

# **An analysis of Wikipedia digital writing**

**dott. Antonella Elia**

Dipartimento di Scienze Statistiche - Sezione Linguistica  
Facoltà di Scienze Politiche - Università degli Studi di Napoli Federico II  
Napoli, Italy

aelia@unina.it

## **Abstract**

This paper is a presentation of a doctoral research in progress focused on a new genre: online encyclopaedias. The introduction to Wikipedia and Encyclopaedia Britannica Online will be followed by a presentation of wiki as a new textual genre. Wikipedia analysis will focus firstly on the investigation of the “WikiLanguage”, the language used in official encyclopaedic articles. Secondly, the “WikiSpeak”, the spoken-written language used by Wikipedians in their backstage and informal community, will be taken into account. The initial findings of this research seem to suggest that, the language of the Wikipedia’s co-authored articles is formal and standardized in a way similar to that found in Encyclopaedia Britannica Online. By contrast, the WikiSpeak, as a new variety of NetSpeak Jargon, can be considered as a creative domain, an independent and individual expression of linguistic freedom of self-representation, characterizing the wiki Computer Mediated Discourse Community.

## **1. Introduction**

The encyclopaedia's structure, either hierarchical or alphabetically ordered, with its evolving nature is particularly adaptable to a disk-based or online format. All major printed encyclopaedias have moved to this method of delivery. Online Encyclopedias can include multimedia (such as video, sound clips and animated illustrations) unavailable in the printed format. They can make use of hypertext cross-references between conceptually related items and, furthermore, they offer the additional advantage of being dynamic: new and frequently updated information can be presented

almost immediately, rather than waiting for the next release of a static format (as with a paper or disk publication).

This research is based particularly on a contrastive linguistic analysis of Wikipedia and Encyclopaedia Britannica Online. The latter is considered one of the greatest examples of general encyclopaedias in the English speaking world. It contains 120,000 articles which are commonly considered accurate, reliable and well-written. Brief article summaries can be viewed for free on the net, while the full text is available only for individuals with monthly or yearly subscription.

On the other hand, Wikipedia is a collaborative authoring project on the web, a repository of encyclopaedic knowledge, an example of a collaborative hypermedium focused on a common project. It is one of the most popular reference websites receiving around 50 million hits per day. It is a social e-democracy environment, designed with the goal of creating a free encyclopedia containing information on all subjects written collaboratively by volunteers. At the time of writing this paper the project has produced over two and half million articles and has been officially recognized as the largest international online community. It consists of 200 independent language editions and the English version is the biggest one with more than 962,995 articles (up to January 2006).

## **2. Wiki as new textual genre**

With reference to the extensive empirical studies of Susan Herring on CMC, wikis and blogs considered as spaces belonging to the second web generation, can be regarded as adding new peculiarities to the existing synchronous and asynchronous tools of the first CMC generation (such as e-mail, mailing list, forum and chat). It is well known in media studies that “the medium is the message” as McLuhan (1964) pointed out in the sixties, and in fact the medium adds unique properties to

the web genre in terms of production, function, and reception which cannot be ignored. Wikis are co-authoring tools which allow collective collaboration. They can be, simultaneously, a repository of information and an asynchronous tool of communication and discussion across the web (see Wikipedia). All wikis have integrated search engines for locating content and are open to anyone since they are considered a public space, even though they can be protected against unauthentic users.

Their main aim is to create documents. Wikis, unlike traditionally designed web sites, encourage “topical writing” by using wiki links and creating a wide network of interconnected pages. The interlinking process becomes simpler to type by just putting the word(s) in square brackets. It simultaneously creates a new topic title (a WikiWord), a new writing space for that topic and a link to that space. Once created, a topic will be available anywhere on the wiki as whenever the WikiWord is typed, it will link to the writing space of that topic (Morgan, 2006).

The writer, the supreme authority in print, is considered the one who transmits content through paper pages, to passive readers, whose role is merely to decode and interpret their message. The electronic writing space, being hypertextual and extremely flexible, changes the landscape. Writers can create multiple structures from the same topics (hierarchy, web, spiral, etc.) and readers can enter, browse and leave text at many points. In the hypertext, the author creates different paths for the reader, although there is neither a canonical path nor a defined page order to follow. The new active readers making their choices, become co-authors of the hypertext (Bolter, 1991). This idea is more pronounced on a wiki than elsewhere, because in an open wiki the reader can (if allowed) really interrupt the process, re-writing, changing, erasing and modifying the original text or creating new topics.

Traditional writing creates a gap between writer and reader. Wiki technology mediates the gap because the two actors assume interchangeable roles in this new open environment. To conclude, wiki text is never static as it is considered revisable, a-temporal as nodes continually change through the collaborative writing process, creating a never ending evolving network of topics. Thus, knowledge becomes webbed, contextualized

though it remains temporary as it can always be changed or vandalized. Luckily, the original version can always, and easily, be recovered by SysOps<sup>1</sup>, through page histories<sup>2</sup> (Morgan, 2006).

Wikis offer two different writing modes. The first one is known as “document mode”. When it is used, contributors create documents collaboratively and can leave their additions to articles. Multiple authors can edit and update the content of documents which gradually become representations of contributors’ shared knowledge (Leuf and Cunningham, 2001). Wikis have two states, “Read” and “Edit”.

“Read state” is by default. In this case, wiki pages look just like normal webpages. When the user wants to edit a page, he/she must only access the “edit state”.

“Document mode” is expository, extensive, monological, formal, refined and less creative than “thread mode”. It is in third person and unsigned. “Document mode” demonstrates that knowledge is collective and that the ideas, not the writers, are the main focus. Writers contribute to “document mode” *refactoring*, reorganizing, incorporating and synthesizing “thread mode” comments in encyclopaedic articles and changing the first to third person (Morgan, 2006).

The second wiki writing mode is “thread mode”. Contributors carry out discussions by posting signed messages in the discussion page connected to the main article. Others reply to the original message and so a group of threaded messages evolves (Morgan, 2006).

“Thread mode” is dialogical, open, collective, dynamic and informal. It develops organically, without a predictive structure. It expresses public thinking, presents multiple positions and is exploratory. Entries are phrased in first person and are signed. Rather than replying to a discussion entry, the writer can *refactor* the page to incorporate suggestions made, then delete the comment.

“Thread mode” demonstrates that knowledge is the result of constructivist collaboration and not a lonely production.

---

<sup>1</sup> SysOp is the abbreviation for "systems operator", and is a commonly used term for the administrator of a special-interest area of an online service.

<sup>2</sup> The page history of all versions of previous pages is available on Wikipedia. It consists of text, date, time and editing authors.

### 3. Research objectives and methodology

#### 3.1. Wikipedia vs Britannica

The first objective of this research has been directed towards the investigation of Wikipedia articles and on what has been defined, in this paper as “WikiLanguage”, the formal, neutral and impersonal language used in the official encyclopedic articles. In this phase, an analysis of randomly selected sample articles has been carried out. The data for this research in progress has been based on two corpora. Up to now, they include a collection of txt files made up of one hundred articles representing topics taken from the Wiki Folksonomy’s<sup>3</sup> eight categories (culture, geography, history, life, mathematics, science, society, technology) and on a contrastive analysis of the same articles found in Encyclopaedia Britannica Online.

The quantitative research, following the methodology proposed by Emigh-Herring (2005), combines corpus linguistics and Biber’s factor analysis. Its purpose has been the empirical measurement of some linguistic features in order to define the degree of formality in the WikiLanguage. The sample articles have been analyzed through the ConcApp Concordancer Program. Different factors have been taken into consideration in order to define the formality of Britannica vs Wikipedia. The first aspect has been articles’ length (total words) as conciseness was found to be a feature of formal written discourse (Chafe 1982, Emigh-Herring 2005). The second, average word length (in letters) as short words have been considered a characteristic of informal genres (Biber 1988, Emigh-Herring, 2005). A high level of lexical density (Halliday, 1985) has been found in formal academic writing. It has been considered the main stylistic difference between speech and writing (Biber, 1988).

Subsequently, the number of unique lexical items in the two corpora has been measured. With reference to the findings of Heylighen

and Dewaele (1999), frequency of word suffixes typical in formal genres (such as *-age*, *-ment*, *-ance/ence*, *-ion*, *-ity*, *-ism*) and impersonal pronouns (it/they) have been calculated (Biber 1988, Emigh-Herring 2005). A contrastive frequency of meaningful keywords has also been investigated. The informality of the language has been measured through the frequency of abbreviations, acronyms, contractions (I’m, don’t, he’s, etc.) and personal pronouns (I, we, you, he/she, they) which have been found to be typical of informal genres, such as face-to-face and phone conversations (Biber 1988, Emigh-Herring 2005). As shown in Appendix A (Fig.1), the first results of this research conducted on one hundred articles have highlighted a number of differences and similarities between Wikipedia and Britannica.

Articles in Britannica have proven to be shorter than those in Wikipedia (average length: 1728 vs 3510 words) and they have shown a higher lexical density (44.9% vs 31.4%). Although the level of total formality is clearly higher in Britannica (50.2% vs 36.6%), the frequency of formal nouns and impersonal pronouns typical of the formal discourse (5.3 vs 5.2) and the average word length (in letters 5.4 vs 5.2) has proven to be very similar. The divergent value is related to lexical density, but if text length varies widely (as happens in the two e-ncyclopedias) the different lexical items will appear to be much higher in the shortest text as their relationship is not linear. Each additional one hundred words of text adds fewer and fewer additional unique words (Biber, 1988). Thus, an interpretation of the collected data seems to suggest that thanks to the collective editorial control, the WikiLanguage of the co-authored articles shows a formal and standardized style similar to that found in Britannica. A table representing a part of the collected data, and their graphical representation, has been provided in Appendix A (Fig. 2,3,4).

#### 3.2 Web analysis

Particular attention has been devoted to Wikipedia digital style due to the importance of the interplay between genre and medium when dealing with web-mediated texts. The layout of sample articles has been investigated (table of content, sections and sub-sections extension) as well as multimodality (tables, graphs, images,

---

<sup>3</sup> Folksonomy is a neologism which indicates a practice of collaborative categorization which makes use of freely chosen keywords. Taxonomy derives from Greek “taxis” and “nomos”. “Taxis” means classification, “nomos” (or nomia) management and “folk” people; so folksonomy means people’s classification management.

audio recordings and videos) and hypertextuality [explicative (internal bookmarks), associative (wikilinks) and explorative links (external weblinks)]. At present Wikipedia does not seem to fully exploit the potential offered by multimodality (and Britannica even less), showing few audio recordings and videos. This is probably due to the feature of Open Source software, keeping with hackers' simple and essential style (i.e. Slashdot and Everything2), to the contributors' average technical skills and to the philosophical choice which grants a privilege to information and content over appearance. One of the prominent properties of Wikipedia is its highly dense hypertextuality when compared to Britannica. The analysis of the articles clearly reveal the abundance of Wikipedia's nodes interlinking and dynamism, made possible by wiki software and, by contrast, the isolation, linearity (page structure) and static nature of corresponding Britannica articles. In this case using Finnemann's (1999) concept of "modal shifts" with reference "to reading mode" and "navigating mode", it is evident that Wikipedia articles actively stimulates the latter allowing the reader to construct his/her own personal pathway, browsing inside and outside the website.

#### 4. WikiSpeak

The second phase of this research will focus on Wikipedia as "Computer Mediated Discourse Community" and on the language, defined in this paper as "WikiSpeak", the language spoken-written by Wikipedians in their informal backstage community. The medium has developed its own wired style and specific glossary, which resembles in some aspects the hackers' Jargon File. The main WikiSpeak distinctiveness lies in the lexicon used.

WikiSpeak is an unofficial and high-context language which can be considered as a new variety of the Netspeak, one of the most creative domains of contemporary English.

Its peculiarity is immediately evident in the "wikilogisms" found in the Community Portal homepage (i.e. *stub*, *NPV*, *wikify*, *backlogs*, *FAQ*, *village pump*, etc.) which can be considered, for its lexical density, a supreme synthesis of WikiSpeak, as well as a political manifesto as the wiki philosophical essence

and its informal community style are clearly disclosed here<sup>4</sup>.

The present investigation has started from its analysis in order to measure the impact of the community front door (content, form, functionality) on the reader, and it will go on analysing the WikiSpeak used in discussion pages connected to the selected articles.

A large number of new words have emerged. WikiSpeak is an informal and colloquial language rich, for example, in acronyms [i.e. NPOV (Neutral Point Of View), COTW (Collaboration Of The Week), IFD (Image For Deletion), etc]. Plenty of abbreviations are also found. They are individual words reduced to two or three letters, [i.e. pls (please), bb ppls (bye bye peoples), etc]. Some abbreviations are like rebuses, as the sound value of the letter, or numeral, acts as a syllable of a word [i.e. B4N (bye for now), CYL (see you later), etc]. Wiki acronyms used in wiki CMC (discussion pages, mailing lists, IRC channels, instant messaging and personal user pages) are not restricted to words or short phrases, but can be sentence-length [i.e. WDYS (what did you say?), CIO (check it out), etc].

Many word processes take place in WikiSpeak, including several ludic innovations. A popular method of creating *wikilogisms* is to combine two separate words to make new *compound words*. Some elements turn up repeatedly, i.e. Wiki (WikiPage, WikiBooks, WikiLink, WikiStress, etc.)<sup>5</sup>. In addition, WikiSpeak makes large use of *blends* (namespace, infobox, quickpoll, etc.) and *semantic shifts* [i.e. orphan, mirror, stub, etc] shown in the wiki glossary available for the newbies.

Distinctive *graphology* is also an important feature of WikiSpeak. All orthographic features have been affected. For example, the status of capitalization varies greatly. There is a strong tendency to use lowercase everywhere on the net. The *lower-case default mentality* means that any use of capitalization is a marked form of communication. Messages wholly in capitals are considered to be shouting and usually avoided. A distinctive

---

<sup>4</sup> In the Community Portal homepage, of 1604 words used, 809 are unique words. The lexical density is very high 50.4%. The keywords are: *help* (19), *you* (16), *article* (16), *collaboration* (8), *free* (7).

<sup>5</sup> In Wikipedia veterans avoid their use as it is considered cliché. However it is tolerated when it refers to technical terms (i.e. *wikilinks*).

feature of Wiki graphology is the way two capitals are used: one initial, one medial.

This phenomenon is called *BiCapitalization* (BiCaps or CamelCase<sup>6</sup>) and is widespread in Wiki community (i.e. MediaWiki, WikiProject, etc.). It is a very interesting example of how a programming language influences the wired style, as BiCaps were used in hackers' communities as a word joiner alternative to the underscore based style and, in the original wiki convention to create links before the invention of [[ \_ ]] square brackets. Now it has become fashionable in marketing for names of products and companies. Outside these contexts, however, BiCaps are rarely used in formal written English, and most style guides recommend against it.

*Spelling practice* is also a WikiSpeak distinctive character. New spelling conventions have emerged, such as the replacement of plural -s by -z. Emotional expressions make use of a varying number of vowels and consonants (yayyyyyyy) and repeated punctuation (WHAT????), but punctuation sometimes tends to be minimalist or completely absent, a great deal depends on the user's personality: some Wikipedians are scrupulous about maintaining a traditional punctuation while some do not use it at all. On the other hand, there is an increased use of symbols not normally part of the traditional punctuation system, such as # , or repeated dots (...), hyphens (---), repeated use of commas (,,,) or asterisks (\*\*). WikiSpeak, as a new variety of the NetSpeak Jargon, can be considered as a creative domain, an independent and individual expression of the linguistic freedom of self-representation in the wiki community of practice.

This research will make use of textual linguistics and corpus linguistics for the investigation of the interactions expressed in the unofficial and informal Wiki CMC.

## 5. Conclusions

In conclusion, this research project has two main focuses: defining the Wikipedia language variations within a dual context of use: official

---

<sup>6</sup> 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.

encyclopaedic entries (WikiLanguage) vs backstage community Speak (WikiSpeak).

Wikipedia, as a new expression for the encyclopedic genre, appears very similar to traditional printed encyclopedias due to its stylistic homogeneity, expressed Neutral Point of View<sup>7</sup> and formal style (Emigh-Herring 2005). The first findings of this research in progress seem to demonstrate how Wikipedia succeeds in reproducing an extant traditional genre even if applied to a collaborative and constructivist scenario. According to Shepherd and Watters (1998), extant subgenres are based on already existing genres in other media forms which have been converted into digital form (i.e. newspaper into electronic news); on the contrary, novel subgenres are entirely dependent on the new medium (i.e. homepages, search engines, webgames, etc.). They stated that when an extant genre migrates to a digital environment, it will initially be faithfully replicated: content and form will be preserved and the capabilities of the new medium will not be fully exploited (see Britannica). At a later stage in the evolution, variant genres are created. This process is driven by the technical capabilities of the new medium. It is the point of view of this study that Wikipedia can be taken as an example of the evolution of an extant traditional genre (encyclopedias) which has been officially preserved in the articles' superficial form, but not in the writing and reading processes (social editing, intertextuality, high informativeness and browsing mechanisms). The articles' textual form seems to suggest that when collaborative users have to respect stylistic established norms (see Wiki Manual of Style<sup>8</sup>) and shared social working ethics (see Wikiquette<sup>9</sup>), diversity and controversy are erased and the official requested style is respected within the open editing system. Nevertheless, technological advantages offered by collaborative software, reinforce the variety, the quick updating and interconnection of the

---

<sup>7</sup> A Neutral Point Of View (NPOV) is writing free from bias. It is generally considered desirable for journalistic and encyclopedic writings. According to the Wikipedia's founder, Jimbo Wales, NPOV is an "absolute and non-negotiable" principle in Wiki Manual of Style.

<sup>8</sup> Manual of Style is a style guide for Wikipedia's contributors. It has the purpose of making the editing easier by following a consistent format.

<sup>9</sup> Principles of Wikiquette are the guidelines on how to work with others on Wikipedia.

information provided by the contributors' multitude. Their voices, even if individually, originally and democratically expressed in the CMC wiki community, are merged and homogenized in the articles' neutrality and formality.

Linguistic analysis cannot be separated from the investigation of the main philosophical and political goals of Wikipedia whose main aim is to pursue freedom of content and information expressed through the Wikipedian "Collective" (Lèvy, 1994) and "Connective" Intelligence (de Kerckove, 1997) in this new acentric rhizomatic environment<sup>10</sup> (Deleuze-Guattari, 1980). Encyclopaedia Britannica is a knowledge compendium without any political meaning hosted by a commercial website (.com). In the 18<sup>th</sup> century, the original French "Encyclopédie" from Diderot and D'Alambert was mainly a political project designed to propagate the ideas of Enlightenment and to establish the reign of reason in Europe (Soufron, 2004). Similarly, Wikipedia in the current I.C.T. age, can be considered as 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. It is aimed at changing the society of the 21<sup>st</sup> 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 wanted the web to be a boundless library of Babel and not a global supermarket as it has become in the *dot.com* era.

## References

Biber Douglas. 1988. *Variation across speech and writing*. Cambridge University Press. Cambridge, UK.

Bolter Jay David. 1991, *Writing Space: Computers, Hypertext, and the Remediation of Print*, Lawrence Erlbaum Associates, N.Y., USA

Chafe Wallace L. 1982. "Integration and involvement in speaking, writing, and oral literature" in D. Tannen (Ed.), *Spoken and Written Language: Exploring Orality and Literacy* (pp. 35-53). Norwood, NJ: Ablex.

---

<sup>10</sup> In *A Thousand Plateaus: Capitalism and Schizophrenia*, Deleuze and Guattari state 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, so the whole is heterogeneous. Deleuze labels the rhizome as a "multiplicity" resistant to structures of domination.

Crystal David. 2001. *Language and the Internet*, Cambridge University Press, Cambridge, UK.

Crystal David 2004. *The Cambridge Encyclopedia of English Language*, Cambridge University Press, Cambridge, UK.

De Kerkhove Derrick. 1997. *Connected Intelligence, the Arrival of the Web Society*, Somerville House Toronto, Canada.

Deleuze Gilles and Guattari Felix. 1980. tr. Eng., 1987, *A Thousand Plateaus: Capitalism and Schizophrenia*, University of Minnesota Press Minneapolis, USA.

Emigh William, Herring Susan. 2005. *Collaborative Authoring on the Web: A Genre Analysis of Online Encyclopedias*. Proceedings of the Thirty-Eighth Hawai'i., International Conference on System Sciences (HICSS-38), IEEE Press, Los Alamitos, USA.

*Encyclopaedia Britannica Online*  
<http://www.britannica.com>

Finnemann Niels Ole. 1999 *Hypertext and the Representational Capacities of the Binary Alphabet*  
<http://www.hum.au.dk/ckulturf/pages/publications/nof/hypertext.htm>

Herring Susan. 1996. *Computer Mediated Communication, linguistic, social and cross-cultural perspectives*, John Benjamins Publishing Company. Amsterdam, Philadelphia.

Heylighen Francis and Dewaele JM. 1999. *Formality of language: Definition, measurement and behavioral determinants*. Internal Report, Center "Leo Apostel", Free University of Brussels, Belgium .

Leuf Bo and Cunningham Ward. 2001. *The Wiki Way: Quick Collaboration on the Web*, Addison-Wesley, New York, USA.

Lèvy Pierre. 1994. *L'intelligence Collective. Pour une anthropologie du cyberspace*, La Découverte, Paris, France.

McLuhan Marshall. 1964. "The Medium is the Message" in *Understanding Media: The Extensions of Man*, Signet, New York, USA.

Morgan M.C. *BlogsandWikis*  
<http://biro.bemidjstate.edu/~morgan/wiki/wiki.ph>

Shepherd Michael, Watters Carolyn. 1998. *The evolution of cybergenres* in International Conference on System Sciences (HICSS-31). Hawai'i, vol. II, p. 97-109 cit. in "Literature Genre & Cybergenre".

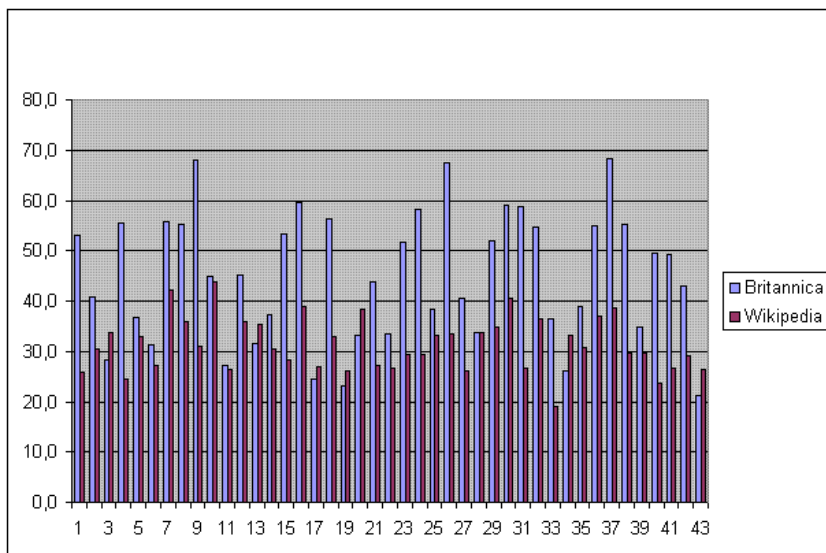
Soufron Jean Baptiste. 2004. *The political importance of the Wikipedia project: Wikipedia toward a new electronic enlightenment era?* <http://soufron.free.fr>

Swales John M. 1990. *Genre Analysis. English in Academic and Research Setting*, Cambridge University Press, Cambridge, UK.

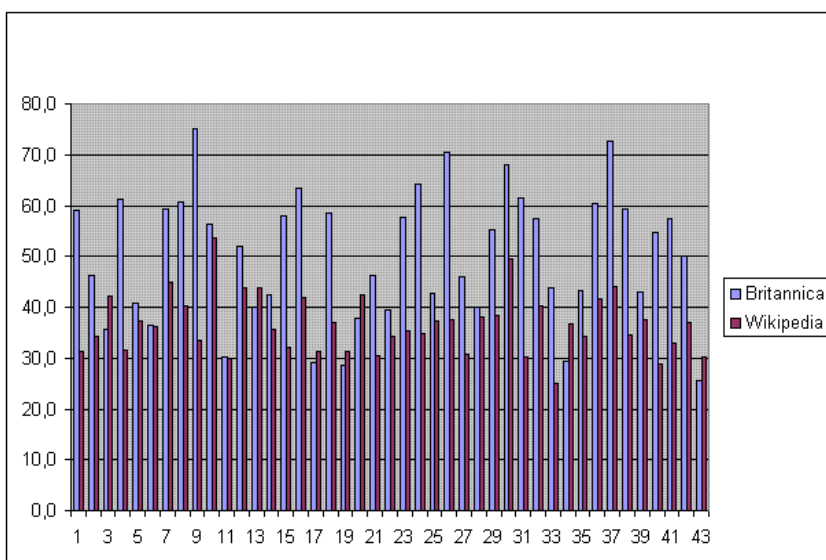
## APPENDIX A

ARTICLE'S TITLE	Lexical	Formal	TOT.	Lexical	Formal	TOT.	Words in articles		Letters in articles		Words' average length	
	density	nouns		density	nouns		BRITANNICA	WIKIPEDIA	BRITANNICA	WIKIPEDIA	BRITANNICA	WIKIPEDIA
		+ P. P.			+ P. P.							
	BRITANNICA			WIKIPEDIA								
Blair Tony	53,2	5,9	59,0	25,8	5,4	31,2	427	9063	2225	44202	5,2	4,9
Walt disney	40,9	5,4	46,3	30,4	3,8	34,2	1716	5311	9054	27355	5,3	5,2
Numerical Analysis	28,4	7,3	35,7	33,7	8,5	42,2	3136	1703	17685	9680	5,6	5,7
Big Bang	55,5	5,7	61,2	24,6	7,1	31,7	389	5284	2094	28450	5,4	5,4
Typewriter	36,8	3,9	40,7	33,0	4,1	37,2	1560	3473	8606	18671	5,5	5,4
Marx	31,3	5,3	36,6	27,3	9,0	36,3	6077	6428	32484	35040	5,3	5,5
Ischia	55,7	3,5	59,2	42,3	2,6	44,8	341	1370	1689	6490	5,0	4,7
Graffito	55,2	5,4	60,6	35,9	4,3	40,2	406	4141	2119	21254	5,2	5,1
Pizza	68,1	6,9	75,0	31,1	2,4	33,5	116	3671	614	18589	5,3	5,1
Jazz dance	45,0	11,2	56,2	43,8	9,7	53,5	411	720	2184	3205	5,3	4,5
Beatles	27,2	3,1	30,3	26,4	3,6	30,0	1852	9084	9859	46474	5,3	5,1
Romanticism	45,2	6,7	51,9	36,0	7,7	43,7	1511	3465	8581	18810	5,7	5,4
Alcoholism	31,6	8,4	40,0	35,3	8,5	43,8	4996	2919	28760	16569	5,8	5,7
Einstein	37,2	5,3	42,5	30,6	5,1	35,7	3867	6391	20042	34148	5,2	5,3
Madonna	53,2	4,7	57,9	28,2	3,8	32,0	615	6145	3255	30753	5,3	5,0
James Dean	59,5	3,8	63,3	38,8	3,1	41,9	442	2342	2268	11500	5,1	4,9
Matrix	24,4	4,8	29,2	27,0	4,3	31,3	1034	1771	4066	8274	3,9	4,7
Quantum Number	56,3	2,2	58,5	32,9	4,2	37,1	135	1009	737	5138	5,5	5,1
London	23,0	5,6	28,6	26,1	5,1	31,2	17138	7942	91560	41239	5,3	5,2
Garibaldi	33,1	4,6	37,7	38,3	4,2	42,5	2916	2203	14570	11238	5,0	5,1
Wars of Roses	43,7	2,6	46,3	27,2	3,3	30,5	796	4298	4065	21390	5,1	5,0
AIDS	33,6	5,9	39,5	26,6	7,8	34,4	3322	5615	17969	31338	5,4	5,6
Bush George	51,8	6,0	57,8	29,3	6,2	35,5	546	7043	2863	37857	5,2	5,4
Berlusconi	58,1	6,2	64,3	29,3	5,6	34,9	227	6748	1260	35990	5,6	5,3
Heart	38,3	4,3	42,6	33,3	3,9	37,2	917	1831	4691	9395	5,1	5,1
Turquoise	67,5	2,9	70,4	33,6	3,9	37,5	277	4258	1550	23131	5,6	5,4
Solar Energy	40,6	5,3	45,9	26,2	4,5	30,7	849	5549	4473	28701	5,3	5,2
Internet	33,7	6,3	40,0	33,7	4,5	38,2	2552	4432	14634	24135	5,7	5,4
Balloon	51,9	3,4	55,3	34,7	3,7	38,4	505	1696	2619	8615	5,2	5,1
Virtual Reality	59,1	8,9	68,0	40,6	8,9	49,5	325	1874	1849	9932	5,7	5,3
U2	58,7	2,7	61,4	26,7	3,5	30,2	438	6433	2313	30752	5,3	4,8
Graph theory	54,8	2,6	57,4	36,5	3,9	40,4	272	1299	1352	6906	5,0	5,3
Boolean algebra	36,5	7,4	43,9	19,0	5,9	24,9	430	2618	2103	9903	4,9	3,8
Himalaya	26,2	3,1	29,3	33,3	3,3	36,6	6808	2710	35642	14579	5,2	5,4
Gobi desert	38,9	4,5	43,4	30,7	3,5	34,2	2172	3850	11529	19312	5,3	5,0
Barcelona	55,1	5,4	60,5	37,0	4,7	41,7	294	3543	1622	18113	5,5	5,1
Bermuda triangle	68,2	4,5	72,7	38,6	5,5	44,1	154	2778	827	14571	5,4	5,2
SARS	55,3	3,9	59,2	29,7	4,8	34,5	456	4826	2485	25611	5,4	5,3
Feminism	34,8	8,2	43	29,6	7,9	37,5	4786	6332	26338	34532	5,5	5,5
Euro	49,6	5,1	54,7	23,7	5,1	28,8	565	7253	3074	37004	5,4	5,1
Racism	49,2	8,3	57,5	26,6	6,3	32,9	689	10320	3818	56183	5,5	5,4
Homosexuality	43,1	6,9	50	29	7,9	36,9	1426	7810	7954	46787	5,6	6,0
Jet engine	21,1	4,4	25,5	26,5	3,6	30,1	5400	5227	27753	27592	5,1	5,3
<b>FORMALITY III %</b>	<b>44,9</b>	<b>5,3</b>	<b>50,2</b>	<b>31,4</b>	<b>5,2</b>	<b>36,6</b>						
				<b>ARTICLES' AVERAGE LENGTH</b>			<b>1728</b>	<b>3510</b>				
				<i>(in words)</i>					<b>WORDS' AVERAGE LENGTH</b>		<b>5,3</b>	<b>5,2</b>
									<i>(in letters)</i>			

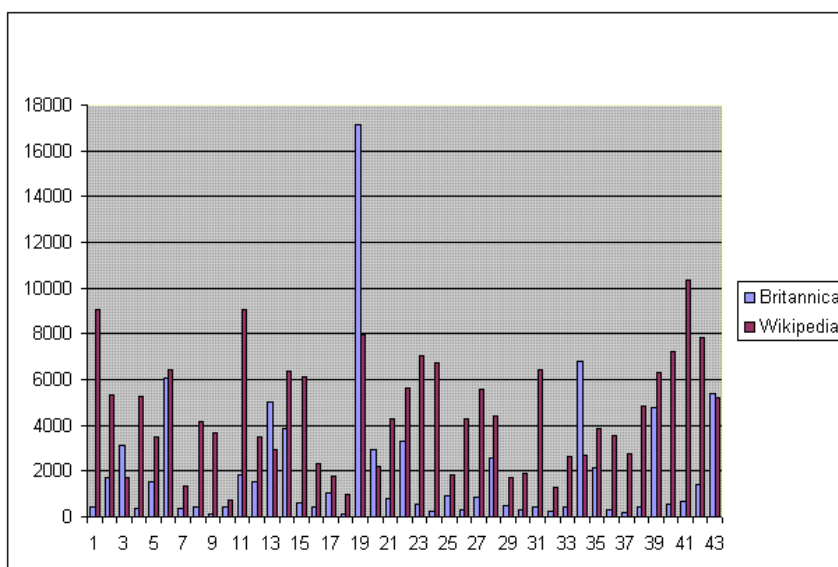
Figure 1. Linguistic formality: Britannica vs Wikipedia



**Figure 2. Lexical density**



**Figure 3. Total formality in percentage**



**Figure 4. Articles' length in words**