TEI Advances for Concordance Compiling:
a case study of Erich Fromm's novel "The Art of Loving"

Nataliia Hrytsiv [1], Ivan Bekhta [1,2], Nataliia Sofiianyk [1], Tetiana Shestakevych [1] and Vasyl Byalyk [3]

[1] *Lviv Polytechnic National University, Stepana Bandery Street, 12, Lviv, 79000, Ukraine*
[2] *Ivan Franko National University of Lviv, Universytetska Street, 1, Lviv, 79000, Ukraine*
[3] *Yuriy Fedkovych Chernivtsi National University, Kotsiubynskyi Street, 2, Chernivtsi, 58012, Ukraine*

Abstract

This paper focuses on the analysis of linguistic markings and coding of prose text by TEI methods. The original and translation of the work were analyzed. In addition, the marking of these texts was carried out and a number of concepts that prevailed in the texts were identified. Therefore, a certain set of tags was created that reflected these dominant concepts. A sample of 120 sentences in the English-language original and 120 sentences in the Ukrainian translation, which contain dominant concepts, was selected for the study. An applied component of the work is the development of a program written in Python - to analyze the representation of the markup and features of the created tags in the original and translation. As a perspective, an intentionally built program called 'Concordance 'Art of loving' can be a platform for analyzing TEI markup of this kind for other authors.

Keywords
Digital humanities, translation studies, TEI, NLP, parallel translation corpus, concordance, Erich Fromm, text mark-up, applied linguistics, translation

1. Introduction

The success of modern intelligent software systems designed for processing texts in natural language requires a large experimental linguistic base [4; 6; 7; 9; 10; 11; 16], which is what this study focuses on. The demand for corpus data corresponds to the emergence of appropriate technical capabilities [1; 3; 6; 14; 15; 17; 21; 22; 25; 31].

Corpora are usually designed for multiple use by many users, so their markup and linguistic support must be unified in some way [20; 23; 24; 28]. Corpora standards usually address the compatibility of markup types.

Difficulties in developing natural language are also predetermined by certain external factors, including the presence of hundreds of natural languages, with typical set of syntactic rules. Depending on the context, words possess multiple meanings within a single language [18; 19; 29]. Complications become obvious even at the level of symbols, digits and graphemes. As machine-readable language, Text Coding Initiative rationalizes texts processing in terms of stressing valid information as found within textness and textuality features [2; 13; 26]. To add, since 1980 Text Coding Initiative [12] has been a global group of written language researchers working in the field of digital humanities.

Resulting from the joint efforts of researchers, the TEI documents such a standard in the TEI Guiding Principles. These recommendations are flexible and fully adaptable to diverse projects; using TEI would advocate the compatibility of the chosen project within other non-digitalized humanitarian projects.

The reason is that for more than two decades TEI has been widely used in projects in a variety of disciplines, fields and subjects internationally. The constant interaction of the wide user community through the TEI organization confirms the importance of the text coding standard and signals about its continuous development and validity.

Within the Translation Studies research, saving and preserving texts in electronic format provides undoubted vistas: high speed information retrieval, easily-conducted editing procedures, multimedia adaptability and localization, hyperlinks options. Scanned texts can be stored in various graphic forms. However, for more efficient use, scanned texts as converted to digital text formats are of paramount advantage. Digital text format reduces the file size and allows other applications to reformat, search, or process text within the dimension of modern digitalized society and academia.

2. TEI parameterization of Erich Fromm's novel
2.1. Justification

The relevance of this study is proven by its theoretical and practical significance for research both in linguistics and in related disciplines: textual linguistics, translation studies, and literary studies. Since the main focus is on fiction text coding with the consequent practical approbation of the results obtained, this study proposes a method of automated search for suggested tags based on sentence selection for both original English material and its Ukrainian translation. Prioritized in the current study is the method of TEI application in relation to a bilingual fiction text.

We opt for (a) describing and characterizing modern TEI methods, (b) applying them to the practical markup of English and Ukrainian texts and (c)preparing materials for concordance on the basis of the results obtained within the scope of tasks conducted.

With this in mind, we concentrate on:

1) cursory analyzes of the main theoretical provisions of corpus linguistics and linguistic markup;

2) preliminary study of the original and translation of the analysed novel;

3) markup process of the original and translation of the novel in accordance with the recommendations of the TEI;

4) creation a sample of sentences, both in the original and in the translation, containing at least one tag from the set created by us;

5) scrutiny of the features, liminations and the definition of a set of tags;

6) choice of information technology tools for creating our software;

7) organizing tags for English and Ukrainian data;

8) analyzes of the results using the created software.

The study was based on the original and the translation of Erich Fromm's novel "The Art of Loving" [27; 32].

The scientific novelty of this work is that the study of the proposed set of tags and sentences in the original and translation has never been automated. The software created helps to distinguish convergent and divergent features between marking up textness peculiarities of the original and the translation, correspondingly.

2.2. Mandatory elements of TEI

The Text Encoding Initiative is the standard for presenting textual material in digital form using text encoding tools. As we have already mentioned, in this paper work we rely on the instructions of TEI, as well as the XML markup language [8]. Due to its advantages and wide popularity, XML has become the metalanguage of choice for expressing the rules of descriptive text encoding in TEI. The

advantage of TEI is obvious since TEI is expressed in terms of extensible markup language (XML) and it provides all the procedures and mechanisms to adapt to specific project needs.

The TEI guidelines define an open standard that can normally be applied to any text and may satisfy any purpose. So, as we have already mentioned, in our work we rely on TEI instructions, as well as XML markup language.

Any XML encoding language consists of five components:
• XML language declaration;
• Elements;
• Attributes;
• Links to objects;
• (P) CDATA.

2.2.1. XML language declaration

XML documents have to enclose with an XML language declaration that indicates the version of XML used. In the markup of both the original novel and the translation, we also cited XML language declarations. For example,

<?xml version="1.0" encoding="UTF-8"?>

The question mark "?" in the declared language of XML signals that this is the so-called instruction for processing this document. The following entry shows that there is an XML that complies with the recommendations of version 1.0; UTF-8 or Unicode character encoding is being used. An XML declaration is optional; however, in case of usage it can only appear at the beginning of an XML file [26].

Another important counterpart of the TEI documents is the title of the TEI. The TEI guidelines state that such a virtual "identity card" is an important part of every TEI document. It must be encoded in the <teiHeader> element before the actual text content in the <text> part. The minimum TEI header for our research looks as follows [26].

A sample of the minimum TEI header of the original text:

```
<teiHeader xmlns="http://www.tei-c.org/ns/1.0">
      <fileDesc>
            <titleStmt>
                  <title> THE ART OF LOVING </title>
                  <author xml:id="EF"> Erich Fromm </author>
                  <editor role="editor"> Harper & Row </editor>
            </titleStmt>
            <publicationStmt>
                  <p>Printed in the United States of America </p>
            </publicationStmt>
            <sourceDesc>
                  <p> Erich Fromm (1956). THE ART OF LOVING </p>
            </sourceDesc>
      </fileDesc>
</teiHeader>
```

Figure 1: Minimal TEI header of the original text

```
A sample of the minimum TEI header of the translation:
<teiHeader xmlns="http://www.tei-c.org/ns/1.0">
        <fileDesc>
                <titleStmt>
                        <title> Мистецтво любові </title>
                        <author xml:id="EF"> Еріх Фромм </author>
                        <editor role = "translator"> Віра Кучменко </editor>
                        <editor  role="editor"> Книжковий  Клуб  «Клуб  Сімейного  Дозвілля
                </editor>
                </titleStmt>
                <publicationStmt>
                        <p> Published by arrangement with HarperCollins Publishers </p>
                </publicationStmt>
                <sourceDesc>
                        <p> Еріх  Фромм  (1956).  Мистецтво  любові.  Переклад  з  англійської
                Віри Кучменко </p>
                </sourceDesc>
        </fileDesc>
</teiHeader>
```

Figure 2: Minimal TEI header of the translation

For a description of the file to be complete, it must contain three components in the subsequent order:

```
<titleStmt>: statement of the title of the electronic text
<publicationStmt>: information about the publication of the text
<sourceDesc>: bibliographic description of the source for electronic text
```

Figure 3: Three compulsory components

2.2.2. Elements and attributes

The two-paragraph document below is an example of a part of an XML document that represents both information and meta-information. Information (plain text) is contained in XML elements separated by opening tags (such as <div>) and closing tags (such as </div>). Additional information about these XML elements can be provided in attributes that consist of a name (for example, @ n) and a value meaning (for example, "I"). Attributes can only be found in the element's initial (opening) tag. XML comments are separated by opening tags (<! -) and closing tags (->). Everything in the comments is ignored by the XML processing software: it is said to be "commented on".

```
<!-- paragraphs go here -->

<div n = 'I'> <p> LOVE, THE ANSWER TO THE PROBLEM OF HUMAN
EXISTENCE </p> </div>

<p> <s> ANY THEORY of love must begin with a theory of man, of human
existence. </s> <s> While we find love, or rather, the equivalent of love, in animals,
their attachments are mainly a part of their instinctual equipment; only remnants of
this instinctual equipment can be seen operating in man. </s> </p>
```

Figure 4: Information and metainformation tagging

The initial and most commonly used element of structural markup is a paragraph. Paragraphs may contain any other elements that we will consider below, as well as those elements that are specific to certain types of texts. It is customary to designate a paragraph by the element *p*. Examples of the markup of the original and translation of Erich Fromm's novel The Art of Loving:

<p> <s> Not that people think that love is not important. <s> </s> They are starved for it; they watch endless numbers of films about happy and unhappy love stories, they listen to hundreds of trashy songs about love — yet hardly anyone thinks that there is anything that needs to be learned about love. </s> </p>

<p> <s> Це трапляється не тому, що люди підтримують думку, ніби любов не є важливою. <s> </s> Насправді вони спраглі до неї переглядають численні фільми про щасливі й нещасливі історії кохання, слухають сотні примітивних любовних пісень, однак водночас мало хто вважає, що, коли йдеться про любов, йому треба чогось учитися. </s> </p>

Figure 5: Structural marking illustration

This excerpt from the original and the translation of the novel demonstrates how other elements of the phrase level (in this case *s* element representing the sentence) can be nested in paragraphs, but not vice versa. Momentous aspect of coding of primary data in the corpus is also the interpretation of editorial markup, which provides additional information about the text, the specifics of lexical items, possibly about its author, etc. Editorial markup usually equals to highlighting certain elements of the text. Visually, the selection (or highlighting) can be conveyed in various ways: say it, by means of bold type and / or italics in lowercase, bold type and / or italics in uppercase, petit, letter spacing, etc. Applying the TEI principle for an uninterpreted typographic selection we use the <hi> element, which marks a word or phrase that is graphically different from the main text and the reason for this selection is unknown or unidentified. The <hi> element uses the *rend* attribute to specify the type of text fragment selection.

The following text with the highlighted words 'object,' faculty, 'love should be coded as follows:

A second premise behind the attitude that there is nothing to be learned about love is the assumption that the problem of love is the problem of an *object*, not the problem of a *faculty*. People think that to *love* is simple, but that to find the right object to love — or to be loved by — is difficult. This attitude has several reasons rooted in the development of modern society.

<p> <s>A second premise behind the attitude that there is nothing to be learned about love is the assumption that the problem of love is the problem of an <hi rend=italic rend=bold> object </hi>, not the problem of a <hi rend=italic rend=bold> faculty </hi>. </s> <s> People think that to <hi rend=italic rend=bold> love </hi> is simple, but that to find the right object to love — or to be loved by —is difficult. </s> <s>This attitude has several reasons rooted in the development of modern society. </s> </p>

Figure 6: Highlighting requirements

However, the Ukrainian translation has not preserved the separation of these words; therefore this passage looks like this:

<p> <s> Другою передумовою, яка криється за ідеєю, що в любові немає чого вчитися, є припущення, нібито проблема любові це проблема об'єкта, а не здібності. </s> <s> Люди гадають, що любити це просто, а от знайти правильний об'єкт для своєї любові чи досягти того, щоб любили тебе, складно. </s> <s> Таке бачення можна пояснити кількома причинами, що вкорінилися в розвитку сучасного суспільства. </s> </p>

Figure 7: Highlighting translated variant

Let us take into consideration an example where the word is enclosed in quotation marks that are not highlighted in italics.

The third error leading to the assumption that there is nothing to be learned about love lies in the confusion between the initial experience of "*falling*" in love, and the permanent state of *being* in love, or as we might better say, of "standing" in love.

<p> <s> The third error leading to the assumption that there is nothing to be learned about love lies in the confusion between the initial experience of "<hi rend=italic> falling </hi>" in love, and the permanent state of <hi rend=italic> being </hi> in love, or as we might better say, of "standing" in love. </s> </p>

Figure 8: Tagging quotations

In translation, this example is transmitted with italics in these words. Therefore, the markup looks like this:

Третє хибне судження, яке формує припущення про те, що, коли йдеться про любов, тут немає чого вчитися, ґрунтується на змішуванні початкового досвіду «*поринання*» в кохання і тривалого *перебування* у стані закоханості.

<p> <s> Третє хибне судження, яке формує припущення про те, що, коли йдеться про любов, тут немає чого вчитися, ґрунтується на змішуванні початкового досвіду « <hi rend=italic> поринання </hi> » в кохання і тривалого <hi rend=italic> перебування </hi> у стані закоханості. </s> </p>

Figure 9: Marking italics

Another example illustrates the situation when tags can be nested; also when not only one word can be highlighted, but an entire expression or sentence (the needed textness feature prioritized).

<p> <s> The first step to take is to become aware that <hi rend=italic> love is an art </hi> , just as living is an art; if we want to learn how to love we must proceed in the same way we have to proceed if we want to learn any other art, say music, painting, carpentry, or the art of medicine or engineering. </s> </p>

Figure 10: Visualizing nesting

Translated segment has it this way:

<p> <s> Перший крок, який слід зробити, це усвідомити, що <hi rend=italic> любов становить собою мистецтво </hi> , подібно до того, як мистецтвом є саме життя. </s> <s> І якщо ми хочемо навчитися любити, то маємо йти тим самим шляхом, що ним зазвичай ідуть, прагнучи оволодіти будь-яким іншим мистецтвом, скажімо, музикою, живописом, теслярським ремеслом, медициною чи інженерною справою. </s> </p>

Figure 11: Visual nesting in translation

Example of marking the original in italics:

But while recognizing their separateness they remain strangers, because they have not yet learned to love each other (as is also made very clear by the fact that Adam defends himself by blaming Eve. rather than by trying to defend her). *The awareness of human separation, without reunion by love — is the source of shame. It is at the same time the source of guilt and anxiety.*

The deepest need of man, then, is the need to overcome his separateness, to leave the prison of his aloneness.

<s> But while recognizing their separateness they remain strangers, because they have not yet learned to love each other (as is also made very clear by the fact that Adam defends himself by blaming Eve. rather than by trying to defend her). </s> <hi rend=italic> <s> The awareness of human separation, without reunion by love—is the source of shame. </s> <s> It is at the same time the source of guilt and anxiety. </hi> </s> </p>

<p> <s> The deepest need of man, then, is the need to overcome his separateness, to leave the prison of his aloneness. </s> </p>

Figure 12: Marking-up the 'italics' format

However, most paragraphs do not have a clear internal structure, but contain prose being coded as a combination of symbols, references to the essence, phrases, marking. They may also contain built-in elements such as lists, figures or tables.   Though, since our markup is made within the framework of a prosaic creative work, there are only a few inserted poetic passages. For example,

<s> It is the polarity of the earth and rain, of the river and the ocean, of night and day, of darkness and light, of matter and spirit. </s> <s> This idea is beautifully expressed by the great Muslim poet and mystic, Rūmi:

<l> Never, in sooth, does the lover seek without being </l>

   <l> sought by his beloved. </l> </s>

<s> <l> When the lightning of love has shot into <hi rend=italic> this </hi> heart, </l>

   <l> know that there is love in <hi rend=italic> that </hi> heart. </l> </s>

<s> <l> When love of God waxes in thy heart, beyond any </l>

   <l> doubt God hath love for thee. </l> </s>

<s> <l> No sound of clapping comes from one hand without the </l>

   <l> other hand. </l> </s>

Figure 13: Poetic lines formatting

In the Ukrainian translation this passage is presented as follows:

<s> Це протилежність землі й дощу, річки й океану, ночі і дня, темряви і світла, матерії та духу. </s> <s> І таку думку експресивно висловив видатний мусульманський поет і містик Румі:

   <l> Воістину, хто любить — то завжди його </l>

   <l> Шукає і кохана, як і він її шукає. </l> </s>

   <s> <l> Як блискавка любові втрапила <hi rend=italic> тобі </hi> у серце — знай: </l>

   <l> У серці <hi rend=italic> іншому </hi> також живе любов. </l> </s>

   <s> <l> Здіймається любов до Бога в твоїм серці? </l> </s>

   <s> <l> То сумніви відкинь — бо ж любить і тебе Він. </l> </s>

   <s> <l> Не буде оплесків, якщо твоя рука </l>

   <l> Не зможе іншої руки торкнутися. </l> </s>

Figure 14: Poetry in translated-product

There are also foreign words and phrases in the analyzed novel that we encode with the <foreign> element. We use the @xml: lang attribute to define foreign languages. Consider the following examples.

Sample 1

<s> Respect exists only on the basis of freedom: "<foreign xml:lang='fr'> l'amour est l'enfant de la liberte </foreign>" as an old French song says; love is the child of freedom, never that of domination. </s>

<s> Повага може існувати, тільки ґрунтуючись на свободі, як співається у старій французькій пісні "<foreign xml:lang='fr'> l'amour est l'enfant de la liberte </foreign>". </s> <s> Любов дитя свободи, а не домінування. </s>

Sample 2

<s> Roman felt pride because he could say "<hi rend=italic> <foreign xml:lang='la'> civis romanus sum </foreign> </hi>"; Rome and the Empire were his family, his home, his world. </s>

<s> Навіть найбідніший римлянин пишався тим, що міг сказати: «<hi rend=italic> <foreign xml:lang='la'> Civis romanus sum </foreign> </hi>»; Рим та Римська імперія були його родиною, домівкою, світом. </s>

Figure 15: Foreign elements

2.2.3. Link (reference) to objects

Another issue with corpus markup is character (symbol) encoding. In previous versions of SGML TEI, special English characters (such as the £ pound), European-accented characters (such as é), and other non-ASCII (US Standard Code for Information Sharing) characters were replaced with links to objects which are separated by "&" and ";". The object name is a line between these two characters.

For example, an entity reference for a "less than" sign < is "<" and an entity reference for an ampersand is "&", "£" which means £, "& eacute;" means é. Object references can be useful in the following cases: symbolic data that cannot be easily typed on the keyboard or not used in XML, shielding reserved characters in XML, providing "template text", i.e. text that is used or can be reused in new contexts or programs. However, since our paper is based on the marking up of the fiction novel, neither in the original nor in the translation are there any special symbols that we have marked in this way.

Although declaring object references can solve the problem of representing special characters (symbols) for languages that have only a small number of such characters, this is not a rational solution for languages that contain many special characters. Therefore, Unicode, which has adopted XML as a mandatory character set for all XML documents, is an attempt to solve this problem.

Unicode is truly multilingual in terms of displaying characters from a very large number of writing systems, and therefore promises to provide a standard for multilingual encoding of corpus characters.

XML Editor allows us to directly enter characters in most writing systems around the world and saves those characters in a way that can be transferred directly between different computer systems. XML support has made it possible to spread these markup schemes more widely in a multilingual context [5].

2.2.4. (P)CDATA

All text in an XML document is usually parsed. When an XML element is parsed, the text between the XML tags is also parsed. The text data that the XML parser will parse is called parsed character data or PCDATA.

An XML document often contains data that does not need to be parsed by an XML parser. For example, characters such as "<" and ">" are not used in XML elements because the parser interprets them as the beginning or end of new elements, resulting in an error message. We can avoid this by looking at them through links to "& lt;" and "& amp;" or CDATA in the document.

The CDATA section starts with <! [CDATA [and ends with]]>. Everything inside the CDATA section is considered by the parser as plain text.

2.3. Marking of 'love' textness by means of TEI

Having analyzed the original and translation of the novel, we also identified a number of special concepts that dominate Fromm's work, namely - motherly love, fatherly love, brotherly love, erotic love, love for God and other types of love. Based on these concepts, a number of tags have been created that correspond to them - <mlove>, <flove>, <blove>, <elove>, <slove>, <Glove>, <olove>.

Let us consider examples of sentences with the concept of maternal love.

Sample 1

<s> But God goes beyond this minimum requirement. </s> <s> On each day after nature and man — is created, God says: "It is good." </s> <s> <mlove> Motherly love </mlove>, in this second step, makes the child feel: it is good to have been born; it instills in the child the <hi rend=italic> love for life </hi>, and not only the wish to remain alive. </s>

<s> Кожен день день сотворення світу, людини завершується словами: «І побачив Бог, що воно добре». </s> <s> <mlove> Материнська любов </mlove> дає дитині змогу відчути: добре, що я народилася. </s> <s> Вона плекає в дитині <hi rend=italic> любов до життя </hi>, а не просто бажання жити. </s>

Sample 2

<s> One may be found in the narcissistic element in <mlove> motherly love </mlove> . </s> <s> In as much as the infant is still felt to be a part of herself, <mlove> her love </mlove> and infatuation may be a satisfaction of her narcissism. </s>

<s> Один із них можна знайти в нарцистичному елементі <mlove> материнської любові </mlove>. </s> <s> Оскільки мати все ще відчуває немовля як невід'ємну частину себе, <mlove> її любов </mlove> і палке захоплення ним можуть бути засобами задоволення нарцисизму. </s>

Figure 15: Sample of taglining

Therefore, the main function of markup is to present textual material in digital form using text encoding. In order for research to be easily shared, analyzed, reused and stored, texts on humanities

field must be coded according to a standard that defines common vocabulary, grammar and syntax. As we can see, these recommendations are fully adapted and adjustable to a specific project, while improving the compatibility of this project with other projects using TEI.

3. Creation and practical use of "Concordance 'Art of Loving'"

In recent years, Python has become increasingly popular due to its efficiency in development areas such as Machine Learning and Data Science. It is so due to its scalability and flexibility, which is so necessary in this segment of programming. The popularity of Python accredited thanks to the fact that this language is able to solve a wide range of problems and its application is possible on all popular platforms today, except for encoding mobile applications. So you can highlight the following benefits of Python:
- Possibility of expansion - there are libraries and frameworks for any type of tasks and needs;
- Simplicity of syntax - everything superfluous has been removed from syntax, the code is pure and clear without superfluous brackets and expressions;
- Interpretability - the Python interpreter exists for all popular platforms;
- PEP - the only standard for writing code that makes code supported and readable even when switching from one programmer to another;
- Open Source - Python interpreter code is open, which allows anyone interested in language development to participate in its development and improvement.

Creating a program for a concordance based on the original and translated novel "The Art of Love" by Erich Fromm is useful, because it will greatly simplify our further translation studies research. Also, taking into account all the advantages of the Python language, the dictionary was created in this programming language. Therefore, an integrated development environment for the PyCharm version 3.7 programming language was used to write the code and run the program. There are numerous advantages of using the created 'Concordance' Art of loving'". To open the program, use open PyCharm, open the code file and press RUN. Our Concordance 'Art of loving' program opens with the following window:
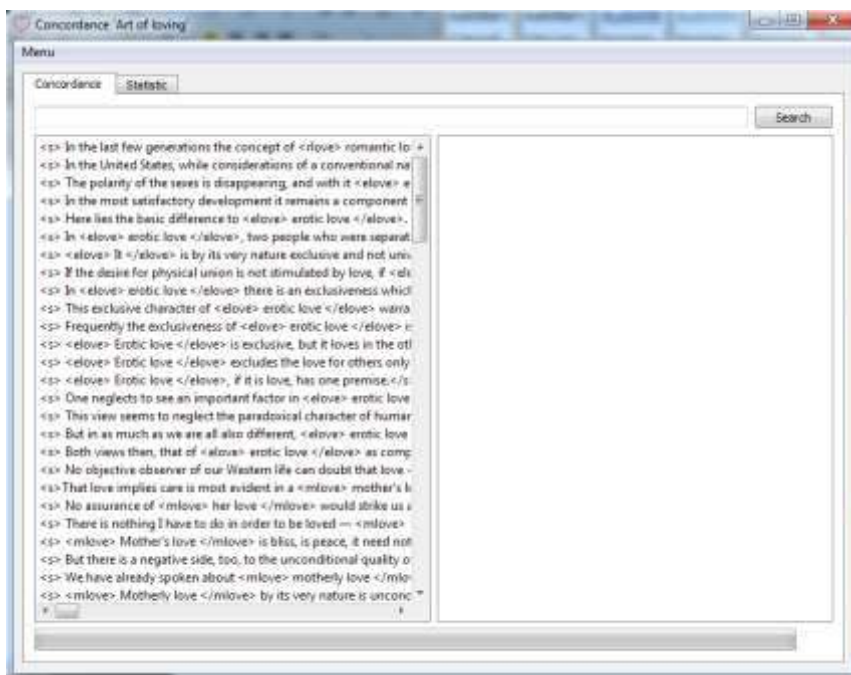

Figure 16: Concordance 'Art of loving' interface

In the window on the left we see a field with a list of sentences that contain <mlove>, <flove>, <blove>, <elove>, <slove>, <Glove>, <olove> tags. When you first start the program, the window on

the right is empty. However, if you enter a word or tag in the search field (Fig.17) and click the 'Search' button, the search results will be displayed in the right window (Fig.18).
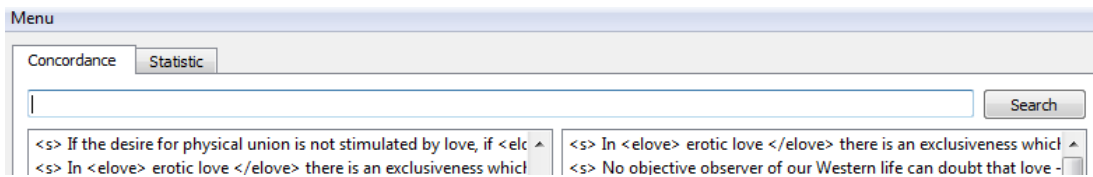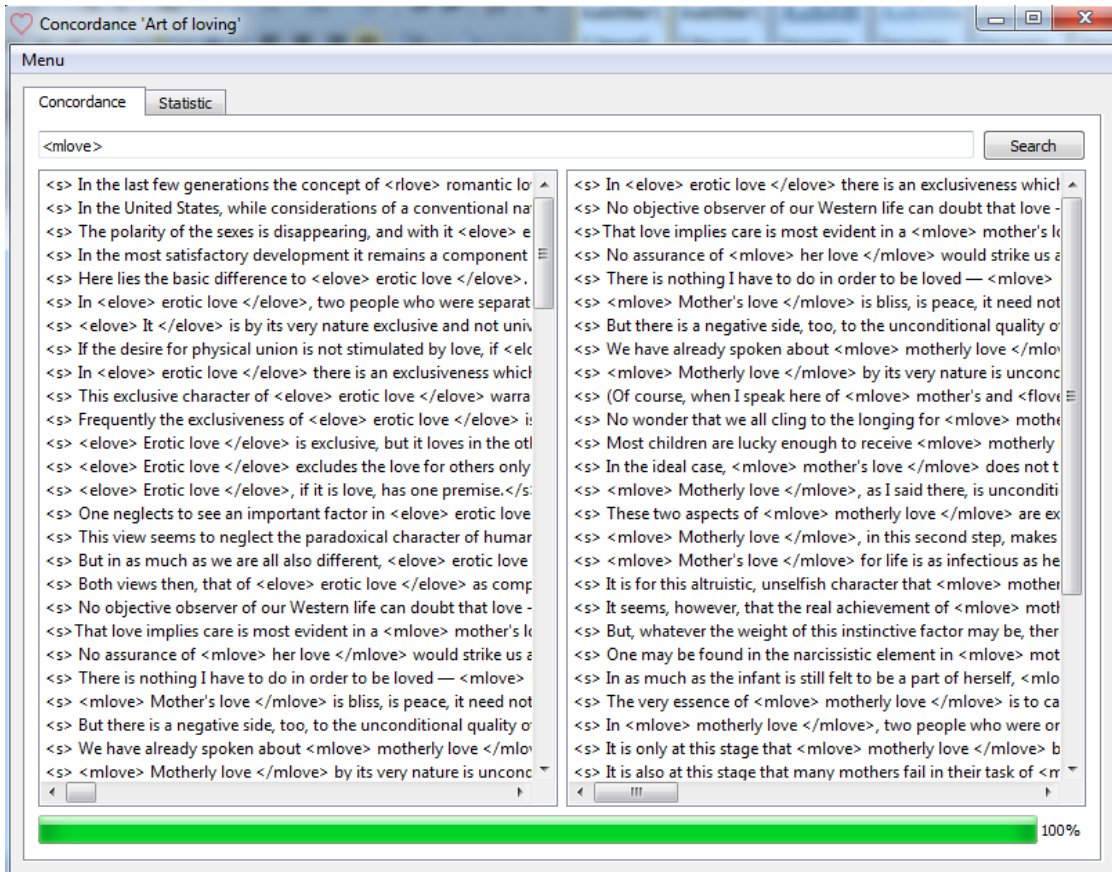

Figure 17: Search box

Figure 18: Example of a program with displayed search results

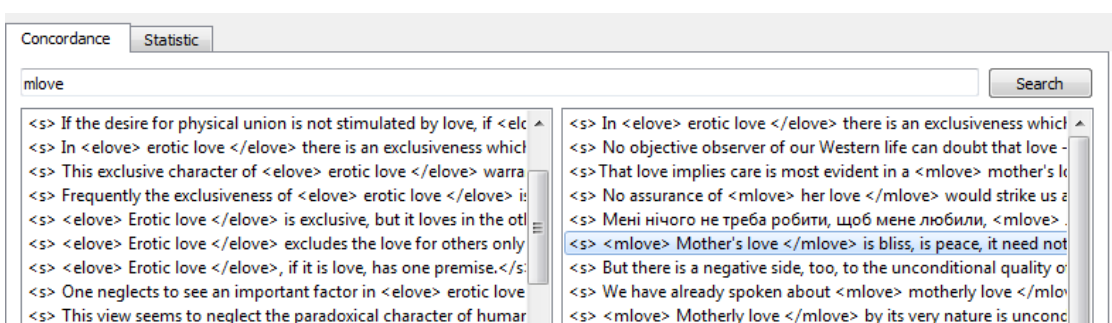When you click on a sentence in the right window, the translation of this sentence will be displayed, as in Fig. 17

Figure 19: Example of a program with the translation of sentences

In addition to tags, you can search by ordinary words or phrases; then the search result will be those sentences that contain these words (Fig. 20). In case of the absence of such words or phrases, the window on the right remains empty (Fig. 21).
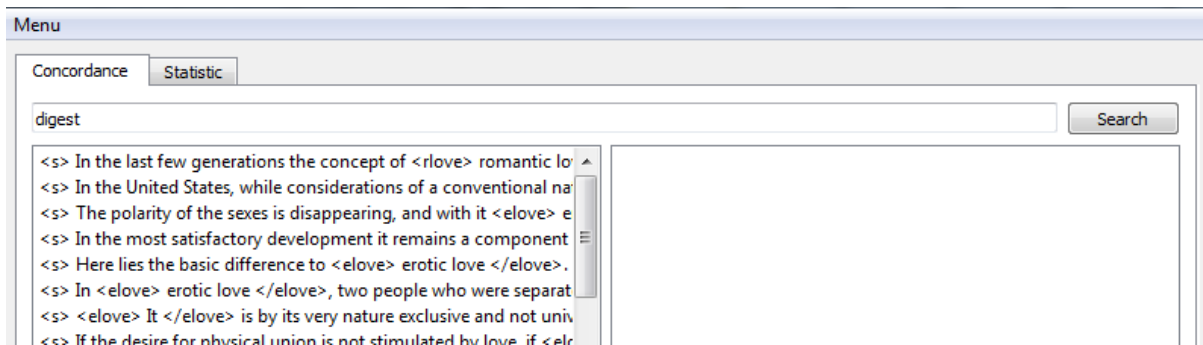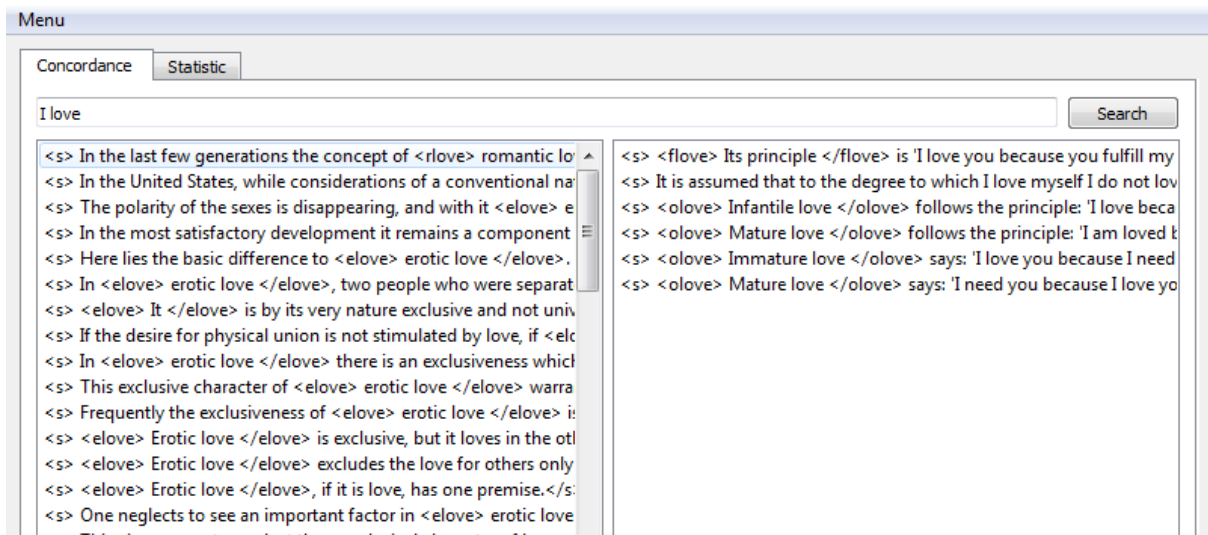

Figure 20: Example of missing word search

Figure 21: Example search by phrase

Apart from this, our program has a second window called "Statistic". This window displays certain statistics. For example, if you enter any of the tags in the search field, the "Statistic" window will display statistics for this tag, namely the number of sentences in which this tag is used (Fig.22).
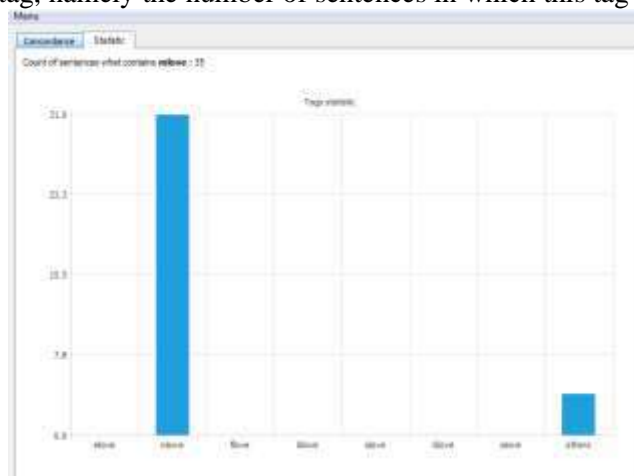

Figure 22: Presentation of statistics for the <mlove> tag in the Statistic window

The graph also shows a separate column "others", which contains sentences that contain other tags in addition to the <mlove> tag. For a better understanding, the following tips have also been developed, given in the column (Fig.23). They show the number of sentences shown on the graph.
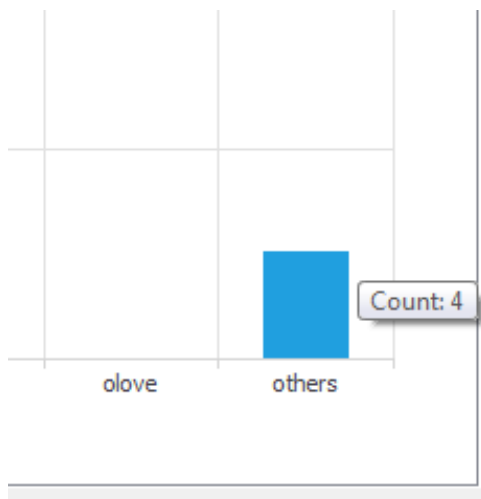

Figure 23: Hint on the column

In order to return to the list of all sentences, i.e. to the starting position, you should delete the text from the search term line, in our case it is <mlove> and press the "Search" button, then the program "Concordance 'Art of loving'" returns to the original pages and a complete list of sentences appears on the left and an empty text box on the right. Like any other program, our Concordance contains certain system functions that are contained in the "Menu" window (Fig.3.2.9).

Figure 24:– "Menu" window

That is, when you click the "About" button, we are shown a dialog box with information about the program and its author with the title "About program" and the text "Concordance based on 'Art of loving' Erich Fromm".

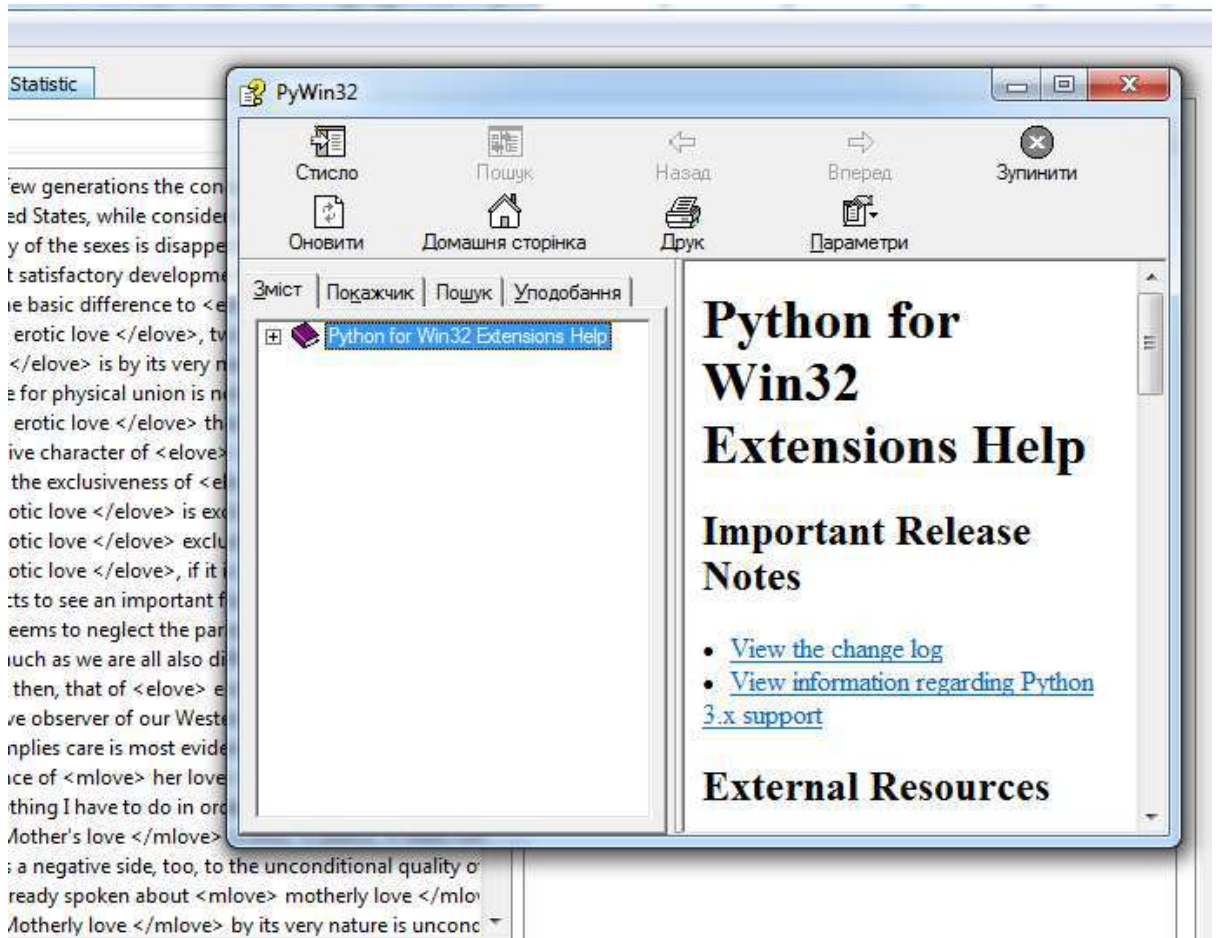In turn, the "Help" button opens a document with system help of using Python (Fig.25).

Figure 25:–«Help» button

When you click on the "Exit" button, a dialog box appears, which allows you to confirm the exit from the program (Fig.26).
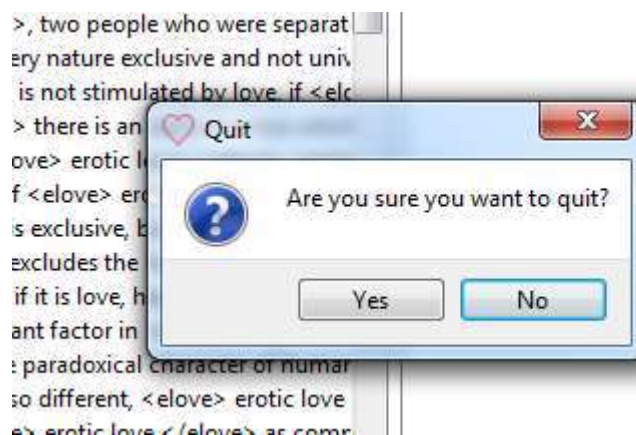

Figure 26: – «Exit» button

Thus, presented here is a detailed and step-by-step description of the Concordance 'Art of loving' functionality written in the Python programming language. The fragment of program code is given below:

```python
import sys

import os

from dictionary import sentences

from PyQt5.QtGui import QIcon, QBrush, QColor, QCursor

from PyQt5.QtCore import Qt

from PyQt5.QtWidgets import (QWidget, QMessageBox, QLabel, QVBoxLayout,
QTabWidget, QMainWindow, QAction, qApp,

                QApplication,     QMenu,      QLineEdit,      QPushButton,
QHBoxLayout,

                QProgressBar, QListWidgetItem,

                QListWidget, QToolTip)

from     PyQt5.QtChart     import     QChart,     QChartView,     QValueAxis,
QBarCategoryAxis, QBarSet, QBarSeries
```

Figure 27:Software code fragment one

```python
        "Count    of    sentences    what    contains    <b>{word}</b>    :
{count}".format(word=self.__last_word, count=count))

        tags = {'elove': 0, 'mlove': 0, 'flove': 0, 'blove': 0, 'slove': 0, 'Glove': 0,
'olove': 0, 'others': 0}

        for row in range(count):

            result = set()

            for tag in list(tags.keys())[:-1]:

                if '/' + tag in self.right_list.item(row).text():

                    result.add(tag)

            number = len(result)

            if number:

                if number > 1:

                    tags['others'] += 1

                else:
```

Figure 28: Software code fragment 2

4. Acknowledgements

The study has been conducted within the complex academic topic "Application of modern technologies for optimization of information processes in natural language" at Lviv Polytechnic National University. The concordance was compiled with the intention to further use this parallel translation corpus for quality assessment of a translation.

5. Conclusions

Thus, the creation of an electronic concordance on the basis of a parallel corpus will greatly facilitate the process of scrutinizing the creative work of German psychologist and philosopher writer Erich Fromm, and in particular the novel 'The Art of Loving'.

One of the most popular programming languages in the world, the Python programming language is also widely used in text studies. Therefore, within the framework of study we developed the Concordance 'Art of loving' program, which contains a list of sentences that contains at least one of a set of tags. However, our program does not have a predefined size limit; this advantage allows adding sentences with new tags or change existing sentences at any time.

Thanks to the Concordance 'Art of loving' program - all sentences that have at least one of the set tags in the original, as well as their Ukrainian counterparts and statistics on them - are gathered in one place, which greatly facilitates the study of the novel The Art of Love Erich Fromm.

All in all, with the help of our program, we determined that there are 31 sentences with the tag <mlove> in the text, <blove> - 12, <flove> - 8, <elove> - 13, <slove> - 18, <Glove> - 18 <olove> - 9. Furthermore, there are sentences in which several tags could be used together. There are 4 such sentences with the tag <mlove> in the text, <blove> - 5, <flove> - 2, <elove> - 5, and the tags <slove>, <Glove> and <olove> were used exclusively separately.

6. References

[1] Aston G., Burnard L. The BNC Handbook. Exploring the British National Corpus with SARA. Cambridge: Edinburgh University Press, 1998. 250 p

[2] Bray, Tim; Jean Paoli, C. M. Sperberg-McQueen, Eve Maler, François Yergeau (September 2006). Extensible Markup Language (XML) 1.0 (Fourth Edition) — Origin and Goals. World Wide Web Consortium. October 29 2006.

[3] Český národní korpus [Electronic recource]: http://ucnk.ff.cuni.cz

[4] Ciotti, F. Tomasi, Formal ontologies, linked data and TEI semantics. Journal of the Text Encoding Initiative (9). 2016.

[5] Corpus markup [Electronic recource]: http://index-of.co.uk/Computational-Linguistics/Unit%203%20Corpus%20markup%20%20.pdf

[6] D. Ivanchyshyn, V. Vysotska and S. Albota, "The Film Script Generation Analysis Based on the Fiction Book Text Using Machine Learning," *2021 IEEE 16th International Conference on Computer Sciences and Information Technologies (CSIT)*, 2021, pp. 68-80, doi: 10.1109/CSIT52700.2021.9648818.

[7] De, N. M., & Veronis, J. Text encoding initiative: background and contexts. Dordrecht, Kluwer Academic Publishers. 1995.

[8] Extensible Markup Language (XML) [Electronic recource]: https://www.w3.org/TR/REC-xml/#sec-well-formed

[9] Francis W. Language Corpora B.C. / W. Francis // In: Jan Svartvik (ed.) Directions in Corpus Linguistics. Proceeding of Nobel 82, Stockholm, 4–8. August, 1991, Berlin –New York:Mouton de Gruyter. 17-35 pp.

[10] I. Bekhta, N. Bondarchuk, O. Marchuk and V. Bialyk, "Processing Modality of British Modernist and Postmodernist Fiction," *2021 IEEE 16th International Conference on Computer Sciences and Information Technologies (CSIT)*, 2021, pp. 81-84, doi: 10.1109/CSIT52700.2021.9648624.

[11] I. Dilai and M. Dilai, "Automatic Extraction of Keywords in Political Speeches," *2020 IEEE 15th International Conference on Computer Sciences and Information Technologies (CSIT)*, 2020, pp. 291-294, doi: 10.1109/CSIT49958.2020.9322011.

[12] Ide N. Introduction. In: Ide N., Véronis J. (eds) Text Encoding Initiative. Springer, Dordrecht. 1995.

[13] Introducing the Guidelines [Electronic recource]: https://tei-c.org/support/learn/introducing-the-guidelines/

[14] Kennedy G. Introduction to Corpus Linguistics. London-New-York: Longman, 1998. 309 p.

[15] Khomytska I., Teslyuk V., Bazylevych I., Shylinska I. Approach for Minimization of Phoneme Groups in Authorship Attribution Attribution // International Journal of Computing. 2020. Vol. 19 (1). P. 55 – 62.

[16] Khomytska I., Teslyuk V., Kryvinska N., Bazylevych I. Software-Based Approach Towards Automated Authorship Acknowledgement—Chi-Square Test on One Consonant Group // *Electronics*. July 2020. Vol. 7:1138.

[17] Korpus Języka Polskiego [Electronic recource]: <http://korpus.pl/index.php?page=welcome>

[18] Kulyna, O. (2021). *Anthropocentrism as Implementation of a Testator/Testatrix's Communicative Goal.* COLINS, https://www.semanticscholar.org/paper/Anthropocentrism-as-Implementation-of-a-Goal-Kulyna/5108d3143f7e1f7c56a04ae204bccf3e5ce81a04

[19] M. Karp, N. y, I. Bekhta and U. Yurlova, "Linhvometriia of inversion and detachment in Ray Bradbury's Dandelion Wine," *2021 IEEE 16th International Conference on Computer Sciences and Information Technologies (CSIT)*, 2021, pp. 13-20, doi: 10.1109/CSIT52700.2021.9648775.

[20] McDonald, David D. and Jxmes Pustejovsky A ComputaLional Theory of Prose Style for Natoral Language Generation', Proceedings oi" the 1985 meeting of the European Association for Computational Linguistics. SAMSON, 1985.

[21] McEnery, T. and Wilson, A. Corpus Linguistics: An Introduction. 2nd Edition, Edinburgh University Press, Edinburgh. 2001.

[22] Meyer Ch. P. English Corpus Linguistics. An introduction. Cambridge University Press, 2004. 168 p.

[23] Michael Stubbs. Text and Corpus Analysis: Computer-Assisted Studies of Language and Culture. Blackwell Publishers (Language inSociety series, edited by Peter Trudgill, volume 23), 1996. 267 pp.

[24] O. Levchenko and M. Dilai, "A Method of Automated Corpus-Based Identification of Metaphors for Compiling a Dictionary of Metaphors: A Case Study of the Emotion Conceptual Domain," *2021 IEEE 16th International Conference on Computer Sciences and Information Technologies (CSIT)*, 2021, pp. 52-55, doi: 10.1109/CSIT52700.2021.9648667.

[25] Slovenský národný korpus [Electronic recource]: http://korpus.juls.savba.sk/index.sk.html.

[26] TEI by Example [Electronic recource]: https://teibyexample.org/

[27] The Art of Loving (1956) by Erich Fromm. New York: Harper & Row, 1956. 148 p.

[28] V. Lytvyn, Y. Burov, V. Vysotska, Y. Pukach, O. Tereshchuk and I. Shakleina, "Abstracting Text Content Based on Weighing the TF-IDF Measure by the Subject Area Ontology," *2021 IEEE International Conference on Smart Information Systems and Technologies (SIST)*, 2021, pp. 1-7, doi: 10.1109/SIST50301.2021.9465978.

[29] What Is Parsing? Definition and Examples in English Grammar [Електронний ресурс]. – Режим доступу :https://www.thoughtco.com/parsing-grammar-term-1691583

[30] What is TEI? [Electronic recource]: https://cdrh.unl.edu/articles/basicguide/TEI

[31] General'nij regional'no anotovanij korpus ukrains'koi movy (GRAK) http://uacorpus.org/Kyiv/ua [Генеральний регіонально анотований корпус української мови (ГРАК)]

[32] Fromm E. Mystectvo lubovi. Per. z angl. V. Kucmenko. Kharkiv: Knyzkovij Klub <<Klub Simejnoho dozvilla>>, 2017. 192 s. [Фромм Е. Мистецтво любові. Пер. з англ. В.Кучменко. Харків: Книжковий Клуб «Клуб Сімейного дозвілля», 2017. 192 с.]