Style Transfer of Modern Hebrew Literature Using Text Simplification and Generative Language Modeling

Pavel Kaganovich[1,†], Ophir Münz-Manor[2,*,†] and Elishai Ezra-Tsur[2,*,†]

[1]Reichman University, 8 Ha'universita St, Herzliya, 4610101, Israel
[2]The Open University of Israel, 1 University Road, Raanana, 43107, Israel

Abstract

The task of Style Transfer (ST) in Natural Language Processing (NLP), involves altering the style of a given sentence to match another target style while preserving its semantics. Currently, the availability of Hebrew models for NLP, specifically generative models, is scarce. The development of such models is a non-trivial task due to the complex nature of Hebrew. The Hebrew language presents notable challenges to NLP as a result of its rich morphology, intricate inflectional structure, and orthography, which have undergone significant transformations throughout its history[1]. In this work, we propose a generative ST model of modern Hebrew language that rewrites sentences to a target style in the absence of parallel style corpora. Our focus is on the domain of Modern Hebrew literature, which presents unique challenges for the ST task. To overcome the lack of parallel data, we initially create a pseudo-parallel corpus using back translation (BT) techniques for the purpose of achieving text simplification. Subsequently, we fine-tune a pre-trained Hebrew language model (LM) and leverage a zero-shot Learning (ZSL) approach for ST. Our study demonstrates significant achievements in terms of transfer accuracy, semantic similarity, and fluency in the ST of source sentence to a target style using our model. Notably, to the best of our knowledge, no prior research has focused on the development of ST models specifically for Modern Hebrew literature. As such, our proposed model constitutes a novel and valuable contribution to the field of Hebrew NLP, Modern Hebrew Literature and more generally computational literary studies.

Keywords

Computational Literary Studies, Modern Hebrew Literature, Natural Language Processing, Style Transfer, Language Model, Hebrew Language

[1]Hebrew orthography has evolved over time, and there are differences between modern Hebrew, biblical Hebrew, and other historical forms of the language. This can make it difficult to create models that are robust across different time periods and genres.

CHR 2023: Computational Humanities Research Conference, December 6 – 8, 2023, Paris, France

*Corresponding author.
†These authors contributed equally.
pavelkag@gmail.com (P. Kaganovich); ophirmm@openu.ac.il (O. Münz-Manor); elishai@nbel-lab.com (E. Ezra-Tsur)
https://www.openu.ac.il/en/personalsites/ofirmanor.aspx (O. Münz-Manor); https://www.openu.ac.il/en/personalsites/elishaiezratsur.aspx (E. Ezra-Tsur)
0009-0007-0485-3798 (P. Kaganovich); 0000-0001-6333-345X (O. Münz-Manor); 0000-0003-1304-8022 (E. Ezra-Tsur)

1. Introduction

Neural ST is a widely used optimization algorithm in computational visual art, which involves leveraging convolutional neural networks (CNNs) to blend a content image with a style reference representation, resulting in a novel visual experience. An example of this process can be observed in the work ”The Face of Art” [22], as shown in Figure 1. In the domain of NLP, ST is employed as a generation task, wherein input sentences are rephrased into a desired target style while ensuring the preservation of the original semantics.

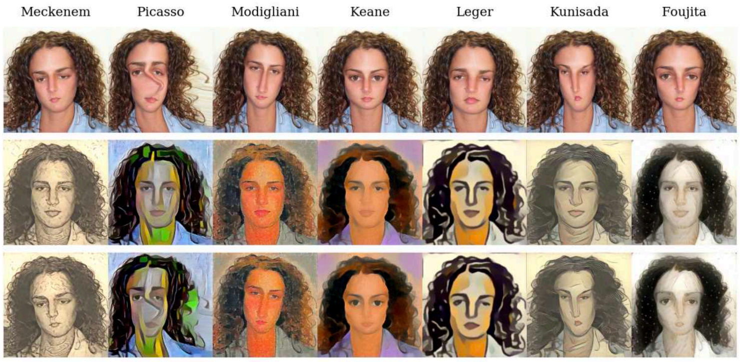


Figure 1: 1st row contains the results of the **geometric** stylization stage. 2nd row contains the results of using the **texture** ST algorithm [6] on the input image, without performing geometric stylization. 3rd row contains the stylization results of **geometry and texture** style transfer application [22].

Our work focuses on applying neural ST to the domain of Hebrew literature, which presents unique challenges even for domain experts. While ST is commonly applied to text sources with distinct styles that are easily recognizable by human readers, such as the Biblical text, Twitter posts, Wikipedia articles, or texts in the style of Shakespeare, literature lacks clear-cut stylistic boundaries. This poses a complex challenge for ST in the literature domain, as determining the appropriate style for a given sentence is often difficult due to the absence of definitive demarcations between various authors' styles. We aim to gain a deeper understanding of the unique style characteristics exhibited by individual authors in Hebrew literature and the intricate relationships between different writing styles. The findings have the potential to significantly contribute to the field of literary studies by illuminating the nuances of authorial styles in Hebrew literature. This could lead to a better comprehension of stylistic choices made by different authors and pave the way for further exploration and analysis of writing styles in Hebrew literature. Ultimately, our research seeks to enrich the understanding of Hebrew literature and its distinct stylistic features, advance the field of ST in Hebrew NLP, and open up new avenues for research at the intersection of literature and computational linguistics.

Previous studies in this area conflates style transfer with the related tasks such as translation [10] , learning latent representations to disentangle style and content from sentences [9], attribute transfer [19] and the relatively simple methodology for controlled paraphrase generation [13] that has achieved state-of-the-art (SOTA) results. Adopting English language style transfer solutions to Hebrew is not a trivial task due to the fact that most previous studies have relied on proposed solutions based on high-quality pre-trained models [1] [4] and datasets [17] that are not commonly available for the Hebrew language.

In this work, we propose a text simplification-based approach for performing ST in the Hebrew language. Our method is unsupervised and does not require parallel data between different styles and proceeds in three simple stages:

1. Create a pseudo-parallel corpus using BT as illustrated in Figure 2a.
2. Fine-tune a pre-trained Hebrew LM as illustrated in Figure 2b.
3. Employing a ZSL approach for ST as illustrated in Figure 3.

The remainder of this paper is structured as follows: Section 2 describes our proposed ST model, outlining its key components and architecture. In Section 3, the evaluation method used for assessing the performance of our model is detailed. Section 4 presents the experimental setup and results, including a comprehensive analysis of the findings. Finally, in Section 5, we present the conclusions drawn from our research, summarizing the key findings, discussing their implications, and suggesting potential avenues for future research.

2. Style Transfer via Text Simplification

It is natural to consider the task of ST as a translation problem that could potentially be addressed using a sequence-to-sequence (Seq2Seq) neural machine translation (MT) model. However, to train such a model for the ST task, it is necessary to collect parallel corpora that are aligned at the sentence level. Given the large number of stylistic categories involved, collecting parallel texts for all or even a substantial number of style pairs is infeasible. Thus, directly casting ST as an MT problem in a standard supervised setting is not viable.

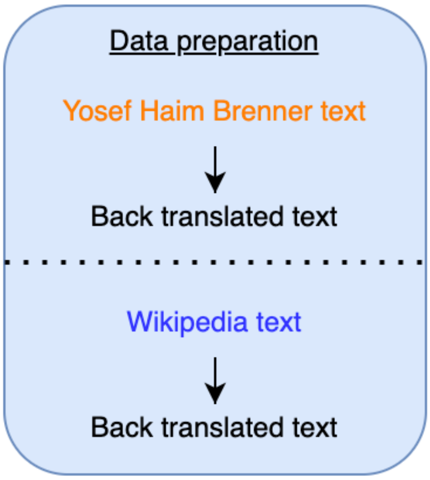
To overcome the lack of parallel texts for ST, we propose creating pseudo-parallel sentence pairs (as illustrated in Figure 2a) using BT. BT is a frequently used technique in NLP for quality assurance in MT and data augmentation[1]. By utilizing BT, we acquire a drier and more colloquial text that preserves semantics but is stripped of specific styles. Subsequently, we fine-tune a pre-trained Hebrew GPT-Neo-small[2] LM [4] to implement our ST model.

The corpus used in this study consists of 35 novels written by four authors from the mid nineteenth century to the present day. These authors were chosen thank to their distinct literary and linguistic styles. The data was gathered from the Ben-Yehuda Project[3], which is a repository of Hebrew literary texts in the public domain. Further details about the corpus, including its construction and filtering operations, can be found in Appendix A.

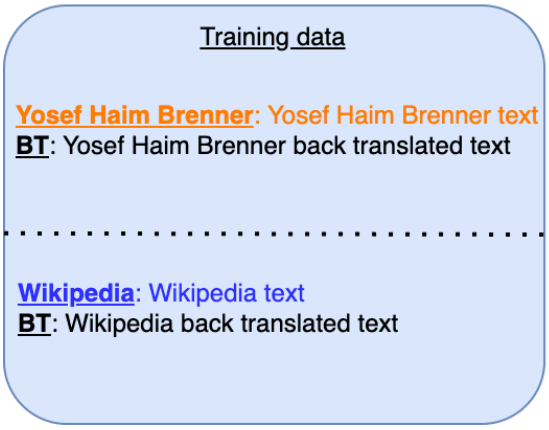
2.1. ST Model Implementation with GPT-Neo

We fine-tune the pre-trained GPT-Neo-small LM [4] to implement our ST model. Utilizing a pre-trained LM as the starting point for our ST model offers several advantages, including improved

[1]Data augmentation is a collection of techniques that manage the process of automatically generating high-quality data on top of existing data.
[2]https://huggingface.co/Norod78/hebrew-gpt_neo-small
[3]https://benyehuda.org/



(a) Back translated sentence pairs are generated by translating a source sentence to a target language and then back to the original language.



(b) The sentence pairs generated by the BT process were utilized in the model training process.

Figure 2: The training corpus was constructed by creating pairs of original text and their simplified versions, and utilizing them in the training process.

output fluency and enhanced generalization to small domain-specific corpus. In our approach, the text and it simplified version sequences are concatenated together using a separator token, as depicted in Figure 2b. To implement our model, we utilize HuggingFace's Transformers library [**HuggingFace's**], further details regarding the architecture and hyperparameters can be found in Appendix C.1.

Following the training process, the trained model is capable of generating simplified versions of literary texts (from one style to back translated text) or vice versa. To achieve ST from style A to style B, we adopt a ZSL approach[4]. ZSL is a machine learning (ML) technique that is commonly employed in the fields of CV and NLP to prompt a model to perform tasks for which it was not explicitly trained. In NLP, ZSL has been applied to tasks such as MT [11] and text summarization [14]. In our case, our model has not undergone explicit training for ST from style A to style B. To achieve ST, we employ prompt engineering, which is a novel approach for leveraging pre-trained LMs to perform tasks without fine-tuning. In this approach, the target task is directly conveyed to the model through a natural language task description that is integrated into the actual input sentence in a specific manner [20]. This task description is referred to as the ”prompt” as it prompts the model to perform the desired task of ST from style A to style B, as illustrated in Figure 3. The prompt for ST consists of the source style, followed by a colon, the source style text, the target style, followed by another colon (Figure 4a). An example of the prompt is provided in Figure 4b, and our expectation from the model is to generate rewritten text in the target style, based on the source text.

[4]ZSL approach demonstrates superior performance compared to a much simpler and more straightforward approach, as shown in Appendices B.2, which involves transition through the BT text.

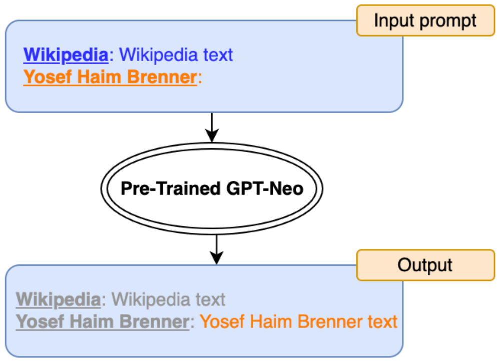
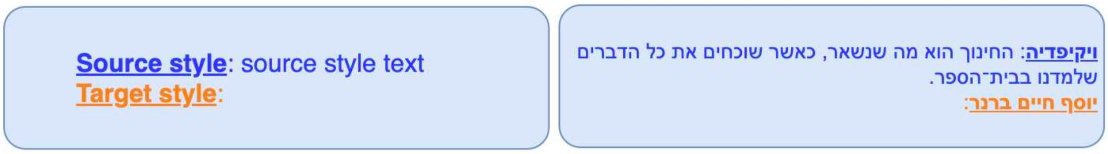


Figure 3: We provide a visual representation of the ZSL process utilized to achieve ST from Wikipedia to Yosef Haim Brenner's style, leveraging a pre-trained LM. The prompt description, integrated into the input sentence, enables the model to perform the task without the need for explicit training, thereby generating the desired output.



(a) Prompt structure

(b) Prompt example. Translated version:
Wikipedia: Education is what remains after one has forgotten what one has learned in school.
Yosef Haim Brenner:

Figure 4: Zero-shot ST prompt.

3. Evaluating ST Model

The evaluation of our ST model is based on the methodology proposed by [13]. In their study, the authors conducted a survey of 23 previously-published ST papers, identifying three common properties on which ST models are typically evaluated.
Given a styles set SS, a target style $j \in SS$ and an output s^j sentence:

1. **Semantic similarity (SIM)** - This property measures the extent to which the semantics of the input sentence s are preserved in s^j. In previous studies it usually done by using metrics like BLEU [16], since BLEU is based on n-gram precision, it aggressively penalizes lexical differences even when candidates might be synonymous with or similar to the reference: if an n-gram does not **exactly** match a sub-sequence of the reference, it receives no credit. An alternative metric to measure semantic similarity is SIMILE [21]. The similarity between s and s^j is obtained by encoding both sentences into vector representations and then calculating their cosine similarity. For this purpose, we trained an unsupervised Hebrew version of SimCSE [5]. Refer to Appendix C.2 for more details.
2. **Transfer accuracy (ACC)** - This property involves identifying the target style $j \in SS$ in

s^j, and requires a classifier for each $\forall i \in SS$ to identify i in s^i and report its accuracy on s^j. We fine-tuned a pre-trained LM, AlephBERT [18], for the classification task. Further details about the architecture and hyperparameters can be found in Appendix C.3.

3. **Fluency (FL)** - This property measures the fluency of the output sentence s^j, as ungrammatical outputs can still achieve high scores on both ACC and SIM, motivating the need for a separate measure. To measure fluency, we used LM perplexity[5] (PPL). For this purpose, we fine-tuned a pre-trained GPT-Neo LM [4]. The decision about the fluency of the output sentence is made based on a PPL threshold derived from PPL calculated on our corpus, where sentences below the threshold are considered fluent, and sentences above the threshold are considered non-fluent. Refer to Appendices C.4 for more details.

Aggregation of Metrics - So far, we have focused on individual implementations of ACC, SIM, and FL. After computing these metrics, it is useful to aggregate them into a single number to compare the overall ST quality across multiple ST model configurations. A good model should jointly optimize all metrics:

$$J(ACC, SIM, FL) = \sum_{x \in X} \frac{ACC(x) * SIM(x) * FL(x)}{|X|}$$

Where x is a sentence from a test corpus X. We treat ACC and FL at a sentence level as a **binary** judgement, ensuring incorrectly classified or disfluent sentences are automatically assigned a score of 0.

4. Experiment

We evaluate our model using methodology proposed by [13], which is described in detail in Section 3.

4.1. Evaluation Setup

The test corpus consists of 500 sentences of each style, this corpus was concealed from the model training process. Given a styles set SS and a source style $j \in SS$, we utilized our model to perform ST from $j \in SS$ to $\forall i \in SS - \{j\}$.
For each sentence pair comprising a sentence in the source style and a sentence in the target style, we calculated the three individual metrics, SIM, ACC and FL, as described in Section 3. Additionally, we calculated the main aggregated metric $J(ACC, SIM, FL)$ to evaluate the overall performance of the ST model.

4.2. Results

The results for each of the individual metrics, as well as the aggregated metric, are presented in Table 1.

The presented Figure 5 illustrates an error analysis that reveals a significant misclassification of samples generated by the ST model. These misclassifications occur when the classifier,

[5]Perplexity is a measure of how well a model fits the test data, low perplexity means better fit.

Table 1
The transfer accuracy (ACC) is determined as the accuracy of the classifier model, as explained in Appendix C.3. The semantic similarity (SIM) is calculated as the average score of the output from the Hebrew SimCSE model, as described in Appendix C.2. The fluency (FL) is calculated as the number of test samples with PPL score below a predefined threshold, using the Hebrew GPT-Neo LM, as detailed in Appendix C.4.

ACC	SIM	FL	$J(ACC, SIM, FL)$
0.63	0.53	0.91	**0.27**

Figure 5: The classifier confusion matrix after ST displays the distribution of classifier labels for sentences that have been transferred to the target style. Each row represents the label distribution for a particular target style (as indicated by the row label). The off-diagonal elements in the matrix reflect mis-classifications, which often occur due to intuitive domain similarities.

using a binary classification approach, identifies styles that share characteristics with the target style but are not actually the target style. This issue arises due to the calculation of the aggregated metric, $J(ACC, SIM, FL)$, which zeros out many sample scores despite their high SIM and FL scores. To address this issue, we propose a hierarchical classification approach. A class-hierarchy tree is constructed based on our domain knowledge, as shown in Figure 6 where the classification decision is made based on different levels of classification resolution.

Table 2
The results for each area in the class-hierarchy tree, as depicted in Figure 6, are as follows.

Number of classes	ACC	F1	$J(ACC, SIM, FL)$
6 (green area)	0.63	0.62	0.27
3 (blue area)	0.75	0.75	0.33
2 (red area)	0.81	0.78	0.35

The fine-grained resolution is the current classification approach, where the classes are very

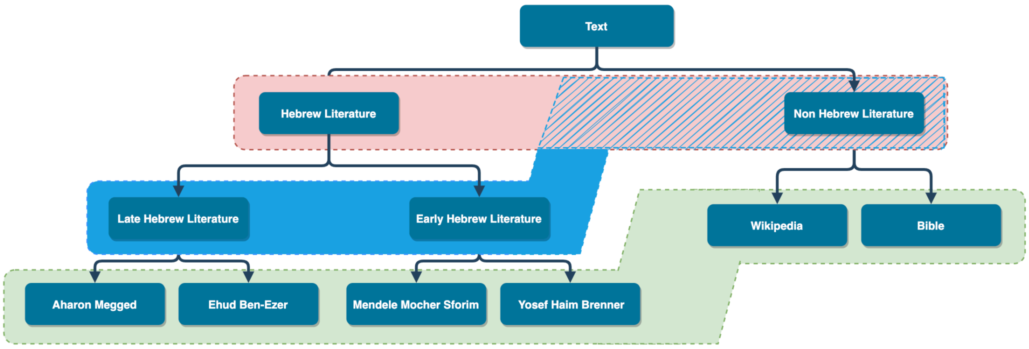
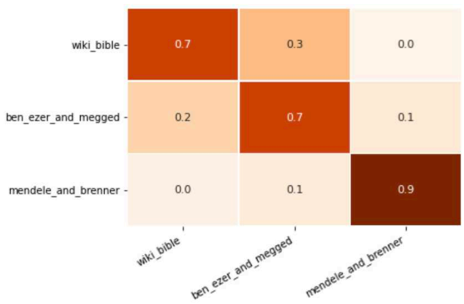
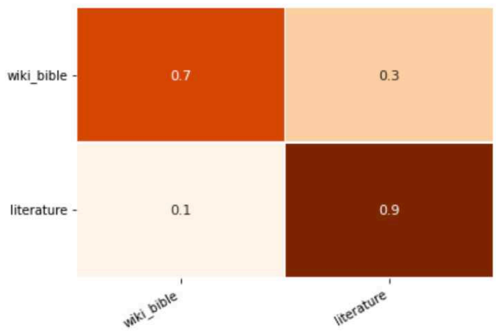


Figure 6: Class labels within the green area are classified using a straightforward flat classification approach, where each example is assigned to its final, leaf-level label. The error analysis of the leaf-level label classification is shown in Figure 5. The red area comprises a *Hebrew literature* class and a *non-Hebrew literature* class. The distinction between labels in the red and blue areas is that the blue area further differentiates between **early** and **late** Hebrew literature. The error analysis for this distinction is shown in Figures 7a and 7b. The classification results and aggregated metrics are presented in Table 2.



(a) Error analysis of levels 2 in the class-hierarchy tree.



(b) Error analysis of levels 1 in the class-hierarchy tree.

Figure 7: The classifier confusion matrix after applying ST (similar to Figure 5) shows an error analysis of levels 1 and 2 in the class-hierarchy tree, as depicted in Figure 6.

specific and detailed, denoted in the green area in Figure 6. The medium-grained resolution defines three classes that are broader than the fine-grained classes but still fairly detailed, denoted in the blue area in Figure 6. The classification results show an improvement in transfer accuracy, with 75% accuracy, as shown in Figure 7a. The coarse-grained resolution defines very broad and general classes, denoted in the red area in Figure 6. All the authors are gathered into a single class - Hebrew literature, and the reference styles are grouped into a non-Hebrew literature class. The classification results demonstrate a further improvement in transfer accuracy, with 81% accuracy, as shown in Figure 7b.

4.2.1. ST Examples

In the analysis of randomly selected test samples, it is often difficult for human readers to determine if the target style has been successfully incorporated into the generated output sentence. This difficulty may persist even when the samples achieve high evaluation metric scores. To provide a clearer illustration of the performance of our ST model, we present several outputs in Figure 8. In these examples, the input text was derived from Wikipedia and assigned a ”neutral” style, while the generated output was created in the style of each of the four authors. The output was examined by a domain-expert (i.e., literary scholar) who validated the results and determined whether they represent accurately the specific style of each author. Unfortunately, it is virtually impossible to provide an English translation of generated results (”poetry is what gets lost in translation”, as the poet Robert Frost wrote and the same holds true for literary style) yet the Hebrew results are striking and help us to better understand and describe the intricacies of each author’s unique style. These examples serve as concrete illustrations of the ST model’s capabilities, allowing literary scholars to gain insight into the unique characteristics and distinctions between various styles. Additional examples are available in Appendix D.

ויקיפדיה	אהוד בן עזר	מנדלי מוכר ספרים	אהרון מגד	יוסף חיים ברנר
מחשב הוא מכונה אלקטרונית המסוגלת לעבד נתונים על פי תוכנה, כלומר על פי רצף פקודות נתון מראש. מערכת מחשב כוללת את החומרה של המחשב, את הציוד ההיקפי הנלווה אליה, את מערכת ההפעלה המנהלת את פעולת המחשב ואת התוכנה המופעלת בו.	מחשב, המכונה גם מחשב אלקטרוני, המסוגל לעבד נתונים על-פי תכנת מחשבים. הרי המערכת היא כלי אלקטרוני שנועד לעיבוד כמויות נתונים על-פי תכנת מחשב, כלומר על-פי רצף פקודות נתונים נתונה מראש. המערכת מיועדת לשרת אנשים. המערכת היא גם מחשב בסיסי מאוד ותוכנה פשוטה שתפקוד אותו מספר קטן של גדולים. אחד התפקידים החשובים שלה הוא מערכת הפעלה.	כח, מחשב הוא. זו היא מלאכת עיבוד הנתונים, כמו שאר דברים. מחשבתו על פי סדר למחשבותיו. זאת אומרת על פי הסדר, לפי התלם מראש. מחשבתו זו היא כל כך חזקה, עד שאין הדעת נותנת שכך צריך להיות.	מחשב“, היא מערכת ממוחשבת המסוגלת לעשת חישוב של נתונים לפי תוכנה, זאת על פי פקודות מובנות מראש של קוד. מערכת ממוחשבת כוללת את החומרה של המחשב, ציוד העזר הצמוד לו, את המערכת הניהוג והפעלה שלה, את מערכת התוכנת תפעול של המערכת וכל דבר הכרוך בה. היא בעלת אופי מתוחכם. היא מהירה מאוד ומפותלת.	מחשב, הוא מכונה מדעית-כוללת המסוגלת לעיבוד נתונים על פי תכנה, שהרי תכונתה היא מהופכת מראש תכנית-מחשב, והיא מכונה מחשבתית הכוללת חומרה של מחשב, שלצרכים יש לה תכונתו, למערכת הפעלה המחסנת את התכנית ומחשבת. הואיל והתכונה היא כמו שצריך להיות.
החינוך הוא מה שנשאר, כאשר שוכחים את כל הדברים שלמדנו בבית־הספר.	- החינוך, היא מה שנשארו, כשנעצמו כל הדברים שלמדנו בבית-הספר.	חנוך, כי נחוץ, כדי שתהא ההשכלה בכל מקום. זה הדבר, שבאנו אל ביתנו ושכח כל אשר עמלנו בבית-הספר.	החינוך הוא מה שנותר, אם שכחו את כל הדברים שלמדנו בבית הספר.	החנוך, זה החלק, אשר לא תשכחו לעולם!.. החינוך היא החלק היותר חשוב, כאשר שכחת את כל אלה, כי בית הספר, אשר ביתכם, אתם הייתם להם המורים והמורות; כי הוא היה בית-ספר של רוח!. כל דבר אחר אין בית הספר צריך להשכיל.
הדבר שהכי חשוב לאיש מדע, זה לא התעודות, לא מספר שנות הלימוד, וגם לא הניסיון שלו, אלא האינטואיציה שלו	הדבר, שאיש מדע לא מתעסק בו. מה שחשוב ביותר בעיני מדעאי, זה לא התעודות, לא מספר שנות הלימוד ולא נסיונות-הניחוש שלו, אבל חושיו הוא-הוא החוש. החושים שלנו חכמים מן המוח שלנו, החושים ים שלנו. אנחנו מבינים אותם וחושים דרכם הם.	דבר, שהרי לא כל אדם נבון וחכם הוא. כלום יש הבדל, שכל אנשי-מדע אינם יודעים? כלום יש הבדל, שכל העוסקים במדע הם שונאי אנשים חמרים-חמרים... ומי יתן ויהיו אנשי-רוח כמונו יודעים את לבבם הטוב ומשתדלים להבין אותו?	זה שהכי חשוב לאיש המדע, ש"הסטודנט" לא לומד בכלל, ואולי גם לא שום נסיון, כי אם אינטואיציה, או אפילו כנות.	הדבר, כי העיקר הוא לאדם המדע, לא תעודות-העתק, גם לא נסיונו זה, אלא האינטואיציות שלו.
התפרס חזבאללה ברחבי דרום לבנון ועל הגבול הבינלאומי, וביסס בה את שליטתו, בבניית ביצורים וצבירת נשק רב, בין היתר טילי נ"ט וארטילריה רקטית ארוכת-טווח, שנועדה לטענתו לקיים מאזן אימה שימנע מצה"ל לפלוש שנית ללבנון, אך למעשה שימשה בידיו כגיבוי בפעולותיו נגד ישראל.	חיזבאללותה, ובה תיפקד היטב. חיזבאלללה היתה רחבה וכובשת לב והיא שלטה בה על רבד של אמצעי לחימה רב ערך, גם טילים מונחי דלק ונשק מונחה, שכנגד טענותיו כי צה"ל אינו מסוגל להתמודד עם הסורים, אך בעצם היא היתה עבורו עוגן למבצעיות נגד מדינת ישראל. היא פעלה רבות בשריון ובפלדותה וגם ים.	שם, ו"להעמידם" עליהם. התמסות של חיזבאללה בדרום לבנון ובשטח הגבול הבינלאומי, וביססה עליה שליטתה, במצודות ובמחנות אימונים ובמצור ובשביה גדולה מאד, כדברי רבי שמעון בן יוחאי, שלא נסתייע לו במלחמתו נגד ישראל. גם בימי הסוכות נהגו אנשי חיל לבנונים להשתמש בתגבורת שהציבו לעצמם כדי להכניעה ולאספקה לבניה מחדש של האומה בארץ.	התפרס, חיזבאללה כבש שטחים נרחבים בתוך דרום לבנון וגדר אותם בחבלי-תיאום ובמבואות עמוקים. על גבולם הוקמו הביצורים והצבתות הצטייד. הוא דחה ככל האפשר את הטענות בדבר מפלתנו, שכן כוונתו היתה לקיים מאזן אימה שימנע מצבא צה"ל לפרוץ ללבנון.	זבב, וכשניטל הביצורים, התחכמו גדודי החזבאללה לכבוש את הדרום כולו ולהכניע בו את מדינתם ולבצר בה את גבורתם הצבאית, בין היתר, כלי-לחימה ונשק ארטילרי ארוכי-הטווח, שנועד לו מטעם ממשלתו להזהיר מפני שפלישתו הצפויה של צבא-הגליל.
אהבה מאופיינת בחיבה, חיפוש אחר קרבת הנאהב, דאגה לו ובקשת טובתו. האהבה יכולה גם להתבטא בצורה רכושנית ולפעמים אף אובססיבית כלפי הנאהב. אהבה לרוב אינה כפופה לרצונו של האוהב או של הנאהב. כאשר האהבה אינה ממומשת, היא יכולה להיות גם מכאיבה ולא נעימה. טרגדיות רבות, כמו גם מחזות, שירים וסיפורים רבים, עוסקים באהבות בלתי ממומשות ותוצאותיהן.	אהבה, אף כי לרוב אין האהבה אלא חלק ממנה. אהבה יכולה להתבטא הן ביחס מר וקשיח והן בעניין טובתו. האהבה עלולה להתבטא בדרך הרכוש והן בהתערטלות כלפי אהובה או להיפך. אהבה לרוב איננה כובלת את רצונותיו של האדם או את האהוב. אהבתם בדרך-כלל לא תלויה ברצונו של אדם. כאשר האהבה לא ממומשת עלולה להיפלט גם המרה-החמה.	אהבה, שהיא מתענגת על חיבה, חיפש קרבת-הנאהבים, דאג לו ובקשו לטובתה. אהבה הרי יכולה להתפרש בדרך רחוקה ולפעמים אובנת גם כלפי אהבה. אהבה יתירה – אף זו אינה תלויה ברצונו של האהוב או של אהובה. כשאהבה אינה מתנתקת אפשר שתהא מסיגה היא גם מכוערה ואינה נעימה. טרגדיות הרבה, כשטירוף-הדעת אינו מביא לידי-כך תוצאות, הרי הן תולדה של צרות רבות.	אהבתה, שהיא רגש עז המצית בקרבה אש של תענוג. האהבה יכולה גם להתבטאה בניכור ולהטעות את אוהביה. אהבה כמעט שאינה נובעת מרוטינתו הטבעית של איש, מן האהוב או את אהובה. כאשר האהבה אינה ממומשת, הרי יש בה גם איזו טינה והינתקות ממנה. אבל לפעמים טינה יכולה להיות ידה ע"י כך שהתעניינות בה תעמיק ותתעורר רגשות כמו תשוקה ואהבה. התאווה היא הדבר החשוב ביותר בחיים. האהבה ממלאת את כל הכוחות של האדם והיא זו שתמלא אותנו עם רצון ותחושת העונג שלה.	אהבתי, אך האהבה הרי היא רגש-אהבה. האהבה עשויה להתפרש כמעט לכל פרטיה; היא מתרגשת על ידי יחס רחשי, הדואג לו ובקש ממנו טובתו. אהבה יכולה להתבטא בכבדותה ופעמים אינה נוגעת כלל אל אהובה. אהבה עלולה לקרוע את לבי-יודע-כל.

Figure 8: The input text to the ST model is denoted by the blue color, while the green color represents the corresponding output text generated by the model, each reflecting a possible style. Additional instances of generated text can be found in Appendix D.

5. Conclusion

The development of Hebrew NLP models is difficult due to the language's unique characteristics. Hebrew is a Semitic language with complex morphology and a rich lexicon that has evolved over centuries of usage. We offer a novel generative ST model for modern Hebrew literary texts, employing a straightforward methodology. Firstly, we generate a pseudo-parallel corpus for text simplification, which serves as the basis for training the model. Secondly, we fine-tune a pre-existing Hebrew LM. Finally, we utilize a ZSL approach to enable the model to perform ST. In addition, our methodology yields a solution for the Hebrew paraphrase generation task, which is a sub-task of the ST process. This approach presents a promising avenue for future research in the field of ST in languages with limited resources. Furthermore, the ability to generate new texts based on their literary and linguistic styles provides a powerful tool for literary scholars. Applying these methods, whether on existing "neutral" texts or on a "prompt", provides an unusual perspective on individual author's style and more generally on the very notion of style, its characteristics and building blocks.

References

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, 2019, pp. 4171–4186. DOI: 10.18653/v1/N19-1423. URL: https://aclanthology.org/N19-1423.

[2] A. Fan, S. Bhosale, H. Schwenk, Z. Ma, A. El-Kishky, S. Goyal, M. Baines, O. Celebi, G. Wenzek, V. Chaudhary, N. Goyal, T. Birch, V. Liptchinsky, S. Edunov, E. Grave, M. Auli, and A. Joulin. "Beyond English-Centric Multilingual Machine Translation". In: *arXiv e-prints*, arXiv:2010.11125 (2020), arXiv:2010.11125. arXiv: 2010.11125 [cs.CL].

[3] A. Fan, M. Lewis, and Y. Dauphin. "Hierarchical Neural Story Generation". In: *arXiv e-prints*, arXiv:1805.04833 (2018), arXiv:1805.04833. DOI: 10.48550/arXiv.1805.04833. arXiv: 1805.04833 [cs.CL].

[4] L. Gao, S. Biderman, S. Black, L. Golding, T. Hoppe, C. Foster, J. Phang, H. He, A. Thite, N. Nabeshima, S. Presser, and C. Leahy. "The Pile: An 800GB Dataset of Diverse Text for Language Modeling". In: *arXiv e-prints*, arXiv:2101.00027 (2020), arXiv:2101.00027. arXiv: 2101.00027 [cs.CL].

[5] T. Gao, X. Yao, and D. Chen. "SimCSE: Simple Contrastive Learning of Sentence Embeddings". In: *arXiv e-prints*, arXiv:2104.08821 (2021), arXiv:2104.08821. arXiv: 2104.08821 [cs.CL].

[6] L. A. Gatys, A. S. Ecker, and M. Bethge. "A Neural Algorithm of Artistic Style". In: *arXiv e-prints*, arXiv:1508.06576 (2015), arXiv:1508.06576. arXiv: 1508.06576 [cs.CV]. URL: https://ui.adsabs.harvard.edu/abs/2015arXiv150806576G.

[7] J. Han, M. Kamber, and J. Pei. “2 - Getting to Know Your Data”. In: *Data Mining (Third Edition)*. Ed. by J. Han, M. Kamber, and J. Pei. Third Edition. The Morgan Kaufmann Series in Data Management Systems. Boston: Morgan Kaufmann, 2012, pp. 39–82. DOI: https://doi.org/10.1016/B978-0-12-381479-1.00002-2. URL: https://www.sciencedirect.com/science/article/pii/B9780123814791000022.

[8] A. Holtzman, J. Buys, L. Du, M. Forbes, and Y. Choi. “The Curious Case of Neural Text Degeneration”. In: *arXiv e-prints*, arXiv:1904.09751 (2019), arXiv:1904.09751. DOI: 10.48550/arXiv.1904.09751. arXiv: 1904.09751 [cs.CL].

[9] Z. Hu, Z. Yang, X. Liang, R. Salakhutdinov, and E. P. Xing. “Toward Controlled Generation of Text”. In: *Proceedings of the 34th International Conference on Machine Learning*. Ed. by D. Precup and Y. W. Teh. Vol. 70. Proceedings of Machine Learning Research. Pmlr, 2017, pp. 1587–1596. URL: https://proceedings.mlr.press/v70/hu17e.html.

[10] S. Jang. *Writing Style Conversion using Neural Machine Translation*. 2017. URL: https://api.semanticscholar.org/CorpusID:20560711.

[11] M. Johnson, M. Schuster, Q. V. Le, M. Krikun, Y. Wu, Z. Chen, N. Thorat, F. Viégas, M. Wattenberg, G. Corrado, M. Hughes, and J. Dean. “Google’s Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation”. In: *arXiv e-prints*, arXiv:1611.04558 (2016), arXiv:1611.04558. arXiv: 1611.04558 [cs.CL].

[12] D. P. Kingma and J. Ba. “Adam: A Method for Stochastic Optimization”. In: *arXiv e-prints*, arXiv:1412.6980 (2014), arXiv:1412.6980. DOI: 10.48550/arXiv.1412.6980. arXiv: 1412.6980 [cs.LG].

[13] K. Krishna, J. Wieting, and M. Iyyer. “Reformulating Unsupervised Style Transfer as Paraphrase Generation”. In: *arXiv e-prints*, arXiv:2010.05700 (2020), arXiv:2010.05700. arXiv: 2010.05700 [cs.CL]. URL: https://ui.adsabs.harvard.edu/abs/2020arXiv201005700K.

[14] P. J. Liu, Y.-A. Chung, and J. Ren. “SummAE: Zero-Shot Abstractive Text Summarization using Length-Agnostic Auto-Encoders”. In: *arXiv e-prints*, arXiv:1910.00998 (2019), arXiv:1910.00998. arXiv: 1910.00998 [cs.CL].

[15] Mila. *Properties of Hebrew Sentences*. 2009. URL: %7Bhttp://yeda.cs.technion.ac.il:8088/segmentation/XMLtokenizerHelp.html%7D.

[16] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu. “BLEU: A Method for Automatic Evaluation of Machine Translation”. In: *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*. Acl ’02. Philadelphia, Pennsylvania: Association for Computational Linguistics, 2002, pp. 311–318. DOI: 10.3115/1073083.1073135. URL: https://doi.org/10.3115/1073083.1073135.

[17] S. Schoch, W. Du, and Y. Ji. “Contextualizing Variation in Text Style Transfer Datasets”. In: *Proceedings of the 14th International Conference on Natural Language Generation*. Aberdeen, Scotland, UK: Association for Computational Linguistics, 2021, pp. 226–239. URL: https://aclanthology.org/2021.inlg-1.22.

[18] A. Seker, E. Bandel, D. Bareket, I. Brusilovsky, R. Shaked Greenfeld, and R. Tsarfaty. “AlephBERT:A Hebrew Large Pre-Trained Language Model to Start-off your Hebrew NLP Application With”. In: *arXiv e-prints*, arXiv:2104.04052 (2021), arXiv:2104.04052. arXiv: 2104.04052 [cs.CL]. URL: https://ui.adsabs.harvard.edu/abs/2021arXiv210404052S.

[19] A. Sudhakar, B. Upadhyay, and A. Maheswaran. ““Transforming” Delete, Retrieve, Generate Approach for Controlled Text Style Transfer”. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics, 2019, pp. 3269–3279. DOI: 10.18653/v1/D19-1322. URL: https://aclanthology.org/D19-1322.

[20] S. Suteera. *Prompt Engineering and Zero-Shot/Few-Shot Learning*. 2022. URL: https://www.inovex.de/de/blog/prompt-engineering-guide/.

[21] J. Wieting, T. Berg-Kirkpatrick, K. Gimpel, and G. Neubig. “Beyond BLEU:Training Neural Machine Translation with Semantic Similarity”. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, 2019, pp. 4344–4355. DOI: 10.18653/v1/P19-1427. URL: https://aclanthology.org/P19-1427.

[22] J. Yaniv, Y. Newman, and A. Shamir. “The Face of Art: Landmark Detection and Geometric Style in Portraits”. In: *ACM Trans. Graph.* 38.4 (2019). DOI: 10.1145/3306346.3322984. URL: https://doi.org/10.1145/3306346.3322984.

Appendices

A. Dataset

The present corpus is composed of 35 novels (as shown in Figure 9) authored by four authors: Aharon Megged (1920-2016),[6] Ehud Ben-Ezer (1936-),[7] Mendele Mocher Sforim (1836-1917) [8] and Yosef Haim Brenner (1881-1921), [9] who exhibit unique literary and linguistic styles. Moreover, these authors represent different stages in the development of Modern Hebrew literature from the mid nineteenth century onward, hence the development of their style is a key component in the development of Modern Hebrew literature. The purpose of including two additional reference styles, namely Wikipedia articles and the Biblical text, was twofold: first, to establish a hierarchy of writing styles (depicted in Figure 6), and second, to introduce a ”neutral” style based on the Wikipedia articles, which can be flexibly adapted to each of the authors’ styles. To achieve this aim, the Wikipedia stylistic references were leveraged to facilitate a seamless transition from the modern Hebrew ”neutral” style to each author’s unique style, as outlined in Table 8.

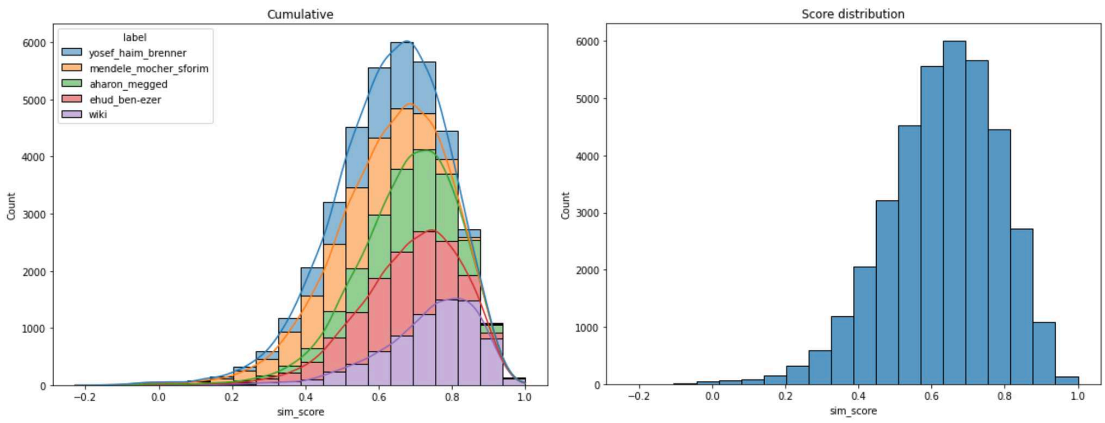
[6] https://en.wikipedia.org/wiki/Aharon_Megged
[7] https://www.ithl.org.il/page_13417
[8] https://en.wikipedia.org/wiki/Mendele_Mocher_Sforim
[9] https://en.wikipedia.org/wiki/Yosef_Haim_Brenner

novels	authors
1. מסביב לנקודה (Around the Point) 2. בחורף (In Winter) 3. שכול וכשלון (Breakdown and Bereavement) 4. פעמיים (Twice) 5. הירושלמי (The Jerusalamite) 6. מן הזווית (From the Corner) 7. שנה אחת (One Year) 8. בימות החמה (In Summer) 9. מכאן ומכאן (From Here and There) 10. מן המיצר (In Distress) 11. עצבים (Nerves)	Yosef Haim Brenner
1. האבות והבנים (Fathers and Sons) 2. בעמק הבכא (In The Valley of Baka) 3. ספר הקבצנים (The Book of Beggars) 4. ספר הבהמות (The Book of Animals) 5. מסעות בנימין השלישי (TheTravels and Adventures of Banjamin the Third) 6. סוסתי (My Horse) 7. מספר הזכרונות (From the Book of Memories) 8. בימי הרעש (In Days of Tumult) 9. שם ויפת בעגלה (Shem and Yefet in a Carriage) 10. לא נחת ביעקוב (No Peace in Israel) 11. בימים ההם (In those days) 12. בסתר רעם (Out of a Thundercloud)	Mendele Mocher Sforim
1. עשהאל (Asahel) 2. געגועים לאולגה (Longing for Olga) 3. הגמל המעופף ודבשת הזהב (The Flying Camel and the golden Hump) 4. היינץ ובנו והרוח הרעה (Heinz, his son and the evel spirit) 5. זבובים (Flies) 6. מיליזילדה היפה (Millizilda the Beautiful)	Aharon Megged
1. אנשי סדום (People of Sodom) 2. בארץ עצלתיים או ספר האופטימיות 3. השקט הנפשי (Peace of Mind) 4. ספר הגעגועים (The Book of Longings) 5. אוצר הבאר הראשון (The Treasure of the First Well) 6. פרשים על הירקון (Riders on the Yarkon River)	Ehud Ben-Ezer

Figure 9: The literary works that comprise the corpus.

The creation of the corpus involved three main steps: (1) Sentence Splitting - the problem of properly locating sentence boundaries in Hebrew text is in many ways less severe than the same problem in English. Properties of Hebrew sentences [15] given by the *'Academy of the Hebrew Language'* makes it relatively easy to identify end of sentence. Each book was decomposed into a sequence of sentences, with an average length of 39 tokens. (2) BT - For the BT task, we employed M2M100 [2], a multilingual seq-to-seq model specifically trained for Many-to-Many multilingual translation. Initially, we selected Arabic as an intermediate language due to its syntactic similarity with Hebrew; however, the BT performance was subpar, prompting us to switch to English as the intermediate language. (3) Semantic Similarity - To assess the quality of the BT process, we utilized an unsupervised Hebrew version of SimCSE [5]. The technical information of this model is provided in Appendix C.2.

The filtering process consisted of several steps. Firstly, we removed sentences that were too short or too long, only retaining sentence pairs with an average token length of 54 tokens, with a range between 20 and 120 tokens. Secondly, given that it is not uncommon to find non-Hebrew text embedded in Hebrew literature, particularly due to the European roots of some authors, we filtered out all such sentences from the corpus. Thirdly, to enhance diversity and prevent copying, we removed back translated sentence pairs with semantic similarity scores lower than 0.4 or higher than 0.95, as described in Figure 10. Fourthly, we removed diacritical



(a) Novels per authors.

(b) The distribution of examples (sentences) across the classes (authors).

Figure 10: Sim score.

marks (niqqud) from sentences in the corpus to ensure corpus uniformity. Ultimately, The final training corpus was composed of 37,000 sentence pairs. The distribution of examples, or sentences, across the different classes is presented in Figure 11.

B. Additional Experiments

B.1. Stylistic Prompts for Controlled Text Generation

The initial stage of our research involved developing a framework for style identification and consistent text generation based on specific stylistic prompts. To accomplish this, we utilized a classifier model as outlined in Appendix C.3 to accurately identify and distinguish between different writing styles. Subsequently, we fine-tuned a Hebrew-based GPT-Neo-small[10] model for 2 epochs with a minibatch size of 4 and a learning rate of 5e-2 to generate text according to a prompted style.

During the training process, the style label (represented by the author name) and corresponding text sequence were concatenated together using a separator token, as illustrated in Figure 13a. For text generation, the model was prompted to generate text in a specific style by providing the model with the style (author name) separator token and a random seed token, as shown in Figure 13b. The generated text (examples of which are presented in Figure 12) was then evaluated by our classifier, which yielded an F1 score of **0.94**, indicating a high degree of similarity with the classification of real data (corpus text).

In order to ensure the originality and novelty of the generated literary text, we conducted an analysis of *n*-gram intersections between the training corpus and a corpus of generated literary text, as presented in Table 3. The purpose of this analysis was to ascertain whether the generated text contained any copied content from the training corpus. Our findings indicate a

[10]https://huggingface.co/Norod78/hebrew-gpt_neo-small

Table 3
The results of the n-gram intersection analysis between the training corpus and a generated literature text corpus assess the novelty and originality of the generated text. The table presents the number of intersecting n-grams for various n sizes and the percentage of the generated n-grams found in the training corpus ($\frac{\#of\_intersection\_n-gram}{\#of\_generated\_n-gram}$).

n	intersection n-grams number	intersection percentage
2	6,663	46%
3	3,515	17%
4	966	4.4%
5	232	1%
6	55	0.26%
7	13	0.06%
8	2	0.02%
9	0	0%

negative linear relationship between n and n-gram intersection, which serves as an indicator of the novelty of the generated text. Specifically, the results demonstrate that the generated text is indeed novel, and does not contain any copied content from the training corpus. Moreover, our attempts to perform ST using a similar method to that described in Section 2.1 were met with unsatisfactory results, indicating limitations in the model's ability to perform this task.

B.2. Transition Through the Back Translated Text

The employment of an intermediary style, such as the back translated text, manifests itself as the most instinctive and straightforward strategy for performing ST in our case. We seek to perform ST from style A to style B through a two-step process, which involves transitioning from style A to the back translated text and subsequently to style B, as visually represented in Figure 14. Unfortunately, the results obtained from this methodology were markedly unsatisfactory.

C. Model Details

All the models were trained on a single NVIDIA V100 tensor core GPU on the Google Colab[11] platform using HuggingFace's [**HuggingFace's**] programming API and a transfer learning method. Transfer learning is a technique in which a pre-trained model is used to enhance the performance of a new model on a related task. This approach saves time and resources, as the pre-trained model serves as a starting point for the new model, enabling it to learn from the pre-existing knowledge. In NLP, transfer learning is implemented via Transformers.

[11]Google colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to ML, data analysis and education. More technically, Colab is a hosted Jupyter notebook service that requires no setup to use, while providing access free of charge to computing resources including GPUs.

C.1. ST Model

This section provides a detailed description of the ST model used in our study. We fine-tune a pre-trained Hebrew GPT-Neo-small[12] LM for 2 epochs. GPT-Neo is an open-source version of the SOTA GPT-3 LM model developed by OpenAI. However, as GPT-3 has not yet been open-sourced, the open-source community has attempted to reproduce its weights and results. One such attempt is the GPT-Neo model developed by eleuther.ai, which has a similar architecture to GPT-3. Hugging Face [**HuggingFace's**] further extended this effort by integrating GPT-Neo into their transformers infrastructure, making it accessible to the NLP community.

In the case of the Hebrew language, two GPT-Neo models are available - GPT-Neo-small and GPT-Neo-xl[13]. Due to limited computational resources, we used the GPT-Neo-small version. We employed the Adam [12] optimizer with a polynomial schedule[14] that includes a warmup period, during which the learning rate increases linearly from 0 to the initial learning rate set in the optimizer, which is 5e-2 in our case. The learning rate then decays as a polynomial function to the end learning rate of 5e-4. We used a mini-batch size of 4 sentences.

For text generation, we employed the top-k [3] and top-p [8] sampling strategies. More specifically, we sampled from the top K tokens, where K refers to the most likely tokens (in our case, we set K to be 50), with a cumulative probability that exceeds P (in our case, we set P to be 0.95).

C.2. SIM Model Details

SimCSE, Figure 15, is a SOTA unsupervised model for learning sentence embeddings. The idea is to encode the same sentence twice with pre-trained transformer based encoder model, AlephBERT [18] model in our case. Due to the used dropout in transformer based models, both sentence embeddings will be at slightly different positions. The distance between these two embeddings will be minized, while the distance to other embeddings of the other sentences in the same batch will be maximized (they serve as negative examples)[15]. The model was trained on our corpus (Appendix A), employing Mean-pooling and cosine-similarity [16] as the similarity metric.

C.3. Classifier Model Details

We employed the AlephBERT [18] model for the task of stylistic classification. AlephBERT is a pre-trained, Transformer-based, large language model specifically designed for Modern Hebrew. This model is trained on a larger corpus with a larger vocabulary and is based on

[12]https://huggingface.co/Norod78/hebrew-gpt_neo-small

[13]https://huggingface.co/Norod78/hebrew-gpt_neo-xl

[14]huggingface - polynomial decay schedule with warmup

[15]from https://www.sbert.net/examples/unsupervised_learning/SimCSE/README.html

[16]Cosine similarity measures the similarity between two vectors of an inner product space. It is measured by the cosine of the angle between two vectors and determines whether two vectors are pointing in roughly the same direction [7]

$$CosineSimilarity(\vec{x}, \vec{y}) = \frac{|\vec{x} \cdot \vec{y}|}{||\vec{x}||||\vec{y}||}$$

the Bidirectional Encoder Representations for Transformers (BERT) architecture introduced by [1]. The results obtained by [18] demonstrate that AlephBERT outperforms previous SOTA models on various Hebrew NLP tasks, including Segmentation, Part of Speech Tagging, full Morphological Tagging, Named-Entity Recognition, and Sentiment Analysis.

To fine-tune[17] AlephBERT for the task of stylistic classification, we followed the BERT authors' recommendations and trained the model for 2 epochs with a learning rate of 2e-5 and a batch size of 4. We employed the Adam [12] optimizer with a linear schedule and a warm-up phase. Specifically, we gradually increased the learning rate from 0 to the initial learning rate set in the optimizer during the warm-up[18] phase and then linearly decreased it to 5e-4. The results of our experiment demonstrate the effectiveness of AlephBERT for the task of stylistic classification.

C.4. Fluency Model Details

We utilized the GPT-Neo-small LM developed by [4] and fine-tuned it for grammatical acceptability judgments task using the LM PPL metric. PPL is a measure of the exponentiated average negative log-likelihood (NLL) of a given sequence:

$$PPL(W) = exp\left(\frac{1}{N}\Sigma_i^N NLL(W_i)\right) = exp\left(-\frac{1}{N}\Sigma_i^N log(P(w_i|w_{<i})\right)$$

W is a tokenized sentence, W contains sequence of tokens ($W = w_1, ..., w_N$) and P is the conditional probability constructed by the LM. To ensure that the model has access to a maximum amount of contextual information, we evaluated PPL using a sliding-window approach, as depicted in Figure 16. This methodology entails repeatedly shifting the context window so that the model can have sufficient contextual information when generating each prediction.

We trained the model for 3 epochs with a learning rate of 5e-2, using a linear schedule with warmup[19] that gradually decreases the learning rate from the initial value set in the optimizer to 5e-5 after a warmup period during which it increases linearly from 0 to the initial learning rate.

The distribution of the PPL values is shown in Figure 17, and the statistical properties of these values are presented in Table 4. We established a threshold of 100 to distinguish between fluent and disfluent sentences, as the majority of PPL values in our corpus corresponded to fluent sentences. Sentences with PPL values lower than 100 were considered fluent, while those with higher values were deemed disfluent.

D. More Example Generations

Additional examples are provided in Figure 18.

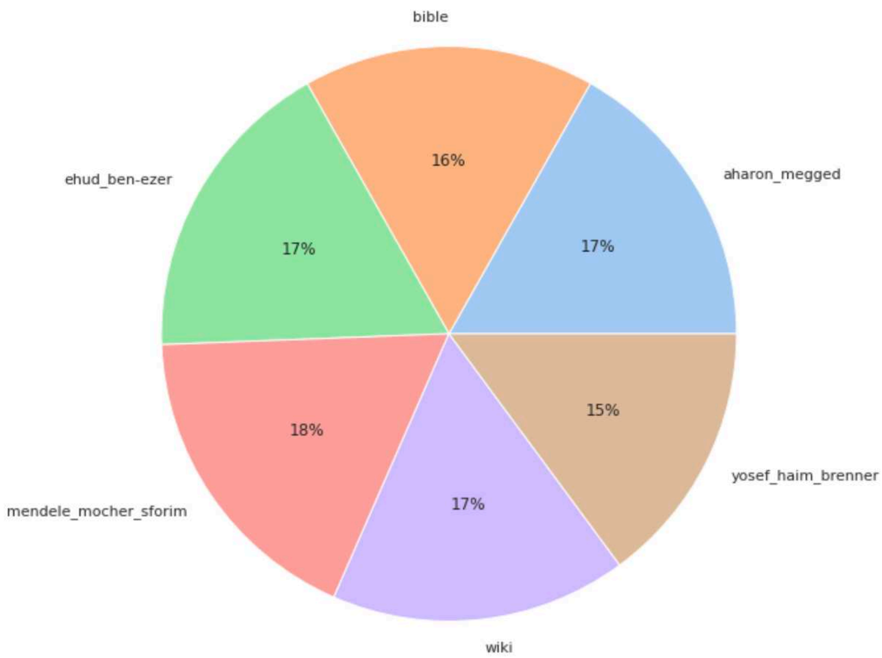
[17] In NLP, fine-tuning is the process of adapting a pre-trained LM to a specific task by updating its weights with new data related to the task. This process involves training the model on a smaller dataset to refine its parameters and incorporate the characteristics of the domain data. By doing so, the model can leverage its pre-existing knowledge and achieve better performance on the given task.

[18] huggingface - linear schedule with warmup

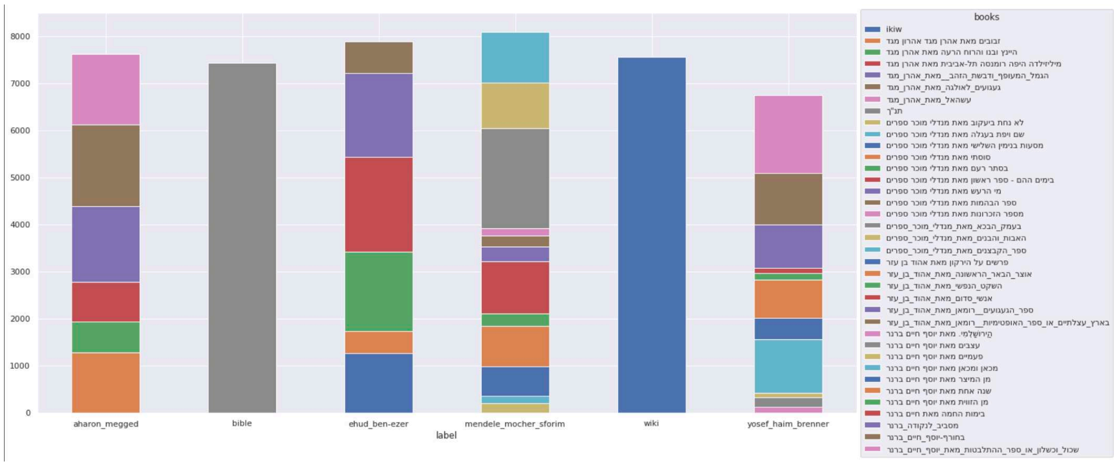
[19] huggingface - linear schedule with warmup

Table 4
Statistical information regarding the PPL values that were calculated for the sentences in our corpus.

Mean	Std	Min	Max	75%
31.5	18.4	1.7	680.8	39.2



(a) The corpus classes, or styles, distribution.

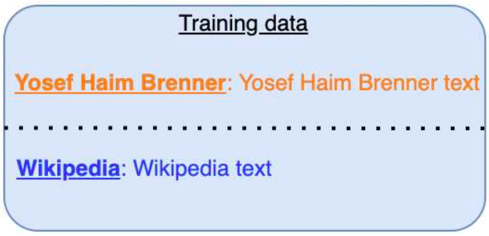


(b) Each bar in the chart corresponds to a different style, and the internal structure of each bar portrays the relative proportion of each book associated with the particular author style.

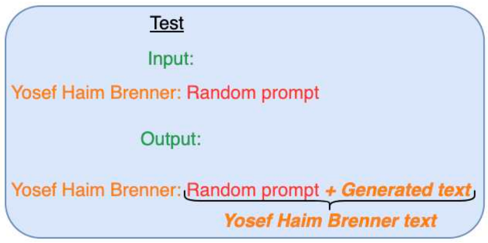
Figure 11: Corpus description.

generated test	author
משמעות החיים היא שאין לאדם עצמו אלא כל אדם באשר הוא. אלא שהטוב כשלעצמו אינו שלם. יש טוב אינסופי, אבל הוא שלם. האם אין האדם יכול להיות שלם רק בגלל שיש לו בחירה חופשית? האם ניתן לומר שזה רע מפני שאין בחירה חופשית?	אהרון מגד
משמעות החיים היא שלוויתם הנאה מכל דבר הואיל ואין הדבר אלא התוודעות של האדם לנבכי נפשו.	אהוד בן עזר
מעות החיים היא כל-כך עמוקה – שאין הדעת יכולה להבין אותה, וכל מה שהיא מבקשת למצוא היא עוקרת ממקומה.	מנדלי מוכר ספרים
משמעות החיים היא: על כל אדם לאהוב, לאהוב את נפשו, לשקוד על קיומו ועל חפצו, להיות תמיד במצב של שחרות, להיות איש אהוב ורצוי, להיות נאמן לאמת הפנימית.	יוסף חיים ברנר
משמעות החיים היא ערך מוסף. משמעות החיים היא ערך המוסף שלנו כבני אדם – הערך המוסף האמיתי שלנו כבני אדם הוא לא רק במה שקורה לנו אלא גם במה שעובר עלינו בכל רגע ורגע. משמעות החיים זה בעצם מה שמתחיל ונגמר. את זה כולנו יודעים ויודעים. אבל מהי מהות החיים?	וויקיפדיה

Figure 12: The initial component (text in green) of the sentence presented to the model serves as the prompt for the text generation process, along with the corresponding author name. The subsequent component (text in blue) represents the text generated by the model.



(a) This figure illustrates the concatenation process used during the training phase of our model. The process involves combining a style label and a text sequence with a separator token. This concatenated sequence is then utilized to train the model to generate text in a specific style.



(b) This figure illustrates the text generation process in our model, which utilizes a style label and a random seed token. The figure shows the concatenation of the style label and a separator token, followed by the random seed token. This concatenated sequence serves as a prompt for the model to generate text in the desired style.

Figure 13: Here, we describe the training process, which involves assigning a style label to each sentence in the corpus and utilizing a style label and random seed token during the text generation process.

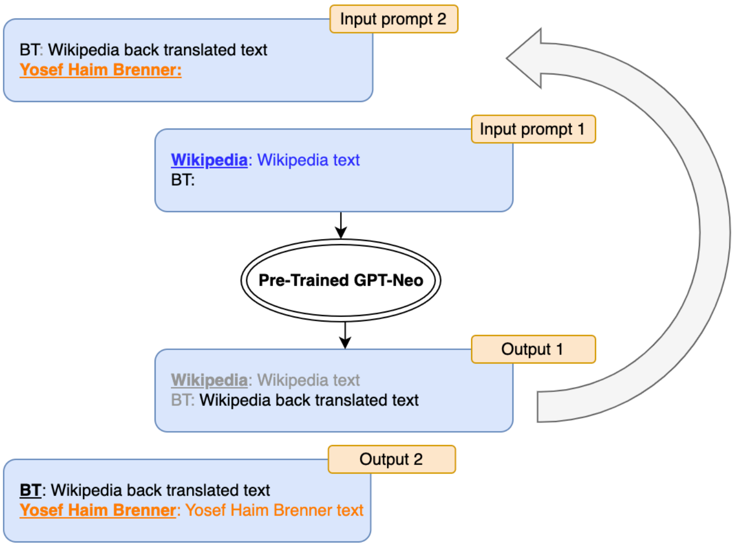


Figure 14: ST in two steps.

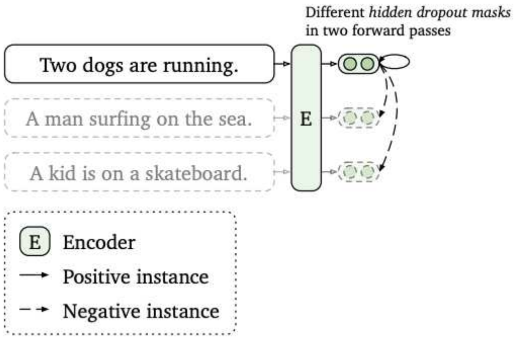


Figure 15: This figure, taken from the work of [5], describes the unsupervised SimCSE. Specifically, the unsupervised SimCSE approach predicts the input sentence itself from in-batch negatives while employing different hidden dropout masks.

Figure 16: Sliding window strategy for calculating the PPL metric.

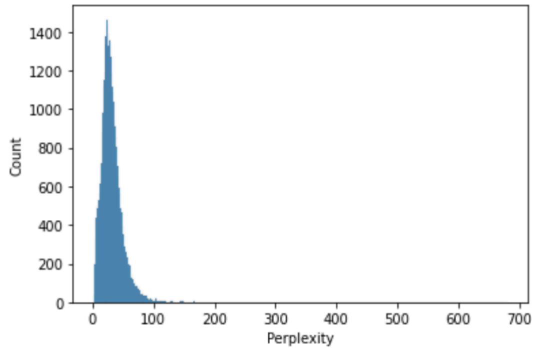


Figure 17: The distribution of PPL values for sentences in our corpus.

Input → output	direction
מוזיקה היא אמנות סידור צלילים בזמן. מרכיביה העיקריים הם גובה צליל (האחראי על מלודיה והרמוניה), קצב (מפעם, משקל וארטיקולציה), דינמיקה, גוון ומרקם. היצירה, הסגנון, החשיבות ואף הגדרת המוזיקה עצמה רגישה למשתנים תרבותיים מורכבים ולהקשרים חברתיים. המוזיקה כוללת יצירות מסוגים שונים, מיצירות מאורגנות בקפידה, דרך אלתור מוזיקלי ועד צורות אקראיות לחלוטין. המוזיקה, שהיא גם אומנות הסידור של הצלילים בעת. חלקיה העיקריים הם גובה הצליל (המחויב למלודיה וברור ההברה) וצורת ההתהבעות והטקסטורה. היצירה, סגנון, חיוניות וקונצרטיות. יצירתה, סגנון, חשיבות ואף הגדרה מדויקת של המוזיקה עצמה. מוזיקה היא גם אמנות הקשב והתפעלות. היא מכילה סוגות שונות, מהיצירות המאורגנות בעמלנות, דרך אלתורים מוזיקליים ועד ליצירות אקלקטיות.	ויקיפדיה←אהרון מגד
בשעה זו, שהייתי בסכנה להפטר מן העולם, נתחבב עלי העולם וכל בריותיו, והייתי מתאמץ לפקוח את עיני ולראותו עוד הפעם. בשעתה, כשהייתי בצרה גדולה לנפשי ולנשמתי, היה מחבבי עולם ומלואו ומתקנא בשכינני-העולם וכל הבריות, ואהיה משתדל, בכל כוחי, להעיר את רוחי עוד פעם הפעם.	מנדלה מוכר ספרים←אהוד בן עזר
כל חיי חלמתי למצוא לעצמי בעל אבל מה לעשות שאני מכוערת ואף גבר לא משתין לאברי. אמנם, כאשר אני ממריאה על מטאטאי לשמיים אני נעשית בחורה בלונדינית מה־זה־מתוקה, ממש מלכת־יופי עולה חדשה מרוסיה (לא זונה) ושמלתי הצחורה נודפת אז ריח כביסה טרייה, גיהוץ ועננים. אבל מה כל זה עוזר לי אם ברגע שאני נוחתת על הקרקע אני חוזרת לצורתי הקודמת, המקומטת והדחוייה? מה לא עשיתי כדי להתקדם בחיים. חיי, ומה גם לחלום שאני חולמת, חלילה, שאני מכוערה וגם איש לא ישים אברי אל לבי. אמנם, בהיותי על כרחי טסה בשמי ונהפכת לבחיר לב הארץ, נעשיתי עוד נערה בלונדינית משונה זו מירושלם (איני הזונה הגויה) ואת שפיפון העבה הייתי מכבדת אז בריח סבון טרי, בגירוד ובהזיות וערפילים. ואולם גם על זה אין לי.	אהוד בן עזר←יוסף חיים ברנר
לפעמים אני עוזר במשהו לאנשים, בדברים קטנים, אבל אני יודע שבמעשים לא די. כי אנשים רוצים שידברו אליהם. והם עצמם רוצים לדבר. לעתים קרובות זה חשוב להם הרבה יותר ממעשים. מלים נכונות, כלומר. כי לעתים גם אנשים המדברים ביניהם, לא מדברים במלים הנכונות. פעם, אבל אני יודע שעושים לא די. כי אנשים רוצים לדבר אליהם. והם עצמם רוצים מאוד לדבר. לעיתים חשוב להם הרבה יותר מעשיהן. מלים נכונות, זאת אומרת. כי לפעמים אנשים שמדברים ביניהם, אינם מדברים בשפה הנכונה.	אהרון מגד←אהוד בן עזר
אין פוחדים, פשוט, כל זמן שמרגישים עצמם עוד רחוקים ממנו, כל זמן שהמוות אינו עומד עוד אצלם על הפרק, כל זמן שכוחותיהם עוד לא עזבום ומר־המות לא סר עוד אליהם, אל הרגשתם הפנימית. מתבייש, פשוט, כי שום פחד לא פקד את נפשי עוד מההתחלה, כל זמן שיש עוד כוח בעולם, שכל כוחותיי עוד לא נמלטו ועלי להמשיך בחיים; שכל כוחי עוד לא נרפא... מפני איזו הרגשה פנימית.	יוסף חיים ברנר←אהרון מגד

Figure 18: Here, we present additional example outputs generated by our model.