

Approaches to Text Simplification: Can Computer Technologies Outdo a Human Mind?

Svetlana Vladimirovna Pervukhina^a
s_pervukhina@mail.ru
Don State Technical University,
Russian Federation

Gyulnara Vladimirovna Basenko
julia.bass@mail.ru
Don State Technical University,
Russian Federation

Irina Gennadjevna Ryabtseva
irina-17@yandex.ru
Don State Technical University,
Russian Federation

Elena Evgenyevna Sakharova
elenasakharova@bk.ru
Don State Technical University,
Russian Federation

ABSTRACT

Narrowly specialized information is addressed to a limited circle of professionals though it provokes interest among people without specialized education. This gives rise to a need for the popularization of scientific information. This process is carried out through simplified texts as a kind of secondary texts that are directly aimed at the addressee. Age, language proficiency and background knowledge are the main features which are usually taken into consideration by the author of the secondary text who makes changes in the text composition, as well as in its pragmatics, semantics and syntax. This article analyses traditional approaches to text simplification, computer simplification and summarization. The authors compare human-authored simplification of literary texts with the newest trends in computer simplification to promote further development of machine simplification tools. It has been found that the samples of simplified scientific texts seem to be more natural than the samples of simplified literary texts since technical background knowledge can be processed with machine tools. The authors have come to the conclusion that literary and technical texts should imply different approaches for adaptation and simplification. In addition, personal readers' experience plays a great part in finding the implications in literary texts. In this respect it might be reasonable to create separate engines for simplifying and adapting texts from diverse spheres of knowledge.

Keywords: Text Simplification; Natural Language Processing (NLP); Pragmatic Adaptation; Professional Communication; Literary Texts

INTRODUCTION

The current stage of human society development is characterized by a tremendous increase in scientific and technological achievements, which in turn has led to a rapid increase in the information flow. A considerable place in it is occupied by narrowly specialized information from various scientific fields, represented by specialized texts. A special text is understood as

^a Main & corresponding author

a text related to a certain branch of human activity (Ozhegov & Shvedova, 1999). The specialized text is based on the scientific picture of the world, which covers the entire “totality of scientific knowledge about the world developed by all private sciences at this stage of the development of society” (Kornilov, 2003). For example, the medical, legal, political and technical texts are distinguished due to their specific spheres of knowledge and peculiarities of text structure. There are a lot of scientific works devoted to the study of specialized (scientific) texts (Antonova, 2011; Drozdova, 2003; Zhidkov, 2014; Mishankina, 2015). Researchers note that specialized texts have a special structure; they use a complex terminological apparatus, factual information, research results, and various kinds of analytical data. Specialized texts are characterized by references, allusions and precedent names, which also transmit specific information with specific knowledge of this sphere. Obviously, such texts are intended for a competent, trained reader. They are addressed to a very specific and rather limited circle of people – specialists or professionals in this field.

However, with the development of computer technology and modern Internet features, specialized information has become easily accessible and is of interest not only to people with specialized education (Perelman, 2008). At the same time, a layman (non-professional) is not able to understand the specialized text completely because they do not have sufficient knowledge for it. Those features of a specialized text, such as terms, links, precedent names, help a professional but become a serious obstacle for an unprepared addressee. Researches are interested in presenting information in a specialized form, and in a simplified version. In (Nygard Larsson, 2018) the researches prove the existence of different forms of information representation depending on the situation of communication – academic or everyday language.

A specialized text is not the only type of text, and we assume that it would be worth outlining the key features of the concept of “text” and its main characteristics. In linguistics, the concept of “text” is recognized as one of the most complex and controversial. First of all, it is recognized as the main unit of communication (Bakhtin, 2000). Depending on the perspective of the study, linguists distinguish different features inherent in the concept of “text”. Yu.M. Lotman gives the text the features of an intellectual personality, considers it as “a complex device that stores diverse codes, capable of transforming the received messages and generating new ones...” (Lotman, 1992). I.R. Galperin, in his classical definition of the text, complements it with such characteristics as informativeness, cohesion, continuum, modality, completeness, subtext (Galperin, 2008). At the same time, he emphasizes as most important such key characteristics as completeness, intentionality (pragmatic attitude), lexicogrammatical and stylistic connectedness. O. P. Vorobyova, L.R., Duskaeva note the addressability and dialogical character of written texts (Vorobyova, 1993); intertextuality, which determines the dynamics and openness of the text (Chernyavskaya, 2009). Depending on the function performed by the text (scientific, technical, publicistic, artistic) these key characteristics make it possible to consider a specific text as a special type of text that performs the function of influencing, regulating and orienting the recipient in the world of events and people.

The aim of our research is to compare human-authored methods of text simplification that have been developed and applied in different discourses so far with those made with machine tools to simplify texts. The human-authored approach includes pragmatic and semiotic aspects that machine tools lack. Such comparison would provide more sophisticated ways for computer text simplification and text summarization. Pragmatic adaptation has received little attention so far, though it comprises a tremendous difference in these approaches. The findings of this study can shed light on more advanced ways of automated computer simplifications of texts.

METHODOLOGY

In the first stage of our research we analyzed simplified human-authored stories. To do that we compared original stories by S. Maugham (1992) and their simplified versions written by G.K. Magidson-Stepanova (Maugham, 2009) in a collection of stories “*The Man With the Scar And Other Stories*”. This collection of stories included 112 pages and was organized as study material for those who study English as a second language. The manual provided simplified texts and tasks for classroom work.

In the second stage of our research we simplified the original S. Maugham’s stories via automated tools on sites <http://rewordify.com/index.php>, and <https://www.simplish.org/>.

In the third stage of our research we compared text transformations made by a human author and by a machine tool. We analyzed structural, compositional, semantic, pragmatic, and semiotic transformations in the adapted texts, focusing on transformations that take into account the reader’s demands and the impression produced on the reader.

RESULTS AND DISCUSSION

There is an objective need to popularize scientific information, T.JU. Vasileva shows the necessity of scientific knowledge dissemination in the society which is profitable for business and self-education. Moreover, it could attract the young to the scientific sphere (Vasileva, 2015). Popularization of science can influence the future development of a country. It solves a number of socially significant tasks: reporting on scientific achievements and discoveries; attracting young professionals to science; forming a constant interest in science among different age groups of the population; teaching schoolchildren and students. This is done by changing trivial facts into accessible and interesting material for the population (Sukhenko, 2016).

BASICS OF HUMAN-AUTHORED TEXT SIMPLIFICATION

The process of scientific knowledge propaganda is carried out by explaining, interpreting incomprehensible information, in other words, by adapting the original special text. Text simplification in the broad sense means its reduction and simplification for the poorly prepared reader. An adapted text is a secondary text, which is built on the basis of the source text with other pragmatic goals and differs from the source text in semantic, structural, and pragmatic terms. In this case, the pragmatic parameters of the standard communicative situation change. Schematically, they can be represented as follows:

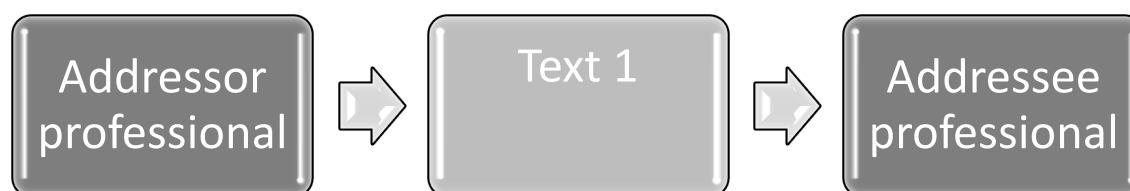


FIGURE 1. Process of communication between professionals

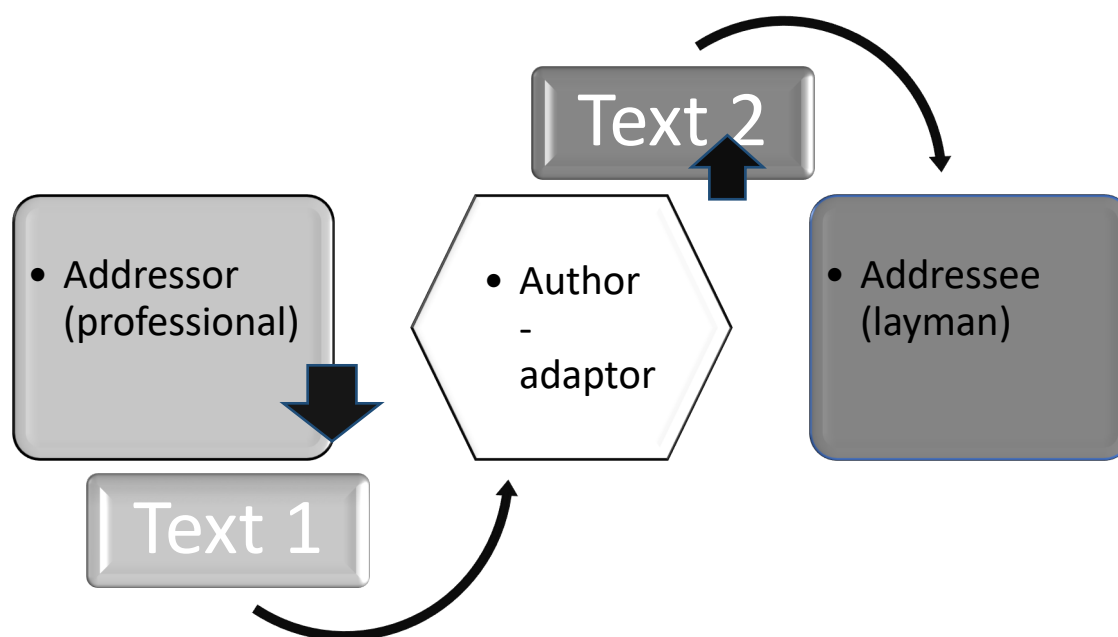


FIGURE 2. Communication process between a professional and a layman with the help of an author-adaptor

Let us make it clear that Text 1 is a source text that is an original specialized text oriented to a competent reader (a professional); Text 2 is an adapted text, its addressee does not have sufficient scientific knowledge. As you can see, during the adaptation, the emphasis is on understanding the text containing special information by another addressee, a layman. The category of addressee in communicative linguistics (in pragmalinguistics) is recognized as one of the most important and is taken into account when creating texts.

Creating an adapted text, one cannot but take into account its pragmatic features. First of all, it is particularly important to focus on the addressee, the potential reader. There are several types of addressees: 1) the addressee who is directly addressed, 2) the quasi-addressee (the imaginary person or subject to which the statement is directed, 3) the addressee-transmitter, who transmits information to the actual addressee, 4) the indirect addressee, the listener, who is present at the act of communication (Pocheptsov, 1986). In case of professional discourse the directly addressed person is also a professional, so Text 1 (in Fig.2) has specific terminology and structure meant for professional communication. The author-adaptor is the addressee-transmitter who transforms the text for a different communicative situation and the layman-addressee, who needs the message of the text or certain information for their practical application. Yet literary texts are aimed at quasi-addressees who are supposed to have mostly linguistic competence and interest on the topic discussed.

In case of an intentional impact, a certain typology of the addressee is set according to the certain criteria: 1) the addressee can be distinguished by the degree of their activity in communication, where the indirect addressee acts as the third party, 2) by the degree of the addressee's inclusion into the communication process, i.e., an active or random addressee (an onlooker, an eavesdrop), 3) by the degree of the addressee's being known to the speaker. The less known the addressee, the more likely they will follow to strictly status communication, 4) based on the speech genre. For example, in literary texts one can observe a critic addressee or an empathic addressee. The difference is shown in the addressee's willingness to be an outside observer or in their emotionally empathizing with the author, 5) on the basis of the number of addressees (in private, in a small or a large group or it may be mass communication), 6) on the

basis of the addressee's social status. These different types of addressees are shown in distancing from the addressor. They can involve the reader into the conversation by means of questions or using personal pronouns, etc.

The criterion of the text personalization is also important for motivating the reader, creating positive emotions from the text. In the last criterion, there can be three more addressees distinguished with different levels of knowledge and social status (Karasik, 2002): 1) "illiterate" (able to write and read, but inactive, uncritical, mainly for entertainment), 2) "pragmatists" (working in public organizations, active, with critical thinking, ambitious), 3) "intellectuals" (primarily interested in philosophical, aesthetic, scientific problems, non-communicative, critical, anti-democratic, ignoring the media).

The types and characteristics of the addressee, with all their diversity, are important information for the author while adapting different texts.

Adapted texts continue gaining popularity among readers in the field of specialized or promotional literature in the foreign language. This may be a group of readers – non-specialists who need special knowledge according to their interests or needs. For example, medical or legal assistance may be required at a definite moment of life. A rather large group of literature readers in the foreign language with a different level of knowledge is observed. In this case the adaptation of the source text is simply necessary. Adapted texts can also be addressed to the general public – potential buyers for promotion purposes.

Adapting the text, the author-adaptor takes into account the age, level of language proficiency, potential gaps in the background knowledge of the addressee. Thereafter it causes significant changes in compositional, communicative and linguistic aspects. In compositional aspect it refers to the level of complexity of the text. In communicative aspect it deals with the manner of referring to the addressee – distance or personal involvement. In the linguistic aspect the level of the addressee's awareness determines the choice of a certain terminology and the need for its explanation.

Depending on the level of background knowledge and the level of language proficiency during adaptation, both amplification and compression of information can occur. Amplification of information involves extending the text in order to explain or emphasize certain points of description. It can include embellishment or technical elaboration. Therefore more information should be added in order to highlight what is more important. Amplification of information makes the text longer but compression of information, on the contrary, makes the text shorter.

Amplification of information occurs if the reader does not have sufficient erudition in the field of the source text. The author-adaptor takes into account the background knowledge of the addressee and seeks to fill in the information gaps. This happens by clarifying abstract concepts or terminology.

Now let us take the example of instructions for the use of this vitamin-mineral complex for children Multi-tabs to consider the amplification of information (1) and an example of compression is the adaptation of the text on the packaging of the vitamin-mineral complex for children Multi-tabs (2).

(1) Amplification: *During a period of intensive growth, the presence of calcium in the body is vital. Calcium deficiency may be caused by stunted growth and poor tooth formation. The Vitamin-Mineral Complex Multi-Tabs Kid Calcium Plus contains 12 vitamins and 6 minerals, including calcium for healthy bone growth and vitamin D3 for optimal absorption of calcium.*

(2) Compression: *30 chewable tablets
raspberry and strawberry flavour
Vitamin and Mineral Complex
MULTI-TABS
Kid
Calcium Plus*

*12 vitamins and 6 minerals
Calcium and Vitamin D3 during Intensive Growth
In case of insufficient use of dairy products*

Example (2) is the description of the medicine composition, recommendations for use, duration of medication, contraindications, etc. As we can see, this adapted text (2) of the instruction is aimed primarily at clarifying the necessary information to the addressee, replenishing their background knowledge about the need to use the vitamin-mineral complex for kids. The explication of the information leads to further development of the adapted text.

Compression is necessary in (2). First, it occurs if the reader is challenged by the language of the presentation in the source text. Then it can also be caused by the lack of time for a detailed study of the source text. Compression is presented at the compositional, lexical, and syntactic levels.

In this case, the phenomenon of compression is simply necessary for the convenience of placing the advertising text on the packaging of the Multi-tabs vitamin-mineral complex. In addition, the main function of the advertising text is to attract the attention of the customers to the product and to induce them to buy it. The adapted text (2) looks like an advertisement which lacks predicates and common homogeneous members of the sentence. We cannot see any abbreviations of words, inscriptions under the pictures or diagrams here because they are not allowed in medical texts. As a result, serious changes take place in the composition of the text, in its pragmatics, semantics and syntax. In this case, the adapted text can significantly increase in volume.

The purpose of using adapted texts is different from the purpose of source texts. Accordingly, the pragmatics of the adapted text has a different focus depending on the type of a text. For example, the adapted legal texts now have the task of clarifying rather than having a legislative force. Adapted medical texts are guided by the recommendations and education of their readers. The adapted literary texts set the task to teach the reader a foreign language.

With the change in the pragmatics of the text, its style also changes. Syntactic constructions are simplified, the length of sentences changes, and the addressness is increased.

During the text adaptation, the author-adaptor focuses on a new readership and thereby rearranges the information in another language. In other words, information is transcoded from one language to another. The purpose of the author-adaptor is not to reduce the new adapted text, but to adapt it to a specific addressee in order to make him understand the information completely.

The study of the process of interpretation of a foreign literary text is quite a relevant object of research. In this case, the text acts as a "verbal form of reflection of culture" and allows tracing the national and cultural specifics of the perception of the surrounding world by a foreign culture, acquaints the students with the historical facts, demonstrates language phenomena in the authentic context. But as such it presents considerable difficulties for students of technical specialities whose knowledge of English is insufficient, which necessitates the adaptation of the source text through its transformation on different language levels: lexical, syntactic-grammatical, stylistic, and compositional. It is also necessary to take into account the age, background knowledge, social, national characteristics of the reader, the level of their language competence.

Thus it would be appropriate to give the definition of an adapted text which is diverse due to their diverse character and types. Since we are aimed at a literary text we together with A.V. Brygina consider that "the adapted literary text is a secondary text produced from the original literary text which has the same with its source theme and plot and which is transformed by a combination of linguistic and non-linguistic principles in order to achieve the goals of successful communication" (Brygina, 2005). We can add two more important factors which specify the secondary text: different pragmatic aims in a different communicative

situation. It is possible to change the author of the text, the circle of readers, historical era and value orientation of text, etc. But cognitive-semantic elements of the text are preserved as a condition for developing the secondary text (or Text 2 in Fig.2). It means that Text 2 is to be based on the actualized concepts of the source text (Text 1), which can be reinterpreted, supplemented by cognitive and emotional meanings. Thus the process of interpretation comes to the fore.

By interpretation we, following V.Z. Demyankov, understand cognitive process in establishing the meaning of speech and non-speech actions (Demyankov, 1999). Cognition suggests here a thorough analysis of all the factors affecting this message, including the identity of the addressee: their background knowledge of the thesaurus (i.e., vocabulary), as well as the ability to perceive and analyze the information received, the ability to interpret or reflect. G.I. Bogin believes the interpretation to be the expressed reflection as it is aimed at finding meanings in this text (Bogin, 2001).

Thus, teaching students of technical specialities a foreign language and developing their ability to interpret and reflect on the 'texts of culture' (literary text) is of paramount importance as misunderstanding of texts' culture may have a negative impact on the work and profession of people, especially engineers who need to communicate with other people which requires the reflective position in a communicative activity (Bogin, 2001: 18). Another fact should be considered that 'texts of culture' have got a cultural context with subtext information; it is very difficult to explicate this cultural context. We support the ideas of those educators who are deeply apprehensive about the level of our national literacy which has been declining greatly and which is important in holding together the texture of our society.

Now let us look at the principles of the literary text adaptation. They are split into non-linguistic (citation, exclusion, etc.) and are associated with a change in the structure of a work of art, and linguistic principles (replacement, reduction, etc.) that affect the sphere of the language proper (Brygina, 2005).

Let us give an illustration of analysis of the text of story "The Verger" by S. Maugham which is included as a reading material in the textbooks for non-linguist students. This story is known as a brilliant example of Maugham's writing which urges students to reflect on such an essential issue of human life as success and happiness. Here are the two versions of the text:

- (3) There had been a christening that afternoon at St Peter's, Neville Square, and Albert Edward Foreman still wore his verger's gown. He kept his new one, its folds as full and stiff as though it were made not of alpaca but of perennial bronze (1), for funerals and weddings (St Peter's, Neville Square, was a church much favoured by the fashionable for these ceremonies) and now he wore only his second-best. He wore it with complacency (2), for it was the dignified symbol of his office, and without it (when he took it off to go home (3) he had the disconcerting sensation of being somewhat insufficiently clad (2). He took pains with it; he pressed it and ironed it himself. During the sixteen years he had been verger of this church he had had a succession of such gowns, but he had never been able to throw them away when they were worn out and the complete series, neatly wrapped up in brown paper, lay in the bottom drawers of the wardrobe in his bedroom".
- (4) "There had been a christening that afternoon at St. Peter's Church, and Albert Edward Foreman still wore his verger's gown. He kept his new gown for funerals and weddings (St. Peter's, Neville Square, was a church often chosen by fashionable people for these ceremonies) and now he wore only his second-best. He wore it with pride, for it was the dignified symbol of his office. He took pains with it; he pressed it and ironed it himself. During the sixteen years that he had been verger of this church he had had a number of such gowns, but he had never been able to throw them away when they were worn out, and all of them, neatly wrapped up in brown paper, lay in the bottom drawer of the wardrobe in his bedroom".

Extract (4) contains a combination of the adaptation principles mentioned above. Some of them were either excluded, replaced or reduced such as the adverbial clause of comparison with the Subjunctive (1), bookish words (2), imbedded clauses (3), which prompts the readers' reflection. Due to the omitted stylistic devices the adapted extract is definitely less expressive. But still it contains enough information which does not distort the main succession of events and allows the students under professional guidance of the teacher to predict that something crucial which is going to change his life is awaiting the verger.

Thus, we may conclude that despite the simplification of the lexical-grammatical structure and the stylistic colouring of the text the plot-forming text components that were preserved keep the text cohesive and available for interpretation.

MACHINE SIMPLIFICATION OF SPECIALIZED TEXTS

Due to the need of specialized texts for different purposes, there have been raised an idea of creation of computer tools to simplify texts of professional content. The urgency of the problem is raised by the following factors. First, there are people with learning disabilities or with aphasia who cannot understand complicated sentence structures (Collados, 2013). Second, self-education in any sphere is very popular, and it creates a demand in specialized texts written in a simple language (not necessarily English). Third, necessity of getting specialized knowledge is acute in a lot of spheres, such as medicine, law, finance, etc. Fourth, learning foreign languages has become a necessity rather than just a hobby.

The scheme of communication of professional texts with the help of text simplification is changed, as Fig.3 shows below. The function of an addressor-adaptor can be performed by computer programs for text simplification:

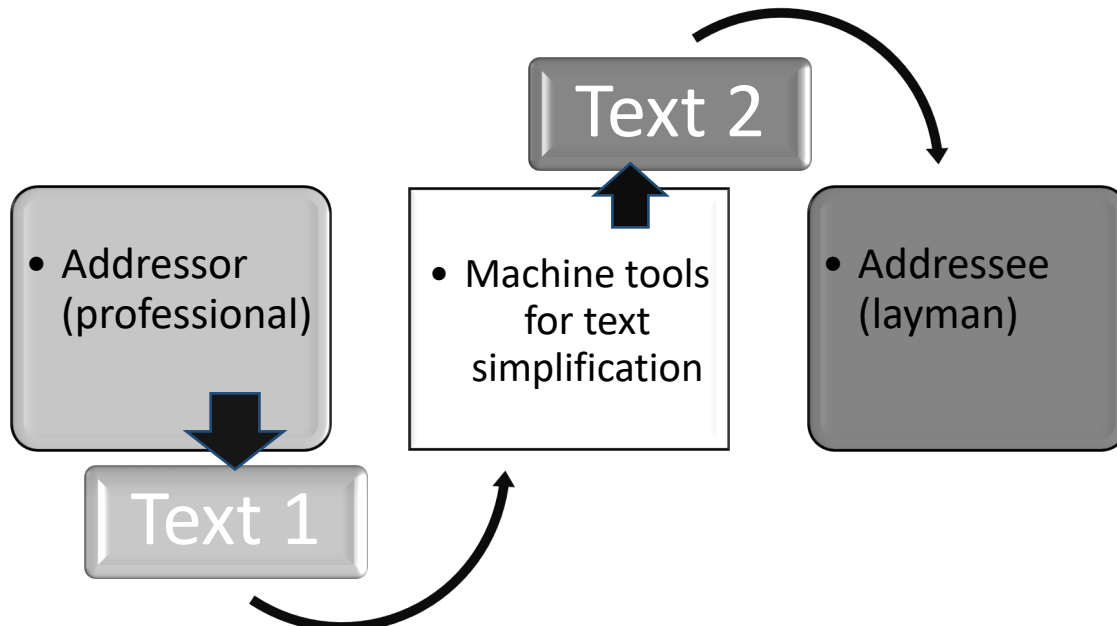


FIGURE 3. Process of communication specialized texts between a professional and a layman with the help of machine tools for text simplification

Table 1 shows different mechanisms that can be used to simplify literary and specialized texts. These mechanisms have been extensively used to adapt complicated texts to the new pragmatic situation and a new reader without certain linguistic or professional competences. We have checked how these mechanisms are exploited in contemporary on-line programs for

text simplification. We suggest here the term ‘adaptation’ with a broader meaning because it includes pragmatic and semiotic transformations as well as all other types of text transformations. The term ‘simplification’ reflects semantic, structural (Collados, 2013), and compositional transformations (Bhargava & Sharma, 2020) on the level of a sentence, complex syntactical unit, and the whole text (Fig.4). Compositional text transformation is done as a summarization process based on “selecting salient sentences from the given text and presenting it to the user” (Bhargava & Sharma, 2020, p. 139).

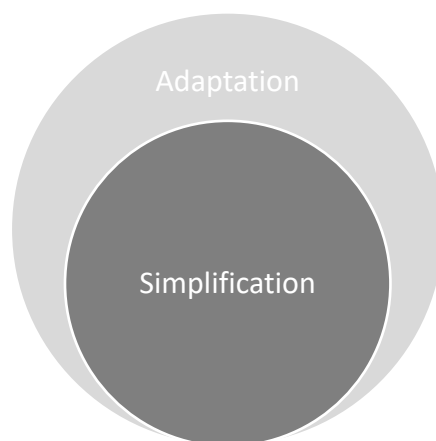


FIGURE 4. Correlation of text adaptation and simplification

Actually, both adaptation and simplification are types of text transformation aimed at making Text 2 more comprehensible for a layman. These types of text transformation have variations shown in Table 1 below:

TABLE 1. Approaches to text adaptation and simplification

| Structural (syntactic) transformations | Compositional transformations | Semantic transformations | Semiotic transformations | Pragmatic transformations |
|---|--|---|---------------------------------|----------------------------------|
| Splitting complex sentences | Text compression (abstractive text summarization) | Using synonyms | Adding pictures | Personal addressing |
| Making presupposition explicit | Text reduction (extractive text summarization) Text extension | Paraphrasing (amplification) | Adding photos | Distancing |
| | | Using hypernyms Using vocabulary from the same semantic and functional field | Adding graphs Adding tables | |

Text compression and text reduction are processes of text summarization. Text compression is seen as abstractive text summarization which is performed on semantic level generating sentences reflecting meanings of bigger parts of the summarized text. Text reduction is extractive text summarization done through choosing salient sentences (Bhargava & Sharma, 2020; Cao & Zhuge, 2020).

Evidently, structural, semantic and compositional transformations overlap. While making presuppositions explicit, the author who simplifies the text makes it longer, so it

contributes to text extension. The same situation is noted with paraphrasing and using a word combination instead of a single word.

Nowadays there are a lot of computer programs for text simplification. Being based on major linguistic principles with understanding how the language system works, such tools opt to simplify the text, check its spelling, and do its summary.

We used <http://rewordify.com/index.php>, and <https://www.simplish.org/> as the computer tools of text simplification.

In our experiment we compared human-authored simplification (and we suggest here using the term ‘adaptation’) according to printed editions, and computer simplification of the same texts. Below we give an example of computer simplification of text (3), giving machine changes in italics made by www.simplish.org:

- (5) “There had been *a giving name* to that *after 12 in the morning* at St Peters, Neville Square, and Albert Edward *manager* still wore his *vergers long overdress*. He kept his new one, its *folds* as full and stiff as though it were made not of alpaca but of taking place every year *metal of copper and tin*, for forms for putting body to *last resting place* and *getting married* (St Peters, Neville Square, was a church much supported by the of high society for *these special public event*) and now he wore only his second-best. He wore it with complacence, for it was the *respectable special sign of his office*, and without it (when he took it off to go home he had the putting at a loss consciousness of being somewhat feebly clothed. He took pains with it; he ironed it and ironed it himself. During the sixteen years he had been verger of this church he had had *a come one after another* of such *long dresses*, but he had never been able to send through air with *force* them away when they were *damaged by use* out and the complete *number, order, group, line, pleasingly simple* for its purpose covered up in brown paper, put in *the lowest part* drawers of the *clothing* in his bedroom.”

There is mostly very neat paraphrasing, explicating possible new knowledge. Unfortunately, such methods lead to damaging the style of the literary text, its rhythm, and its suspense. In some cases, the changes do not reflect the original idea. Contextual synonymy can be noted only in variants of human text transformation (as in (4)). It is clear that (4) chooses synonyms according to the context, which lacks in (5).

Though the tested programs show a little bit unmatched versions, still some evident tendencies can be marked, and they are shown in Table 2.

TABLE 2. Comparison of human-authored adaptation and computer text simplification

| | | Human-authored adaptation | | Computer text simplification | |
|---|---------|---------------------------|-----------------------------|------------------------------|-----------------------------|
| | | Literary texts | Specialized technical texts | Literary texts | Specialized technical texts |
| Structural (syntactic) transformations | | | | | |
| Splitting sentences | complex | + | + | + | + |
| Compositional transformations | | | | | |
| Text compression (abstractive summarization) | text | + | + | - | + |
| Text reduction (extractive summarization) | text | + | + | - | + |
| Text extension | | + | + | + | + |
| Semantic transformations | | | | | |
| Using synonyms | | + | + | + | + |
| Paraphrasing (amplification) | | + | + | + | + |
| Using hypernyms | | + | + | + | + |

| | | | | |
|--|---|---|---|---|
| Using vocabulary from the same semantic and functional field | + | + | + | + |
| Semiotic transformations | | | | |
| Adding pictures | + | + | - | - |
| Adding photos | + | + | - | - |
| Adding graphs | - | + | - | - |
| Pragmatic transformations | | | | |
| Personalization | - | + | - | - |
| Distancing | - | + | - | - |

Some approaches to text simplifications connected with sentence length were researched in (Fry, 1968) and are used by humans and computers. We can observe abstractive and extractive text summarization only as a special function of a machine tool. Text extension is a natural consequence of paraphrasing, but it may cause difficulties and misunderstanding in adaptation of literary texts due to mismatching the context. There have been noted some problems with using vocabulary from the same semantic and functional field connected with choosing proper synonyms matching the context. Graded readers usually have pictures and photos, for literary texts are often creolized, and so are texts of modern educational discourse (Platonova & Tarasova & Golubinskaya, 2015). The creolized (non-verbal element of the text) part of the text performs such functions as visualization, sense production, providing social background, and emotional impact. It also helps for text compression because it uses the image as the first signal system for transmitting information. It is critically important for some technical spheres (medicine, biology, geography, etc.). There are no such machine tools yet that would organize text creolization.

Another issue that is different in adaptation is personal addressing and distancing. Personalization is involving the addressee in the process of reading the text by taking into account personal characteristics. Distancing is choosing between using 2nd person or 1st person plural, or not addressing the reader at all. Usage of 2nd person when addressing your reader is a widely exploited tool in non-literary texts made by a human author-adaptor. The purpose of such addressing is to attract reader's attention, creating atmosphere of a real conversation, building up relationship (Dumitrache & Dumitraşcu, 2014). In (Chouchani & Abed, 2020) the personalization is studied as a tool that helps to target the advertising content to a definite group of users according to their interests, and they consider the technical procedures of finding out such interests. In human adaptation such a thing has already been done, and the group of users is usually defined by the situation of communication that is clear for the author-adaptor (it can be a doctor who looks through an instruction for medicine or a patient without medical background knowledge; in educational discourse it can be schoolchildren at a certain age, etc.). Actually, the technical means to consider personalization in simplifying (and adapting) specialized texts are still to come. Meanwhile, personalization can be especially fruitful for adaptation of texts for people with learning difficulties and lack of motivation for studies.

Text extension has been researched by (Man, 2014) with the aim to find out means of activating background knowledge to help understanding of the text. Our material shows that it is necessary to find out tools to determine the essential amount of information as in the following example:

| Source text | Simplified text |
|--|---|
| <p>This syndrome is characterized by a recurrent functional collapse of the airways during sleep, causing a substantial reduction and, in some cases, a total ceasing of airflow, regardless of any breathing efforts in progress. This leads to periodic perturbations in the exchange of air. The collapse or opening of the pharynx during sleep depends on two opposite forces: one is the "dilating" force which tends to keep the pharynx opened through the genioglossi muscles and pulmonary volume. The other is the "collapsing" force which tends to close the pharynx. These include negative intraluminal pressure and the positive extramural pressure which the soft tissues apply.</p> | <p>This disease is seen as a repeating functional collapse of the tube (from the mouth to the lungs), during sleep, causing a big reduction and, sometimes, a total stopping of airflow, (without any concern about/having nothing to do with) of any breathing efforts in progress. This leads to occasional (disturbing or causing small changes in things) in the exchange of air. The collapse or opening of the throat during sleep depends on two opposite forces: one is the "dilating" force which tends to keep the throat opened through the genioglossi muscles and lung-related sound level. The other is the "collapsing" force which tends to close the throat. These include negative intraluminal pressure and the positive extramural pressure which the soft tissues apply.</p> |

FIGURE 5. Example of Text Simplification of a specialized text by <http://rewordify.com>

The analysis of current studies in the field of text simplification has shown that the researchers have done a huge leap in the direction of facilitating understanding of complicated texts. Most of these researches pay attention to sentence structure and semantic level. Still it is far from natural communication and NLP (natural language processing) because the pragmatic aspect should also be counted. All the samples of adapted texts would be more natural if they were addressed to the reader's needs and situation of communication.

The samples of simplification of scientific (specialized) texts seem to be more natural since technical background knowledge is more likely to be misunderstood and processed with machine tools. Literary texts can also be misinterpreted because of vocabulary, linguistic and cultural concepts, but explication of presuppositions in order to explain the unknown might break the category of quantity in communication. Besides, literary texts are about implications and subtext which is very difficult to explicate. The thing is that personal readers' experience plays a great part in finding the implications in literary texts. In this respect it might be reasonable to create special engines for simplifying and adapting texts from different spheres.

We have come to the conclusion that literary and technical texts should apply different approaches for adaptation and simplification. Literary texts have got their intertextuality, and they preserve pragmatic value that is very important for their interpretation. The word choice here reflects the intention not only for accuracy of the sense but for the beauty of the language. Meanwhile, technical texts are aimed at precision and practical value.

Contextual synonymy depends on the specific topic of the text. Besides, some methods of text summarization also depend on the sphere of knowledge as it is shown in (Moradi, 2018) who presented a Biomedical Summarizer. The idea of this summarizer consisted in matching the terminology of the input text (the source text) to the database (the Unified Medical Language System). This work proves the idea of importance of the same background knowledge and the mutual system of shared concepts in understanding texts.

The ability to deal with the text on a specific topic is connected with addressee's knowledge in the professional sphere, and as a consequence, the amount of terminology. That is why the topic is of paramount importance. It might be difficult to find exactly the reader's needs; in this case the visuals may be a straightforward approach to clarify the topic. Visuals

reflect the first signal system and give necessary details about the part of the world image described in the text. The graphic means (bold, italics, etc.) should be employed to pinpoint important details and facilitate the understanding of the text structure. Using adapted texts in the foreign language classroom presupposes certain language skills and the technical background, therefore it is connected also with the constructing a new scientific world image in a foreign language.

CONCLUSION

The article has explored human-authored and computer-made adaptation of texts, both technical and fictional. It concentrated on structural, compositional, semantic, semiotic and pragmatic aspects of text transformation. The findings showed evident differences in methods used for human-authored texts and computer-made texts. The human-authored texts use not only linguistic information for adaptation, but also take into account semiotic and pragmatic aspects. The choice between basic strategies of amplification and compression is also made by a human adaptor for pragmatic reasons. Computer adaptation can be performed by Text Simplification and Text Summarization programs, but the choice for each of them is still for the human reader. So, at the moment, human text adaptation meets the requirements better because it takes into account the pragmatic aspect. Personalization and visualization attract reader's attention, the latter provides another semiotic code for a reader. Investigating comprehension of texts with various semiotic codes has not been done in this research, but would bring interesting results for further studies.

We have found out that literary and technical texts make use of dissimilar transformations for adaptation and simplification. We took medical texts as a specific technical material, but taking other spheres could give a different perspective on the issue. Studying transformations on broader literary material, using graded readers written by different authors-adaptors would provide more information on effective strategies in human-authored adaptation that could be used in creation of computer tools for text simplification and summarization.

Our analysis compares natural language processing and computer language processing by studying text transformations aimed at making texts easier to understand. Such analysis contributes to the creation of more sophisticated Text Simplification and Text Summarization programs. Despite a very big step in technological text processing, pragmatic adaptation is a hard nut to crack for Information Technologies.

Adapted texts contribute greatly to self-education and educating the public in general. Creation of an adapted text should be based on general rhetorical principals and on the reader's peculiarities, which should also be a question for further research.

REFERENCES

- Antonova, A. (2011) Empathy as Speech Manipulation Target in Pre-election Discourse of Great Britain. *GEMA Online® Journal of Language Studies*. 11(3), 97-107. <http://journalarticle.ukm.my/2764/>
- Bakhtin, M.M. (2000) *The problem of speech genres. Author and hero. To the philosophical foundations of the humanities*. St. Petersburg: Azbuka.
- Bhargava, R., & Sharma, Y. (2020). Deep Extractive Text Summarization. *Procedia Computer Science*, 167, 138–146. URL: <https://www.sciencedirect.com/science/article/pii/S1877050920306566>
- Bogin, G.I. (2001). *Obtaining Ability to Understand: Introduction to Hermeneutics*. Tver': Tver' State University.

- Brygina, A.V. (2005) About Some Methods of Text Adaptation on Content and Meaningful Levels. *RUDN Journal of Language Studies. Linguistics.* 7, 84-88. <https://cyberleninka.ru/article/n/o-nekotoryh-printsipah-adaptirovaniya-teksta-na-soderzhatelno-smyslovom-urovne>
- Cao, M., & Zhuge, H. (2020). Grouping Sentences As Better Language Unit for Extractive Text Summarization. *Future Generation Computer Systems.* 109, 331–359. <https://www.sciencedirect.com/science/article/abs/pii/S0167739X19318989>
- Chernyavskaya V. E. (2009) *Linguistics of the text. Polycode content. Intertextuality. Interdiscursivity.* Moscow.
- Chouchani, N., & Abed, M. (2020). Automatic Generation of Personalized Applications Based on Social Media. *Procedia Computer Science.* 170, 825–830. <https://www.sciencedirect.com/science/article/pii/S1877050920306050>
- Collados, J.C. (2013). Splitting Complex Sentences for Natural Language Processing Applications: Building a Simplified Spanish Corpus. 5th International Conference on Corpus Linguistics. *Procedia – Social and Behavioral Sciences.* 95(2013), 464–472. <https://www.sciencedirect.com/science/article/pii/S1877042813041906>
- Demyankov, V.Z. (1999) Interpretation As a Tool and As an Object of Linguistics. *Issues of Philology.* 2, 5 – 13. <http://www.infolex.ru/Interpret.html>
- Drozдова, T.V. (2003). *The Problem of Understanding a Scientific Text.* Astrakhan: AGTU. <http://cheloveknauka.com/nauchnyy-tekst-i-problemy-ego-ponimaniya>
- Dumitrache, I.-C., & Dumitraşcu, V. (2014). The Principle of Personalization – The Basis for an Efficient Educational Process. *Procedia – Social and Behavioral Sciences.* 128, 463–468. <https://www.sciencedirect.com/science/article/pii/S1877042814022800>
- Duskaeva, L. R. (2004) *Dialogic nature of newspaper speech genres: Autoref. diss. ... dr. philol. nauk.* St. Petersburg.
- Fry, E. A (1968). Readability Formula That Saves Time. *Journal of Reading.* 11(7), 265–271.
- Galperin, I. R. (2008). *Text as an object of linguistic research.* Moscow.
- Karasik, V.I. (2002). *Language of Social Status.* Moscow: ITDGGK “Gnozis”.
- Kornilov, O.A. (2003). *Linguistic World-Image as Derived National Mentalities.* Moscow: CheRo.
- Lotman, M. Yu. (1992). *Selected articles in 3 volumes. - Vol. 1 Articles on semiotics and typology of cultures.* – Tallinn.
- Man, Y. (2014). Feature Extension for Short Text Categorization Using Frequent Term Sets. *Procedia Computer Science.* 31, 663–670. <https://www.sciencedirect.com/science/article/pii/S1877050914004918>
- Maugham, S. (2009). *A Man with The Scar and Other Stories.* Adapted by G.K. Magidson-Stepanova. Tasks made by L.T. Dobrovol'skaya. Moscow: Airis-press.
- Maugham, S. (1992). *Collected Short Stories.* Penguin Twentieth-Century Classics.
- Mishankina, N.A. (2015). The Pragmatics of Scientific Discourse. *Vestnik NGPU.* 2(24), 126–133. <https://cyberleninka.ru/article/n/pragmatika-nauchnogo-diskursa/viewer>
- Moradi, M. (2018). CIBS: A Biomedical Text Summarizer Using Topic-Based Sentence Clustering. *Journal of Biomedical Informatics.* 88, (December 2018), 53–61. <https://www.sciencedirect.com/science/article/pii/S1532046418302156>
- Nygaard Larsson, P. (2018). “We’re Talking about Mobility”: Discourse Strategies for Promoting Disciplinary Knowledge and Language in Educational Contexts. *Linguistics and Education.* 48, 61–75. <https://www.sciencedirect.com/science/article/pii/S0898589818301268>
- Ozhegov, S.I., & Shvedova, N.Y. (1999). *Explanatory Dictionary of the Russian language.* Moscow: Azbukovnik

- Perelman, Y.I. (2008). To the Methodology of Scientific Popularization. *Vestnik MGU. Part 20: Teacher Education*. 3, 122–126. <https://cyberleninka.ru/article/n/k-metodike-nauchnoy-populyarizatsii>
- Platonova, Iu., & Tarasova, E., & Golubinskaya, A. (2015). Creolized Text as a Form of Modern Educational Discourse. Worldwide Trends in the Development of Education and Academic Research, 15-18 June 2015. *Procedia – Social and Behavioral Sciences*. 214, 788–796. <https://www.sciencedirect.com/science/article/pii/S1877042815060747>
- Pocheptsov, G.G. (1986). About Communicative Typology of the Addressee. *Speech Acts in Linguistics and Methodology: Coll. of scient. papers*. Pyatigorsk: Pub.house TGPIIY.
- Sukhenko, N.V. (2016). The Specifics of the Popularization of Science in Russia. *Vestnik NGTU after R.E. Alekseev. Part: Management in Social Systems. Communicative Technologists*. 4, 18–22. <https://cyberleninka.ru/article/n/spetsifika-populyarizatsii-nauki-v-rossii>
- Vasileva, T.YU. (2015). *Tasks for the Popularization of Science, Innovative and Technological Development, Programs for Promoting Expert Knowledge in the Media. Russia: Trends and Development Prospects: Collection of Articles*. Moscow: INION RAN. <https://cyberleninka.ru/article/n/zadachi-populyarizatsii-nauki-innovatsionnogo-i-tehnologicheskogo-razvitiya-programmy-prodvizheniya-ekspertnyh-znaniy-v-sredstva>
- Vorobyova, O. P. (1993) *Linguistic aspects of the addressability of a literary text: monolingual and interlanguage communication*. Moscow.
- Zhidkov, A.V. (2014). Scientific and Technical Language and Scientific and Technical Translation. *Science Time*. 5, 67–71. <https://cyberleninka.ru/article/n/nauchno-tehnicheskii-yazyk-i-nauchno-tehnicheskii-perevod>

ABOUT THE AUTHORS

Pervukhina Svetlana Vladimirovna, Doctor of Philology, Professor of Department “World Languages and Cultures”, Don State Technical University. Her interests are linguistics, text interpretation, professional discourses, applied linguistics. Has published a number of scientific articles on these topics and a manual “Structural, semantic, discourse and pragmatic characteristics of adapted texts”.

Basenko Gyulnara Vladimirovna, PhD in Linguistics, Associate Professor of Department “World Languages and Cultures”, Don State Technical University. Her interests are Linguistics, Pragmalinguistics, Psycholinguistics and Political Discourse. Has published a number of scientific articles on these topics and a monograph on the ideological speech clichés in the Constitutions of Russia and Germany.

Ryabtseva Irina Gennadjevna, PhD in Linguistics, Associate Professor of Department “World Languages and Cultures”, Don State Technical University. Her interests are communicative-pragmatic characteristics of the texts of political discourse. Has published scientific articles on the topic of research and quite a few textbooks on teaching English Grammar, Writing and Communication.

Sakharova Elena Evgenyevna, PhD in Linguistics, Associate Professor of Department “World Languages and Cultures”, Don State Technical University. Her interests are Linguistics, Pragmalinguistics and Theory of Discourse. Has published a number of articles on pragmatics of discourse and on pragmalinguistic study of speech behaviour of the sender of a message.