File Not Found: Error 404 as an Example of a Spontaneous Web Genre

Grzegorz Cebrat
Państwowa Wyższa Szkoła Zawodowa w Tarnowie
gcebrat@poczta.okay.pl

Abstract

The development of the Internet has led to the emergence of new digital genres, also known as cybergenres or web genres. The existing research into diverse web pages has revealed that the new medium not only generated changes in traditional genres (reproduced or adapted) but also created a number of novel genres. The creation, development and nature of spontaneous genres are connected with the evolution of the web since they have no counterparts in traditional media. The present study deals with one of them: it concentrates on the genre analysis of a small set of 35 web pages representing the 404 error message genre, which are automatically generated to indicate that a server cannot find the page requested by a web user who follows a dead or broken link. The paper presents the origins of the genre in question, its structure, content and communicative purposes. The research has been carried out by means of the methodology of genre analysis proposed by John Swales (1990). Four moves that express the communicative purposes of 404 error have been identified and the commonest rhetoric choices have been analyzed. Additionally, the paper presents a selection of unusual, ingenious or funny 404 error pages.

Key words: Internet, genre, web pages

File Not Found: Błąd 404 jako przykład spontanicznie powstałego gatunku internetowego

Streszczenie

Rozwój Internetu doprowadził do powstania nowych gatunków mowy – gatunków internetowych. Dotychczasowe badania nad różnorodnymi stronami WWW ujawniły nie tylko zaistnienie zmian, które dokonały się w tradycyjnych gatunkach mowy za sprawą nowego medium (gatunki reprodukowane i zaadoptowane) ale powstanie szeregu całkowicie nowych gatunków, ściśle powiązanych z ewolucją i właściwościami Internetu, a niemających swoich odpowiedników w tradycyjnych mediach. Niniejsza praca stanowi przyczynek do badań nad

jednym z nich - stroną Błąd 404, automatycznie generowaną przez serwer, który nie może odnaleźć i wyświetlić strony żądanej przez użytkownika. W oparciu o zbiór składający się z 35 stron Błąd 404, autor analizuje strukturę, cele komunikacyjne oraz zawartość stron. Analiza zgromadzonego materiału przy zastosowaniu metodologii badań nad gatunkami mowy Johna Swalesa (1990) wykazała obecność czterech posunięć (moves) wyrażających cele komunikacyjne oraz najczęściej występujące i najbardziej charakterystyczne sformułowania. Dodatkowo zaprezentowano wybór nietypowych, oryginalnych oraz zabawnych przykładów stron Błąd 404

Słowa kluczowe: Internet, gatunek, strony internetowe

1. Introduction

The study of genres goes back hundreds of years to Greek philosophers and Roman orators, who first understood that not only content but also form, structure and context play an important role in producing and understanding information. However, only recently do modern researchers into the theory of speech genres have a first-hand chance to observe and reflect on the rapid evolution of traditional genres and appearance of new ones in the new environment, that is the World Wide Web. The Internet revolutionized not only the ways of communication but also influenced, shaped or constrained the traditional and conventional system of genres of speech, their forms and features, which has opened the new area of research into the issue how the Internet can shape knowledge dissemination by means of new genres, called web genres, digital genres or cybergenres, most of which were usually drawn from previously existing written genres (see Crystal 2004: 95–127, Grzenia 2008: 150–178). Thus, online encyclopedias, like Wikipedia, commonly replace multivolume printed works (Tereszkiewicz 2010) and so do online dictionaries (Burke 2003), cybermemorials expand the functions and range of press obituaries (Cebrat 2016) and online advertisements supplement traditional classified ads (Ziębka-Białożny 2009). Yet, the advent of the new medium also led to creating and implementing a number of new web genres one of which is the subject of the present paper.

2. Web genres and their classifications

The concept of the web genre was introduced by Joanne Yates and Wanda Orlikowski (1992), who in their study of evolution of the late 19th century business letter into the memo genre in electronic mail, show that genres evolve over time in response to institutional changes and social pressures. In some cases, the changes to an existing genre are so extensive that they lead to the emergence of a new genre. They point out that one of the reasons for the emergence of variants

of existing genres or of new genres is the introduction of a new communications medium. The idea has been developed by Shepherd and Watters (1998: 97), who claim:

> The combination of the computer and the Internet has been such a powerful trigger that it has resulted in the emergence of a new class of genre, which we call cybergenre, existing in this new medium […] Cybergenre can be characterized by the triple, <content, form, functionality>, where functionality refers to the capabilities afforded by this new medium (Shepherd, Watters 1998: 97).

Therefore they identify the cybergenre as two main classes of subgenres: extant and novel. The extant class consists of those based on genres existing in other media, such as paper, radio or television, which have migrated to this new medium, yet they closely resemble existing genres in both content and form with little new functionality. However, they may then gradually evolve into a variant cybergenre as they incorporate functionality, afforded by the computer and the Internet. Thus, extant cybergenres can be subdivided into replicated genres, including most digitized text documents, being relatively faithful reproductions of the genres as they appeared in their source media. Their content and form are virtually the same and there is little new functionality based on the new medium as texts are often created with word processing software offline and simply uploaded to the web (the present paper would exemplify the case if published online). In the process of genre evolution, variant genres, based on existing genres begin to exploit the capabilities afforded by the new medium: their form and content may be somewhat different; however they are characterized by substantial new functionality. Thus, although they tend to preserve the basic genre conventions of their offline precursors, they offer a wide range of adaptations, such as user commenting and hyperlinking. They can be exemplified by hyperlinked documents with images or video components, such as e-book or e-zines.

On the other hand, the class of novel subgenres consists of those genres that have developed in this new medium and have no real counterpart in another medium. They are also subdivided into two subcategories: spontaneous and emergent. The former are completely new genres, which means they have never existed in other media; they arise with no clear antecedent genres such as interactive pages (e.g. search engine interface or online order form). The latter are based on genres originally replicated in the new medium but which have evolved so far from the original that they are classified as being new genres (for instance, the blog, the hotlist, or the homepage). Thus, the primary distinguishing feature between the extant and novel genres is a

level of functionality that makes the novel genres fully dependent for its existence on the new medium as they cannot exist in the traditional media.

Other researchers into cybergenres (also referred to as digital genres, web genres, electronic genres or genres of digital documents) propose different taxonomies. Crowston and Williams (2000) identify four types of web genres: reproduced genres (i.e. replicated), adapted genres (i.e. variant), novel genres and unclassified pages; Santini (2007) observes the existence of the category of emerging genres, i.e. genres in the stage of development, which therefore lack specified genre conventions; that is the category that roughly corresponds to Crowston and Williams' (2000) category of unclassified pages. More recently, Herring (2012) proposes three-fold division into familiar (instead of reproduced), reconfigured (instead of adapted), and emergent genres.

As can be seen, all abovementioned taxonomies propose at least basic two-way classification by distinguishing between a wide range of genres resulting from the evolution of traditional genres in the new medium and displaying varied degree of affinity with their predecessor, and entirely original genres characteristic of and encountered exclusively in the World Wide Web. The genre of 404 Error to be discussed in the following sections represents the latter type, and therefore, following Shepherd and Watters' (1998) taxonomy, can be classified as an example of a spontaneous web genre, as it never existed in any traditional media.

3. The characteristics and origin of 404 Error

Technically speaking, 404 Error (also known as HTTP 404, 404 Not Found, File not found, or simply, 404) is an error message in a Hypertext Transfer Protocol (HTTP) standard response code, which is generated by the website hosting server in order to indicate that a client was able to communicate with a given server, but the server could not find the page that was requested because s/he attempted to follow a link that is either dead or broken. It means that the requested page has been moved or deleted, or, possibly the client has mistyped its Uniform Resource Locator (URL), that is, colloquially termed, its web address.

The category of error pages to which 404 Error belongs is intended for situations in which the error seems to have been caused by the client. Their three-digit code begins with 4, like 401 Unauthorized (displayed when authentication is required) or 403 Forbidden (displayed when the web server denies access to the requested resource). For the first time, they were classified as novel web genres and referred to as 'Pages About Web Servers' by Crowston and Williams (2000: 209), who state:

We found many pages that conveyed information regarding the functioning of Web servers and that can be considered to be novel genres in that they have recognized communicative purposes and distinctive forms. These genres are novel to the Web because their purposes are tied to the functioning of the Web infrastructure (Crowston, Williams 2000: 209).

Like all other HTTP status codes, Error 404 was established by the World Wide Web Consortium in 1992, as a part of the HTTP 0.9 spec. Their status codes were defined by Tim Berners-Lee, the man who invented the World Wide Web in 1989, although an urban legend claims that 404 was named after a room in CERN where the original web servers were located; yet, no such room has ever existed there. ('A brief history of 404')

4. Web pages collection and methodology

The collection of 35 web pages analyzed in the paper comes from the *404 Research Lab*, a web site, started in Minneapolis, MA in 1996. It is exclusively dedicated to searching for, collecting and cataloging 404 Error pages, and provides a gallery of some of the best pages on the web, arranged into twenty thematic categories. Since the present study aims to be qualitative and not quantitative, the choice of web pages to be selected for analysis depended on the preselection done by the *404 Research Lab* and subjective choice of the author of the paper. It should be emphasized that the number of 404 Error pages present in the WWW is theoretically unlimited since each web site owner may create his/her personalized 404 Error page, replace it with a new one, or delete it completely (which means that the default 404 Error page will be displayed).

As Rodrigo (2014: 37f.) claims, web genres can be approached from traditional genre theories although the Internet as a medium has its own intrinsic and unique features. She observes that since the late 1990s a number of scholars investigating particular web genres, for instance, Crowston and Williams (2000), Yates and Orlikowski (2002), Askehave and Nielsen, have followed the analytical model as proposed by Swales (1990), who defined genre as "a class of communicative events, the members of which share some set of communicative purposes," which "constitute the rationale for the genre." Consequently, this rationale "shapes the schematic structure of the discourse and influences and constraints choice of content and style" (Swales 1990: 58). Therefore, it is essential that the communicative purpose(s) of the genre in question should be defined and analyzed to see how they are realized by the move structure of a text, which in turn, determines the choice of rhetorical strategies. In the case of Error 404

pages not only verbal choices should be considered but visual ones as well, since many of their creators purposefully and skillfully combine the text and the image.

5. The analysis of the genre

5.1. Analysis of the communicative purposes of Error 404 page and its components

When an Internet user accesses an Error 404 page s/he does not do it on purpose (unless s/he is a researcher into that genre), but just the opposite: s/he is usually in search of information, which for some reason cannot be accessed. Therefore, the primary purpose of the Error 404 page is to inform him/her about the problem that s/he has encountered and explain why it has occurred; therefore, the commonest or most probable reasons for its occurrence should be provided. Since the user, while reading such a message, displayed on his/her computer screen, might (and usually does) experience negative emotions such as anger, disappointment or dissatisfaction, the owner of the website or creator of the Error 404 should not only explain why the required page was not accessed but also apologize for any inconvenience, comfort the visitor, and offer help in solving the problem, and prevent him/her from discontinuing the search. Thus, the page should contain some tips that can help him/her in finding solution to the problem, which should ultimately help him/her access the page that s/he demanded.

In order to achieve the abovementioned communicative purposes the web page should contain the following components:

- The status code (e.g. 404, 404 Not Found, Error 404) or some other information about the type of problem;
- The web site owner's name and/or logo, which enables his/her identification;
- A (more or less detailed) explanation of why the user is seeing this page;
- A list of common mistakes that may explain the problem;
- The link(s) back to the home page and/or other pages that might be relevant;
- The email link to web site owner so that visitors can report the missing pages, or other problems, especially those concerning the local web site navigation (i.e. concerning links within the text of a given web page, linking to other pages within the website).

5.2 Analysis of functional units (moves and links) of Error 404 page and their rhetorical strategies

Although Error 404 is always a single-page web genre it can contain as many as four functional units (moves). They are usually distinct since boundaries between them are marked off by means

of changes in the type and/or size of the font used, shifts in colors, use of frames, or a combination of those techniques. Additionally, the information can be presented in different sections of the web page, and illustrated with visuals. Its functional structure consists of the following moves:

Move 1: informing the visitor about the problem s/he has encountered.

The shortest and simplest way to do requires providing the number of the error:

> *404*
>
> *Error 404*
>
> *404 – Not found*

Alternatively, instead of providing the number, the problem can be named:

> *Page not found.*
>
> *This page does not exist.*
>
> *Document not found.*
>
> *We cannot find the page or file you are looking for.*
>
> *The requested* [page name/address] *was not found on this server.*
>
> *You've requested a page on* [website name] *that cannot be found.*

This move in the only obligatory functional unit as it identifies the problem. It can be found in all 35 analyzed websites; thus it may be regarded as the distinctive feature of the genre in question; yet, if only this move was present, the message might sound disappointing, impolite and unhelpful were it not followed by three remaining (theoretically, from the point of computer technology, optional) moves.

Move 2: providing reason(s) for its occurrence.

A non-specialist user may not know why Error 404 page has been displayed on his/her computer screen and what it means. Therefore, it is suggested that the most probable reason or reasons for and a short, simple explanation of the problem should be provided. As the examples below reveal, its language is generally simple and easy to comprehend, as it is addressed to non-specialists, therefore computer terminology tends to be avoided:

> *Some files or pages may no longer be in the same place.*
>
> *The page cannot be found because the URL you typed is incorrect.*
>
> *The resource you are looking for has been removed, had its name changed, or is temporarily unavailable.*
>
> *The page you've requested does not exist or is no longer available.*
>
> *It's likely the file has been moved to a new location.*

Move 2 has been found in six of the analyzed pages (17%).

Since the explanation of the problem provided in Move 3 may be not sufficient for an angered or frustrated user, it is advisable to implement Move 3.

Move 3: apologizing, comforting and offering help.

The rarest of all the moves, with only four instances (9%, see examples below). The web page owner may feel responsible for user's problems and expresses his/her apology, which tends to be expressed in a formal style:

> *We regret the misdirection and will redirect you momentarily to our home page.*
>
> *We apologize for any inconvenience this may have caused you.*

Additionally, s/he may comfort the disappointed or confused user and offer his/her help in solving the problem; that reassuring message may become more direct and colloquial to lower the distance between the page owner and the user:

> *Do not worry, though. You have lots of options.*
>
> *Need additional help? Please contact us at* [correct URL].

This leads to Move 4, where more specified hints are provided.

Move 4: suggesting solution(s).

The concluding Move 4 provides at least one solution to the problem the use has encountered. Again, the advice tends to be expressed in direct informal style and avoids unnecessary computer jargon. It is stated by means of either a simple imperative sentence:

> *Try using the search box above to find the page or file.*
>
> *Try to navigate to the page or file by clicking one of the images below.*
>
> *Check that the URL or web address above is correct (if you typed it in).*
>
> *Please check the* [owner's name] *home page to find the new location.*

or a complex one, consisting of an imperative main clause and dependent conditional clause:

> *If you think there is an error or broken link on the* [owner's name] *website, report it to us.*
>
> *If you are not redirected within the next 10 seconds, please Click here.*
>
> *If you feel you've reached this page by mistake, verify the web address and try again!*

As can be seen from the examples above, web page owners tend to provide one solution; hardly ever do they provide alternative solutions, as happens in the following case:

> *If you have time, please could you let us know what you were looking for by getting in touch. Alternatively, please use our site search which is available at the top right of all of our pages.*

This move occurs more frequently than Move 2 and 3, and it has been found almost in every third move (31%).

5.3 Visual imagery and language humor

Since reading such a message, displayed on the user's computer screen, may (and usually does) evoke negative emotions of anger, disappointment or dissatisfaction, and even makes him/her discontinue the search, page creators tend to use a variety of gimmicks, employing original verbal and visual tricks so as to appease or pacify him/her. Creators of Error 404 page frequently employ images of animals, supernatural creatures, film characters, which comment on the problem, for instance, the page of FindMeAGift shop presents a fat panda, that has just swallowed something and says: "I didn't eaten it… I swear." (https://www.findmeagift.co.uk/gifts/coffee-or-wine-glass); similarly, Walt Disney Word page displays one of its cartoon characters, claiming: "Someone Ate the Page! Or for some other reason, the page you're looking for cannot be located." (https://disneyworld.disney.go.com/404/). The page of a social forum, Twingly, shows a ninja fighter with a sword shouting: "Oh no! The Angry Ninja drop kicked one of our servers" (https://www.twingly.com/errors/500.html?aspxerrorpath=/404). Alternatively, the text may point at user's absurdity of his/her pointless search, as in the case of the McLeans' family homepage, where a picture of rabbit accompanies the statement: "I have no idea what you're talking about … so here's a bunny with a pancake on its head" (https://www.michevan.id.au/404), or it includes a play on words, as on a Will Wilkins' home page, where two workers carry a traffic sign "Error 404: Road Not Found" (in computer jargon it should be a path or directory, not a road) (http://iamww.com/404/).

Many a time pages contain images that refer to web owner's profile, profession or activity, and an appropriate jocular comment, for instance the page of a car repair center, Engine-New-Ity, contains a photo of compressed old vehicles in auto scrap and the following lengthy monolog:

> It looks like you're lost. This is Lot 404. This is where we keep the vehicles of the annoying customers. What are you looking for? Engine repairs? Or did your vehicle end up in this lot? Have you been a naughty customer? Remember, Engine-New-Ity is the place for your automotive repairs. You're almost there. Just navigate your way back to where you want to be by using the navigation buttons above. (http://www.enginenewity.com/404.html)

However, other Error 404 page creators refrain from verbal and/or visual jokes; their message is more subtle and may require extra-linguistic knowledge to be fully comprehended by a user, which is exemplified by Error 404 pages of several museums. Their content displays a high degree of intertextuality. For instance, the page of the Prado Museum includes a fragment of a famous painting in its collection (Rogier van Weyden's *The Descent from the Cross*), presenting a head of a weeping woman – in this case she seems not to be crying over the death of Christ but rather a lost page user (https://www.museodelprado.es/404); similarly, the page of Swedish Vasa Maritime Museum comments an old painting of a sinking sailing boat with the statement: "Oops! Något gick galet!" [Oops! Something went wrong!] (https://www.vasamuseet.se/404), and the page of Smithsonian National Air and Space Museum displays a photo of starry sky with a message: "404 Page not found. Houston, we've had a problem", the latter sentence being a quote of the famous Apollo 13 report to the base after an explosion crippled their spacecraft.

6. Conclusion

The conducted analysis has proved that the Error 404 page is an example of novel, spontaneous genre, whose origin and evolution is undoubtedly linked with the development of the World Wide Web. Its rationale is connected with its primary purpose: informing its reader that the requested page cannot be accessed, which is of fundamental significance for any web user. Taking a single page form, the page aims at providing the user with the information about the problem s/he has encountered and, optionally, its possible reasons and practical solutions. In its minimal form, the structure of the genre is limited to informing the user solely about the problem (obligatory Move 1). Although the content of its four-move form tends to be restricted by employing one of typical sentences in each optional move (Moves 2-4), presented in Section 5, a number of Error 404 page creators attempt to make their pages less schematic and formulaic. Thus, taking the linguistic content, recurrent motifs and visual imagery of some analyzed pages into consideration in Section 5, the research has revealed their authors' ingeniousness, creativity and, last but least, humor. The existence of web sites dedicated to the genre, like *404 lab*, also proves vivid interest of non-specialists into most unconventional or creative examples of the genre.

Bibliografia

404 lab [web page] Retrieved from http://queenofsubtle.com/404/ July-December 2017.

"A brief history of 404" *404 lab* [web page]. Retrieved from
http://queenofsubtle.com/404/?page_id=1921 July 6, 2017.

Askehave, Inger, Anne Ellerup Nielsen (2005) "Digital Genres: A Challenge to Traditional Genre Theory." *Information, Technology & People 18*: 120–141.

Burke, Sean Michael (2003) "The Design of Online Lexicons." [In:] Sterkenburg, Piet van (ed.) *A Practical Guide to Lexicography.* Amsterdam: John Benjamins.

Cebrat, Grzegorz (2016) Death Notice as a Genre: An Analysis of The New York Times Online Edition. [Unpublished Ph.D. thesis. Uniwersytet Śląski w Katowicach]. Available at: https://www.sbc.org.pl/dlibra/show-content/publication/edition/220974?id=220974.

Crowston, Kevin, Marie Williams (2000) "Reproduced and Emergent Genres of Communication on the World Wide Web." *The Information Society 16*: 201–216.

Crystal, David (2001) *Language and the Internet.* Cambridge: Cambridge University Press.

Grzenia, Jan (2008) *Komunikacja językowa w Internecie.* Warszawa: Wydawnictwo Naukowe PWN.

Herring, Susan C. (2012) "Discourse in Web 2.0: Familiar, Reconfigured, and Emergent." [In:] D. Tannen, A. M. Tester (Eds.), *Georgetown University Round Table on Languages and Linguistics: Discourse 2.0: Language and new media*. Washington, DC: Georgetown University Press. Retrieved from
http://info.ils.indiana.edu/~herring/GURT.2011.prepub.pdf December 10, 2017.

Rodrigo, Isabel Herrando (2014) "Genre Evolution in Research Communication in English." *Revista Canaria de Estudios Ingleses 69*; 35–49.

Santini, Marina (2007) "Characterizing Genres of Web Pages: Genre Hybridism and Individualization." *Proceedings of the 40th Hawaii International Conference on System Sciences.* Retrieved from
https://pdfs.semanticscholar.org/c6e7/beff68eb8f03523d8a64a7f97bc9b915ee5b.pdf November 11, 2017.

Shepherd, Michael, Carolyn Watters (1998) "The Functionality Attribute of Cybergenres." [In:] HICSS Proceedings of the Thirty-second Annual Hawaii International Conference on System Sciences. Vol. 2. Washington, DC: IEEE Computer Society; 97–109. Retrieved from https://web.cs.dal.ca/~shepherd/pubs/evolution.pdf December 3, 2017.

Swales, John (1990) *Genre Analysis: English in Academic and Research Settings.* Cambridge: Cambridge University Press.

Tereszkiewicz, Anna (2010) *Genre analysis of online encyclopedias: the case of Wikipedia.* Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.

Yates, Joanne, Wanda Orlikowski (1992) "Genres of Organizational Communication: A Structurational Approach to Studying Communication and Media." *Academy of Management Review 17*; 299-326.

Yates, Joanne, Wanda Orlikowski (2002) "Genre Systems: Structuring Interaction Through Communicative Norms." *Journal of Business Communication 39*, No. 1; 13–35.

Ziębka-Białożny, Justyna (2009) *Gender and Discourse - a Contrastive Analysis of Polish and English Singles Ads.* Katowice: Wydawnictwo WSZMiJO.