

ПРАГМАТИЧЕСКИЕ АСПЕКТЫ ИНТЕРНЕТ-КОММУНИКАЦИИ: К РАЗРАБОТКЕ ЖАНРОВЫХ МОДЕЛЕЙ ВЕБ-САЙТОВ

Кононенко И. С. (irina_k@cn.ru)

Институт систем информатики им. А. П. Ершова

СО РАН, Новосибирск, Россия

Ключевые слова: интернет-жанр, веб-сайт, праксиологический параметр, коммуникативный параметр, жанровая структура, жанровый маркер

PRAGMATIC ASPECTS OF INTERNET COMMUNICATION: TOWARDS WEBSITES GENRE MODELS¹

Kononenko I. S. (irina_k@cn.ru)

A.P. Ershov Institute of Informatics Systems, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

A two-level multifaceted genre classification is proposed to cover pragmatic aspects of communication on the Web. Genre categories of websites and genre types of site constituents (pages and structural blocks) are represented as vectors of relevant pragmatic features. Praxeological parameters (activity subject, beneficiary, product, environment) are involved to represent human activity that underlies communication and manifests itself in the site structure, content and form of site constituents. Communicative parameters encompass the hierarchy of communicative tasks (including anticipated reactions of the target audience), functionality of site constituents, and the affordances of communication channel (interactivity, multimodality, and dynamics of content). Functions of site constituents together with medium features are exemplified to determine genre types of pages. The type of a textual page corresponds to a certain genre schematic structure composed of content blocks. The extraction of genre schemata is possible using the so called genre markers (cue words and constructions) that are formalized as lexico-grammatical patterns provided with format conditions.

Key words: web genre, website, praxeological parameter, communicative parameter, genre schematic structure, genre marker

¹ The research was supported by Russian Foundation for Basic Research, project No 13-07-00422.

1. Introduction

Recently, special attention is paid to communication practices realized in the digital environment via numerous web pages and sites. These phenomena are investigated within new research fields such as Internet linguistics and digital genre studies [Santini et al. 2010, Shchipicina 2010]. At the same time the widespread communication on the Web stimulates site constructors to greater effort to support the web-based intercourse with technological and software solutions aimed at automatization of website development [Site Technologies Inc.]. Within this context the more general task is discussed: to support automatic generation of the website structure and design as well as certain elements of the site content.

Possible decision may be based on common features, relatively stable stylistic and compositional types observed in the wide variety of websites. All these features correspond to the classic definition of text genre, according to which genre is a type model for the speech unit construction [Bahtin 1986]. So, any website should be in compliance with some genre model that presents its “standard repeated genre form”.

The wide range of modern approaches to web genres is represented in [Mehler et al. 2010]. A very useful summary is provided in introduction [Santini et al. 2010], whose authors emphasize limitations of purely topical classification of web resources. Really, presentations of the same topic may be essentially distinct, for example, the implantation problem described on the website of a dental office, as opposed to that discussed in the dental health forum. Genre categories should be topically neutral as much as possible, though specific genre forms may be more or less closely related to topics. Yandex catalog of Russian websites allows for the opposition of topic and genre by using the branched and extensive topical hierarchy, along with the simple classification by information types (goods and services proposals, advice and instruction, reference works, forums, and events) [Yandex catalog]. A. A. Kibrik considers genres in close relation with functional styles, which reflect spheres of human activity [Kibrik 2009]. [Crowston et al. 2010] point out that genre is a medium for participation in a communicative act; so, in identifying and labeling genres “the gestalt of the various components of the communicative act” is to be captured.

An intrinsic multifaceted nature of web genre implies functional, formal (compositional and lexico-grammatical), and content aspects, which correspond to at least two levels of physical representation: website as a whole and site constituents (site section as a number of web pages, individual web page, and structural block, or move). Website is a manifestation of a communicative act addressed by the author to the target audience and performed as a part of some human activity. Site genre is a conventional way to perform the communicative act on the Web, so an attempt to systematize and formalize site generation requires considering two types of pragmatic parameters that may form the base of multifaceted classification of web genres:

- praxeological parameters, which are facets to classify spheres of human activities,—to represent activities context of communication (section 2);
- communicative parameters that represent communicative context proper: communicative tasks and characteristics of communicative situation brought about by the medium (channel) of communication (section 3).

- Thus obtained genre categories correlate with site genre models considered as patterns for structuring information with respect to several aspects, in particular, by putting it into compositional scheme and lexico-grammatical shape. In section 4 it is described how site genre pragmatics is reflected in the compositional scheme of a web page text. This structure is identified with the help of genre markers represented as lexico-grammatical patterns.

2. Praxeological Parameters

It is common to describe human activity as a process of purposeful interaction of a subject with an object, this process involving such components as a product (realized goal), resources, and conditions.

In the activities ontology all the concepts are considered as entities that are classified as Objects or Processes, on the one hand, and Agents and Non-Agents, on the other. The Process determines a number of Roles (role relations), each of which represents certain aspect of the process and puts the constraints on the potential role-fillers (participants). Agent entities, as opposed to Non-Agents, are conscious, volitional, and able to intentional activity, hence, they may fill the role of the activity Subject. Other roles that may be filled by Agentive entities are Object, Counter-agent, and Beneficiary. Non-Agent entities are physical, social, and mental objects and processes, as well as temporal and locative objects. Their roles in the activity process are Object, Product, Beneficiary, Time, Location, etc. Ontological concepts can be refined within some specified taxonomy. Concepts, both entities and relations, are characterized by attributes that describe conceptual properties.

Proceeding from these ontological considerations, a set of praxeological parameters has been introduced to build a multifaceted categorization of activities. Some of these parameters and corresponding genre based manifestations are exemplified below.

Subject: {*Individual, Group, Institution, State*}. *Group* can be refined as a *Social group* (particularly *Family*) or a *Community* (such as interest group). Different values of the parameter are reflected in the genre form: first person singular pronouns are specific to personal pages (personal communication) and quite impossible on the website of organization (institutional communication); first person plural pronouns are frequent on institutional commercial and advertising pages but not on the official site of the state authority.

Beneficiary: {*Individual, Group, Institution, Society*}. Benefactive relation corresponds to the participant supposed to make use of the activity results (Product). Different values distinguish benefits (and corresponding activities) that are used individually, collectively, or by the whole society.

Product Separability {*Goods, Services*}. It differentiates between *production of goods* and *rendering services*, as the process of rendering a service overlaps or coincides with consumption of its results. Inseparability can explain, for example, the introduction of constituent block “making an appointment” in the structure of services website.

Product Substantiality: {*Substantial, Unsubstantial*}. *Unsubstantial* benefits are mental or physical properties of a person (health, knowledge, etc.).

Product Type {*Material, Financial, Informational, Spiritual, Vital*}. The combination of “substantiality” and “type” values represent benefit varieties:

<*Substantial, Material*>—the results of *material production* (equipment, buildings, furniture) or *material services* (public utilities, freight services);

<*Unsubstantial, Vital*>—the state of physical or mental state as a result of *vital* (medical or recreational) *services*;

<*Substantial, Spiritual*>—the results of *spiritual activities* in their material embodiment: works of art, literature, socio-cultural events.

Let’s consider but one example—that of the design services. The products of design activity are material by form, but have spiritual aesthetic value, which is reflected in the website content and form. Though words of positive evaluation are characterized with high rate of usage in all services proposals, the design sites are notable for high frequency of affective lexemes, in contrast to the rational evaluations of material goods and services. Moreover, the description of previous activity of the designer (put on the “about us” page) usually includes “honours” content block to present the Subject’s activity advance with relation to some evaluative scale: getting rewards, taking part in prestigious creative contests.

Product Form: {*Digital, Physical*}. By this parameter digital or virtual products that are delivered electronically (like e-text, graphics, audio and video files, software, financial instruments, etc.) are differentiated from real physical products. The “download” block in the website structure is a typical example.

Environment: {*Virtual, Real*}. The *Virtual* value corresponds to activities that are carried out in the virtual world of the Web or another electronic medium. “Product type” and “environment” combinations differentiate activities by “virtuality extent”, e.g., in the sphere of *e-commerce* there are electronic shops that sell physical goods and electronic libraries delivering digital books by download. This distinction has reflections in the website structure: the delivery of physical goods in real environment may cause content blocks that discuss shipping method or present locative and temporal conditions of the activity.

Possible combinations of values of praxeological parameters form feature vectors differentiating activities that underlie websites and corresponding site models. Specific features contribute to the website structure, composition and choice of language means for the constituent pages and their structural blocks.

3. Communicative Parameters

The website considered as a communicative act (CA) addressed by the author to the target audience (TA) is performed in the context of some type of human activity (Ac). Then the author is a Subject, and TA is a Counter-agent of the activity. Information conveyed by means of CA may concern some topic T as well as certain aspects of Ac.

3.1. A site is targeted on realization of **communicative tasks** (CT) and corresponding **reactive tasks** (RT) for TA.

Communicative tasks

1. Phatic: establish the contact and keep in touch with TA;

2. Illocutionary: convey information about T and/or Ac (in particular, for the purpose of getting information that is essential for Ac);
3. Perlocutionary: to ensure the desirable reaction RT of TA:

Reactive tasks

1. Phatic: TA enters into communication and keeps up the exchange/contact;
2. Cognitive:
 - i. TA accepts/learns information about T/Ac;
 - ii. TA activates/develops certain attitude to T and/or Ac (with a pragmatic task in mind, the author is seeking for favourable evaluation of Ac by TA);
3. Communicative: in response to CA, TA performs a communicative act that conveys information about TA (knowledge, beliefs, or attitudes) with respect to T or Ac;
4. Pragmatic: TA performs certain non-communicative act Ac', which is presupposed by Ac and embedded in its structure (order/purchase, donation, application for participation, etc.).

Communicative tasks form the hierarchy, in which the performance of the higher-level tasks is presupposed by accomplishment of the lower-level ones. Task structure is determined by ultimate goals and motives, so it may serve a basis for subdivision of websites into three genre groups.

Informative site: CT structure includes RT1—RT2. Site is aimed at modification of beliefs and attitudes of TA by providing information on the topic T, starting from assumption of the general interest of the audience to T. This genre category is somewhat similar to monologue as no correction of the author's knowledge about TA occurs.

Communicative site: CT structure includes RT1—RT3. Site is intended to facilitate human communication and interactions (blogs, forums, social media sites). It provides dialogue or conversation exchange in the course of which mutual beliefs of the participants may be clarified and corrected.

Business site: CT structure includes RT1—RT4. Site is directed toward organization of joint activity. This genre category is somewhat intermediate between informative and communicative ones. The communication is highly stereotyped and the beliefs about the target audience are more or less general (similarly to informative sites). Still, the site structure ensures a feedback for TA on those aspects of the activity that are crucial for effective accomplishment of pragmatic task RT4.

With the view of specifying genre models of websites, it might be useful to combine the obtained genre categorization with a topical one for a unified hierarchy, or a rubricator, of websites to be constructed. Then the feature vector representing specific genre category of, say, the dentist office website would look as follows:

*<Business, Institutional Subject, Individual Beneficiary, Services,
Unsubstantial, Vital, Physical, Real, Stomatology >*

3.2. The instances of different genre categories are structured differently, i.e. the website genre correlates with certain composition(s) of typified structural

constituents: individual pages, sections (groups of pages), and constituent blocks (moves) within the pages.

Genre typing of block/page/section involves functional aspect as well as considerations of communication medium such as interactivity, content dynamics, and modality. Table 1 illustrates this classification in reference to the components of business site. It is significant that many web pages are multi-functional, for example, *Main/Home page* is intended to welcome TA and inform them of the overall purpose of the site, i.e. to present the Activity with focus on most important aspects (Subject as a site owner, Benefits, Products, News and Events, etc.) and give links to their detailed descriptions (navigation function). Prevailing functionality of the page is resultant of multiple functions of its constituent blocks.

Table 1. Communicative parameters for genre typology of website constituents

Interactivity				
interactive		non-interactive		
<i>Forum</i> <i>Questions and Answers</i> <i>Search form</i> <i>Shopping cart</i> <i>Registration form</i>		<i>FAQ(s)²</i> <i>About us page/section</i> <i>Main/Home page</i> <i>Contact info (Subject)</i> <i>Article</i>		
Dynamics				
invariable content		variable content		
<i>About us page/section</i> <i>Main/Home page</i> <i>Contact info (Subject)</i> <i>Article</i>		<i>Forum</i> <i>Questions and Answers</i> <i>Commented page</i> <i>News</i>		
Modality				
text	image	video	audio	
<i>Article</i> <i>Text portfolio</i>	<i>Photo gallery</i> <i>Before and After</i>	<i>Video gallery</i>	<i>Audio library</i>	
Functionality				
Presentational (Ac, Subject)	Informative (Ac, T)	Informative (Product)	Contact	Directive
<i>Main/ Home page</i> <i>About us page/ section</i>	<i>Article</i> <i>Gallery</i> <i>FAQ</i> <i>News</i> <i>Staff</i>	<i>Catalogue</i> <i>Goods info</i> <i>Services info</i> <i>Price list</i>	<i>Contact info</i> <i>Registration form</i>	<i>Questions and Answers</i> <i>Commercial page</i> <i>Shopping cart</i> <i>Registration form</i> <i>Search form</i>

² The FAQ page is structured as a succession of question-answer pairs, but unlike *Questions and Answers* pages FAQs are non-interactive as they are created by site developers on base of preliminary analysis of possible informational needs of the audience.

4. Genre Schemata and Genre Markers

Website genre corresponds to a number of variants of site structure, which involves the choice of site constituents (sections, pages, and blocks) with their genre types, layout, and hyperlinks. On the page level genre types correspond to **genre based schematic structures** composed of functionally determined content blocks.

In case of a textual page, the compositional scheme is described as a succession of text blocks³, each being a relatively independent and semantically coherent text fragment that represents certain **content aspect** of the website pragmatics. Consider the site genre defined as <Business, Services, Institutional>. For this category of sites the presentational *About us* page usually includes the description of the Activity and its Subject and may look like the following succession of content blocks:

<About_Preamble>, <History>, <Advantages>, <Licences>, <About_Coda>.

The proposed formalization of genre schemata of textual pages is based on the earlier works on summarization of scientific papers by the so called “indicator method” [Bljumenau et al. 1981]. According to this method, identification of content aspects is supported by special lexicons of non-topical words used to design and organize scientific text narrative.

The analysis of business websites of Runet has discovered a wide range of characteristic words, distinctive set expressions, verbal clichés that indicate pragmatic content aspects and provide the clues for determining the boundaries of their presentation in the text (text fragments). These words and expressions are specific for particular genres and may be considered as **genre markers**. For example, the presentational pages of *Services* sites (*About us* or *Home*) usually focus on the “Advantages” aspect that describes favourable features of the proposed services in order to demonstrate the superior position of the activity Subject over the competitors. The corresponding content block could be detected on the page with the help of genre markers “why us”, “our advantages”, “five reasons” (see table 2 for formal descriptions of Russian markers and counterpart English examples).

We are developing the inventory of Russian genre markers, which are formalized as lexico-grammatical patterns (cf. [Bol’shakova et al. 2006]). The pattern expression may include words, punctuation marks, and slots that are bound with lexico-semantic and grammatical constraints (grammatical class, features, agreement). The pattern elements may be optional (shown in square brackets) and variable (curly brackets). In addition, each pattern is accompanied with text fragment presentation conditions partially based on HTML text formatting tags. They specify the format type (heading, title, list, paragraph, etc.) and position (beginning, end, inside, next to, etc.) of text fragments to be identified (marker fragments) and extracted (aspect fragments). So for any marker fragment identified, say, in the heading, the scope of the aspect fragment may be defined as the list following the marker fragment, or the text following the marker fragment as far as the next marker.

³ In [Kibrik 2009] it is proposed that the linguistic definition of genre and genre categories should be based on the study of types of “passages” (i.e. blocks of genre schemata).

Table 2. Genre markers of the content aspects: “Advantages”

Aspect	Marker pattern	Text presentation		Genre	
		Marker fragment	Aspect fragment	Page type	Site category
Advantages	<p>“why us” почему {мы [лучше]; именно мы; именно (наш X); Y} Lex-sem X—<company_type> компания; фирма; центр; клуб; магазин... Y—<company_name> Gramm X <Noun, Case=nom, Number=sing> (наш, X) <Agree (Gender, Number, Case)></p>	Head Par_begin	List_next Text_next	About us Home	<Business, Services, Institutional>
	<p>English examples: <i>Why us? Why choose us; Why our company?</i></p>				
	<p>“our advantages-1” [наши] преимущества</p>	Tit_in	Text	Advantages (in About us section)	
	<p>English examples: <i>Advantages; Our Competitive Advantages</i></p>				
	<p>“our advantage-2” [в чем] {наши преимущества; преимущества {работы; сотрудничества} с {нами; (наш X); Y} Lex-sem X—<company_type> компания; фирма; центр; клуб; магазин... Y—<company_name> Gramm X—<Noun, Case=instr, Number=sing> (наш, X) <Agree (Gender, Number, Case)></p>	Head Par_begin Par_in	List_next Text_next	About us Home	
<p>English examples: <i>Some of the advantages of our company</i></p>					
<p>“five reasons” ([N] причина) [<...>] {{ купить; покупать} [именно] {у нас; в (наш X); в Y} Lex-sem N—number/Numeral X—<company_type> компания; фирма; магазин... Y—<company_name> Gramm X—<Noun, Case=loc, Number=sing> (наш, X) <Agree (Gender, Number, Case)> (N, причина) <Coord (Gender, Number, Case)></p>	Head Par_begin Head_in Par_in	List_next Text_next	About us Home	< Business, Commerce, Institutional >	
<p>English examples: <i>Top 5 Reasons to Buy Direct from Knipf; A few reasons why you should order from us; Top 10 reasons to purchase from Nissan of McKinney</i></p>					

5. Conclusion

In the study reported here an attempt has been made to look at the issue of identifying genres on the Web for the purposes of Web generation. The pragmatics based approach to the development of a multifaceted web genre classification is two-level: website genre categories considered with regard to praxeological and communicative context and page genre types specified in terms of functional and medium-related features.

The genre type of a textual page corresponds to a variant of genre schematic structure that is composed of blocks, each representing some content aspect. Genre schemata could be extracted from texts of web pages by using the inventory of genre markers, which is now under development. Genre markers are formally described by common lexico-grammatical patterns, additionally supplied with text presentation (format and position) features. The full-scale repertoire of genre markers described in this way will be useful to not only analyze web pages and identify their genre based schematic structures but also to generate such a structure, expand, and partially populate it with standard content blocks.

References

1. *Bahtin M. M.* (1986), The Problem of Speech Genres [Problema rechevyh zhanrov], in *Aesthetics of Verbal Creation [Jestetika slovesnogo tvorchestva]*, Iskuststvo, Moscow, pp. 250–296.
2. *Bljumenau D. I., Gendina N. I., Dobronravov I. S., Lahuti D. G., Leonov V. P., Fedorov E. B.* (1981), Formalized Summarization by Using Verbal Clichés (markers) [Formalizovannoe referirovanie s ispol'zovaniem slovesnyh klishe (markerov)], *Scientific and Technical Information. Series 2. Information Processes and Systems [Nauchno-tehnicheskaja informatsija. Ser. 2. Informatsionnye protsessy i sistemy]*, № 2 pp. 16–20.
3. *Bol'shakova E. I., Vasil'eva N. E., Morozov S. S.* (2006), Lexicosyntactic Patterns for Automatic Processing of Scientific and technical texts [Leksiko-sintaksicheskie shablony dlja avtomaticheskogo analiza nauchno-tehnicheskikh tekstov], *Proc. of the 10th National Conference on Artificial Intelligence with International Participation CAI-2006 [Desjataja Nacional'naja konferencija po iskusstvennomu intellektu s mezhdunarodnym uchastiem KII-2006]*, Vol. 2, Fizmatlit, Moscow, pp. 506–514.
4. *Crowston K., Kwaśnik B., and Rubleske J.* (2010), Problems in the Use-Centered Development of a Taxonomy of Web Genres, in *Mehler A., Sharoff S., Santini M.* (eds), *Genres on the Web. Computational Models and Empirical Studies*, Springer, Dordrecht, pp. 69–84.
5. *Kibrik, A. A.* (2009), Mode, genre, and other parameters of discourse classification [Modus, zhanr i drugie parametry klassifikacii diskursov], *Problems of Linguistics [Voprosy jazykoznanija]*, № 2, pp. 3–21.
6. *Mehler A., Sharoff S., Santini M.* (eds) (2010), *Genres on the Web. Computational Models and Empirical Studies*, Dordrecht: Springer.

7. *Santini M., Mehler A., and Sharoff S.* (2010), Riding the Rough Waves of Genre on the Web, in Mehler A., Sharoff S., Santini M. (eds), *Genres on the Web. Computational Models and Empirical Studies*, Springer, Dordrecht, pp. 3–31.
8. *Shchipicina L. Ju.* Complex description of genre of computer-mediated communication (on the material of the news agencies web pages) [Kompleksnaja harakteristika zhanra komp'juterno-oposredovanoj kommunikacii (na primere veb-stranic novostnyh agentstv)], available at: <http://www.pags.ru/science/conferences/E-Conference/Shipitina.doc>
9. *Site Technologies Inc*, available at: <http://www.veloxsites.com/>
10. *Yandex catalog*, available at: <http://yaca.yandex.ru/>