


Toward an Empirically-generated Typology of Weblog Genres

Maciej Maryl
Polish Academy of Sciences

Krzysztof Niewiadomski
Polish Academy of Sciences

Maciej Kidawa
University of Warsaw

Follow this and additional works at: <http://docs.lib.purdue.edu/clcweb>

 Part of the [American Studies Commons](#), [Comparative Literature Commons](#), [Education Commons](#), [European Languages and Societies Commons](#), [Feminist, Gender, and Sexuality Studies Commons](#), [Other Arts and Humanities Commons](#), [Other Film and Media Studies Commons](#), [Reading and Language Commons](#), [Rhetoric and Composition Commons](#), [Social and Behavioral Sciences Commons](#), [Television Commons](#), and the [Theatre and Performance Studies Commons](#)

Dedicated to the dissemination of scholarly and professional information, [Purdue University Press](#) selects, develops, and distributes quality resources in several key subject areas for which its parent university is famous, including business, technology, health, veterinary medicine, and other selected disciplines in the humanities and sciences.

CLCWeb: Comparative Literature and Culture, the peer-reviewed, full-text, and open-access learned journal in the humanities and social sciences, publishes new scholarship following tenets of the discipline of comparative literature and the field of cultural studies designated as "comparative cultural studies." Publications in the journal are indexed in the Annual Bibliography of English Language and Literature (Chadwyck-Healey), the Arts and Humanities Citation Index (Thomson Reuters ISI), the Humanities Index (Wilson), Humanities International Complete (EBSCO), the International Bibliography of the Modern Language Association of America, and Scopus (Elsevier). The journal is affiliated with the Purdue University Press monograph series of Books in Comparative Cultural Studies. Contact: [<clcweb@purdue.edu>](mailto:clcweb@purdue.edu)

Recommended Citation

Maryl, Maciej; Niewiadomski, Krzysztof; and Kidawa, Maciej. "Toward an Empirically-generated Typology of Weblog Genres." *CLCWeb: Comparative Literature and Culture* 18.2 (2016): [<https://doi.org/10.7771/1481-4374.2846>](https://doi.org/10.7771/1481-4374.2846)

This text has been double-blind peer reviewed by 2+1 experts in the field.

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

This is an Open Access journal. This means that it uses a funding model that does not charge readers or their institutions for access. Readers may freely read, download, copy, distribute, print, search, or link to the full texts of articles. This journal is covered under the [CC BY-NC-ND license](#).

CLCWeb: Comparative Literature and Culture, the peer-reviewed, full-text, and open-access learned journal in the humanities and social sciences, publishes new scholarship following tenets of the discipline of comparative literature and the field of cultural studies designated as "comparative cultural studies." In addition to the publication of articles, the journal publishes review articles of scholarly books and publishes research material in its *Library Series*. Publications in the journal are indexed in the Annual Bibliography of English Language and Literature (Chadwyck-Healey), the Arts and Humanities Citation Index (Thomson Reuters ISI), the Humanities Index (Wilson), Humanities International Complete (EBSCO), the International Bibliography of the Modern Language Association of America, and Scopus (Elsevier). The journal is affiliated with the Purdue University Press monograph series of Books in Comparative Cultural Studies. Contact: <clcweb@purdue.edu>

Volume 18 Issue 2 (June 2016) Article X
Maciej Maryl, Krzysztof Niewiadomski, and Maciej Kidawa

"Toward an Empirically-generated Typology of Weblog Genres"
<<http://docs.lib.purdue.edu/clcweb/vol18/iss2/X>>

Contents of **CLCWeb: Comparative Literature and Culture 18.2 (2016)**
Thematic Issue **New Work in the Empirical Study of Literature**. Ed. Aldo Nemesio
<<http://docs.lib.purdue.edu/clcweb/vol18/iss2/>>

Abstract: In their article "Toward an Empirically-generated Typology of Weblog Genres" Maciej Maryl, Krzysztof Niewiadomski, and Maciej Kidawa propose a typology of weblog genres based on empirical data, namely on the analysis of metadata and the study of blogs' content. In Study 1 they explore 287 categories used by Polish bloggers to classify their blogs. The analysis shows that most categories are topical, but some could be useful for genre analyses. In Study 2 they analyse "syntagma" combinations of 2-3 categories assigned to 88 252 blogs on one of the Polish blog platforms. Through quantitative analysis and clustering 3 main groups are identified: public sphere, private sphere, leisure-time activities. And in Study 3 the authors present an alternative take on findings conducting a qualitative, non-topical analysis of 322 Polish blogs aimed at uncovering weblog genres according to their communicative purpose. In consequence the following genres emerge: diaries, reflection, criticism, information, filter, advice, modelling, and fictionality.

Maciej MARYL, Krzysztof NIEWIADOMSKI, and Maciej KIDAWA

Toward an Empirically-generated Typology of Weblog Genres

Defining a weblog genre is not an easy task. In a 2008 research project Polish bloggers were asked about their attitudes towards and definitions of the cultural activity they are practicing. Researchers decided to entitle their paper using a puzzling definition delivered by one of the respondents, namely that "a blog is ... a blog" (Gumkowska, Maryl, Toczyski 291). Thus, we may travesty Gertrude Stein's famous verse and claim that blog is a blog is a blog is a blog. And we can identify a weblog when we see one similar to Stanley Fish's students as described in *Is There a Text in This Class?* by referring to its formal features. Blogs are "frequently modified web pages in which dated entries are listed in reverse chronological sequence" (Herring, Shedit, Writh, Bonus 142). When we encounter something like that we could qualify it as a weblog. Yet, when blogger are asked under what circumstances would they consider blog a literary genre, responses indicate that a blog is literature, when it is "literary," that is when it contains textual forms we could assign to literary tradition like poems, songs, short stories, dramas (Gumkowska, Maryl, Toczyski 295 **page**). And his brings us to one of the most fundamental questions of internet genre theory which concerns the level of description on which we should operate. In other words, if a "blog is a blog" how could we describe it as a genre? "Does a new medium automatically make for a new genre?" ask Janet Giltrow and Dieter Stein while establishing foundations for such a theory and suggest that "The general characteristic of Internet genres appears to be a greater fluidity and pragmatic openness. There is a constant and fast proliferation of genres—or of forms of communication that are candidates for being a genre" (9). Giltrow and Stein discuss this issue on the example of e-mail which enables several genres, but is not a genre itself; rather, it is a "hypergenre," a superordinate category (9-10). E-mail, as a "technologically-induced form is dispersed amongst several or many genres, recruited to very diverse functions or 'social actions'" (10). Thus it serves as a higher-level category, encompassing multiple (sub)genres. While embracing these distinctions, we propose an approach based on the New Rhetorics approaches to genre assuming a pragmatic perspective and treating a genre as "social action" in Carolyn Miller's sense: "Genre, in this way, becomes more than a formal entity; it becomes pragmatic, fully rhetorical, a point of connection between intention and effect, an aspect of social action" (21). Hence, we speak not about a blog as a genre, but about blog genres in plural, that is, certain forms of communication that employ the "hypergenre" of blog. A similar path was taken by Jill Walker Rettberg, who treats blogs as a medium within which one might "identify different genres and sub-genres such as the diary-style blog, the filter blog or the political blog" (20). The differences would lay in the content not in the form.

We suggested elsewhere that "blog" as a term and concept is similar to "book," a general and formal description of a certain medium of communication which varies across various usages (see Maryl and Niewiadomski 85). We know how a book supposes to look like (usually it has a cover, a table of contents, title page, etc.), but it still can contain various forms like a monograph, magazine, fiction, guide, poem, drama, cooking recipes, etc. It is impossible to describe the book as a genre, yet what we can do is to explore its various forms. Likewise, in the case of blogs, we aim at analyzing lower levels in order to create a typology of weblog genres. Hence, in this study a blog as such is understood as a format, a "hypergenre," a superordinate genre or a medium for diverse genres. All blogs share to some extent following characteristics that distinguishes them from other digital genres:

"the discrete post as fundamental organizing unit; date- and time-stamping of posts; the appearance of posts in reverse chronological order; hyperlinking to external sites; the archiving of posts and references to posts with permalinks and trackbacks; the reference to other likeminded or otherwise interesting blogs through the provision of a blogroll; the capacity for reader comments on posts; and the organization of posts by keywords into separate browsable categories" (Morrison <<http://www.digitalhumanities.org/companionDLS/>>).

We are asking what purposes blogs may serve. As Amy Devitt puts it: "Perhaps blogs are part of a subset of genres that can be defined by format alone, denying our action-based definition of genres. More likely, I think, is that the form in context becomes substantive ... Reverse chronology may be the most visible form in blogs, but it still requires context and combining with other forms, variable as they are, to create meaningful generic action" (40). Our study is aimed at identifying such instances in which certain context determines a distinctive blog genre. This approach shifts the research perspective from investigating genre features of a blog towards the analysis of particular genres the blog format may contain. Blogs are hence approached as textual or rhetorical hybrids build from various competing genres, fulfilling certain aims (see Jamieson and Campbell; Miller and Shepherd). Genre scholarship of weblogs may be distinguished into two strands: one assessing the relationships between blogs and other forms of computer-mediated communication (e.g., Herring, Shedit, Writh, Bonus; Devitt) and the other concentrating on the offline predecessors of weblogs (e.g., Miller and Shepherd; Serfaty). Aimeée Morrison provides a comprehensive overview of existing blog genres typologies which she situates on the spectrum between "those dedicated to journalism, politics, and tabloid-style celebrity gossip" and "the personal blogs, generally considered to be HTML inheritors of paper-based genres of life writing: autobiography, memoir, journal, scrapbook, and diary ... In between these poles lies the vast and shifting middle ground, occupied by aspiring writers seeking both audiences and book contracts with provocative and literate projects, and professionals of all stripes creating classic digest-style blogs on their areas of expertise designed to appeal to a specific core of similarly trained readers"

(<<http://www.digitalhumanities.org/companionDLS/>>). We explore the middle ground and propose a typology of weblog genres that could be based on available empirical material. Although our method

differs across three presented studies, what we address in greater detail is informed by two main methodological assumptions. First, we draw from empirically-grounded construction (see Kluge) as it combines empirical analysis with theoretical knowledge in the process of creating a typology. It is especially visible in Studies 1 and 3 where principles of grounded theory are incorporated into the coding process (see Lonkila). Second, our approach is inspired by the framework of complementary explorative data analysis (CEDA) in that it tries to combine quantitative and interpretive approaches into a single research design in order to standardize hermeneutical findings (see Sudweeks and Simoff). In Study 1 we explore the categories available on Polish blogging platforms and used by authors to classify their blogs. Our analysis shows that most of categories are topical, but others could be useful for genre analyses. Study 2 looks at "syntagmas," that is, combinations of categories on one of the Polish blogging platforms. Through quantitative analysis and clustering we managed to both replicate Morrison's spectrum and complicate it by dividing the core set of genres into four groups. Thus, in Study 3 we present an alternative take on those findings, conducting a qualitative analysis of 322 Polish blogs and establishing a more detailed and non-topical typology of weblog genres.

Study 1: exploration of blog metadata as genre indicators. The digital form of the text gives researchers unprecedented possibilities to analyze new types of extra-textual materials (e.g., links, tags, meta-data, etc.), which should not be omitted in interpretation of the new forms of writing. The aim of this study is to construct a typology of blog genres based on such data, that is on the categories assigned to blogs by their authors. Although most of the blogging platforms nowadays use tags and free keywords, some of them still operate with predefined categories. Our research was conducted on four major Polish blogging platforms blox.pl, bloog.pl, blog.interia.pl, and blog.onet.pl. We chose these platforms owing to their size, heterogeneous content, and the use of categories, while we disregarded smaller blogging platforms dedicated to narrow issues. The analyzed material was downloaded in January and February 2013. Some of the platforms may have changed. The analysis was conducted on 287 categories from the above-mentioned websites, including 125 categories listed as main ones and 162 subcategories. Additional 12 categories from Blog Roku 2012 were analyzed. Given the aim of the research, we did not separate between the categories and subcategories and analyzed them together (it was also dictated by the fact that a category in one platform could be a subcategory in the other). After the preliminary bottom-up interpretation of the list of categories, the research team members created a classification that was later elaborated in the group discussion. The following classes of categories were agreed upon: "topical" (referring to the dominant blog topic), "communicational" (defining the partners of communication), "formal" (hinting on the blog format), "genre" (referring to the common and recognized genres), "perspective" (marking the author's specific take on the world). It should be noted that resulting classes serve as prototypes and are not mutually exclusive, that is each blog may fall under more than one class. Let us discuss them in more detail.

Topical categories: this group contains categories, which either directly (and usually precisely) describe the main topic of the blog (e.g. "Computers," "Foreign languages," "Motorcycles and Scooters," "Comic Books" or hint at a broader body of topics (e.g., "Relationships," "Leisure," "Kids." The latter is the case of supracategories and the former subcategories with specific content (e.g., "Hobby," "Knitting"). We also distinguish between the categories referring to different spheres of human activity: private ("Gardening," "Kitchen and Diet," "Beauty," "Sex and Eroticism") or public ("Television," "Politics," "Journalism," "Law"). Yet, such an analysis would require research into the actual content of the blogs and in this particular study we remain at the metadata level. Communicational categories: this class consists of categories which convey information about communication participants: a writer (e.g., "Company and Professional") blogs written by experts (e.g., "Blogs by Famous People"), a desired reader (e.g., "For Men," "Blogs for Everyone"), a certain community to which writer and readers belong (e.g., "30+," "Women Matters," "Teen," "Kids"). The most problematic issue was to find a consequent and unquestionable difference between topical and communicational categories. We decided to treat a category as communicational only when the category conveyed more than just a simple fact she/he is interested in. For example, "Emigration" category (blox.pl) that first sight it denotes the topic (i.e., everything connected with emigration), yet this category is a subcategory of "Diary" and other subcategories include "Teens" and "Family." Hence we decided that it denotes a subgroup of personal journal writers on emigration rather than the overall topic itself.

Genre categories: this class contains categories which refer to well-established and recognized genres of speech and writing from various domains: literary (e.g., "short stories", "poetry"), journalistic (e.g., "columns," "reflections"), personal writing (e.g., "diary," "journal"), speech (e.g., "gossip," "jokes"). Genres evoke the way a certain topic is approached and presented in a blog. If we take into consideration the category "news," it refers to both the topic (i.e., latest information), but also to the mode of writing (i.e., evoking journalistic forms of presentation). Formal categories: this class of categories refers to those features of blogs which correspond to its format. They often serve as a signal that a non-textual medium is used (e.g., "vlogs") or refer to communication based on movie clips and "pictures" suggesting the priority of images or the visual. This category also informs us that blogs under this category are dedicated to the presentation of various graphic or ICT applications which assist in the content display (e.g., "templates," "blog add-ons"). Perspective: this class consists of categories which evoke a particular and distinctive approach of the blogger to the topic and the dominant mood of the blog. It also offers hints on style and linguistic register of the blog. An example of such a category is "Dark": it does not tell us anything specific about the topic (everything could be dark) nor about the communication participants, genre, or form. What blogs under this category have in common is a dark layout and pictures, dominant poetics of sadness in writing together with a melancholic and pessimistic worldview. Similar—although from an opposite emotional register—would be a category "Crazy" which also denotes a different and more colloquial style. Subsequently, all 287 categories were assigned to

the classes created by the team. The categories were assigned to those classes by three independent coders, who achieved a satisfactory level of intercoder reliability (86% average pairwise percent agreement; Cohen's kappa of .722). Here are our results:

Table 1. Classes of categories. N=287

Class	Number of categories	Percentage of categories
Topical	201	70%
Communicational	37	12.9%
Formal	18	6.3%
Genre	18	6.3%
Perspective	13	4.5%

Note: intercoder reliability is calculated with ReCal3 <<http://dfreelon.org/utis/recalfront/recal3/>>; see also Freelon).

We see that topical categories are dominant in bloggers' self-descriptions of their blogs, so it is hard to distinguish any genres on the base of single category. Yet, it could be possible if we looked at the combination of categories assigned to particular blogs. For instance, the topical category frequently appears with categories from other classes as we present in Study 2.

Study 2: syntagmatic analysis (metadata-based typology of weblog genres): on certain blogging platforms like blogg.pl, users can choose more than one category to describe their blogs. Such sequences of 2-3 categories offer more information about the content and the style (e.g., "Film, Stars, Music"; "Stars, Music, Gossip"; "Leisure, Women Matters"). Such a descriptive strand of categories we call a "syntagma" and we investigate such combinations as a possible key to metadata-generated typology of blog genres. We conducted the study on blogg.pl, the only platform that made available the list of blogs with combinations of categories. We chose blogg.pl for two reasons: first, it published a list of blogs together with its metadata (title and short description, number of visits, creation date, number of posts and categories). Second, it claimed to host over 1 million blogs at the time of the study. Given the fact that according to Internet Live Stats

(<<http://www.internetlivestats.com/internet-users/poland/>>) around 25 million inhabitants of Poland use the internet, that would mean that 4% not only blogs per se, but uses blogg.pl for this purpose. The lists of blogs at blogg.pl are available under particular categories. We downloaded the entire lists from respective categories (55,931 list pages with data on 1,116,934 blogs). Yet, once we removed the repetitions it turned out that blogg.pl contains only 200,319 blogs. Repetitions were caused by putting the same blogs under different categories (e.g., if a blog was assigned three categories it was listed three times on various lists) and some pages on the list are repeated creating the false impression that there are more blogs than in reality. For example, if we follow <<http://blogg.pl/kid,241,knazwa,Pami%EAtniki,sort,o,stroma,2447,katalog.html>> we find a catalogue page with the list of blogs. After clicking "następne" ("next") one sees exactly the same list of blogs on the next page. As seen in Table 2, only 34% of them have more than 5 entries, which is roughly the same amount as in the case of blogs with only one post.

Table 2. Number of posts in blogs on [Blogg.pl](http://blogg.pl). N=200,319

Posts	Blogs	%
1	70 799	35%
2	25 505	13%
3	15 160	8%
4	11 112	6%
5	8 856	4%
More than 5	68 887	34%

On the average, a blog in our sample consisted of 14 posts, yet the mean was only 3 posts per blog. Hence, it is clear that in this sample most of the blogs were short-lived or accidental and only some of them have been maintained for a longer period. Basing on those numbers and subsequent rough analyses of blogs with small number of posts we can distinguish three types of blogs in this sample: fake blogs, test blogs, and actual blogs. The first category consists of blogs created automatically by bots or humans wanting to set up a free website to post links to other websites in order to affect their page rank. Tests blogs are the ones which were set up by some individuals either for learning purposes or to get acquainted with this form of communication. After a couple of initial posts such blogs are abandoned. For instance, many blogs were set up by youths for relational purposes—to share their life stories and interact with friends through comments, and the lack of this feedback resulted in huge disappointment and abandoning the blog after posting a couple of entries. Those relational blogs were quite popular in the pre-Facebook era but this function was taken over by most recent and popular social networks (see Miller and Shepherd 267-71). Finally there are actual blogs whose authors posted content for a longer time. Yet, it is hard to say when should we call a blog "actual." For instance, in our sample of 43,679 (22%) out of 200,319 blogs have more than 10 posts, 24,413 (12%) and more than 20 only 9305 (5%) more than 50 and the just 3989 (2%) blogs exceeded 100 posts with the modest number of 88 (0,04%) beyond the threshold of 1000 posts. We assumed that even if a blog was short-lived, the author had an idea about how to categorize its content. Therefore, we chose to rule out the blogs with a single post treating them as probably fakes.

Study 3: from the group of remaining 129,520 blogs we selected the ones which were classified under more than one category. Hence, the final analysis was conducted on syntagmas from 88,252

blogs from bloog.pl out of which 16,247 had two categories and the remaining 72,005 three categories. Beginning the analyses, we acted under the assumption that the order of categories is random and we put all categories within syntagmas in the same order, i.e syntagmas A-B-C, C-B-A, and B-C-A were treated as one type. The ordering yielded 7689 combinations consisting of 2 and 3 categories as presented in Table 3:

Table 3. 20 most popular syntagmas from bloog.pl. N = 88252

	Syntagma	Occurrences	%
1	politics work psychology	969	1,10%
2	film celebrities music	524	0,59%
3	women matters diaries various	433	0,49%
4	celebrities music gossip	430	0,49%
5	diaries poetry	420	0,48%
6	leisure women matters diaries	388	0,44%
7	women matters diaries	382	0,43%
8	hobby pets	379	0,43%
9	music pictureblogs diaries	379	0,43%
10	diaries various	371	0,42%
11	film photography celebrities	370	0,42%
12	pictureblogs diaries various	359	0,41%
13	leisure diaries various	356	0,40%
14	diaries poetry various	353	0,40%
15	celebrities music	353	0,40%
16	pictureblogs diaries poetry	350	0,40%
17	photography celebrities music	341	0,39%
18	leisure hobby jokes&fun	324	0,37%
19	women matters diaries psychology	306	0,35%
20	women matters diaries poetry	304	0,34%

One should note the overrepresentation of the category of diaries which gives a special genre signal marking the subjective or more private writing. We may also observe some recurrent combinations (e.g., connections between women matters and diaries, as well as between music and gossip). In order to deepen those observations and to examine the entire range of possibilities we decided to look at all these combinations and perform a cluster analysis by transforming syntagmas into cases described by categories as nominal and binary variables. Hence if a certain category appeared in the syntagma, it scored 1 for the relevant variable and 0 for the remaining ones. Working with nominal data, we used the Russel and Rao algorithm for hierarchical clustering and SPSS's within-groups linkage as clustering method to achieve as small as possible distance between the cases in the cluster (see Finch 87; Tan, Steinbach, Kumar 519-24). The results are presented on the dendrogram (Figure. 1).

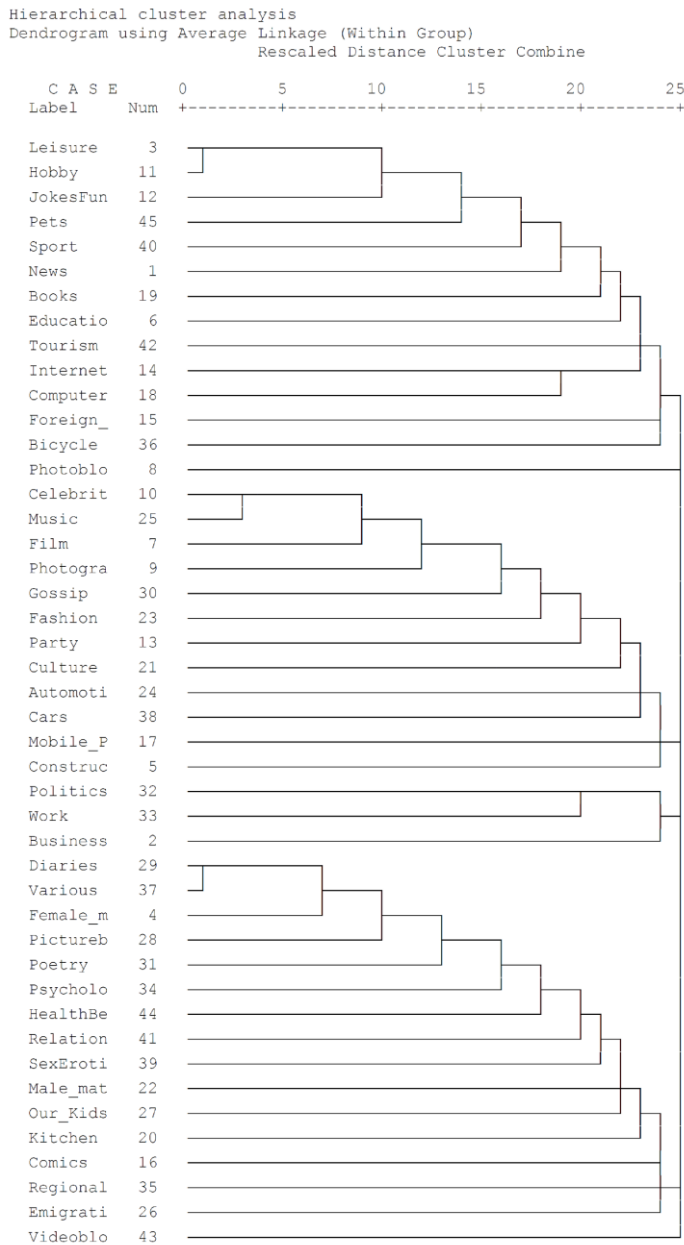


Figure 1. Dendrogram of relationships between categories based on Bloog.pl syntagmas.

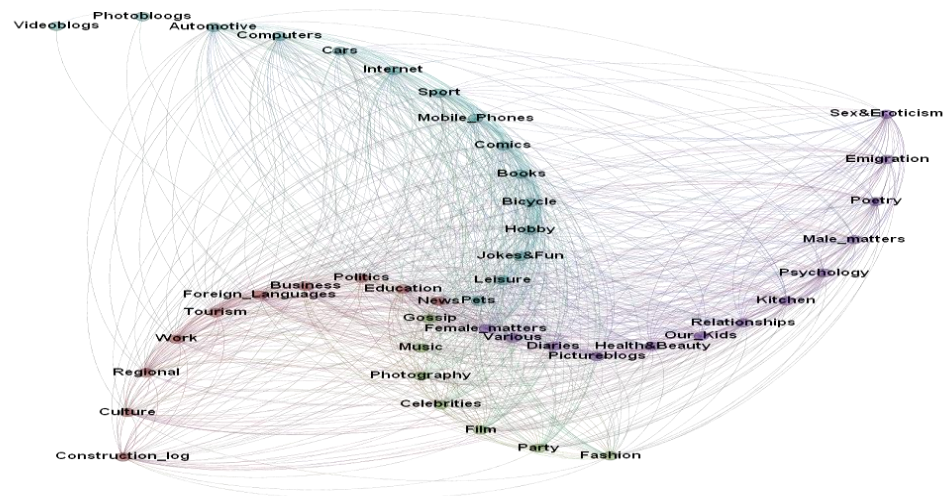
The algorithm clustered the variables into four groups we interpreted as follows:

1. Public sphere and leisure (*News, Leisure, Education, Photoblogs, Hobby, Jokes&Fun, Internet, Foreign Languages, Computers, Books, Bicycle, Sport, Tourism, Pets*)
2. Public sphere (*Business, Politics, Work*)
3. Private sphere (*Female matters, Comics, Kitchen, Male matters, Emigration, Our Kids, Pictureblogs, Diaries, Poetry, Psychology, Regional, Various, SexEroticism, Relationships, Videoblogs, Health&Beauty*)
4. Leisure and culture (*Construction log, Film, Photography, Celebrities, Party, Mobile Phones, Culture, Fashion, Automotive, Music, Gossip, Cars*).

Although the dominant field of each cluster could be identified, there were some outliers making it harder to create a better classification. In order to show the relationships better we conducted a subsequent visualization using Gephi. Since Gephi visualizes relationships (edges) between two elements (nodes), we transformed the syntagmas into pairs (e.g., ABC translates into AB and BC). Every edge is assigned a certain weight, which represents the strength of connections between particular nodes counted as a number of such connections. The resulting network consists of 45 edges (categories) and

1799 nodes. To cluster our network we used Louvain modularity algorithm supported by Gephi (see Blondel, Guillaume, Lambiotte, Lefebvre). The algorithm yielded 4 communities (Q = 0,124), presented in Figure 1: blue on top (33,33%), green on the bottom (28,89%), red on the left (22,22%) and purple on the right (15,56). The results were grouped by modularity class using Gephi's Radial Axis Layout as presented in Figure 2:

Figure 2. Four communities obtained from blog.pl syntagmas



The results seem to be consistent with the ones generated through previous clustering. Although they differ at some points with classifying particular categories, we can distinguish main groups such as public sphere (News, Education, Politics, Business, etc.), private sphere (Diaries, Relationships, Poetry, Sex & Eroticism), leisure-time activities (Leisure, Pets, Hobby, Sport, Computers, Automotive) and cultural activities (Gossip, Music, Photography, Celebrities, Film, Fashion). Both latter categories, as the graph depicts, seem to be intermediaries between the public and the private. Hobby, for instance, denotes certain activities of general interest which are applied in the private (i.e. non-professional) life, whereas the last group seems to refer to the public activities with some relevance for the everyday life (e.g. Fashion or Gossip). The results are consistent with other studies in the field, especially with genre typology based on the spectrum between the public and the private, as proposed by Morrison. Moreover, it proposes a higher level of granularity and distinguishes the public sphere into three groups. It is worth noting that those for groups correspond with print market culture. We could find the antecedents of those genres in quality newspapers, hobby and lifestyle magazines, tabloids, autobiographical writings. It seems that such classifications simply replicate the traditional print market categories. However, the study of metadata has two important shortcomings. First, the typology presented here is mostly topic-oriented, as are most of the categories, what was described in greater detail in Study 1. Therefore it is difficult to assess to what extent the blogs actually represent any distinctive rhetorical features. Second, and most importantly, the quality of the material we used is disputable. Although we operated on over eighty thousand cases, we are still dealing with metadata which are assigned by bloggers arbitrarily. The main problem here is the fact that these categories are projected specifically and given by the platform administrator. Moreover, new categories had been added over the years.

Next, we present a short historical reconstruction based on the Internet Archive's Wayback Machine (<<http://www.archive.org/>>). The platform was launched in the end of 2005 and January 2006 and it had 35 categories. Later in 2006 a new one was added (Celebrities) followed by three more in 2007 (Gossip, Politics, Pets), two in 2008 (Construction, Business), one in 2009 (Fashion), two in 2010 (Emigration, Our Kids), and two in 2011 (FotoBlogs, VideoBlogs). Although the novelty effect is visible in the case of two last categories (the lists were downloaded soon after they were introduced, hence they are underrepresented in the sample), the others seem to be merged with other categories. Yet the fact that users had at their disposal different sets of categories over time may have influenced their choices. Further, it is difficult to judge to what extent their selections resulted from the actual consideration given to their own writing or whether they were made mechanically and in haste to finish the registration procedure (hence the popularity of such syntagmas as "Politics Work Psychology" which in Polish are next to each other on the alphabetically ordered list from which a blogger chooses categories). Finally, we cannot guarantee that they reflected the content of the blog since the author may have changed the original idea, but had never reassigned the categories.

Given the fact that metadata allowed us to establish logical categorization, we decided to conduct a follow-up qualitative study focusing on "actual" blogs' communicative purpose and functions. By doing so we wanted to bring the genre discussion beyond the constraints of the blog topic as we explain in the next study.

Study 3: in order to gather a material for interpretations, we needed to sample a set of actual blogs, excluding the ephemeral or the fake ones. Thus we used the listings of the most popular blogs from the major Polish platforms (blox.pl, blog.pl, blog.interia.pl, bloog.pl), weblogs listed in various industry rankings, current reports, and contests for the best blog of 2012. We ended up with the list of 322 popular Polish blogs, which were downloaded for analyses. The blogs were at first analyzed individually and tagged. Afterwards we discussed our initial classifications in detail during the team meetings in order to reach the maximum agreement over the working definitions which were subsequently applied to blogs. We adopted Miller's approach treating genre as social action and concentrating on the following aspects: communicative purpose (aim and intention), blog's target domain described on the spectrum between public and private spheres, the role of the speaker, dominant features of author's language, discursive dominants of the blog, and methods applied by the author in order to maintain blog's linguistic and topical coherence. We identify the following genres: diaries, reflection, criticism, information, filter, advice, modelling, fictionality, filter, and unblog. The genres we single out and define should be considered as prototypes as an empirical embodiment of the particular genre's general features: they describe the dominant communicative purpose of the blog as a whole, yet not many blogs would qualify only to one of those genres.

Diaries: a narrator speaks about herself/himself and his/her experiences and thus the blog contains descriptions of events from private life. The author is identical with the narrator. The discourse dwells on traditional journal forms (e.g., travel writing, log, diary). Events are described chronologically in a journal manner. The style is confessional and emotional. Most authors choose a dominant theme connected with their work, interests, or, more generally, social role. The roles writers accept are often similar, so there can be different subtypes of diary blogs distinguished. There are travelers' diaries including diaries of parents to seriously ill children or professional diaries (e.g., written by a truck driver, midwife, probation officer, writer, in-line skater, cyclist, midwife).

Reflection: a narrator speaks about "how things are" in the world, formulating universal truths about life and its meaning. The discourse is subjective, yet expressed in the objective terms, usually without the reference to current events employing such genres as aphorism, poem, essay. The style is philosophical or poetical, sublime. The author appears to be eager to express his/her feelings. Therefore, the comments section of the blog is often used as a chat room not only to praise the writer's poetic sense, but also to share emotions the log has aroused. The number of replies is usually larger than in other blog types and the same readers tend to take part in the discussions, creating a community around the writer.

Criticism: a narrator makes a point about the matters regarding the external world and acts as an expert in the particular (usually broad) field of the public sphere. The discourse is similar to journalistic genres (commentary, column, review). The main aim of the author is to express her/his position on the given issue and acts as an expert, who evaluates some social and cultural phenomena and whose opinions are valid for the readers' community. Thus, in many blogs from this category an inner perspective of the author, her/his opinions, thoughts, and feelings equally important as the phenomena she/he tries to evaluate. Authors of such blogs like to emphasize their sincerity and self-reliance as the main virtues of the blogger. It is worth to highlight that also the readers of such blogs tend to express their own views and judgments through comments.

Information: a narrator speaks about the facts in the objective world. Subjectivity is not important here (although always present in blogs). The discourse is factual, descriptive, and objectivizing. It is similar to such genres as a press release or agency news and report. It is mostly concerned with current events which aims at informing about current issues connected to weather and climate, including physical phenomena. We note that this blog genre is underrepresented in our sample which leads us to a hypothesis that objectivity and factuality are peripheral features of blogs.

Advice: a narrator acts as an expert on how we should be doing something. This genre conveys instructions concerning particular activities in private life. The style is transparent, descriptive, and technical at times. These blogs adapt such genres as recipes or Do It Yourself books. On the one hand, it can be argued that these blogs are focused on general issues, they explain how to do something. On the other hand, many of such online handbooks tell the story of the narrator who is experiencing doing something. The advice is given in the everyday context as if it were a small chapter of her/his autobiography.

Modelling: a narrator wants to serve as a role-model for readers and they present themselves as an example. This genre is autobiographical and usually dedicated to life style or fashion. The premise of this blog type is that an attribute such as someone's looks express the entire person, his/her personality, character, ideas, position in society, etc. It may be the reason why such blogs usually contain more advertisements than others. Apart from self-presentation, modelling is also about serving as an inspiration for readers who can adopt parts of blogger's lifestyle in their daily routines.

Fictionality: a narrator presents fictional events and is not identical with the author. Discourse using traditional narrative genres (e.g., short story). Fictional blogs work as 1) platforms for presentation of literary texts published elsewhere or 2) original, born digital texts matching the conditions of an online publication (e.g., the story divided into short episodes, one per post).

Filter: a narrator filters web content providing links to external materials with little or no commentary. Often the narrator publishes photos or videos founded elsewhere in the web (it is especially the case of pornographic blogs). This genre is similar to digest, collection, or anthology. The narrator mediates between the world and the readers by filtering the content.

Unblog: this category was created for blogs only by name, that is, "vortals" run collectively by a group of people writing about current affairs. Although as a genre it could be placed close to the category of "criticism," we treat them as collective online magazines.

In conclusion, various events in the history of the internet had impact on blogs revolutionizing their functions, forms and purposes: better compression algorithms for multimedia, and online video services including YouTube or social media such as Facebook. Yet, despite continuous changes, blogs have already become an indispensable part of our communication sphere and the term "blogger" seems to be a sort of a digital synonym for a journalist and writer. In order to account for these changes and to assess the place and the role of blogs in contemporary culture, we need to understand the genre diversity of blogs. Our typology should be considered polythetic based on similarity between elements rather than on belonging exclusively to a single class (see Bailey 7). Our typology is combined, both conceptual and empirical and we use empirical material. In Studies 1 and 3 we apply an inductive approach and obtain classes through pragmatic reduction of polythetic groups formed through the interpretive process. In Study 2 the polythetic classification is created through the quantification of relationships between elements. Typologies formed in Studies 1 and 2 should be considered conceptual and empirical as they are based on the existing categories and are in fact typologies of natural typology. We use metadata, that is, categories applied on various blogging platforms to classify the content. First, we created a conceptual typology of popular typologies to show which concepts are most prevailing in popular classifications of blogs (Study 1). Second, we performed a quantitative classification of this conceptual typology in order to find empirical patterns in the use of actual typologies (Study 2). Thus we are able to distinguish four groups of blogs dedicated to various spheres of interest (public life, leisure, culture, private matters), which correspond to similar divisions in the contemporary press market. Although our results seem to be consistent with previous research, they describe the topical spectrum of blogs owing to the primarily topical nature of those categories. And third, in order to account for the communicative aspect of blog genres, the follow-up qualitative study was conducted. The typology in Study 3 is empirical and conceptual and it accounts for the communication purpose irrespective of the topic of the blog. Last but not least, we note difficulties with assigning particular blogs to a single class. Hence, coding and subsequent quantification should take into consideration single blog posts which should be coded for genre class separately what at the time of the study exceeded the possibilities of the research team.

Note: above research was supported by the Polish National Science Centre (2011/03/N/HS2/06232). The full description of Study 1 was first published in Maryl, Maciej and Krzysztof Niewiadomski. "Blog = Książka? Empiryczne badanie potocznej kategoryzacji blogów" ("Blog = Book? Empirical Research on the Popular Categorization of Blogs"). *Przegląd Humanistyczny* 4 (2013): 83-90. Copyright release to the authors.

Works Cited

- Bailey, Kenneth D. *Typologies and Taxonomies: An Introduction to Classification Techniques*. Thousand Oaks: Sage, 1994.
- Blondel, Vincent D., Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. "Fast Unfolding of Communities in Large Networks." *Journal of Statistical Mechanics: Theory and Experiment* 10 (2008), P10008.
- Devitt, Amy J. "Re-fusing Form in Genre Study." *Genres in the Internet: Issues in the Theory of Genre*. Ed. Janet Giltrow and Dieter Stein. Amsterdam: John Benjamins, 2009, 27-48.
- Finch, Holmes. "Comparison of Distance Measures in Cluster Analysis with Dichotomous Data." *Journal of Data Science* 3.1 (2005): 85-100.
- Fish, Stanley. *Is There a Text in This Class? The Authority of Interpretive Communities*. Cambridge: Harvard UP, 1980.
- Freelon, Deen. "ReCal: Intercoder Reliability Calculation as a Web Service." *International Journal of Internet Science* 5.1 (2010): 20-33.
- Giltrow, Janet, and Dieter Stein. "Genres in the Internet: Innovation, Evolution, and Genre Theory." *Genres in the Internet: Issues in the Theory of Genre*. Ed. Janet Giltrow and Dieter Stein. Amsterdam: John Benjamins, 2009, 1-26.
- Gumkowska, Anna, Maciej Maryl, and Piotr Toczyski. "Blogs oczyma blogerów" ("Blogs from the Bloggers' Perspective"). *Literatura. Społeczeństwo. Komunikacja* (Literature, Society, Communication). Ed. Anna Gumkowska. Warszawa: Wydawnictwa Akademickie i Profesjonalne, 2009. Vol. 2, 285-309..
- Herring, Susan, Lois Ann Shedit, Elijah Writh, and Sabrina Bonus "Weblogs as a Bridging Genre." *Information, Technology & People* 18.2 (2005): 142-71.
- Jamieson, Kathleen Hall, and Karlyn Kohrs Campbell. "Rhetorical Hybrids: Fusions of Generic Elements." *Quarterly Journal of Speech* 69 (1982): 151-67.
- Kluge, Susann "Empirically Grounded Construction of Types and Typologies in Qualitative Social Research." *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 1.1 (2000): <<http://nbn-resolving.de/urn:nbn:de:0114-fqs0001145>>.
- Lonkila, Markku. "Grounded Theory as an Emerging Paradigm for Computer-assisted Qualitative Data Analysis." *Computer-aided Qualitative Data Analysis: Theory, Methods and Practice*. Ed. Udo Kelle. Thousand Oaks: Sage, 1999, 41-51.
- Maryl, Maciej, and Krzysztof Niewiadomski. "Blog = Książka? Empiryczne badanie potocznej kategoryzacji blogów" (Blog = Book? Empirical Research on the Popular Categorization of Blogs). *Przegląd Humanistyczny* 4 (2013):83-90..
- Miller, Carolyn R. "Genre as Social Action." *Genre and the New Rhetoric*. Ed. Aviva Freedman and Peter Medway. London: Taylor & Francis, 1994, 20-36.
- Miller, Carolyn R., and Dawn Shepherd. "Blogging as Social Action: A Genre Analysis of the Weblog." *Into the Blogosphere: Rhetoric, Community, and the Culture of Weblogs*. Ed. Laura Gurak, Smiljana Antonijevic, Laurie Johnson, Clancy Ratliff, and Jessica Reymann. blog.lib.umn.edu (2004): <<http://blog.lib.umn.edu/blogosphere/>>.
- Miller, Carolyn R., and Dawn Shepherd "Questions for Genre Theory from the Blogosphere." *Genres in the Internet: Issues in the Theory of Genre*. Ed. Janet Giltrow and Dieter Stein Amsterdam: John Benjamins, 2009, 263-290..
- Morrison, Aimée. "Blogs and Blogging: Text and Practice." *A Companion to Digital Literary Studies*. Ed. Ray Siemens and Susan Schreibman. Oxford: Blackwell, 2008 <<http://www.digitalhumanities.org/companionDLS/>>.
- Rettberg, Jill Walker. *Blogging*. Cambridge: Polity P, 2008.
- Serfaty, Viviane. *The Mirror and the Veil: An Overview of American Online Diaries and Blogs*. Rodopi: Amsterdam, 2003.
- Sudweeks, Fay, and Simeon J. Simoff. "Complementary Explorative Data Analysis: The Reconciliation of Quantitative and Qualitative Principles." *Doing Internet Research: Critical Issues and Methods for Examining the Net*. Ed. Steve Jones. Thousand Oaks: Sage, 1999,29-55..
- Tan, Pang-Ning, Michael Steinbach, and Vipin Kumar. *Introduction to Data Mining*. Boston: Pearson Addison Wesley, 2005.

Author's profile: Maciej Maryl works at the Institute of Literary Research of the Polish Academy of Sciences where he is also founding head of the Digital Humanities Centre. His interests in scholarship include sociology of literature and relationships between technology and culture. **E-mail: Maciej.Maryl@ibl.waw.pl**

Author's profile: Krzysztof Niewiadomski works towards his PhD at the Institute of History of the Polish Academy of Sciences. His interests in scholarship include historiography and Polish literature of twentieth century. **E-mail: kniewiadomski@o2.pl**

Author's profile: Maciej Kidawa works towards his PhD in Polish philology at the University of Warsaw. His interests in scholarship include twentieth-century Polish émigré(e) literature and the poetics of the novel. **E-mail: maciej.kidawa@gmail.com**