

# Content Analysis with Textual Visualization Tools: Example of Turkey's Key Events in 2016

Güllü Yakar

Necmettin Erbakan University, Faculty of Fine Arts, Meram. 42040 Konya, Turkey  
gyakar@konya.edu.tr

**Abstract:** *Data mining and information design together enable exploration of unknown data. Especially qualitative researches - like social media analysis, linguistic or communicational studies that statistical methods cannot be performed directly- necessitate usage of special techniques to extract patterns and structures. Textual-based visualization techniques allow to analyze the unstructured corpus, explore the inner relations and provide meaning to these texts. In this study, textual visualization techniques have been used to show the different styles of reporting the news.*

**Keywords:** *Textual data visualization, word cloud, word tree*

## 1. Introduction

Textual data such as political declarations, scientific publications, news articles, novels and poems hold special codes in their own structures. Words and phrases constitute the base of these texts. Selected words, their relations and frequency in which they are used emphasize the narrator's intended statement. Decoding this structure can be achieved with content analysis method, and findings can be represented by using matrices or diagrams. However, cases of multidimensional data requires advanced visualization techniques.

Text visualization methods vary according to their qualities as analytic and visualization task, data source and property, visualization dimensionality, visual representation and alignment. In addition to advanced visualization methods that appeal experts such as researchers, scientists, library managers, and so on; simple visualization methods are available for non-expert users, also. Researches show that there is a gap on the decision processes involved in creating, designing, and justifying visual representations for non-expert audiences, especially for the research intermediary [1] [2]. Visualizations that appealing to non-expert users, which described as "popular, attractive, fun ways to display text data in graphical form" by DePaolo ve Wilkinson [3], are also ideal for mass media readers. Viegas, Wattenberg and Feinberg [4] asserted that major newspapers and magazines like the Washington Post, the Boston Globe, the Guardian make use of these visualization systems to depict political and social issues for their communicative and illustrative aspects.

### 1.1. Visual Features

Textual-based data visualization, or text visualization, provides data mining process with graphical and interactive representation. Besides the basic visual elements and principles; additional factors like time, motion, user-centered interactivity also shape the result. Typographic elements are the representative form of linguistics and basis of textual visualization. Font size, density, texture or color's properties like hue, saturation and lightness create hierarchy between words and highlight the importance. Direction and position determine the inner- contextual relations within the text. Visualization techniques have also applied to design aesthetic layouts or give information about an issue. In these cases, color and font selection gains additional importance. Warm or cool colors have different effects and visualization elements should be colorized with an appropriate color palette.

There are some variables that should be consider when using textual visualization. The first of these is length of word. "Long words occupy more space than small ones do, thus resulting in a misconceived impression of weight, given that we tend to associate size with importance. The issue is even more prominent when other

visual variables, such as color and weight, are added to the typographic system, because they influence the perception of hierarchy in the graphic” [5]. The other one is that, length of the text might require extraction and elimination of less frequent words.

## 1.2. Textual Visualization Types

- Relational decoding of text: renders relationships between words, including the direction of the connection, that is the word order. These diagrams proportion the words in accordance with their frequencies and relate them with arrows. Elements like color saturation, size, direction and position are essential to show the hierarchy between words.
- Relational decoding of a word: Selected word constitutes the root of visualization and this root diverges to branches enabling the analysis of word with the text. That word tree connects the root word and related sentences with arrows. Word frequencies affect the sequence of branches, their position and size. Punctuation, repeated words and conjunctions together constitute the context, for this reason reducing the content of the text creates negative results.

“Word Trees are a powerful way to understand context around a term, however they are limited in that they only represent context in terms of strict adjacency of terms” [6].

“The area of the word is very roughly proportional to the frequency...it leaves sufficient blank space above and below that the overall tree structure is visually obvious. Branches of the tree continue at least until they define a unique phrase used exactly once. Instead of stopping at the first unique phrase, the tree continues until a period is reached. To distinguish between the main tree of unique phrases, and the additional context, the former is colored black and the latter is drawn in gray. One somewhat counterintuitive design choice is that we do not discard stopwords or even punctuation. The rationale is that prepositions and commas are often critical to understanding the meaning of a text. Leaving them out might put together phrases that mean very different things [7].

- Word’s frequency: shows words in a block, word cloud or cirrus, in accordance with their proportional size depending on their occurrence in text. For word cloud or Cirrus, the determining element is the font size of words. Words are categorical data and unlike quantitative data, they can not be classified with color values. Categorical data is separated and displayed with color hues. Typeface and composition provide additional effects. The number of words can adversely affect legibility, therefore eliminating less frequent words can be considered. “Word clouds are therefore only useful when you are trying to get a quick and rough sense of some of the dominant keywords used in the text. They can be an option for working with qualitative data during the data exploration stage, more so as a means for reporting analysis to others” [8].

TABLE 1. General features of word tree and word cloud visualizations [9]

	Word Tree	Word Cloud
<b>Analytic task</b>	Lexical/ syntactical analysis	Text summarization / topic analysis/ entity extraction
<b>Visualization task</b>	Region of interest	Clustering/classification/ categorization
<b>Data property</b>	Networks	Typographic composition
<b>Domain</b>	Literature/poems	Broad spectrum
<b>Visualization dimensionality</b>	2D	2D
<b>Visual representation</b>	Node-link/Diagrams	Clouds/tags/galaxies
<b>Visual alignment</b>	Metric-dependent	Linear/parallel
<b>Display type</b>	Interactive	Statical

## 2. Method

In this study, textual visualization techniques are used to analyze Turkey's key events in 2016. Related news articles in national, international and foreign media websites have been reached through literature review. Population is limited to English textual resources in order to prevent data loss during the reporting process. Samples were selected from an international news broadcasting corporation, BBC's website and an international organization, Human Rights Watch's website: "Turkey Profile – Timeline: a Chronology of Key Events" [10] and "Turkey: Events of 2016" [11].

In two-worded phrases, the tilde (~) function have used so that the words seem as single. The texts were transferred to web-based visualization tools after manuel editing on word processor. Web-based text reading and analysis environment Voyant Tool have been used to analyze the documents' general inventory.

Sample 1. "Turkey Profile – Timeline: a Chronology of Key Events", BBC

2016 key events section of this text has 295 total words and 187 unique word forms. Most frequent words in the text: Erdogan (5); president (5); attack (4); kills (4); Russian (4).

Sample 2. "Turkey: Events of 2016", HRW

2433 total words and 922 unique word forms had been used in this text. Most frequent words were "government" (22); "Turkey" (22); "rights" (18); "coup" (16); "human" (13) before extraction. Words "Human" and "rights" were connected with the tilde (~) function so that these two words seem as a single word. Most frequent words in the document after extraction: Turkey (22); government (22); coup (16); human rights (13); journalists (10). First two keywords were defined as stopwords and main keywords are reached. Terms coup (16); human rights (13); journalists (10) were taken as meaningful terms according to the aim of study.

### 2.1. Word Cloud Generation and Features

A web-based text visualizing tool, Wordle, has been used to explore texts by identifying key terms being used. "...to ensure the underlying text being used is carefully prepared in advance to reduce the noise [12]; default removing functions of Wordle for numbers and common English words have been used. The limit of the maximum words determined as 30 on word cloud to highlight the important terms. Because of deceptive effects on visualizing; "Turkey", "Turkish", "government", "İstanbul", "Ankara", "including", "many", "also", "UN", "members" words were defined as stopwords in visualizing process and removed with Wordle's manuel removing function. Since there are proper names, upper and lower cases have been left as it is in the text.

Cloud consist of horizontally aligned words and straighter edges. Colors, font and composition have been chosen by taking appropriateness and legibility criterions into consideration.

### 2.2. Word Tree Generation and Features

To display the hierarchical structure of the words in the context, Word Tree diagrams were used. Texts have not been intervened in order to preserve their semantic and syntactic form. The term "president" which is the most frequented word of Sample 1, visualized in web-based Word Tree tool (Figure 3). The most frequented word of Sample 2, the term "coup" visualized twice: In Figure 4, the term "coup" has been displayed at the beginning of the sentences. In Figure 5 and 7, word tree visualization tool's "reverse" function have been used. This visualization shows the sentences that finish with term "coup".

## 3. Results

The results of word cloud visualizations below show the points of departure between BBC and HRW reports.

"Coup", "media", "authorities" are common terms of two texts. While BBC report, Sample 1, is using the term "TAK" about terrorist organizations, HRW uses the term "PKK". Terms "seekers", "detainees", "detention", "asylum", "journalists" are only seen in HRW's report, Sample 2.



Fig. 1. Word cloud visualization of Sample 1



Fig. 2. Word cloud visualization of Sample 2

Figure 3 displays the word tree visualization of Sample 1 with the term “president”. Since the term “president” is associated with the terms “Erdogan” and “says”, text can be asserted to be a report which compiles statements.

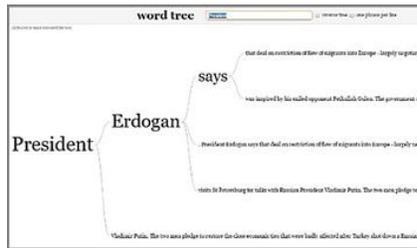


Fig. 3. Word tree visualization of Sample 1 with the term “President”

Figure 4 and 5 presents the two visualization example of Sample 2’s second most common term “coup” at the beginning and end of the sentences. Although the term “coup” seems to be associated with the term “attempt”; since it is tied with unique phrases, it has been left separate.



Fig 4. Word tree visualization of Sample 2 for word “coup”



Fig 5. Reverse word tree visualization of Sample 2 for word “coup”

Although terms “human” and “rights” used in conjunction in word cloud visualization, the term “right” is separated in word tree visualization since this word is also linked with other words. Following the term “human” in frequency ordering, the term “rights” is also linked with the terms “women”, “children” and “civil and political” (Figure 6). After the first search for the term “rights”, word “human” added to the initial search (Figure 7).

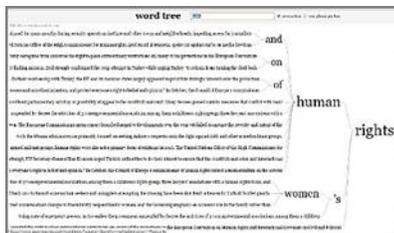


Fig 6. Reverse word tree visualization of Sample 2 for word “rights”



Fig 7. Reverse word tree visualization of Sample 2 for words “human rights”

## 4. Conclusion and Further Research

Data types could be quantitative, spatial, temporal or textual kind. Visualization could be designed to serve publishing or sharing the data on social media as well as it could serve to explore and analyze the context and its structure. Visualization techniques could be used in scientific –especially qualitative- researches, linguistic, cultural analyzes and data journalism as an exploratory tool in addition to its explanatory function. Textual visualization tools can be used to evaluate the points of departure between different sources of news reporting the same topic and allows to create content analysis. Textual visualization techniques are believed to provide an alternative approach to the typical representation method used in qualitative data analysis.

## 5. References

- [1] Gough, P., Ho, X., Dunn, K., Bednarz, T.: Art and Chartjunk: A Guide For NEUVis. In : 7th International Symposium on Visual Information Communication and Interaction, pp 171-177. Sydney (2014)
- [2] Allen, W. L.: Making Corpus Data Visible: Visualising Text with Research Intermediaries, *Corpora*, 12, 3. (2017)
- [3] DePaolo, C. A., Wilkinson, K.: Get Your Head into the Clouds: Using Word Clouds for Analyzing Qualitative Assessment Data. *TechTrends*. 58, 3. 38-44. (2014)
- [4] Viegas, F. B., Wattenberg, M., Feinberg, J.: Participatory Visualization with Wordle. In *IEEE Transactions on Visualization and Computer Graphics*, 15, 6, 1137-1144 (2009)
- [5] Meirelles, I.: Design for Information: an Introduction to the Histories, Theories and Best Practices Behind Effective Information Visualizations. Rockport, Massachusetts (2013)
- [6] Diakopoulos, N., Elgesem, D., Salway, A., Zhang, A., and Hofland, K.: Compare Clouds: Visualizing Text Corpora to Compare Media Frames. In: *IUI Workshop on Visual Text Analytics*. (2015)
- [7] Wattenberg, M., Viégas, F. B.: The Word Tree, an Interactive Visual Concordance. *IEEE Transactions on Visualization and Computer Graphics*, 14, 6, 1221-1228 (2008)
- [8] Kirk, A.: *Data Visualization: a Handbook for Data Driven Design*. Sage, UK (2016)
- [9] ISOVIS Group, <http://textvis.lnu.se/>
- [10] BBC Europe, <http://www.bbc.com/news/world-europe-17994865>
- [11] Human Rights Watch, <https://www.hrw.org/world-report/2017/country-chapters/turkey>
- [12] Kirk, A.: *Data visualization: a Successful Design Process*. Packt, Birmingham (2012)E