ISCKMC 2020
International Scientific Congress «KNOWLEDGE, MAN AND CIVILIZATION»

FEATURES OF COMPRESSED SCIENTIFIC TEXT: INCREASING INFORMATION DENSITY

Rashid Sultanovich Alikaev (a)*, Irina Nikolaevna Makhova (b),
Arkadiy Aleksandrovich Borisov (c), Mirrat Rashidovna Toguzaeva (d),
Zalina Olegovna Dotkulova (e), Daniya Saidinovna Kozhokova (f)
*Corresponding author

(a) Kabardino-Balkarian State University named after Kh.M. Berbekov, 173, Chernyshevsky str., Nalchik, 360004, Russia, ralikaev@mail.ru
(b) Stavropol State Agrarian University,12, Zootechnicheskiy Ln., Stavropol, 355017, Russia, zheltova.ira71@mail.ru
(c) North-Caucasus Federal University, 1, Pushkin str., Stavropol, 355017, Russia, arka.dij@mail.ru
(d) Kabardino-Balkarian State University named after Kh.M. Berbekov, 173, Chernyshevsky str., Nalchik, 360004, Russia, mirra_alikaeva@mail.ru
(e) Kabardino-Balkarian State University named after Kh.M. Berbekov, 173, Chernyshevsky str., Nalchik, 360004, Russia, zalina-dotkulova@mail.ru
(f)Kabardino-Balkarian State University named after Kh.M. Berbekov, 173, Chernyshevsky str., Nalchik, 360004, Russia, lovedaniya888@mail.ru

Abstract

The comparison of possible ways of presenting the general abstract content in primary scientific and secondary encyclopedic and reference texts provides an opportunity to understand a pragmatically conditioned mechanism of compression of basic information. Given the recent emphasis on comprehensive consideration of various compression units in the process of preserving and increasing the information density of a text, the question of the specificity of representation of components of a single information and knowledge continuum in the space of scientific communication is particularly important. The dual essence of a scientific text is caused by the need to present the broadest argumentative base of a certain sphere of the information-knowledge continuum in the minimum volume of a scientific, encyclopedic or dictionary entry. This contradiction is resolved by the intensional implicit-associative and expletive-compressor method of information components expansion. Based on a comparative analysis of multi-level compressor units, within the framework of the application of definition, component and contextual analysis, the paper identified key methods used to compress information and simplify the structure of a scientific text. The methods of compressive presentation of cognitive and expert information were described. Possible occurrence of information insufficiency in the secondary explication of an original content of the main components of the generalized content extracted during intentional reflection is often leveled in expert opinion through approximation, thematic annotation, intensely structured referencing and prospective thesis presentation and comment-persuasive review.

1. Introduction

The issues of information compression in scientific texts of various types are most relevant in terms of representing the maximum number of components of information and knowledge continuum in a minimum text space. However, to date, there has not been a single definition of the term and a consistent classification of criteria and language means of compressing verbal material in the process of presenting a wide information field.

2. Problem Statement

The complex cognitive process of "repeating" the information-knowledge continuum involves many operations primarily aimed at taking into account the addressable focus factor – approaching a text to a recipient, structuring and facilitating the perception of cognitive components – these are mechanisms of adaptation, approximation, generalization, as well as compression. Thus, it is necessary to emphasize the inherent need to compress lengthy reasoning in argumentative scientific discourse when explicating a generalized content. As part of the knowledge components compression defined in the information, regardless of their primary or secondary nature, the compression ratio or content deployment can be represented in the form of a graded scale, characterized, inter alia, by the use of standard cliched optimizers and generators in the text.

3. Research Questions

The very concept as a linguistic term – "compression" (lat. Compressio) came from the theory of communication to describe the processes of compressing a speech signal without losing the amount of information contained in it (Kharkevich, 1957). Many scientists understand compression as a certain principle of creating the maximum of meaning in the conditions of the minimum of material verbalizers. The others advocate the understanding of this term as a certain activity to reduce surface structures restored under the conditions of the need to distribute the semantic space (Glukhov & Komarova, 2004). We are more in favor of this definition and in the study we will try to analyze the possible ways of implementing this implicit representation of knowledge components of the information space in scientific texts of various genre implementation.

This paper mainly focuses on the analysis of ways and means of presenting cognitive information on the material of scientific texts of English, Russian and French languages as part of the implementation of the dominant functional-pragmatic task of a scientific text; clarification and description of compression methods in relation to the content of a scientific-encyclopedic text as part of the application of approximation and generalization strategies.

4. Purpose of the Study

The main purpose of the study is to identify and describe ways to represent the maximum number of components of information and knowledge continuum in the minimum form of a scientific text.

5. Research Methods

The study utilized such methods as the comparative analysis (to identify general and specific in texts in different languages), the component analysis (to describe the range of special lexical units most fully), the definition analysis (to clarify the content of special and general scientific terms and concepts); contextual analysis and quantitative counting techniques were also used.

6. Findings

In many different texts, we can distinguish the principle of preserving the textual space as a single, inextricable whole, acting on the basis of the synergy of components. The text as a whole-formed space should not lose its dominant meaningful unity, its continuality, otherwise its interpretive and accomplishing potential is steadily declining. Thus, Umerova (2011) interprets compression as ensuring the principle of language economy, determined by genre and addressable focus, simplification of structures at various levels of the language system in superficial explication providing for increase or preservation of the information content of units. Taking into account the object of our analysis, we emphasize that the given definitions note such significant compression characteristics as the reduction in the volume of the primary text (reduction of the form according to certain rules) and its preservation without losing the quality of information ("subjective meaning") in the secondary text.

Within the framework of the generalized classification of compressor tools, the main principles of compression and simplification of the text structure can be distinguished: 1) reduction of units, which are divided into specialized, contextual-variational and usual-normative; 2) acronymization according to normative models; 3) elypsis of communicative syntactic structures; 4) simplification of the structure of statements; 5) implementation elements (brackets, references, intertextual inclusions); 6) reduction of aspectual-temporal diversity of initial predicative structures; 7) generalization of terminology systems; 8) graph-font variation (spacing, italics, petit, punctuation variation, etc.).

Compressiveness is one of the key characteristics of scientific and scientific-reference texts, often it subordinates two other trends in the formation of a single secondary empirical verification of a text as part of presenting the information and knowledge continuum – approximation and generalization of content.

Text compression has a wide range of applications: it is required in the process of presenting an abstract, thesis, summary, source content, while the types (empirical or reflexive) of the original source are not determinative factors. A deep and multidimensional understanding of the source content is the primary stage that records intentional reflexion over the depth content and surface forms of representation of components within the information and knowledge continuum (Bredikhin, 2015). Verbalization of knowledge components of basic and relevant fragments extracted from the whole sonma is not limited to reduction but represents a linguocreative cognitive-communicative process designed to facilitate the recognition of elements implemented in the secondary structure.

Compression of any kind is not just compression of the text of the original source, but it involves intentional "text within the text" rethinking and explication of key elements of the information and knowledge continuum in a sphere set by the source text. The verbalization of knowledge components is

carried out in the framework of strategies selected taking into account the targeted focus and tactics by specific language means. Markers of expert information in the secondary text perform the function of binding basic theses, but banal citation does not form an adequate reflexive space of awareness, which is designed to realize not only the goals of inspection, but also offers a recipient the necessary "schemes of action" – models of text interpretation adapted for implementation and consolidation of information of increased density (Alikaev & Bredikhin, 2015). It should be noted that the encyclopedic texts we study often represent the basic provisions of several, sometimes contradictory opinions on one phenomenon.

Due to the difficulty of extracting and compressing information when creating an encyclopedic text, it is conditionally possible to distinguish several mandatory stages of cognitive processes. The primary of these is active reading of primary sources in order to adequately interpret the basic components explicating cognitive information. This stage is represented by several operations on 1) the actual division of text space, the identification of argumentative markers of axiomatics and proven positions and their relation to new knowledge presented, 2) the analysis of the structure and composition of the source text in order to determine the basic verbalizers of cognitive information (terminology system, types of syntactic constructions that stimulate cognitive activity, paragraph-thematic division, etc.) at various levels of text hierarchy. The logical sequence in the representation of known and introduced elements of the terminology system (following the principles of continuity and evolutionary complication of knowledge components) itself creates a certain algorithm for distributing new content within the framework of inductive-deductive reflection. Individual intensely introduced terms of new knowledge forming the so-called "sense wells", stimulate reflexive efforts on search and understanding. The methods of deploying active-sirconstant relations in complicated periods also contribute to the creation of a certain type of perception of a scientific text, form a special thinking that allows performing operations of "internal background argumentation". Thematic-paragraph division in a rhematic sense followed by a key and commenting paragraph phrase with indirect presentation of new components of knowledge and final output structures broadcasting direct transitions is one of the most effective ways to compress information. Elimination of elements occurs mainly when the commenting part of the source text is compressed.

The newly introduced generalized content of a text within original texts is reflected in microthemes that emphasize one or another aspect of phenomenon consideration. This division and, accordingly, a detailed representation of each point of view within a compressive narrative is not possible, therefore, the basic paragraph-thematic formulas are combined in a single superphrase unity, which includes the expert opinion with allusive and direct references to primary sources. *Some people constantly use them as a serious educational tool, aiming to improve their own or their childrens word power'. Others use them only for fun – as the arbiter in a game of Scrabble, for instance. Others do not use them at all and do not replace them…* (Crystal, 1997, p. 64).

The next step in creating a secondary text is to implement the expert opinion into a text that introduces the components of cognitive information. Expert information not only dictates a certain model perception of submitted information to a reader, but also exposes the opinion of the author of the secondary text. Thus, we cannot talk about the unconditional objectivity of the secondary text in relation to all primary sources. The creation of an image of expert information in relation to certain cognitive

components occurs at the previous stage of its active desobjectivation. Having studied the entire body of authoritative texts available THE author of the encyclopedic text on the basis of his own views on the issue under consideration forms a list of opinions relevant for confirming or refuting a particular opinion. Verbalization at this stage, as a rule, is characterized by the presence of cliched phrases of scientific axiology (positive or negative), for example: *Typological classification of languagesn based on any one feature of a language structure, which is recognized as leading...* (Yartseva, 1990, para. 14). In this case, there is direct explication of the positive attitude to a certain approach in the classification of language systems.

Conversely, the devaluation of significance of a particular point of view can also occur without an argumentative basis, however, always within the framework of recognizing the effective components of the criticized theory, which corresponds to the ethics of presenting scientific information: *Sepir's work is distinguished by systemic approach, orientation to the functional aspect of typology, desire to cover phenomena of different levels of language, but the very concept of class in it turned out to be unclear, as a result of which the grouping of languages is not obvious* (Yartseva, 1990, entry 443b).

The data of installation of expert information into the information and knowledge continuum occurs both on the basis of the use of lexical means with the semantics of non-integration of national scientific nominees of positive connotation *unclear, unobvious*; and with the help of comparative components of positive and negative in the described point of view: *systematicity – uncertainty*. Besides, the verbalization of criticism always requires an expansion of the context by presenting a positive alternative declaration, for example: *this entailed the emergence of ... based on clearer and more verified classification criteria...* (Yartseva, 1990, entry 443b).

The logical connections of a compressive secondary text sometimes transform the logic of presentation of the source due to the change in structure itself. The leveling of this contradiction is also carried out on the basis of standardized bonds corresponding to the genre-stylistic characteristics of the target secondary text, determinable addressable focus (special encyclopedia, encyclopedia for a wide range of average recipients, children's encyclopedia, etc.). For example: *As part of the explanation of this fact, different theories were put forward...* (Yartseva, 1990), *to explain this you have to refer to...* (Panov, 1984).

Further development of these cliches, as a rule, is carried out only in the form of truncated enumerations of various points of view in some open list: *... from the reconstruction of static to the reconstruction of dynamic, etc.* (Yartseva, 1990). In this case, expanded paragraph phrases can be reduced to key concepts of the source text.

General scientific terms of a generalized content are the basis for creating a primary understanding of the informationally overloaded text of encyclopedias and encyclopedic dictionaries. Such lexical units are easily recognized by a reader and structure his perception of new information in a compressed form in three algorithm-driven schemes.

1. Composition-structural content of the information-knowledge continuum (*views, processes, positions, theories, opérations, modalité, coupd'oeil, method, stepladder,* etc.). The absence of such verbalizers in the source text does not affect their frequency in the secondary.

2. The following group is represented by verbalizers of argumentative division (*hypothesis, description, conditions, facts, interpretation, evidence, argument, synthèse, renseignements,* etc.). These lexical units are designed to delimit the cognitive information presented in the original source and are included in the expert description. For example: *Sociolinguistics studies the etiquette rules* (Yartseva, 1990), *Many authors strived to prove this preliminary proposal* (International Encyclopedia of Linguistics), *This point of view is interpreted by the authors...* (Yartseva, 1990).

3. The third group of verbalizers includes characterological or axiological components of the expert assessment of cognitive components. These lexical units are found only in secondary texts or in primary sources of the analytical type. Given the compression of the original text based on review and assessment of the verifiability of various theories, many difficulties arise with the transfer of this aspect of expert information. Among the most common are such lexical units as features, specifics, criticism, meaning. More abstract, and therefore less frequent in the texts of encyclopedias are such verbalizers as tendency, necessity, regularity, etc.

For example: *unlike the terms correlated with this concept* (Yartseva, 1990), *The more widespread trend...* (Longman Dictionary of Language Teaching and Applied Linguistics, 2010), *The set of said goals gives an idea of..., La critique de la notion considérée est fondée sur…* (CNRTL, 2012). The inclusion of such lexical units in cliched expressions of a secondary text performs the function of updating a particular cognitive component.

The mechanisms of distributing and determining the cognitive components of the generalized semantic structure of the secondary encyclopedic text are closely related to the inclusion of the described phenomenon of constructions of generalized-abstract semantics in the terminological system with components of superiority of the exposed expert opinion: *At the present stage of development of linguistic knowledge..., It is an admitted fact that..., le plus efficace esimateur*. These structures mark the logical development of an "erroneous" opinion in a point of view that deserves a positive (according to the author of the secondary text) assessment.

The inferential knowledge of each microtheme in a compressive text is verbalized in an immediate determinative form: *it was proved, confirmed, the totality of facts gives reason to argue, totality of the science provides reasons enough to, prouver l'utilité, A terme, un filet sera immanquablement endommagé*. At the same time, generalization lexical units *a set*, *universally recognized* etc. make it possible to reduce valence potentials of an expression and eliminate lengthy explanations thus achieving the maximum density of cognitive information.

Thus, the compilation of each of the compressive secondary texts should be based not only on the analysis of the semantic content of the primary sources, but also on a functional-style analysis of the form of presentation of the primary sources. The creation of a complex intellectual-selective and analytical-synthetic forms serves not only to transfer the basic components of the information and knowledge field of a certain topic, but also forms models of perception and further use of obtained cognitive information, creates a form of expert knowledge.

As elimination techniques that realize the completeness of a content in syntactic structures, one can call the specific structure of a statement in a nominative generalized-abstract format, for example, the inclusion of multiple genitival nomination and concretization of the basic concept in Russian, structures

with prepositions of a formal genitive in English and French. The personalized predication, which also explicates abstractness and universality of represented components, is typical for the scientific text in its compressive forms, at the same time verbal forms can be replaced on formal deputies *represents, being, is*, etc., for example: *A text is a text* (Panov, 1984). Such a closed non-aggregated representation method can have two forms: 1) uncomplicated closed statement, 2) distributed by nominative groups explaining and clarifying the second component of a circular statement.

Nominal compound predicates with brief participles like *can be used* are widely used in scientific style. For example: *Experimental techniques are widely used in linguistics*... (Crystal, 1997). In this case, the use of an indeterminately personal passive tense with an approximator makes it possible to avoid listing both the areas of application of experimental methods and the specific indication of these methods, which contributes to the generalization of the general meaning of a statement and the possibility of using verbalized components in various areas of linguistic knowledge. Besides, the universalist nature of this declarative statement increases its argumentative applicability, despite the decrease in the degree of objective verification.

For qualitative and circumstantial characterization of phenomena, adverbs ending with *-o* (in Russian) are usually used: *convincingly, interestingly, significantly, -y: widely, trenchantly, essentially, theoretically*, or *-ant/ent: convaincant, probant, légitimement*. For example: *The topic of the conversation and the characteristics of the social setting (e.g. the level of formality) are also highly relevant as are the personal qualities of the informants (e.g. their fluency and consistency)* (Crystal, 1997, p. 217).

Thus, the exposed expert information affects the construction of specific "action schemes" in the reception of a scientific text and allows compensating for the implication of cognitive components.

7. Conclusion

So, in the form of reflecting the information-knowledge continuum specific to a certain sphere of implementation, the compressive scientific texts realize the dominant goal of forming a certain type of thinking and model of cognition, while the ordering and categorization of content takes place in compressed extremely reflexive forms, which means that surface structures vary not only depending on the addressable focus, but also on the degree of understanding of a deep content.

The compressiveness of information in scientific and scientific-encyclopedic texts is one of the basic characteristics of new knowledge taking into account the addressable focus. This increases the information density of a text within the framework of thematic annotations, intensely structured referencing and prospective thesis presentation and comment-persuasive review. Each of these methods of "compressing" the information flow is realized by abstract-nominative presentation of judgments.

References

Alikaev, R. S., & Bredikhin, S. N. (2015). Action schemes as a discoursivity marker of scientific text: formal logic vs. hermeneutics. *Bull. of Volgograd State Univer. Ser. 2: Linguist., 2*(26), 121–127.
Bredikhin, S. N. (2015). General principles of text construction as an object of meaning desobjectivation. *Cognit. Stud. of lang., 20*, 634–640.

CNRTL (2012). Trésor de la langue française: dictionnaire du XIXe & XXe siècles, plus de 100000 mots: définition, étymologie, citations, synonymes, antonymes (+ audio) (ouautre version)*.* http://www.cnrtl.fr/

Crystal, D. (1997). The Cambridge Encyclopedia of Language. In *International Encyclopedia of Linguistics.* Cambridge University. Press. http://www.oxfordreference.com/view/10.1093/acref/9780195139778.001.0001/acref-9780195139778

Glukhov, G. V., & Komarova, S. S. (2004). Linguistic compression and implication. *Word – statement – discourse, 1*, 46–51.

Kharkevich, A. A. (1957). *Theoretical foundations of radio communication.* Theortechizdat.

Longman Dictionary of Language teaching and applied linguistics (2010). Pearson Education Limited.

Panov, M. V. (1984). *Encyclopedic dictionary of a young philologist*. Pedagogy.

Umerova, M. V. (2011). Language compression: types and levels of implementation. *Topical iss. of modern sci., 17*(1), 260–269.

Yartseva, V. N. (1990). *Linguistic encyclopedic dictionary*. http://tapemark.narod.ru/les/