Content analisys as a part of the study of semiotic and communicative nature of the title of a specialised text

Victoria CRAVCENCO
Valentin CIJACOVSCHI
State University of Moldova, Chisinau (Republic of Moldova)

Résumé: L'analyse de contenu a une longue histoire: les premiers cas du comptage de fréquence simple des termes étant mentionnés au 18ème siècle. Avec le développement de tous les aspects des activités humaines, l'apparition des nouvelles branches interdisciplinaires de la science, des technologies et des industries, l'analyse de contenu a progressivement évolué, elle aussi, en utilisant des techniques de plus en plus complexes pour travailler des volumes immenses de toutes sortes d'information, en incluant la littérature scientifique, technique, politique, sociale, économique et d'autres. Pour économiser le temps, les efforts humains, l'argent et rendre l'analyse de contenu plus efficace, les savants cherchent de nouvelles techniques qui pourraient être appliquées pour travailler des textes, même automatiquement.

Content analysis is a standard methodology widely used in social sciences and the humanities, e.g. in sociology, communication, linguistics, and psychology since 1950-ies. It is a technique allowing methodical, systematic, objective, and quantitative analysis of the content of some of textual or visual documents in order to classify and interpret their constituent elements. [1]

Earl Babbie defined content analysis as "the study of recorded human communications, such as books, websites, paintings and laws." Harold Lasswell formulated the main questions of content analysis in the following way: "Who says what, to whom, why, to what extent and with what effect ?." Ole Holsti defined content analysis as "any technique for making inferences by

BDD-A17858 © 2007 Editura Universităţii din Suceava
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:50:57 UTC)

objectively and systematically identifying specified characteristics of messages."

The main stages of content analysis are the following:

1. Pre-analysis and constitution of a corpus: in order to compose a corpus. It includes selection of textual or visual documents (books, administrative reports, transcription of interviews or speeches, conversations, film dialogues, images etc.) made according to a predetermined research question or in an inductive approach, by having a question concerning the object, about which the researcher already has a general idea.

2. Reading and categorization: in order to classify the elements of the corpus by themes or thematic categories (applicable, exhaustive, exclusive, and objective categories). The documents can be processed by a person or computer.

3. Classification by coding and numbering of the units: in order to distinguish the units dividing the content: registration unit, numeration unit, context unit. During the first and subsequent reading, the researcher makes a classification of the documents by creating some categories or assigning some codes to the documents, differentiating them in such a way. These categories or codes can be concerned with the content of the document (for example, with the semantic fields) or to the context of the text (for example, source, date, theme).

4. Interpretation of the results: in order to draw conclusions concerning the studied corpus. Interpretation is generally made during the reading and classification of documents.

Usually, content analysis marks out the social determinants of the texts as it assumes that the reality of the discourse exists outside of the text. Thus, traditional content analysis has a tendency to minimize the importance of linguistics. However, the difference between discourse analysis, which regards the importance of linguistics, and content analysis has a tendency to disappear.

Rapid development of different aspects of human activities, industries, science, and technologies require new approaches of processing the immense volumes of ever-increasing volumes of

BDD-A17858 © 2007 Editura Universităţii din Suceava
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:50:57 UTC)

the documented information. It has become clear that such information must be analysed, systematised, categorised, and only then can be further processed: rewritten, summarized, translated etc.

In this regard, content analysis has been developed to satisfy many needs of all categories of specialists dealing with texts. It has been defined in different ways:

1. Statistical (quantitative) semantics;

2. Techniques for objective quantitative analysis of the message content;

3. Techniques for drawing conclusions by means of objective and systematic finding of characteristics of the messages.

4. A broader definition of content analysis is "a research method for finding certain words or concepts in a text or a series of texts".

Quantitative content analysis studies the occurrence of certain characteristics (variables) of the contents in the text.

Qualitative content analysis allows drawing conclusions even on the basis of unique presence or absence of certain characteristics of the contents.

Simple frequency provides less information than relative frequency, the latter being calculated as a relation of absolute frequency to the length of the analyzed text.

In content analysis, certain conceptual formations are presented by categories. As a category may serve a series of words belonging to the same group. Thus, the category of economy would include the words: economy, unemployment, inflation etc. The frequency of occurrence of categories, and not that of separate words, allows drawing a conclusion about the message.

The quality of content analysis depends on the categories found in the text. The main areas where content analysis is applied nowadays are the following:

1. There are two or more texts to be compared by certain categories. For example, the purpose is to find out what attention is paid to certain topics in two different newspapers. If these newspapers are designed for the same target group, essential

BDD-A17858 © 2007 Editura Universităţii din Suceava
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:50:57 UTC)

difference in frequencies will allow drawing a conclusion about differences in their policies.

2. Another problem that can be solved by means of content analysis is tracking the changes in occurrence of certain categories. For example, the frequency of certain scientific themes and proper names of scientists allows drawing conclusions about the success in certain fields.

New functions studied in content analysis are presence of archetype and aggression in the text. For example, comparing the frequency of aggressive words in a text read by a deputy to the frequency of the general population of the region, it is possible to judge the degree of the deputy's aggression. Different norms of the degree of aggression will be applied to people of different professions, i.e. not a deputy but a military man.

Since the middle of 50ies of the last century, a greater attention is paid not simply to the presence or absence of some categories in the text but to the relations between these categories. Special attention is paid to co-occurrence of the words belonging to various categories. It can be found out that some categories have a tendency to be used together, while others - separately.

Collocations meaning contexts of the use of words can be also studied by content analysis, and allow attributing the text to a certain topic.

Thus, formal methods applied in context analysis make it possible to find essential information in the texts making it a powerful and valuable technique used for text analysis [2].

Collocations and categories, in their turn, can be analysed in order to find out simple frequencies of categories, relative frequencies, assessment of categories concerning the norm etc. In other words, content analysis allows determining several themes in the text and analysing them separately.

At present, content analysis is widely used for processing huge volumes of information. Automatic categorisation is used for this purpose. In this context, the biggest issue is to determine the categories within the limits (i.e. the terms) hereof the content analysis will be performed.

BDD-A17858 © 2007 Editura Universităţii din Suceava
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:50:57 UTC)

Some investigations have shown that the title / subtitle of the text is the key to the understanding of the meaning of the text. It is of great interest for both linguists and IT specialists, taking into account the possibility of automated processing of the titles / subtitles of all kinds of texts, including technical and scientific, political and social, economic and other types of literature [3].

Thus, we have valued the semantic, semiotic and communicative nature of titles / subtitles and of the terms / syntagms they are composed of in a series of articles from "Business Week" magazine. In our view, the contents of these articles reflect the reality of the economic views, tendencies, policies, and activities characteristic of the modern society and the vocabulary (thesaurus) used by its members.

This magazine is published 50 times a year, contains 20-25 sections, ca. 70-90 articles each issue. We have chosen the articles belonging to Economic Viewpoint, Economics, and Economic Trends sections from 2003-2007 issues. Regarding all the texts as a closed and reduced system known as a text corpus [4], we have carried out the linguistic, semiotic and semio-linguistic analysis of the texts, first of all, based on the titles and subtitles.

We have used the notion of a title as a title of introduction of the article believing that it contains information that can be immediately revealed / retrieved by the specialist reading the article. The lexical units of the title / subtitle can be perceived in different ways: the author may give us the perspective of the topic covered in the article; the reader may predict the contents.

As a proof of this, we have received letters from "Business Week" editors saying that:

1) The author usually indicates which section his / her article fits best.

2) The editor decides in which section to print the article.

Still, they say, there might be a situation when there are too many articles written by authors for some sections, and very few for the others. In such a case, the editor might decide to publish the article in a section different from the one it was initially

BDD-A17858 © 2007 Editura Universităţii din Suceava
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:50:57 UTC)

intended for, as one article may cover several topics, for instance, social, political, and economic life of the community.

The semantic nature of each term used in the title / subtitle is manifested due to its semantic characteristics that must be interpreted as components of the whole notion. At the same time, these semantic characteristics are the elements of the whole complex construction represented by our knowledge. These elements help us understand how the objects, phenomena and events existing in the world are presented, adjusted and operated. Namely semantic characteristics of the term make human brains distinguish one group of terms from another. They ensure our ability to recognise objects, phenomena and events, i.e. ensure our intellectual activity which allows us to identify the essence of the real state of things on the basis of some characteristics. This ability is stimulated by our intelligence, association ability, professional, and general knowledge.

The author of [3] affirms that structural and lexical, semantic and syntactic particularities of titles / subtitles make it possible to use computer for automated processing of huge volumes of information.

Each linguistic sign in the system of language operates as a sign of language, while in the text it becomes a sign of speech. In the case of language, it is responsible for classification and expression; in the case of speech – for communication. Title/subtitle, as a special type of a complex sign, possesses the same qualities as a compound word or expression, and such a compound sign as a sentence:

1) A simple sign represents one or two - three lexical units, while the meaning of the title / subtitle is represented by a chain of words.

2) The referent of a simple sign is represented by a separate object or a simple situation, while that of a title / subtitle is represented by a complex situation.

3) On the contrary to a simple sign, the communication function of a title / subtitle, its generation and perception are assured by the templates of language construction developed for

BDD-A17858 © 2007 Editura Universităţii din Suceava
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:50:57 UTC)

most of the languages, including a sing – theme ? (t), a sign – rheme ? (r) and, if necessary, a sign – determinative ? (d): ? (r) + ? (t) ? (d).

4) While the valence of a simple sign is restricted by the limits of the sentence, the valence of such a complex sign as a sentence is determined by its relation to the sentences situated to the left and to the right, the valence of a title / subtitle has a wider range being determined by its relation to the whole text.

An efficient communication model would be the following: when the information perceived from the title / subtitle coincides with the information included into the title / subtitle by the author, i.e. Title1=Title2. But in reality this happens very rarely because of the paradox of communication – while the message is being transmitted, it can express three types of meanings:

1) Independent "collective" or invariable meaning characteristic of any text as an element of the semiotic system of the natural language used for communication;

2) Author's meaning with his / her own motivation and environment, using his / her own way of linguistic expression;

3) Meaning attributed to the text by each reader separately.

The paradox of communication consists in the fact that, without the invariable meaning, there would be no author's meaning, and no reader's meaning, on the one hand; while the existence of the universal independent meaning is not possible without undertaking an investigation by the members of communication process, on the other hand.

Thus, it has been pointed out to the fact that in order to generate, transmit and perceive a title / subtitle are used the same principles as those used for verbal communication, while the schemes used to generate and decipher a title / subtitle have some particularities, on the contrary to a simple text:

1) The author of the title / subtitle assumes that the reader understands the facts and the situations illustrated by the title / subtitle in the same way as he / she does.

2) The explicit expression of the contents of the theme and rheme generated by the author.

BDD-A17858 © 2007 Editura Universităţii din Suceava
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:50:57 UTC)

3) The text of the title / subtitle has restricted possibilities to introduce the contents of the article to the reader; the title / subtitle contains some new information that completes the reader's thesaurus.

4) The title / subtitle of a specialised text contains the same lexical units that the reader uses as a specialist in the field, i.e. the author and the reader have the same thesaurus, and make the same interpretation of the message contained in the title / subtitle.

So, our assumption that the article titled 'New Drugs Cut Costs, And Medicare Can Help' (by Gary S. Becker, published on March 22, 2004 in the section Economic Viewpoint of Business Week magazine) deals with the state policy on the health-care costs of the population was confirmed by the following words and expressions found in the text of the article: medical care, health-care costs, drugs, health, Medicare, health-care plans, medical spending, aging of the population, health economist, Organization for Economic Cooperation & Development, aging population, total medical costs, the quality of life, Journal of Clinical Psychiatry, antidepressant expenditures, depressed person, hospital stays, quality of patients' lives, depression, adult mortality, raising life expectancy, effective drugs, Alzheimer's disease, cardiovascular disorders, cancers, AIDS, diabetes, nerve disorders, Parkinson's disease, medical problems, DNA, genome, development of treatments, share of drugs, full cures, onset and severity of major diseases, hospitals, nursing homes, the care of family members, costly medical care, system of drug coverage, Medicare system, patient spending on drugs, drug expenditures, donut in drug coverage, additional drug costs, drug deductible, coverage for expensive drugs, hospital and nursing-home stays, drug outlays, additional drug expenditures.

Thus, we can draw a general conclusion that various techniques can be used for carrying out content analysis, some of them being more precise, others being rather vague; the choice depending on the necessities. With the development of all aspects of human activities, appearance of new interdisciplinary branches of science, technologies and industries, content analysis has

BDD-A17858 © 2007 Editura Universităţii din Suceava
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:50:57 UTC)

gradually evolved, too, using more and more complex techniques for processing immense volumes of all kinds of information, including scientific, technical, political, social, economic and other types of documents. In order to save time, human efforts, money and to make content analysis more efficient, researchers search for new techniques that could be applied to process texts, even automatically.

Bibliography

Charaudeau, Patrick; Maingueneau, Dominique, *Dictionnaire d'analyse du discours*, Seuil, Paris, 2002, p. 670.

Cijacovschi, V.A., *Semiotico-communikativnye aspecty avtomaticeskoi pererabotki zagolovka naucno-tehniceskogo texta. Avtoreferat diss. d-ra filolog. nauk: LGU*, Leningrad, 1989, 31 str.

Robert, André D., Bouillaguet, Annick, *L'analyse de contenu*, 2e éd. mise à jour, Paris, Presses Universitaires de France, 2002, p.128.

Shalak, Vladimir Ivanovich, *Content-analiz. Prilozhenia v oblasti: politologyi, psihologyi, sotsiologyi, culturologyi, economiki, reklami / Rossiyskaya akad. nauk; In-t filosofyi*, M. Omega-L, 2004, 272 str.

BDD-A17858 © 2007 Editura Universităţii din Suceava
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:50:57 UTC)