

# Text-to-Lyrics Generation with Image-based Semantics and Reduced Risk of Plagiarism [ISMIR2023]

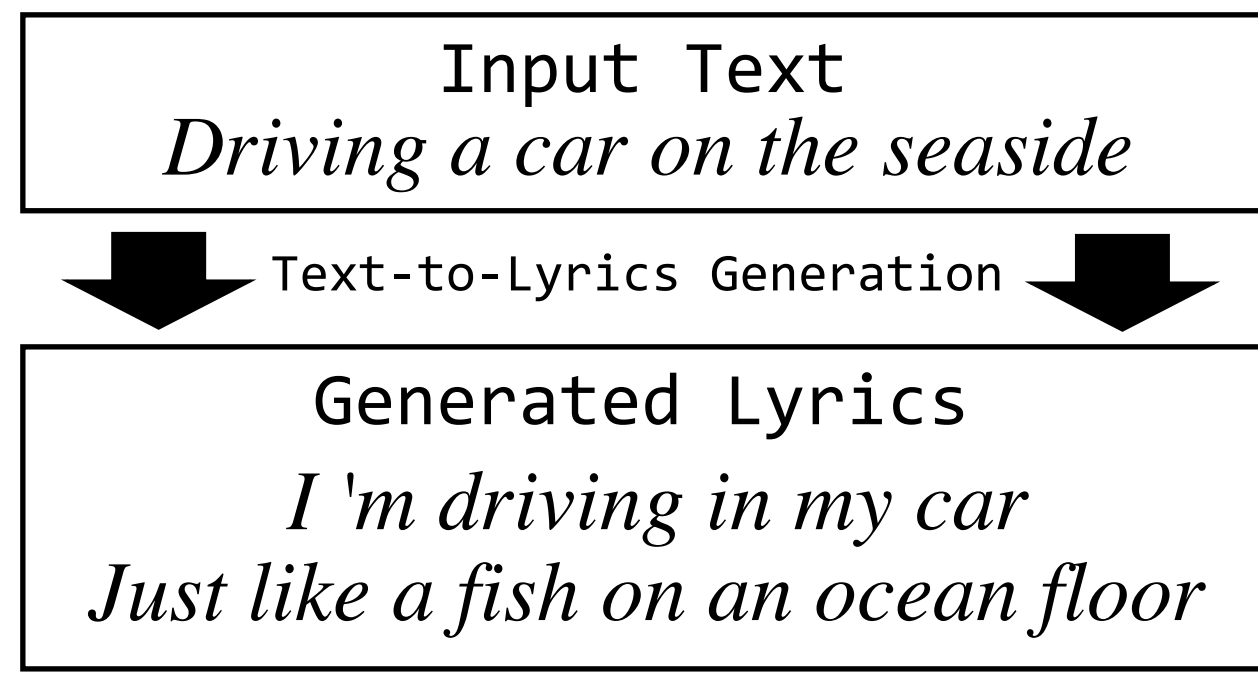
Kento Watanabe and Masataka Goto. National Institute of Advanced Industrial Science and Technology

## Overview

■ **Goal:** Text-to-lyrics generation system by suggesting the generated lyrics to users who struggle to find the right words to convey their message.

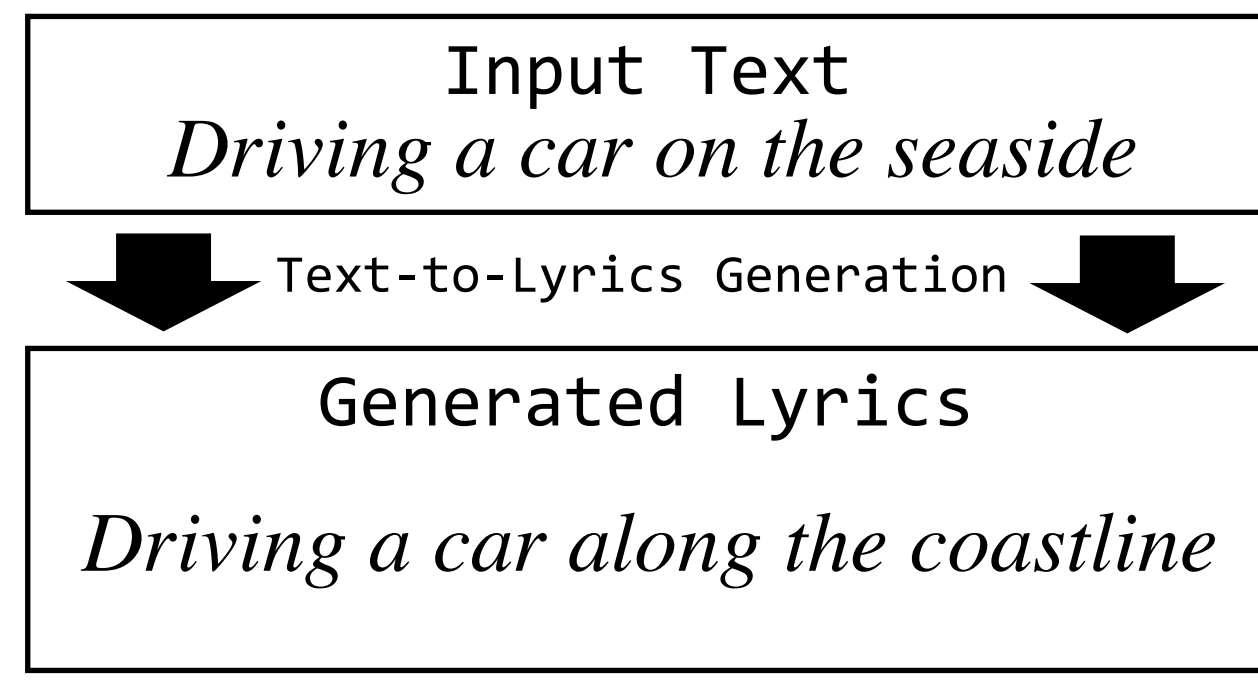
### Desirable System

- Generate lyrics that convey a message similar to the input text.
- Generate lyrics as a source of new inspiration.



### Undesirable System

- Generate lyrics that are similar in wording to the input text.
- Generate lyrics that may not provide enough inspiration.



### Technical Problem:

- (1) Training a text-to-lyrics encoder-decoder is not feasible since there is no text-lyric paired data.
- (2) Machine learning methods may unintentionally reproduce training data (plagiarize existing lyrics).

### Contributions:

- (1) Propose a two-step pipeline that enables training of a text-to-lyrics generation system even without paired text-lyrics data.
- (2) Propose a lyric generation method that reduces the risk of plagiarizing lyrics from the training data.

## Text-to-Lyrics Generation

■ **Key Idea:** Generating lyrics from input text by routing through an intermediate image.



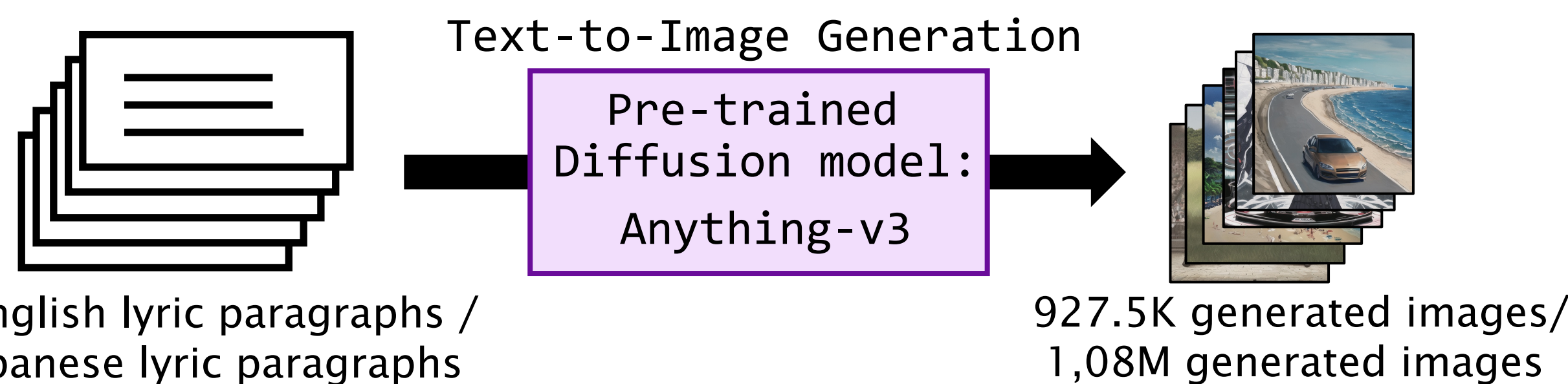
The wording information from the input text is removed. Only the semantics of the input text are conveyed.

### Advantages of using the intermediate image:

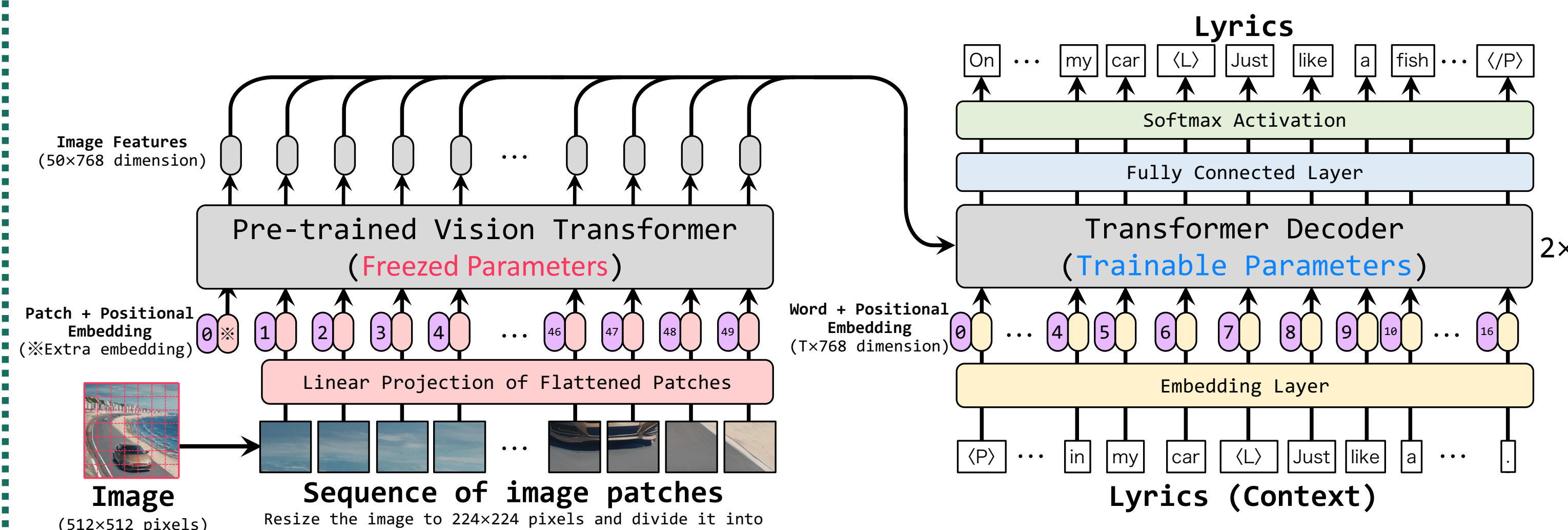
- The image represents the semantics of the input text without holding its wording information.
- The style of the input text can be any word, phrase, sentence, or paragraph.

### But, we must train the Image-to-Lyrics Enc-Dec:

- ☹ To train our Enc-Dec, image-lyric data is required.
- 😊 We therefore generate images from existing lyrics by using a diffusion model to create training data.



■ **Model Structure:** A model that connects a pre-trained Vision Transformer with a Transformer Decoder.



## Anti-Plagiarism for Lyrics Generation

■ **Question:** How to check if lyrics are plagiarized?

### Phrases that are safe to generate:

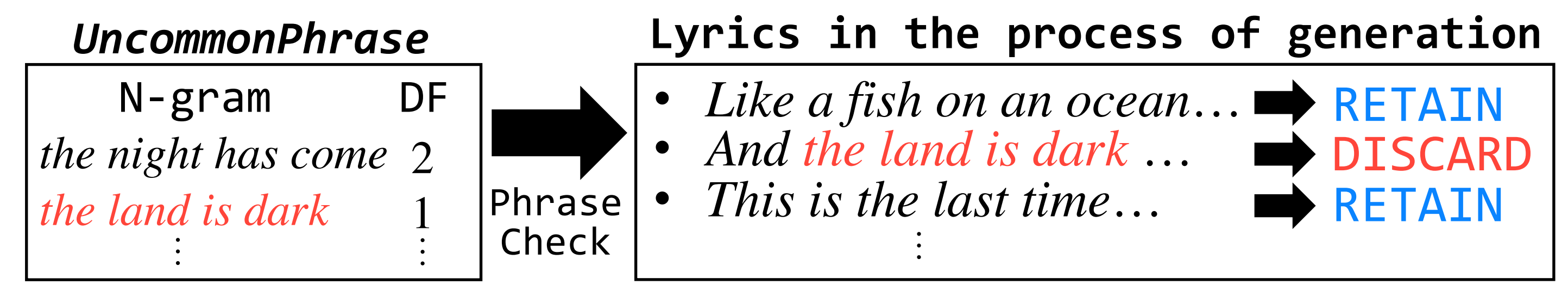
- New phrases that do not exist in the dataset.
- Commonly used phrases, such as "I love you".

### Phrases with a risk of plagiarizing existing lyrics:

- Uncommon phrases used in only a few songs.

■ **Proposed Method:**

- (1) Collect all word N-grams with DF (Document Frequencies) 1-3 in the training data and add them to a list of uncommon phrases, *UncommonPhrase*.
- (2) During the lyrics generation process, if any phrases within the generated lyrics are found in *UncommonPhrase*, those lyrics are discarded.



## Experiments

To test our lyrics generation, we used plot summaries of 20 Disney films from Wikipedia and their theme song lyrics, assuming the lyrics reflect the film's content.

### Examples of Generated Lyrics

Input Text (plot summary)	Image	Generated Lyrics
In a park in England, a young girl named Alice with her cat, Dinah, listens distractedly to her sister's history lesson, and begins daydreaming of a nonsensical world. She spots a passing White Rabbit...		Come to me, come to me I want a little rabbit and it's the same as you
Frozen. Princess Elsa of Arendelle possesses magical powers allowing her to control ice and snow, often using them to play with her younger sister Anna. After Elsa accidentally injures Anna ...		We're so bright and shining On our way to the ice

😊 We succeeded in using images to capture moods of the input text and generating lyrics accordingly.

### Quantitative Evaluation:

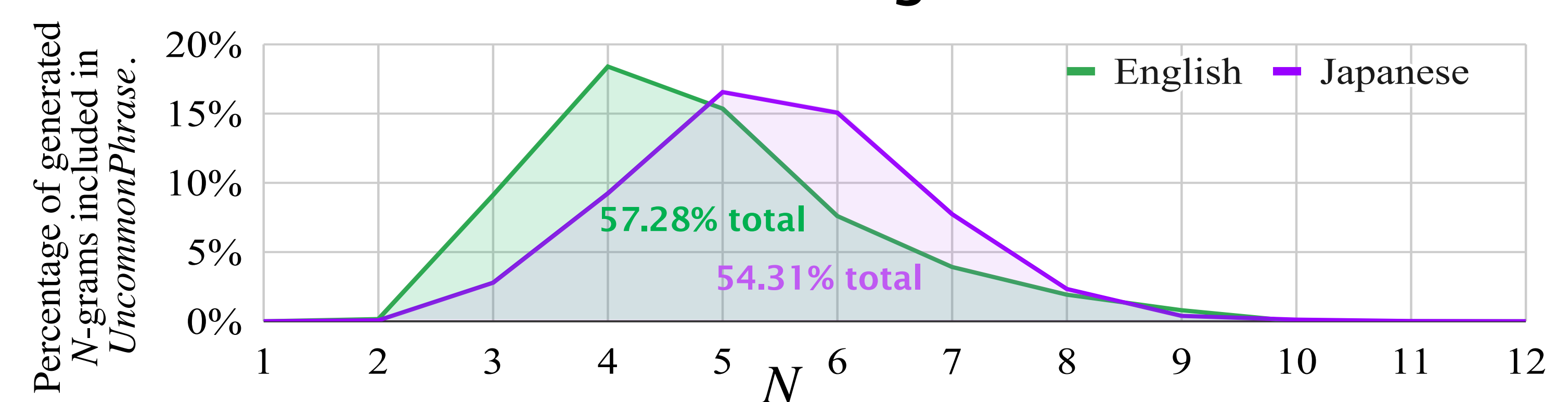
**Perplexity (PPL):** Predictability of the phrasing in the original lyrics in the test set.

**Normalized Edit Distance (NED):** Evaluate whether the proposed method generates lyrics that differ in wording from the input text.

Training paired data for Enc-Dec: [X]-to-Lyrics Generation	English		Japanese	
	PPL↓	NED↑	PPL↓	NED↑
Image-to-Lyrics (proposed)	<b>84.86</b>	<b>0.78</b>	<b>231.49</b>	<b>0.92</b>
Summary-text-to-Lyrics	346.73	0.69	306.19	0.86
Back-translated-lyrics-to-Lyrics	544.21	0.71	1051.58	0.66
First-half-lyrics-to-Second-half-lyrics	163.98	0.68	583.13	0.90

😊 Image-lyric pairs are more effective than other paired data sets as training data for Enc-Dec generating lyrics that are semantically related to the input text but differ from it in wording.

### Ablation Test for the Anti-Plagiarism Method



☹ Without our proposed anti-plagiarism method, approximately 57.3% of the generated N-grams (where N ranges from 1 to 12) matched our *UncommonPhrase*, indicating the risk of plagiarism.

😊 Our method effectively reduces this risk of plagiarism from 57.3% to 0%.