

## APPLYING PYTHON TO DETECT COHESION IN AMY TAN'S *THE JOY LUCK CLUB*

**Karp M. A.**

### INTRODUCTION

The research's importance stems from the fact that there are few studies in the field of text linguistics that focus on textuality and cohesion in particular. The goal involves solving the following tasks:

- to identify lexical cohesion, anaphoric references, and conjunctions in the text in English using a program written in Python;
- to analyze cohesive devices in the text of Amy Tan's *The Joy Luck Club* in English and its version *Клуб веселощів і удачі* in Ukrainian.

In the investigated novel *The Joy Luck Club*, cohesion as a standard of textuality reflects the grammatical connectivity of the text. This is the answer to the question of how the components of the text are interrelated. It is important to note that the components of a text's surface structure are linked via grammatical forms and relations. Cohesion is considered the most text-oriented criterion suggested by R. A. de Beaugrande and W. Dressler within the concept of textuality.

### 1. Related works

Within the field of text linguistics, R. A. de Beaugrande, W. Dressler, R. Fowler, M. Ghadessy, M. Giuffrè, H. G. Widdowson, and others<sup>1,2,3,4</sup> have extensively studied textuality and cohesion theory. In their 1981 work *Introduction to Text Linguistics*, Robert-Alain de Beaugrande and Wolfgang U. Dressler proposed that textuality comprises seven standards: cohesion, coherence, intentionality, acceptability, informativity, situationality, and intertextuality<sup>5</sup>. They emphasize that cohesion and coherence are the primary text-oriented criteria, demonstrating how text elements interact to create overall meaning.

---

<sup>1</sup> Fowler R. *Linguistics and the Novel*. London : Routledge, 1989. 160 p.

<sup>2</sup> Ghadessy M. *Text and Context in Functional Linguistics*. Amsterdam : John Benjamins B.V, 1999. 340 p.

<sup>3</sup> Widdowson H. G. *Text, Context, Pretext. Critical Issues in Discourse Analysis*. Oxford : Blackwell, 2004. 196 p.

<sup>4</sup> Giuffrè M. *Text Linguistics and Classical Studies*. Dressler and de Beaugrande's Procedural Approach. Springer International Publishing AG, 2017. 192 p.

<sup>5</sup> Beaugrande R.-A., Dressler W. *Introduction to Text Linguistics*. London : Longman, 1981. 270 p.

In *Cohesion in Literary Texts* by Waldemar Gutwinski, cohesion is integrated into the language structure, not as isolated grammar category. This study reveals insights into discourse structure, participant line realization, sentence ordering, and semologic units<sup>6</sup>. *Cohesion in English*, by Michael Halliday, is a seminal work exploring cohesion's role in discourse coherence. The book analyzes linguistic mechanisms, like reference, substitution, ellipsis, conjunction, and repetition, fostering smooth information flow. It influences linguistics, and computational research, remaining vital for understanding language's organization and coherence in communication<sup>7</sup>.

According to Jan Renkema's opinion, cohesion is a connection that occurs when the interpretation of one textual element is dependent on other textual elements<sup>8</sup>. Furthermore, Deborah Schiffrin claims that cues are used by speakers and listeners to find the deeper meaning of language in the surface structure of the text<sup>9</sup>.

*Language, Cohesion and Form* by Margaret Masterman, highlights her pioneering role in computational linguistics. Emphasizing meaning over grammar, she believed machines could decode sentence meanings<sup>10</sup>. In *Collaborating towards Coherence*, Sanna-Kaisa Tanskanen examines how individuals use cohesive strategies to convey meaning within and between messages. The study finds that cohesion varies based on communicative conditions<sup>11</sup>.

*Lexical Cohesion and Corpus Linguistics*, by John Flowerdew, explores how words and phrases in text, known as lexical cohesion, enhance coherence and information flow. It employs corpus linguistics, using computational methods to analyze language datasets, revealing patterns of lexical use and their impact on textual cohesion.<sup>12</sup> *Languages and Social Cohesion* by Gabriela Meier explores how languages promote unity, understanding, and inclusivity in diverse societies. The book draws on interdisciplinary research

---

<sup>6</sup> Gutwinski W. *Cohesion in Literary Texts: A Study of Some Grammatical and Lexical Features of English Discourse*. The Netherlands : Mouton, 1976. 184 p.

<sup>7</sup> Halliday M.A.K., Hasan R. *Cohesion in English*. London : Longman, 1976. 400 p.

<sup>8</sup> Renkema J. *Discourse Studies : An Introductory Text Book*. Philadelphia : John Benjamins Publishing Company, 1993. 363 p.

<sup>9</sup> Schiffrin D., Tannen D., Hamilton H. *The Handbook of Discourse Analysis*. Massachusetts; Oxford : Blackwell Publishers Ltd, 2001. 874 p.

<sup>10</sup> Masterman M., Wilks Y. *Language, Cohesion and Form*. The United States of America : Cambridge University Press, 2005. 324 p.

<sup>11</sup> Tanskanen S.-K. *Collaborating towards Coherence: Lexical Cohesion in English Discourse*. Amsterdam ; Philadelphia : John Benjamins Publishing Company, 2006. 208 p.

<sup>12</sup> Flowerdew J., Mahlberg M. *Lexical Cohesion and Corpus Linguistics*. Amsterdam ; Philadelphia : Benjamins Current Topics, 2009. 132 p.

and case studies to show how linguistic diversity strengthens communication, empathy, and collaboration<sup>13</sup>.

In summary, the study of cohesion in literary texts has developed over time, progressing from individual grammar categories to a crucial aspect of language structure. These investigations have impacted fields like linguistics, education, computational research, and philosophy, illuminating complex language element relationships. Ultimately, this research highlights cohesion's importance in shaping communication, discourse structure, and social integration.

## 2. Methods

Linguistic processors are used to perform text analysis and synthesis<sup>14</sup>. At present there are various approaches to automatic text analysis.

Python's convenience comes from the fact that it is a high-level language with a set of structural programming constructs and modularity support. Python's flexibility and versatility ensure its widespread use<sup>15</sup>. Another advantage of Python is the ability to divide program code into modules that can then be used in other software products. Python comes with a large library of standard modules that can be used to build new programs. Python can help you save a significant amount of time that would otherwise be spent on compilation.<sup>16,17,18</sup>

This study is based on Amy Tan's novel *The Joy Luck Club* in English and its Ukrainian translation *Клуб веселощів удачі*. *The Joy Luck Club* is essentially an episodic novel. Amy Tan has structured the novel as a series of sixteen interwoven and interconnected stories, divided into four sections with

---

<sup>13</sup> Meier G., Smala S. *Languages and Social Cohesion: A Transdisciplinary Literature Review*. London; New York : Routledge Focus, 2021. 199 p.

<sup>14</sup> Когут В. С., Тарасова О. В., Комп'ютерна лінгвістика як галузь мовознавства. *Сучасний стан та перспективи розвитку науки* : матеріали Міжнародної студентської наукової конференції, 2020. С. 82–84.

<sup>15</sup> Яковенко А. В. *Основи програмування. Python. Частина 1 : підручник для студентів, які навчаються за спеціальністю 122 «Комп'ютерні науки» спеціалізації «Інформаційні технології в біології та медицині»*. Київ : КПІ ім. Ігоря Сікорського, 2018. 195 с.

<sup>16</sup> Karp M., Kunanets N., Yaremchuk T. Paragraphemic component of multimodal text analyzed in software product written in Python. *Proceedings of the 2nd International workshop IT project management* (ITPM 2021). Slavsko, Lviv region, Ukraine, 2021. Vol. I. P. 356–365.

<sup>17</sup> Karp M., Burtnyk A., Bekhta I., Kunanets N., Melnychuk O., Shainer I. Study of linking words in political speeches of Bill Clinton using Python. *IEEE 17th International Conference on Computer Science and Information Technologies* (CSIT 2022). Lviv, Ukraine, 2022. P. 77–82/83.

<sup>18</sup> Khomytska I., Teslyuk V., Bazylevych I., Kordiyaka Yu. Machine learning and classical methods combined for text differentiation. *Proceedings of the 6th International Conference on Computational Linguistics and Intelligent Systems* (COLINS 2022). Gliwice, Poland, 2022. Vol. I. P. 1107–1116.

four stories in each. Each chapter is prefaced by a short fable, proverb, or vignette that summarizes the essence of the stories that make up the book.

In the text of the novel *The Joy Luck Club* cohesion as a standard of textuality reflects the grammatical connectivity of the work. In order to detect some cohesive relations in the text, namely lexical cohesion, anaphoric references, conjunctions, and connecting words, we worked out a program in the programming language Python.

To do this, we framed an application in Python. The first of the application's additional modules is a set of NLTK libraries designed to analyze natural language texts. The Colorama and termcolor modules are needed to display color text in the console. We also used our own auxiliary modules: transitionalWords – which contains an array with a list of conjunctions (conjunctions and connecting words); cohesiveDevices – which contains functions for searching for cohesion methods in the text. Let us take a closer look at the cohesiveDevices module (see Fig. 1).

```

1 from nltk.corpus import stopwords, wordnet
2 from nltk.tokenize import word_tokenize, sent_tokenize
3 from nltk.tokenize.treebank import TreebankWordDetokenizer
4
5
6
7 def getReferenceWords(text):
8     words = word_tokenize(text)
9     stopwords = nltk_stopwords(stopwords)
10    return getWordsByTags(['PRP', 'PRP$', 'WP'], tagged(words))
11
12
13 def getAnaphoricReferences(tagged_text):
14     words = []
15     for (word, tag) in tagged_text:
16         if tag in tag_prn and word.lower() not in words:
17             words.append(word.lower())
18     return words
19
20
21 def getRepeatedWordsAndSynonyms(text):
22     words = word_tokenize(text)
23     tagged_words = nltk_pos_tag(words)
24     nouns = getWordsByTags(['NNP', 'NNPS'], tagged(words))
25
26     wordCounts = {}
27     for w in nouns:
28         if w in wordCounts.keys():
29             wordCounts[w] += 1
30         else:
31             wordCounts[w] = 1
32
33     repeatedWords = []
34     for key in wordCounts.keys():
35         if wordCounts[key] > 1 and key in nouns:
36             repeatedWords.append(key)
37
38     synonyms = []
39     for word in nouns:
40         if word in repeatedWords:
41             synonyms = nltk_stem_tagged_synonyms(word)
42             synonyms = getReferencesFrom(['NNP', 'NNPS'], tagged(synonyms))
43             for s in synonyms:
44                 if s in words and s != word and s in nouns:
45                     synonyms.append(word)
46             synonyms.append(word)
47
48     return set(repeatedWords + synonyms)
49
50
51 def getAnaphoricReferences():
52     synonyms = []
53     for key in wordCounts.keys():
54         for i in getReferencesFrom(key):
55             synonyms.append(i)
56     return set(synonyms)

```

**Fig. 1. The module cohesiveDevices**

To search for anaphoric references, we use the getReferenceWords method, in which the tokenized text is tagged using the post\_tag function, which is necessary for further searching for pronouns using the getWordsByTags function.

To detect lexical cohesion (repetition and synonyms) in the text, we use the getRepeatedWordsAndSynonyms method. Thus, we count the number of repetitions of all nouns and then select those words that are repeated more than once. For nouns, we also use NLTK to search for synonyms and add the word

along with the synonym to the resulting array if the synonym is found in the text.

The next step is to combine the arrays of synonyms and repeated words. As a result, we get an array of unique words using the set method and return the value from the function. Now let us look at the main module, main.py (see Fig. 2).

```
1 import sys
2 from nltk.corpus import stopwords
3 from nltk.tokenize import word_tokenize
4 from nltk.tokenize.treebank import TreebankWordDetokenizer
5
6 def highlightCohesiveDevices(text, referenceWords, repeatedWordsAndSynonyms):
7     formattedText = ''
8     numWords = 0
9     numReps = 0
10    numCoh = 0
11    numTrans = 0
12    formattedText = ''
13
14    for word in word_tokenize(text):
15        if word.lower() in (referenceWords for w in transitionalWords):
16            formattedText += word + ' (transitional)'
17            numTrans += 1
18        elif word.lower() in (referenceWords for w in repeatedWords):
19            formattedText += word + ' (repeated)'
20            numReps += 1
21        elif word.lower() in (referenceWords for w in repeatedWordsAndSynonyms):
22            formattedText += word + ' (repeated/synonym)'
23            numCoh += 1
24        else:
25            formattedText += word + ' '
26
27    detokenizedText = TreebankWordDetokenizer().detokenize(formattedText)
28
29    for w in repeatedWords:
30        numReps += len(w.lower()) * len(w.lower())
31        if len(detokenizedText.split(w.lower())) > 1:
32            detokenizedText = detokenizedText.replace(w, colored(w, 'red'))
33    numCoh += 1
34
35    for w in repeatedWordsAndSynonyms:
36        numCoh += len(w.lower()) * len(w.lower())
37        if len(detokenizedText.split(w.lower())) > 1:
38            detokenizedText = detokenizedText.replace(w, colored(w, 'green'))
39
40    print('Total words: ', len(words))
41    print('Total (transitional) words: ', numTrans)
42    print('Total (repeated) words: ', numReps)
43    print('Total (repeated/synonym) words: ', numCoh)
44    print('Total (unique) words: ', len(words))
45
46    return detokenizedText
47
48 def main():
49     text = sys.argv[1]
50     referenceWords = getReferenceWords(text)
51     repeatedWordsAndSynonyms = getRepeatedWordsAndSynonyms(text)
52     highlightCohesiveDevices(text, referenceWords, repeatedWordsAndSynonyms)
```

Fig. 2. The main module

In this module, we read the text to be analyzed from the text.txt file. We call the getReferenceWords and getRepeatedWordsAndSynonyms functions and pass them to the highlightCohesiveDevices method along with the text. This method is designed to search for cohesive devices in the text and highlight them with appropriate colors using auxiliary libraries, as well as count the number of devices used. Anaphoric references will be highlighted in green, conjunctions in red, and lexical cohesion tools in white.

The next step in the for loop is to go through the elements of the tokenized text, mark the words with the appropriate color, and add them to the resulting array formattedText. After the loop completes, we detokenize the text using the detokenize function of the TreebankWordDetokenizer class contained in the NLTK library. To search for conjunctions consisting of two or more words, we additionally loop through the elements of the transitionalWords array and mark the phrases found in the text with the appropriate color.

### 3. Materials

It is worth noting that cohesion is a complex criterion of textuality and requires significant human resources to fully analyze it in the text. In this

study, we have used examples from the novel *The Joy Luck Club* to demonstrate each of the cohesive devices.<sup>19,20</sup>

We have considered partial repetition, which involves using the same basic components of a word but transferring them to another word class. For example:

[1.] *As is the Chinese cook's custom, my mother always made disparaging remarks about her own cooking.* І за китайською кухарською звичкою моя мама завжди робила зневажливі зауваження щодо її кухаріння.

[2.] *But mostly I slept. I took the sleeping pills Ted had left behind in the medicine cabinet.* Але в основному мене тягнуло на сон. Я приймала сподійне, яке Тед залишив у аптеці.

[3.] *Promise! Promise! Honey-sweetheart, my promise is as good as gold.* Обіцяю! Обіцяю! Милий, моя обіцянка на вагу золота.

Parallelism involves the repeated use of the same structural patterns but filled with different expressions. Examples from the novel *The Joy Luck Club* clearly demonstrate such repetition:

[4.] *And if you could float down this river and eat the moss for food, you would be strong enough to climb the peak. If you slipped, you would only fall into a bed of soft moss and laugh. And once you reached the top, you would be able to see everything...* І якби ти могла спуститись вниз цією річкою і поїсти мох, то ти була б достатньо сильною, щоб піднятися на вершину. Якби ти послизнулася, ти б впала на ліжко з м'якого моху і лише засміялась. І як тільки ти досягла б вершини, ти б змогла побачити все...

[5.] *No scar, no skin, no flesh. Ні шраму, ні шкіри, ні плоти.*

Another characteristic of cohesion is the use of paraphrasing. A paraphrase is a repetition of the content with a change of expression. The concept of paraphrasing is identified with the concept of synonymy. For example:

[6.] *I'm drawn first to a large platter of chaswei, sweet barbecued pork cut into coin-sized slices.* Спочатку мене приваблює велика таця часвею, солодкої смаженої свинини, нарізаної скибочками завбільшки з монету.

[7.] *She talked of clever peddlers who served every kind of simple food: steamed dumplings, boiled peanuts, and my mother's favorite, a thin pancake with an egg...* Вона розповідала про спритних торговців, які подавали будь-яку просту їжу: пельмені, варені на пару, варений арахіс і мамин улюблений тонкий млинець з яйцем...

Another cohesive device that contributes to the compactness and efficiency of the text is the ellipsis. As a literary stylistic figure, an ellipsis consists of omitting a specific member of a sentence or phrase that is easily understood from the meaning of the poetic speech. For example:

---

<sup>19</sup> Тан А. *The Joy Luck Club*. New York : G. P. Putnam's Sons, 1989. 288 p.

<sup>20</sup> Тан Е. *Клуб веселощів і удачі*. Київ, 2005. 285 с.

[8.] *I wanted to tell you ... Rich and I are getting married.* Я хотіла сказати тобі... Ми з Річем одружуємося.

[9.] *We need to think about what our marriage is really based on ... not this balance sheet, who owes who what.* Нам потрібно думати про те, на чому насправді базується наш шлюб... не бухгалтерський баланс, хто кому що повинен.

Cohesion should also be considered through the concepts of time and aspect. In the text of Amy Tan's novel, different temporal forms are used in different aspects, as the characters tell their stories in several temporal frames at once.

[10.] *Auntie Lin explains how mad she got at a store clerk who refused to let her return a skirt with a broken zipper.* Тітонька Лін розповідає, як вона розлютилася на продавця магазину, який не дозволив їй повернути спідницю зі зламаною блискавкою.

Thus, as a result of processing the text of Amy Tan's novel *The Joy Luck Club* with a program written in Python, we got a graphical display of the found means of cohesion, such as lexical cohesion, anaphoric references, and conjunctions in the text. We also performed additional text analysis to find and highlight all types of cohesive relationships.

#### 4. Results

It is important to note that lexical cohesion in Amy Tan's *The Joy Luck Club* is manifested by repetition and synonyms. The most commonly used anaphoric deixis in the text are pronouns that function instead of nouns or noun phrases with which they correlate.

Conjunctive plane in the text of Amy Tan's *The Joy Luck Club* is represented by the following conjunctions: *and* – *i/ї/я, moreover* – *більше цього, also* – *також, in addition* – *на додаток, besides* – *окрім, furthermore* – *крім цього, and others.* A conjunction is an additional relation that is used to connect two interdependent events or situations mentioned in a sentence.

Disjunction in the novel is expressed by the conjunction *or* – *або* (sometimes extending to *either...*, *...or* – *або...*, *...або, ні...*, *...ні*). In the text, disjunction expresses an alternative.

Contrajunction is most often expressed by the conjunction *but* – *але*, less often with conjunctions *however* – *однак, yet* – *проте, nevertheless* – *тум не менше*, and others. The function of contrajunction is to cause problematic transitions in places where seemingly unlikely combinations of events or situations occur. Subordination is represented in the text by conjunctions such as *because* – *тому що, since* – *оскільки, as* – *тому, thus* – *таким чином, while* – *тимчасом як, therefore* – *унаслідок цього*, and others. Subordinating conjunctions make general types of cohesive relations explicit.

As a result, we have obtained a program that can be used to identify the means of cohesion in an English text (see Fig. 3).

```

C:\Projects\cohesion\cohesion\script\main.py
total words: 4001
lexical cohesion are highlighted in white, found 16 elements
morphemic references are highlighted in green, found 98 elements
conjunctions are highlighted in red, found 68 elements

... they arrived in San Francisco, ... mother made ... hide these shiny ... were the same brown-checked Chinese ... until the brief
... Malcom Society gave ... two hand-me-down ... all ... large in size for American ... . The society was composed of a group of white-
... American missionary ... from the First Chinese Baptist Church ... because of their gifts, ... haven't could ... refuse ... invitations
... to join the church ... nor could ... ignore the ... practical advice to succeed ... English through table ... on weekends in
... , ... , through their practice on Saturday mornings ... this was how ... around set the bus, the jeep, ... the 16, ... , ...
... could sense that the word of these ... had unspeakable tragedies ... had left behind in China ... hopes ... couldn't begin to express
... fragile English ... of ... lost ... recognized the memory in three ... faces ... saw how quickly ... eyes moved ...
... told ... for the Joy Luck Club ... Joy Luck was an ... someone from the ... of her first marriage in Hawaii, ...
... the Japanese case ... that's why ... think of Joy Luck as ... well in ... was the ... would always talk ... was bored, ...
... there was nothing to do ... every ... had been washed ... the Formula Table had been signed down twice ... sitting ... cut reading ... the ...
... of ... smoking one roll ... cigarettes ... another, ... warning ... to ... this ... would ... a ... of ...
... sent to ... by unson relatives from ... would ... the bottom of ... would ... pull out a lively ... of ...
... of ... began to ... with one ... rhythm ... would start ... store, ... over the ... told ... the same
... except for the ... which grow ... cutting ... into ... eventually ... site ... dressed about ... before ...
... to ... had ... been ... speaking Chinese ... I ... dressed ... a ... with ... the ...
... of these ... were white mists ... if ... could float ... than this ... out the ... for food, ... would be strong ... to climb
... if ... slipped, ... would only fall into a ... of soft ... laugh ... reached the top, ... would be able to see anything
... feel such happiness ... would be enough to ... here were ... over ... The ... dressed about ...
... I ... realized he ... dress ... the poor ... thought ... saw the ... I ... could see ...
... looked like giant ... heads trying to jump out of a vat of oil ... behind each ... could see ... of another ...
... and ... the clouds would ... just a little ... the ... would suddenly become ... elephants ... slowly toward
... can ... see their ... at the root of the ... were ... rocky ... little ... hanging ... rock ... in the ...
... melody, ... tunes, ... and ... these were ... strange ... beautiful ... can't ... ever ... didn't ... come to ...
... beautiful ... was ... the ... as ... husband brought ... the babies to ... thought ... would be safe ... was an ...
... with the ... a ... in ... a small room in a two-story ... went off to the ... to ...
... Japanese were ... even after ... the ... every ... every four ... thousands of ... poured into ...
... looking for places to live ... came free the fast, west, ... south ... were rich ... poor, Shanghai, Canton, ...
... , ... just Chinese, ... foreigners ... missionaries of every religion ... there was, of course, the ... any ...
... thought ... were by ... in ... was a ... of ... together ... didn't ... in the ...
... would have been plenty of reason for fighting to break out among these different ... can ... see ... Shanghai ... with north ...
... bankers with bankers, ... with ... refugees ... looked down on ... also ... didn't ... that ...
... of the case ... to ... on ... suffered the same ... all had the same ... contained ...
... like the ... I ... have the ... air force ... said ... hands ... pushed ... around ... gave ... dirty diseases,

```

Fig. 3. The result of program execution

In the second part of the case study, we took a closer look at the cohesive devices in Amy Tan's *The Joy Luck Club* and gave examples from the novel to illustrate how the author used a particular device. We studied partial repetition, parallelism, paraphrase, and ellipsis. It was also important to consider the text in terms of time and aspect because the novel uses a huge number of different tense forms. This is due to the fact that the story is told in several time frames, so the sentences are constructed in a logical way to reflect the period in which a certain event takes place.

## 5. Discussions

Taking into account the theories proposed by various linguists as well as the tendencies in the study of cohesion and its division into grammatical and lexical, we have created a program for searching and highlighting both types. The practical significance of this work is that the program written in Python is useful for studying the cohesion criterion in English texts of various genres.

In this study, we have examined cohesion as a criterion of textuality in Amy Tan's *The Joy Luck Club* in English and its Ukrainian version *Клуб веселощів і удачі*. The stability of the text as a system is maintained through the continuity of linguistic units. The notion of *continuity* used here is based on the assumption that various language units in a text and their situational use are interconnected. From a cognitive point of view, this means that each



linguistic unit is a tool for accessing at least several other units. The most obvious illustration of this phenomenon is the language syntax system, which imposes organizational models on the surface structure of the text. By using the term *cohesion*, we aim to emphasize the function of syntax in the communicative aspect.

The basic units of syntax are patterns that clearly indicate dependencies. These units include a phrase, a part of a complex sentence, and a clause. For large amounts of text, there are language tools that show how structures and constructions that have already been used can be reused, modified, or compressed. These tools contribute to stability and economic efficiency in terms of both the use of materials and the effort required to process them. Among these devices are repetition, partial repetition, parallelism, paraphrase, pro-form, and ellipsis. Repetition has a disadvantage, namely, a decrease in the level of information content. As a result, methods in which forms are repeated with slightly different content or content is repeated with different forms are frequently used. Parallelism involves the reuse of surface formats but filling them with different expressions.<sup>21</sup>

In the text of Amy Tan's *The Joy Luck Club*, cohesion is often used to shorten and simplify the surface text, even though it results in a certain loss of definition. One obvious technique is the use of pro-forms: short words devoid of their own specific meaning that can be found on the surface of the text instead of more deterministic expressions that activate meaning. These pro-forms save text users from having to repeat everything to keep the current content active.<sup>22</sup>

Different tense forms are used in different aspects, as the characters tell their stories in several time frames at once. That is why cohesion should also be considered through the concepts of time and aspect. These categories are organized very differently in different languages. Some of these differences arise mainly from the point of view of the users of the text at that moment, while others have to do with the organization of situations or events in the textual world with each other. If the verb systems do not make the distinction explicit, modifiers or conjugations should be used.

A clear means of signaling connections between events or situations is through conjunctions and the use of connecting expressions. Therefore, in this study, we have considered four main types: conjunction, disjunction, contrajunction, and subordination. Thus, as a result of the study, we have come to the conclusion that cohesion is a key criterion within textuality, as it is responsible for the structural and grammatical coherence of the text.

---

<sup>21</sup> Beaugrande R.-A., Dressler W. Introduction to Text Linguistics. London : Longman, 1981. 270 p.

<sup>22</sup> Beaugrande R.-A., Dressler W. Introduction to Text Linguistics. London : Longman, 1981. 270 p.

## CONCLUSIONS

To summarize, cohesion is a key criterion of textuality that refers to how the actual words we can hear or see are structurally and grammatically related. The components of a text's surface structure are dependent on one another based on grammatical forms and properties. As a text-oriented criterion it is defined as a set of linguistic means that can be used to create a specific texture in a text while also contributing to the human ability to interpret a text holistically. The most important principle and criterion of textuality is cohesion. There are linguistic means for large volumes of text that demonstrate how previously used structures and constructions can be reused, modified, or compressed. Repetition, partial repetition, parallelism, paraphrase, pro-form, and ellipsis are examples of such devices. We analyzed the text of the novel from the point of view of cohesion using a program written in Python programming language. For a deeper linguistic analysis, we conducted an additional analysis of cohesive devices in the text of Amy Tan's novel *The Joy Luck Club* in English and its version *Клуб веселощів і удачі* in Ukrainian.

## SUMMARY

Textuality is the primary property that distinguishes a set of sentences as a text which in its turn is regarded as a communicative phenomenon that meets its seven standards, according to Robert-Alain de Beaugrande and Wolfgang Dressler's textual theory. Textual cohesion as the primary standard of textuality is a text-oriented criterion, as it is responsible for the grammatical connectivity of text as well as the coherence of internal textual worlds. It is a link that appears when the interpretation of one element is dependent on the other elements in the text. This standard is related to "semantic connections" in a text, which are created when elements that combine to form a certain meaning are dependent on one another. The purpose of this paper is to explore cohesion as a textuality standard, which refers to the components of the text's surface structure in Amy Tan's "*The Joy Luck Club*" in both its English and Ukrainian versions. A program written in Python was created to identify and highlight lexical cohesion, anaphoric references, and conjunctions in the researched text. It can be used for other studies of cohesion, both from the perspective of textuality and for searching and analyzing cohesive meaning in English texts of various genres. The program's primary benefit is its simplicity and the wide range of its capabilities. To demonstrate and confirm the theoretical data, excerpts from Amy Tan's novel "*The Joy Luck Club*" in English and its version "*Клуб веселощів і удачі*" in Ukrainian were used.

## Bibliography

1. Когут В. С., Тарасова О. В. Комп'ютерна лінгвістика як галузь мовознавства. *Сучасний стан та перспективи розвитку науки* : матеріали Міжнародної студентської наукової конференції, 2020. С. 82–84.
2. Тан Е. Клуб веселощів і удачі. Київ, 2005. 285 с.
3. Яковенко А. В. Основи програмування. Python. Частина 1 : підручник для студентів, які навчаються за спеціальністю 122 «Комп'ютерні науки» спеціалізації «Інформаційні технології в біології та медицині». Київ : КПІ ім. Ігоря Сікорського, 2018. 195 с.
4. Beaugrande R.-A., Dressler W. Introduction to Text Linguistics. London : Longman, 1981. 270 p.
5. Flowerdew J., Mahlberg M. Lexical Cohesion and Corpus Linguistics. Amsterdam ; Philadelphia : Benjamins Current Topics, 2009. 132 p.
6. Fowler R. Linguistics and the Novel. London : Routledge, 1989. 160 p.
7. Ghadessy M. Text and Context in Functional Linguistics. Amsterdam : John Benjamins B.V, 1999. 340 p.
8. Giuffrè M. Text Linguistics and Classical Studies. Dressler and de Beaugrande's Procedural Approach. Springer International Publishing AG, 2017. 192 p.
9. Gutwinski W. Cohesion in Literary Texts: A Study of Some Grammatical and Lexical Features of English Discourse. The Netherlands : Mouton, 1976. 184 p.
10. Halliday M.A.K., Hasan R. Cohesion in English. London : Longman, 1976. 400 p.
11. Karp M., Burtnyk A., Bekhta I., Kunanets N., Melnychuk O., Shainer I. Study of linking words in political speeches of Bill Clinton using Python. *IEEE 17th International Conference on Computer Science and Information Technologies (CSIT 2022)*. Lviv, Ukraine, 2022. P. 77–82/83.
12. Karp M., Kunanets N., Yaremchuk T. Paragraphemic component of multimodal text analyzed in software product written in Python. *Proceedings of the 2nd International workshop IT project management (ITPM 2021)*, Slavsko, Lviv region, Ukraine, 2021. Vol. I. P. 356–365.
13. Khomytska I., Teslyuk V., Bazylevych I., Kordiiaka Yu. Machine learning and classical methods combined for text differentiation. *Proceedings of the 6th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2022)*. Gliwice, Poland, 2022. Vol. I. P. 1107–1116.
14. Masterman M., Wilks Y. Language, Cohesion and Form. The United States of America : Cambridge University Press, 2005. 324 p.

15. Meier G., Smala S. Languages and Social Cohesion : A Transdisciplinary Literature Review. London ; New York : Routledge Focus, 2021. 199 p.
16. Renkema J. Discourse Studies : An Introductory Text Book. Philadelphia : John Benjamins Publishing Company, 1993. 363 p.
17. Schiffrin D., Tannen D., Hamilton H. The Handbook of Discourse Analysis. Massachusetts; Oxford : Blackwell Publishers Ltd, 2001. 874 p.
18. Tan A. The Joy Luck Club. New York : G. P. Putnam's Sons, 1989. 288 p.
19. Tanskanen S.-K. Collaborating towards Coherence : Lexical Cohesion in English Discourse. Amsterdam; Philadelphia : John Benjamins Publishing Company, 2006. 208 p.
20. Widdowson H. G. Text, Context, Pretext. Critical Issues in Discourse Analysis. Oxford : Blackwell, 2004. 196 p.

**Information about the author:**

**Karp Marta Andriivna,**

Candidate of Philological Sciences, Associate Professor,  
Associate Professor at the Department of Applied Linguistics  
Lviv Polytechnic National University  
12, Stepan Bandera st., Lviv, 79013, Ukraine