	9,3	2,2	2,9	9,5	9,6
4,9	5,6	20,9	19,1	167	1851	52,1	60,7	1,2	4,4	15,0	9,3	0,6	1,8	1,2	1,4	2,4	2,4	12,6	10,4	1,8	3,7	7,2	10,9
5,3	5,4	22,2	26,5	7576	2384	34,2	32,7	4,6	4,3	10,7	9,4	2,5	2,6	2,1	1,5	2,5	2,3	11,0	11,5	2,1	3,1	9,6	11,8
5,8	5,4	18,3	18,4	950	4856	44,4	45,8	5,5	4,5	6,9	8,2	2,1	2,3	1,9	1,7	3,1	3,7	10,8	9,9	2,9	2,9	11,3	11,7
6,0	5,7	23,3	27,0	372	674	53,5	45,2	4,3	1,9	9,4	10,7	1,6	1,9	1,3	1,0	4,8	3,6	11,8	10,8	1,1	1,3	12,7	11,7
5,4	5,3	24,0	22,7	456	4826	55,3	53,2	4,4	4,6	9,6	9,4	1,8	2,6	4,2	1,7	3,5	0,5	11,8	12,1	3,3	3,6	12,7	12,0