

---

# A Generic Tool for Visualizing Patterns in Poetry

**Onur Musaođlu**  
onurmusaolu@gmail.com  
Bogazici University, Turkey

**Özge Dađ**  
dagozge@gmail.com  
Bogazici University, Turkey

**Özge Küçükakça**  
ozge.kucukakca@boun.edu.tr  
Bogazici University, Turkey

**Alkım Almila Akdađ Salah**  
almilasalah@sehir.edu.tr  
Sehir University, Turkey

**Albert Ali Salah**  
salah@boun.edu.tr  
Bogazici University, Turkey

---

## Introduction

Visualization of text can be a useful exploration tool for looking at the corpus of a poet, especially when dealing with a prolific author with a large body of output over the years. In this work, we describe a flexible and extensible tool for analyzing the corpus of a poet, and make a case study of Nâzım Hikmet Ran. Since poetry has its own challenges over plain text, we have developed novel ways of visualizing the structure, the rhythm and affective tone of each poem, as well as ways of looking at the continuities (or discontinuities) of features in the entire corpus over the years. The designed system integrates a database for holding meta-information, and a website for creating and linking interactive, parameterized visualizations.

Nâzım Hikmet Ran is one of the most famous poets of Turkey. Although he was a great patriot, he has spent many years in prison and in exile due to his communistic political views. His poetry is translated into more than fifty languages. We believe this tool can be

particularly useful to compare different translations of the poet's work, to see how certain stylistic or semantic features are retained (or lost) during translations.

## Related Work

Most poetry visualizations focus on the aesthetics of information rather than the functional aspects. An example is [Diana Lange's visualizations](#) that transform individual poems into beautiful visual displays, resembling flowers. A similar project is Boris Müller's [Poetry on the Road](#), which turns a text into an image through an arbitrary transformation function, for instance by treating every word as a location and creating a heat map of the entire text. The outcomes of these projects do not tell us much about the poets or the œuvres in question.

In contrast to such artful renderings of poems, there are studies that focus solely on the grammatical and structural problems of poetry writing. Such studies rather try to find quantitative ways to analyze poems, enabling a computational approach for the evaluation of technical quality and subtlety of the rhymes (Opara, 2014; Dalvean2015). In the same line of research, there are visualization tools such as Graphwave, SentimentGraph, SentimentWheel and Ambiances (Meneses and Furuta, 2015). Such visualization examples constitute the starting point of our explorations for devising a new visualization system that is both scalable and modular in nature, i.e. a tool that would accommodate different natural language processing (NLP) tools, as well as new visualization techniques.

## Methodology

In this section, we briefly describe the database structure, as well as the software tools used to create the system.

### Database

After his death, Nâzım Hikmet's collected works appeared in a single volume (Nâzım, 2007). The digital version of this volume is not publicly available, but we received a special permission from the publishers to use this volume.

Nâzım is a poet who paid special attention to the visual structure of his poems and it is imperative to retain this structure as accurately as possible. Consequently, line indentations were kept intact for each line, as well as the fonts of the individual words. We also paid attention to the fact that the collected works included some text written in prose. Thus, the data-

base structure, depicted in Fig. 1, is entirely hierarchical and ordered according to books, works in a book, lines in a work, words in a line, and characters in the words. This may seem to be an extensively elaborate representation, but it allows detailed structural analysis, as well as the analysis of visual and rhythmic features of each work.

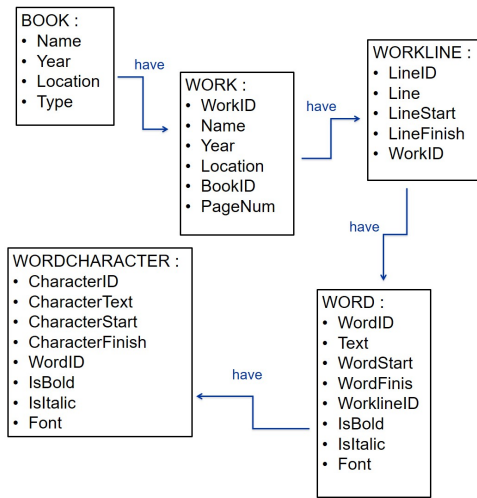


Fig. 1: Database structure.

## Software

Since the project involves a dynamic, parametric and interactive system, many software technologies were used. To keep structured data, user data and web page related content, a MySQL database was used. A Turkish-based affect analysis tool was integrated with the system, and Perl was employed to read and parse data for the affect analysis tool. The main programming language of the project is Java and all back end code is also developed in this language. The Spring Framework was used to create the model-view-controller (MVC) structure of the application. In the front end of the application, AngularJS was used to create the MVC structure and to create a single page application. Moreover, to make the application responsive, CSS3 and Bootstrap technologies were used for mobile phone support.

## NLP Challenges

In order to parse Nâzım's corpus, a Turkish Morphological Parser and Disambiguator was used (Sak et al., 2008). With the help of this tool, we get part of speech tags of the words, as well as some grammatical

information about verbs (i.e. conjugation and tense) and about words' grammatical number. For certain instances, the morphological parser suggests more than one form or number. To solve such problematic instances, an off-the-shelf disambiguator was used. The results of this disambiguation tool suggests the most appropriate form for a given context, which helps in making a decision on the preferred form of a given verb, noun, or pronoun.

The system was enhanced with a text-based affect analysis tool, which returns valence, arousal and dominance values for a given sentence and each word in that sentence (Aydın Oktay et al., 2015). One of the challenges in parsing poems versus prose texts is the lack of a specific notation for indicating the end of a sentence. For the sake of simplicity, each line of a given poem was treated as a sentence, and valence, arousal and dominance values were computed for every word and line individually. These values are stored in the database for fast retrieval.

## The System Interface

The generated system works as an interactive visualization tool with a web interface. For the user experience of the web system, a responsive interface is prepared that can even be reached via smart phones. Also, to keep data alive and to allow flexible operations, a single page application is created, with which users can surf between different tabs without losing information. The system design is modular and expandable, as each work unit is separated such that new visualizations can be easily added to the system. The system can also be tailored to visualize a new database easily. The only requirement is that the work of the artist be parsed in the same hierarchical way, and placed on a SQL-capable database.

## Visualizations

One of the motivations behind the visualizations is to give information about the analysis on the corpus of Nâzım, and other poets when the database is expanded by the addition of new authors. The tool incorporates a search function, and allows different visualizations to be prepared from the results of the search. Most of the prepared visualizations are interactive charts. They can be used for showing a term's usage over the years, over geographic locations and over publications. The search can be conducted on a collection of works, or in a single work. A separate page was created for searching for a sequence of words, and to prepare comparative visualizations.

We briefly describe two visualizations here to serve as examples. The first visualization is called the “poetry barcode” (see Fig. 2). In this visualization, one poem is visualized, and each line of the poem is represented by a horizontal line. The length and the color of the lines are set according to NLP and affect analysis results, and the lines form six different columns, which show the change of line lengths, usage of active/passive phrases in a line, inflections of words in a line according to person information, as well as valence, arousal, and dominance values of each line.

Nâzım has a lot of stylistic features in his poems. To be able to analyze and extract these features, we have prepared visualizations about the usage of alliteration and his unique verse structure. Alliteration is a stylistic device, in which a number of words, having the same consonant sounds, occur close together in a series. To quantify alliteration, a measure was developed that uses the background frequency of each letter in the poet’s corpus. By using a sliding window based evaluation, letter frequencies are calculated, and compared with the base frequency of that letter in the corpus. Fig. 3 illustrates the automatic alliteration detection. Fig. 4 shows a number of additional visualizations in a bird’s eye view, including a visualization about passive/active voice usage.

## Conclusions

A complete web page opened to the wider public is in construction, as it requires some security features due to copyrights of the works in the database. But the system is operational in the offline mode, and already provides many visualization options.

Since the database contains composition years and places for poems (where available), it is possible to search for words that are historically relevant for Nâzım. To give an example, a search for the words “hürriyet,” (freedom) and “hapis” (jail) restricted to 1919-1938 and 1938-1963, we can easily see that these words’ usages are significantly increasing after 1938, when he was arrested for the first time. Other possible uses include the visualization of words associated with different colors, prominent in his poetry, over the years.

The proposed system keeps data and visualizations separate, but well-connected. This enables the addition of new artists to the proposed system. The tool also can be used as a platform to evaluate poetry translation. We show grammatical and affective features of the words in some visualizations like the “poetry barcode”. It is possible to use these visualizations to get

an idea about translation quality in literal translations, where the emphasis is on word-for-word translation.



Fig. 2: A poem’s barcode, visualizing the structure of the poem together with verb conjugations, passive/active verb usage, and emotional tone.

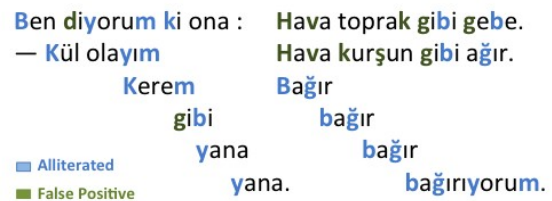
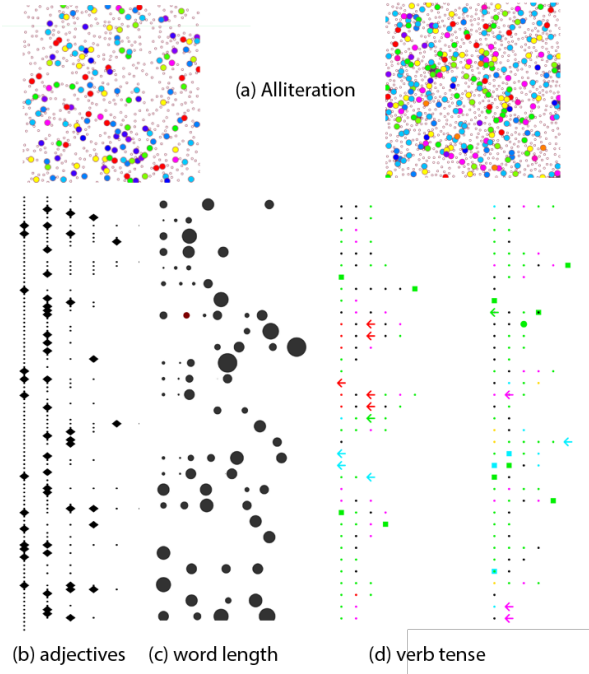


Fig. 3: Visualizing alliteration in the poem. Best viewed in color.



Sak, H., Güngör, T., and Saraçlar, M. (2008). Turkish language resources: Morphological parser, morphological disambiguator and web corpus. In *Advances in natural language processing*, pages 417–427. Springer.

Fig. 4: A bird's eye view of several visualization options in the system. Best viewed in color.

## Acknowledgments

We would like to thank Yapı Kredi Yayınları for giving us permission to use the digital version of Nâzım's works. This study is realized in collaboration with Nâzım Hikmet Kültür ve Sanat Araştırma Merkezi.

## Bibliography

- Aydın Oktay, E., Balcı, K., and Salah, A. A. (2015). Automatic assessment of dimensional affective content in Turkish multi-party chat messages. In *Proc. Int. Workshop on Emotion Representations and Modeling for Companion Technologies*, pages 19–24. ACM.
- Dalvean, M. (2015). Ranking contemporary American poems. *Digital Scholarship in the Humanities*, 30(1):6–19.
- Meneses, L. and Furuta, R. (2015). Visualizing poetry: Tools for critical analysis. *paj: The Journal of the Initiative for Digital Humanities, Media, and Culture*, 3(1).
- Nâzım, H. (2007). *Bütün şiirleri*. İstanbul: YKY.
- Opara, K. R. (2014). Grammatical rhymes in Polish poetry: A quantitative analysis. *Digital Scholarship in the Humanities*, page fqu029.