

TARİHSEL OLAYLARI ÇÖZÜMLEMEDE METİN MADENCİLİĞİ: NUTUK ÜZERİNE BİR UYGULAMA

H. Ekrem BAYAR*
Doğan YILDIZ**

ÖZ

Tarih, insanlığın geçmişten günümüze yaptığı eylemler ile yaşanan olayları, kişileri, toplumları, ülkeleri, mekan ve zaman gibi kavramları bir araya getirerek nedensellik içerisinde anlatan bilim dalıdır. Tarihsel olayların doğru ve objektif bir biçimde incelenmesi ve yorumlanması ile insanlığın geleceğine büyük katkıları olabileceği gibi tarih bilimine yeteri kadar önem verilmemesi, insan topluluklarının durumlarının kötüye gitmesine neden olacaktır. Mustafa Kemal Atatürk'ün de dediği gibi: "Tarihini bilmeyen milletler, yok olmaya mahkumdur!". Bu sebeple tarih biliminin işlevselliğini arttırmak adına istatistiksel yöntemler ile tarih bilimine katkıda bulunulması amaçlanmıştır. Metin madenciliği ve doğal dil işleme yöntemleri ile tarihsel metinler analiz edilerek tarihsel olaylar daha objektif bir biçimde açıklanabilir. Metin Madenciliği ise, veri analitiğinin alt dallarından bir tanesidir ve temel amacı metinleri yapısal bir veri formatına dönüştürdükten sonra metinlerden anlamlı bilgiler çıkarma sürecine dayanır. Araştırma kapsamında tarihsel metin ve veri kaynağı olarak Mustafa Kemal Atatürk'ün Nutuk'u ele alınmıştır. Literatür araştırması sonucu İngilizce metinlerin analiz edilme kolaylığı ve kullanılan yöntemlerin ve yazılım kütüphanelerinin daha uygun olduğu görüldüğü için veri kaynağı olarak İngilizce'ye çevrilmiş Nutuk üzerinden tüm analizler yapılmıştır. Milli Mücadele Dönemi'nde en çok tekrar eden kelimeler, kelimelerin

* Yıldız Teknik Üniversitesi İstatistik Bölümü, h.ekrembayar94@gmail.com, (ORCID: 0000-0002-9657-5786).

** Dr. Öğr. Üyesi Yıldız Teknik Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümü, dyildiz@yildiz.edu.tr, (ORCID: 0000-0001-7430-2368).

birbiriyle olan iliřkisi, duyu analizi, konu sınıflandırma ve benzerliđi gibi analizler araştırma kapsamında yapılmıřtır.

Anahtar Kelimeler: *Nutuk, Metin Madenciliđi, Milli Mücadele, Mustafa Kemal Atatürk, Veri Bilimi.*

TEXT MINING ON HISTORICAL EVENTS: AN APPLICATION ABOUT THE GREAT SPEECH OF ATATURK

ABSTRACT

History is a branch of science which bands together people, societies, concepts such as time and explains them by dealing with a cause-effect relationship. Whereas the accurate and objective analysis and interpretation of historical events can make great contributions to the future of humanity, not paying enough attention to the science of history may cause the conditions of human societies to deteriorate. For this reason, as Mustafa Kemal Atatürk also stated; this study is aimed to contribute to the science of history by using statistical methods in order to increase the functionality of the science of history. Text mining and NLP methods help to explain historical events by analyzing more objectively. Text Mining is one of the sub-branches of data analytics and its main purpose is to analyze meaningful information from texts after converting texts to a transformed data format. In the scope of the research, Mustafa Kemal Atatürk's Nutuk is considered as a historical text. As a result of the literature search, it was assumed the English texts were analyzed easily and the methods and software libraries used were more appropriate, so all analyzes were carried out on Nutuk, which was translated into English. During National Struggle, the most repeated words, the connections of words with each other, sentiment analysis, subject classification and similarity of the events were analyzed within the scope of the research.

Key Words: *Ataturk's Great Speech, Data Science, Mustafa Kemal Ataturk, Text Mining, Turkish Independence War.*

Extended Abstract

The data, which is seen as today's oil, shows its effect in every field with the advancement and widespread use of technology. The rapid advancement of technology has also facilitated the applicability and use of statistics and computer science in all areas. Ease of storage and processing of data in every field creates various opportunities and accelerates the development of the fields. If world history is considered as the largest data source, artificial intelligence methods should also be used in historical science and historical events should not be handled only with traditional methods. By using artificial intelligence methods such as text mining,

natural language processing and machine learning, historical science should also adapt to the digital age. In the scope of this study, it has been tried to show that a historical event can be handled with text mining methods.

Text mining is the process of transforming unstructured text into a structured format to identify meaningful patterns and new insights. Text mining combines different disciplines such as linguistics, statistics, data mining and machine learning. Text mining techniques are basically divided into four categories: Classification, clustering, association analysis and information extraction. In this context, text mining methods include some problems such as named entity recognition, idea mining, text classification, sentiment analysis, keyword-title extraction etc.

The content of the study will include a neutral evaluation and text mining applications on historical events. The main text source of this project is The Great Speech (Nutuk), and it includes text mining methods and observation of historical events, inferencing and predicting, comparing, relating and visualization. The text is converted from unstructured data format to structured data format in order to carry out these studies and to apply analytical methods. In the Great Speech, the events are described in a continuous flow with a chronological order starting from May 19, 1919 to 1927. The main purpose of the Great Speech is to clarify the events that took place during the Turkish War of Independence, to contribute to the history and to create a resource for the future of the Turkish nation. In the last part of the Great Speech, the assigned the duty of Turkish youth and told the Turkish youth to protect and defend the Turkish independence and the Turkish Republic forever. The Great Speech is the manifesto of protecting the Turkish Independence and the Turkish Republic.

In the data analytics, there are two types data format as structured and unstructured. Image, voice and text data sources are unstructured. These data sources need to transform structured format to analyze. That's why, the biggest problem of text mining is that the data to be analyzed is not structured. The first step of text mining is pre-processing the text data. Therefore, stop-words, symbols, punctuation marks etc. expressions that do not have a meaning alone were removed. In addition, all words in the text have been translated into lowercase letters. The text that was wanted to be analyzed was divided into sentences and words and turned into a regular text format. Since text mining and natural language processing applications do not yet show sufficient development for the Turkish language. In this study, English version of the Great Speech was preferred for analysis. All of analyzes in the research were conducted with the help of R programming language.

Contributing to the science of history with statistical methods will enable historical events to be interpreted more clearly. To understand the cause-effect relationships between historical events can be successful by using correlation analysis, sentiment analysis, topic modelling, network analysis, word frequencies, text classification.

In summary, word frequencies and tf-idf methods were used to obtain a general information about text. It has been inferred that the frequent use of words in a text can be an indicator for explaining events and identifying people. Correlation analysis proves that there is a relationship between the periods and the strength of this relationship calculated. In addition, with the help of correlation analysis, the situations and relationships between words were revealed. The association of randomly selected words from the text with other words has been examined. With the help of sentiment analysis, a text gives objective information. As a result of the sentiment analysis, The Great Speech was classified as a positive text considerably. The text classification method does not work in such an example because the periods and historical events in the Great Speech are interconnected.

Giriş

İnsan toplulukları, geçmişten beri varlığını sürdürme iç güdüsüyle bilgi üretmiştir. Ürettiği bilgiyi de kullanarak modern toplumlar için gerekli değişimleri yakalamaya çalışmışlardır. Günümüzde bu gelişim, veri odaklı yapılmaktadır. Geçmişten günümüze üretilen en büyük bilgi ise insanlık tarihidir.

Bilgisayar, internet gibi konularda teknolojik ve bilimsel gelişmelerin hızlanması, gelişmesi pek çok alanı olumlu yönde etkilemiştir. Geçmişte bir bilginin işlenmesi ve o bilgiden değer üretilmesi çok maliyetli iken günümüzde büyük verilerin işlenmesi, makine öğrenmesi ve yapay zeka uygulamalarını daha kolay hale gelmiştir. Bu uygulamaların yaygınlaşmasıyla veri çağında hayatta kalmak için güncel olmak, teknolojiyi takip etmek ve dijital çağa ayak uydurmak gerekir. Artık alanında uzman hiç kimse dijital çağ, veri çağı veya bilgi çağı adına ne denirse denilsin teknolojik gelişmelere, araçlara ve yöntemlere uzak kalmamalıdır. Yapay zeka uygulamaları metinleri işleyebilir hale gelmişken tarihçilerin de artık bu uygulamalara kayıtsız kalmadan dijital tarih kavramı ile araştırmalarını daha ilgi çekici hale getirip dijital ortamlara sunması tarih öğrenimi ve farkındalığını kolaylaştıracaktır.

Dijital tarih, tarihçiye yardımcı olma yoluyla işini kolaylaştırmakla kalmayıp araştırma metodunu da değiştirmektedir. Öncelikle, tarihçi, araştırma sorusuna cevap bulmak için daha büyük veri gruplarını sorgulama

ve farklı metotlar uygulama fırsatı bulmaktadır. Değişen teknoloji yöntemlerini kendi avantajına kullanmak için iki noktada dikkatli olmalıdır; dijital tarihçinin düşmanı değildir, öyle görmemeli ve uygun biçimde kullanılmalıdır. Bunun için, araştırmasının bir parçası olarak kullandığı dijital araç-gerecin eleştirisini yapmalıdır.¹

Tarihçilerin dijital çağda başvurabileceği en önemli araçlardan bir tanesi metin madenciliğidir. Bu araştırma metin madenciliği yöntemlerinin tarih bilimine katkısı olabileceği yönünde düşünülerek yapılmıştır. İstatistiksel yöntemler kullanılarak, tarihsel olaylar dışında olduğu gibi metinler, objektif bir biçimde açıklanabilir, sınıflandırılabilir ve konu anlam bütünlüğü bakımından çeşitli yargılarda bulunulabilir.

Metin madenciliği süreci, yapılandırılmamış metin verilerinden anlamlı bilgi çıkarmayı sağlayan çeşitli aktivitelerden oluşur. Metin madenciliği tekniklerini uygulamadan önce, metin verilerini temizleme ve kullanılabilir bir biçime dönüştürme uygulaması olan metin ön işleme aşaması ile çalışmalara başlanır. Analizler için verileri uygun bir formata getirmek adına dil tanımlama, sözdizimi ayrıştırma, parçalama, konuşma parçası etiketleme gibi tekniklerin kullanılmasını içerir. Bu işlemler tamamlandığında verilerden bilgi elde etmek için metin madenciliği, doğal dil işleme, makine öğrenmesi ve yapay zeka algoritmaları kullanılabilir. Analiz çıktıları, veri görselleştirme teknikleri kullanılarak bulguları daha geniş kitlelere ulaştırabilir.

Metin madenciliği tekniklerinin uygulanması için Mustafa Kemal Atatürk'ün Türk İstiklal Harbi'ni ve Türkiye Cumhuriyet'inin kuruluşunu adım adım kaleme aldığı Nutuk kullanılmıştır. Türkçe metinler işlemenin zorluğu nedeniyle Nutuk'un İngilizce çevirisi üzerinden tüm analizler gerçekleştirilmiştir. Literatürde metin madenciliği, doğal dil işleme, makine öğrenmesi gibi yapay zeka uygulamalarının kullanıldığı araştırmalara rastlanmamıştır. Ancak Nutuk üzerine yapılan terim ve söz gruplarının analizi, içerik analizi, dizin analizi gibi çeşitli yaklaşımlarla Nutuk, istatistiksel bakış açılarıyla incelenmiştir. Bu araştırmada Nutuk, sosyal bilimlerden bağımsız olarak, istatistik bilimi ve bilgisayar bilimine ait olan yöntemlerin uygulanmasıyla ele alınmıştır. Araştırma, sosyal bilimlerde çalışma alanı olan olaylar hakkında iddialarda bulunmamaktadır. Burada amaç, yapay zeka uygulamalarının sadece sayısal verilerle değil; aynı zamanda sözel verilerle de uygulanabilirliğini göstermektir. Bu yöntemler sadece Nutuk özelinde değil tüm yazılı kaynaklarda faydalı olabilecek analizler sunmasına yardımcı olacaktır. Yapılan araştırmada da metin

¹ Fatma Acun, "Dijital Tarih ve Dijital Tarihçiliğin Tarihyazımına Etkisi Üzerine", **Tarihyazımı**, 2 (1), 2020, Journal of Historiography, s. 66-90.

madenciliğinin temel konuları ele alınıp Nutuk üzerinde uygulanmıştır. Yapay zeka yöntemleri kullanılarak daha kapsamlı analizler yapılması mümkün olmakla birlikte, bu analizleri sosyal bilimcilerle birlikte yapmanın literatüre eşsiz faydaları olacaktır.

Araştırmanın gerçekleştirilmesi için gereken çalışmalar sırasıyla şu şekildedir:

- İngilizce çevirisi elektronik metin formatında elde edilmiştir.
- Analizlerin yapılması için veri kaynağının R programlama dili tarafından okunması sağlanmıştır.
- Metindeki yazım hataları düzeltilmiştir ve noktalama işaretleri kaldırılmıştır.
- Tek başına anlamı olmayan ve çok fazla tekrar eden duraklama kelimeleri, metin içerisinden çıkarıldıktan sonra düzenli bir metin verisi oluşturulmuştur.
- Veri düzenli hale getirildikten sonra metin madenciliği uygulamalarına geçilmiştir.
- Analiz sonuçları veri görselleştirme teknikleri kullanılarak açıklanmıştır.

Çalışma sonucunda temel olarak ulaşılmak istenen amaçlar aşağıdaki gibidir:

- Tarihsel belgelerin hazırlanması, işlenmesi ve raporlanması çok fazla maliyet ve zaman gerektirir. Bu sebeple tarih biliminin, kurum ve kuruluşların, araştırmacıların metin madenciliği yöntemleri ile söz konusu çalışmalarında yarar sağlamak.
- Metin madenciliği yöntemlerini kullanarak tarihsel olayları ayırt etmede, karşılaştırmada, sonuç çıkarmada başarılı sonuçlar elde etmek.
- Dijital çağda eğitim, ekonomi, siyaset, uluslararası ilişkiler gibi güncel olayları açıklamada teknolojinin ve diğer bilimlerin kullanılmasını yaygınlaştırmak.

1. Metin Madenciliği

Enformasyon sistemleri 1960 yıllarında, metin madenciliği çalışmaları 1980 yıllarında başlamış olmasına rağmen, teknolojik yetersizliklerden kaynaklı çalışmalar 2000'lerden sonra hızlanmıştır. Metin madenciliği farklı disiplinleri bir araya getirerek dilbilim, istatistik, veri madenciliği ve makine öğrenmesi konularını birleştirir.

Metin madenciliği yöntemlerinin temelinde matematiksel ve istatistiksel yöntemler bulunmaktadır. Metin madenciliği teknikleri temel olarak dört

kategoriye ayrılmaktadır. Bunlar; sınıflandırma, kümeleme, birliktelik analizi ve bilgi çıkarımıdır. Bu teknikler altında metin madenciliği uygulamaları: Yazar tanıma, fikir madenciliği, metin sınıflama, duygu analizi, anahtar kelime-başlık çıkartımı biçimlerinde sıralanır.²

“Metin madenciliği çalışmaları, metin kaynaklı literatürdeki diğer bir çalışma alanı olan Doğal Dil İşleme (Natural Language Processing, NLP) çalışmaları ile çoğu zaman beraber yol yürümektedir. Doğal dil işleme çalışmaları daha çok yapay zeka altındaki dil bilim bilgisine dayalı çalışmalarını kapsamaktadır. Metin madenciliği çalışmaları ise daha çok istatistiksel olarak metin üzerinden sonuçlara ulaşmayı hedefler. Metin madenciliği çalışmaları sırasında çoğu zaman doğal dil işleme kullanılarak özellik çıkarımı da yapılmaktadır.”³

Metin madenciliği uygulamaları veri madenciliği, iş zekası, güvenlik, yayıncılık, kayıt yönetimi, ayrıca bilimsel metinlerin işlenmesi, duygusal analiz araçları, doğal dil işleme / anlam bilimsel araç veya hizmetler, otomatik reklam yerleştirme, arama / bilgi erişimi gibi konuları içermektedir.⁴

Bu uygulamaların alt uygulamalarında ise pazar araştırmaları, sosyal medya analizleri, müşteri ilişkileri, sahtekarlık tespiti, metinlerden anlam ve bilgi çıkarımı, doküman özeti, sınıflandırma, yazar tanıma ve soru-cevap sistemleri gibi uygulamalar gerçekleştirilmektedir.

Çalışma içeriği yansız bir değerlendirme ile tarihsel olaylar üzerinde metin madenciliği uygulamalarını kapsayacaktır. Bu projenin ana metin kaynağı Nutuk olup, metin madenciliği uygulamaları ile tarihsel olayları gözlemlene, çıkarımda ve tahminde bulunma, karşılaştırma, ilişkilendirme ve görselleştirme çalışmalarını içerecektir.

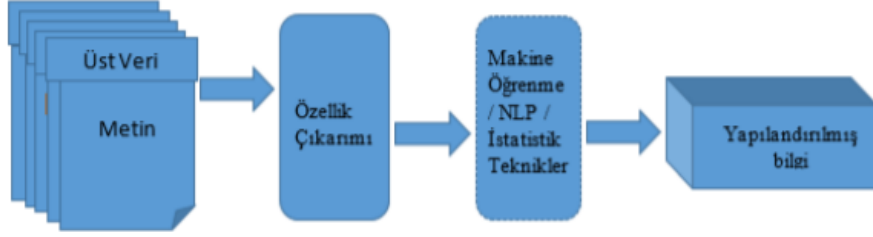
Metin madenciliğindeki amaç analitik yöntemler veya tabii dil işleme yoluyla, metni analiz edilebilmesi için yapılandırılmış veriye dönüştürmek olduğundan temel olarak Şekil 1’deki gibi özetlenebilir.⁵

² Volkan Tunalı, **Metin Madenciliği için İyileştirilmiş Bir Kümeleme Yapısının Tasarımı ve Uygulaması**, Doktora Tezi, Marmara Üniversitesi, Fen Bilimleri Enstitüsü, 2011.

³ Rıfat Aşlıyan, vd., “Metin İçerikli Türkçe Doküman Sınıflandırılması”, **Akademik Bilişim Konferansı Bildirileri**, Muğla Üniversitesi, 10-12 Şubat 2010, s. 659-665.

⁴ Wikipedia, Metin Madenciliği, (https://tr.wikipedia.org/wiki/Metin_madencili%C4%9Fi) adresine bulunan Metin Madenciliği tanımlarından faydalanılmıştır. (Erişim Tarihi: 14 Mayıs 2019).

⁵ Ramazan Acun, “Tarihi Metinlerden Cevher Çıkarmak: Metin Madenciliği”, **Avrupa Tarihinde Türk Eli Doç. Dr. Gümeç Karamuk Armağanı**. Ankara, 2017, s. 27-38.



Şekil 1: Metin madenciliği aşamaları

1.1. Metin Madenciliği Yöntemleri

Enformasyon Getirimi: Metin verisi hakkında bilginin toplandığı ilk aşamadır. Örneğin, web üzerindeki veri kaynakları baz alındığında dosyaların tarihleri, dosya isimleri, kullanıcı bilgileri, izin bilgilerinin toplandığı aşamadır.

Doğal Dil İşleme Aşaması: Metinde bazı anlamsal bilgilerin elde edilmesinde sıklıkla kullanılan bir aşamadır.

Adlandırılmış Varlık Tanıma: Metin içerisinde istatistikî bazı özelliklerin elde edilmesi için kullanılır. Örneğin; kısaltmalar, semboller, yer isimleri vb. bu yaklaşımla bulunur.

Örüntüsü Tanımlı Varlıkların Bulunması: Metnin içerisinden alınan bazı bilgilerin (e-posta adresleri, tarihler, telefon numaraları, adresler vb.) özel olarak alınmasıdır.

Eş Atıf: Bir varlığa işaret eden (atıf eden) isim kelime gruplarını ve diğer terimlerin bulunması/ayrılmasını hedefler.

İlişki, Kural, Olay Çıkarımları: Çeşitli amaçlarla metnin içerisinden bazı bilgilerin çıkarılması istenebilir.⁶

Duygu Analizi: Metin içerisinde geçen ifadelerin duygularını ortaya çıkarır. Temel olarak metinleri olumlu-olumsuz şeklinde iki gruba ayırır. Olumlu-olumsuz duyguyu özelleştirerek daha farklı duygular da tespit edilebilir.

2. Nutuk'un Tarihsel Niteliği ve Önemi

Nutuk, yazıldığı dönemde Cumhuriyet Halk Fırkası Genel Başkanı olan Mustafa Kemal Atatürk'ün 15-20 Ekim 1927 tarihlerinde, yerli ve yabancı basın mensuplarının da katıldığı partisinin 2. yılı olan 15-20 Ekim 1927

⁶ Şadi Evren Şeker., vd., "TimeML and Turkish Temporal Logic", **Proceedings of International Conference on Artificial Intelligence**, ICAI 2010, s. 881-887.

tarihleri arasında kendisinin ve silah arkadaşlarının faaliyetlerini özetleyen konuşmadır.⁷ Sunumu günlük beşer altışar saatlik oturumlarla altı günde tamamlanmış ve toplam 36 saat 31 dakika sürmüştür.⁸

Nutuk'ta olaylar, 19 Mayıs 1919'dan başlayarak tarihsel olayları kronolojik bir sıra içinde ve kronolojik sıraya ek olarak Mustafa Kemal Atatürk'ün kendi görüşlerini de tarihsel olaylarla birlikte açıkladığı, ancak diğer kitaplarda olduğu gibi konular bölümlere, alt bölümlere ayrılmadan, sürekli bir akış içinde anlatılmaktadır.

Nutuk'un yazarı yalnız olayları belgelerle inceleyerek sonuçlarını ortaya koymaya çalışan bir tarih yazarı değil, aynı zamanda eserdeki olayların da doğrudan kahramanı olarak bizzat kendisidir. Bu eserde tarih yapan ile tarih yazan aynı kişi olduğu için üstün değere sahip bir eser niteliği taşımaktadır. Bu haliyle eser tarihi olayların yanında siyasal olayları da anlatması bakımından çok ciddi bir yapıt olup tarih yöntemiyle kaleme alınmıştır.⁹

Atatürk'ün Nutuk'u yazmasındaki amacı; Milli Mücadele sırasında meydana gelen olayları açıklığa kavuşturmak, tarihe katkıda bulunmak ve Türk milletinin istikbali için bir kaynak bırakmaktır. Bu kaynağı da Gençliğe Hitabe ile bitirerek Türk gençliğini vazifelendirmiş ve Türk gençliğinin Türk bağımsızlığını, Türk Cumhuriyeti'ni, ilelebet, muhafaza ve müdafaa etmelerini söylemiştir. Nutuk, Türk istiklalini ve Türk Cumhuriyeti'ni korumanın manifestosudur.

Atatürk, Nutuk'ta Gençliğe Hitabe'den önce, Nutuk'un amacını bizlere anlatmaktadır.

“Muhterem Efendiler, sizi günlerce işgal eden uzun ve teferruatlı beyanatım, en nihayet mazi olmuş bir devrin hikâyesidir. Bundan, milletim için ve müstakbel evlâtlarımız için dikkat ve teyakkuza davet edebilecek bazı noktalar tebarüz ettirebilmiş isem, kendimi bahtiyar addedeceğim”.

Efendiler, bu beyanatımla millî hayatı hitam bulmuş farz edilen büyük bir milletin istiklâlini nasıl kazandığını ve ilim ve fennin en son esaslarına müstenit, millî ve asrî bir devleti nasıl kurduğunu ifadeye çalıştım.

⁷ Sinan Meydan, **Nutuk'un Deşifresi**, Truva Yayınları, önsöz.

⁸ İsmail Arar, “Büyük Nutuk'un Kapsamı, Niteliği, Amacı”, **Atatürk'ün Büyük Söylevi'nin 50. Yılı Semineri – Bildiriler ve Tartışmalar**”, TTK, Ankara, 1980, s. 126.

⁹ Zeynep Korkmaz, Ergün Aybars, “Nutuk”, **70. Yılında Ulusal ve Uluslararası Boyutlarıyla Atatürk'ün Büyük Nutku ve Dönemi Sempozyumu**, Yay. Haz. Gül E. Kundakçı, Ankara 1999, s. 8.

Bugün ulaştığımız sonuç, asırlardan beri çekilen milli felaketlerin yarattığı uyanıklığın eseri ve bu aziz vatanın her köşesini sulayan kanlarıdır bedelidir.

Bu sonucu, Türk Gençliğine emanet ediyorum”¹⁰



Resim 2: TBMM’de Nutuk’u okurken, Ankara, 15-20 Ekim 1927.

Yurdumuzun parçalanıp işgal edildiği günlerden başlayarak, Türk tarihinde bir dönüm noktası olan İstiklal Savaşı’nı, Türkiye Cumhuriyeti’nin kuruluşunu ve inkılapların yapışım anlatan Nutuk, siyasi ve milli tarihimizin birinci elden, çok değerli bir kaynak eseridir. Bu nedenle Nutuk bir bir mesajdır, mirastır. Atatürk’ün yaptıkları, yapacakları, düşünceleri, tecrübe ve birikimleri bu eserde. Nutuk ayrıca bir şahitlik belgesidir. Bu şahitlik, bizzat Türk Milleti’nin huzurunda, Anadolu’daki son Türk Devleti’nin ve Türk Milleti’nin gelecekteki güzel günleri için atılan adımlar ve imzalar içindir.¹¹

¹⁰ Mustafa Kemal Atatürk, **Nutuk**, Alfa Yayınları, s. 735.

¹¹ Hasan Cicioğlu, Atatürk’ün Büyük Nutuk’ta Adları Geçen Muhalifleri, Muhalefet Gerekçeleri, Atatürk’ün Cevabı, **75. Yılında Nutuk’u Anlayarak Okumak, Bilgi Şöleni**, 17-18 Ekim 2002, s. 111-155

Kısaca Nutuk, Mustafa Kemal'in Türk İstiklal Harbi'ni anlattığı tarihi ve siyasi bir metindir. Hasan Ali Yücel ise Nutuk'u şöyle ifade etmiştir; "*Bize bıraktığı çağdaş Türkiye'nin kurtuluş tarihi.*"¹²

Nutuk sadece ulusal bir eser değil aynı zamanda uluslararası bir değere sahiptir. Türklerin emperyalist devletlere karşı bağımsızlık mücadelesi ve başarısı tüm mazlum milletlere örnek olmuştur. Nutuk, tüm dünya halklarına, ulusal kurtuluşu ve çağdaş devlet kurmanın yönlerini anlatan bir öğreti kitabıdır.¹³ Tüm uluslara bağımsızlık savaşı, modern bir toplum kurma ve toplumsal devrimleri anlatan bu öğreti kitabı farklı dillere çevrilmiştir. Nutuk'un 1927 yılı basımı 1928 harf devriminden önce olduğu için Osmanlıca'dır. Türkçe olarak birinci cildin ilk baskısı 1928 yılında basılmıştır.¹⁴ Nutuk'un basıldığı diğer diller ise; İngilizce, Fransızca, Almanca, Rusça, Farsça, Türkmence, Kazakça ve Japonca dillerine çevrilmiştir. Arapçaya da çevirisinin yapılması yönünde talepler olmasına rağmen, Arapça çevirisi yapılmamıştır.¹⁵

Mustafa Kemal Atatürk'ün kaleme almış olduğu başka eserleri de olmasına rağmen, Nutuk, içerik, konu anlam bütünlüğü ve anlatış tarzı olarak diğer eserlerine göre daha çok ön plana çıkmıştır. Nutuk, başta sosyal bilimciler olmak üzere farklı alanlardan pek çok uzmanın dikkatini çekmiş, inceleme ve araştırma konusu olmuştur. Nutuk'un önemi, amacı, kapsamı gibi konularda araştırmanın yanı sıra ve Nutuk'un tarih eseri olup olmadığı bağlamında tartışmalar bulunmaktadır.

Literatürde karşılaşılan analizlerden biri olan terim ve söz gruplarının analizinde, Nutuk'ta kavramların nasıl ifade edildiği ve Atatürk'ün bu kavramları nasıl kullandığını gösteren kavram ve terim örneklemlerinden yararlanılarak çeşitli istatistiksel sonuçlara ulaşılmıştır.¹⁶ Bir diğer çalışmada ise içerik analizi ile Nutuk'ta geçen kurum ve kişilerle ilgili olumlu olumsuz ifadeler, bu ifadelerin frekansları, yıllara göre düşünce farklılıkları gibi

¹² Hasan Gürkan, "Atatürk'ün Söylev'i (Nutuk) Üzerine Bir Bibliyografya Denemesi", **Çağdaş Türkiye Tarihi Araştırmaları Dergisi**, XIII/27 (2013-Güz), s. 123-136.

¹³ Cavit Orhan Tütengil, "Söylev'in Öğretisi", **Türk Dili Dergisi**, (Söylev Özel Sayısı), Yıl: 27, Cilt: XXXVI, Sayı: 314, 1 Kasım 1977, s. 373.

¹⁴ Mehmet Kılıç, "Büyük Nutuk'un Cumhuriyet Tarihindeki Yeri ve Önemi", **Türk Dünyası Araştırmaları** 121 (2019): s. 113-136

¹⁵ Bilal N. Şimşir, Atatürk'ün Büyük Söylevi Üzerine Belgeler, Ankara, TTK, 1991.

¹⁶ Hüseyin Ağca, "Atatürk'ün Bütük Nutuk'unda Geçen Milli Kimlikle İlgili Kavram, Terim ve Söz Gruplarının Analizi", **75. Yılında Nutuk'u Anlayarak Okumak, Bilgi Şöleni**, 17-18 Ekim 2020, s. 106-110.

konular incelenmiştir.¹⁷ Literatürdeki bu araştırmalara ek olarak, metin madenciliği, doğal dil işleme, makine öğrenmesi gibi yapay zeka uygulamalarının Nutuk üzerinde işlenmesiyle daha çok araştırma sorusu ve problemi ortaya çıkabilir.

3. Metin Madenciliği Uygulamaları

3.1. Veri Ön İşleme ve Düzenli Metin Formatının Oluşturulması

Araştırmadaki analizler R programlama dili yardımı ile yapılmıştır. Metin madenciliğinin en büyük problemi, analizi yapılacak olan verinin yapısal olmamasıdır. Bu yüzden metin verisi üzerinde veri ön işleme yapılarak anlamsızlıklar ortadan kaldırılmıştır. Bu anlamsızlıklar, metin içerisindeki simgeler, noktalama işaretleri gibi tek başına anlamı olmayan ifadelerin çıkarılması sağlanmıştır. Ayrıca metin içerisindeki tüm kelimeler küçük harflere çevrilmiştir. Daha sonrasında analiz edilmek istenen metin, cümlelere ve kelimelere parçalanarak düzenli metin formatı haline getirilmiştir. Metin madenciliği ve doğal dil işleme uygulamaları Türkçe dili için henüz yeterli gelişmişliği göstermediğinden, analizler için İngilizce Nutuk tercih edilmiştir.

```
library(dplyr) # Veri Manipülasyonu
library(tidytext) # Düzenli Metin Araçları
library(stringr) # Metin Manipülasyonu
library(ggplot2) # Veri Görselleştirme
library(tidyr) # Veri Seti Manipülasyonu
library(scales) # Veri Görselleştirmesi için Standardize Fonksiyonları
library(wordcloud) # Kelime bulutu
library(wordcloud2) # Fotoğraf üzerinde Kelime Bulutu
library(reshape2) # Kelime Bulutu
library(igraph) # Ağ Analizi
library(ggraph) # Ağ Analizi
library(widyr) # Frekans ve Korelasyon İşlemleri
library(tm) # Matris İşlemleri
library(Matrix) # Matris İşlemleri
library(quanteda) # Metinsel Veriler için Kantitatif Analiz
library(topicmodels) # Konu Modellemesi
library(methods) # Resmi Metodlar ve Sınıflar
```

Şekil 4: Metin madenciliği uygulamaları için R programlama dilinde kullanılan kütüphaneler.

Nutuk ayrıca dönemlere, bölümlere ve yıllara göre parçalara bölünüp, karşılaştırma yapmak için hazırlanmıştır.

¹⁷ Hakan Uzun, “Atatürk’ün Nutuk’unun İçerik Analizi”, **CTAD: Hacettepe Üniversitesi Atatürk İlkeleri Ve İnkılâp Tarihi Enstitüsü Cumhuriyet Tarihi Araştırmaları Dergisi**, Yıl 1, 2, 125–143, 2005, s.137.

Nutuk dönemselsel olarak “Meclis Öncesi Dönemi (19 Mayıs 1919 – 23 Nisan 1920)”, “I.Meclis Dönemi (23 Nisan 1920 - 11 Ağustos 1923)” ve “II.Meclis Dönemi (11 Ağustos 1923-1 Ekim 1927) olarak üçe; yıllara göre 1919-1925 arasında bölümlere ayrılmıştır. Ayrıca Nutuk’ta olaylar, 19 Mayıs 1919’dan başlar ve ardı ardına anlatılmaktadır. Veri kaynağında Nutuk, diğer kitaplar gibi bölümlere veya alt bölümlere ayrılmamıştır. Milli Savunma Üniversitesi Kara Harp Okulu’nun internet sitesinde Nutuk yer almaktadır. Kara Harp Okulu’nun sitesi kaynak olarak alınıp konu başlıkları veri setine yeni bir değişken olarak eklenmiş ve Nutuk bölümlere ayrılmıştır. Bu ayırma işlemi ile bölümler arasında karşılaştırma yapmak kolaylaşacaktır.¹⁸

THE.GREAT.SPEECH	book	linenumber	chapter
1 CHAPTER-1	Nutuk	1	1
2 Gentlemen,	Nutuk	2	1
3 I landed at Samsun on 19 May 1919. This was the situati...	Nutuk	3	1
4 The group of Powers which included the Ottoman Gove...	Nutuk	4	1
5 The Army had been deprived of their arms and ammuniti...	Nutuk	5	1
6 The Entente Powers did not consider it necessary to res...	Nutuk	6	1
7 Christian elements were also at work all over the countr...	Nutuk	7	1
8 Certain information and authentic documents that fell i...	Nutuk	8	1
9 (Document I), established by the patriarchate in Istanbul...	Nutuk	9	1
10 The Armenian Patriarch, Zaven Efendi, also worked in co...	Nutuk	10	1

Şekil 5: Nutuk, dönemler ve bölümler – Düzenli metin formatı

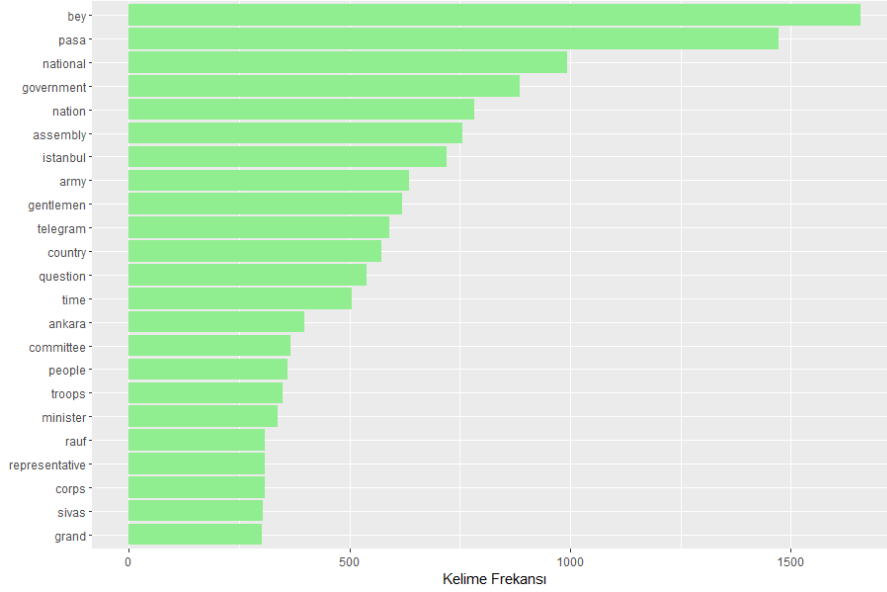
Şekil 5’te görüldüğü gibi veri setindeki “THE.GREAT.SPEECH” değişkeni metin içerisindeki her bir paragrafı temsil etmektedir ve bu paragraflar da “linenumber” değişkeni ile temsil edilmiştir. Veri setinde hem Nutuk hem de dönemler yer aldığından “book” değişkeni içerisinde “Nutuk, Meclis Öncesi, I.Meclis ve II.Meclis Dönemleri” yer almaktadır. “chapter” değişkeninde ise Kara Harp Okulu’ndan alınan Nutuk bölümleri gösterilmiştir. Bu işlemin nedeni, yapılan analizlerin karşılaştırmasını yapmayı kolaylaştırmak içindir.

3.2. Kelime Frekanslarının Oluşturulması

Düzenli metin madenciliğine göre metin içerisinde saklanan bilgi genelde tek bir kelimedir. Ancak bu bilgi bir n-gram (n sayısına göre metnin

¹⁸ Kara Harp Okulu, Nutuk, (http://www.kho.edu.tr/hakkında/harbiyeli_ataturk/nutuk/nutuk.html), adresinden Nutuk’un bölümlendirilmesi konusunda yararlanılmıştır. (Erişim Tarihi: 14 Mayıs 2019).

kelimelere parçalanması), cümle veya paragraf da olabilir. Bu yüzden, Nutuk n-gramlar'a yani birer, ikişer (bigram), üçer (trigram) kelimelere parçalanmıştır. Bu araştırma içerisindeki çıktılar birer ve ikişer kelimelere parçalama işlemi sonrası yapılan analizler ve görselleştirmelerini içermektedir. Parçalama işleminden sonra analizlerin gerçekleştirilmesi için kelime frekanslarına bakılır. Kelime frekansları çıkartıldığında metin içerisinde çok fazla tekrar eden fakat tek başına bir anlamı olmayan kelimelere duraklama kelimeleri (stop words) denir. Çalışma, İngilizce metin üzerinden yapıldığından İngilizce'deki duraklama kelimeleri; the, of, at gibi metin içerisinde çok fazla tekrar eden ancak tek başına bir anlamı bulunmayan kelimeler, analizlerin anlamlılığı için metinden çıkartılmıştır. Şekil 6'da ise Nutuk içerisinde en fazla tekrar eden kelimeler gösterilmiştir.



Şekil 6: Nutuk içerisinde en fazla tekrar eden kelimeler.

Nutuk karakterlere, kurumlara ve olaylara dayanmaktadır. Metin içerisinde “bey” ve “paşa” kelimeleri en fazla tekrar eden kelimelerdir. Nutuk içerisinde çok fazla sayıda kişinin ismi geçmektedir. Literatürde Nutuk üzerine yapılan dizin ve içerik analizlerinde tekil kişi sayısının farklı hesaplandığı ve bir belirsizliğin olduğu görülmüştür¹⁹. Ayrıca metin

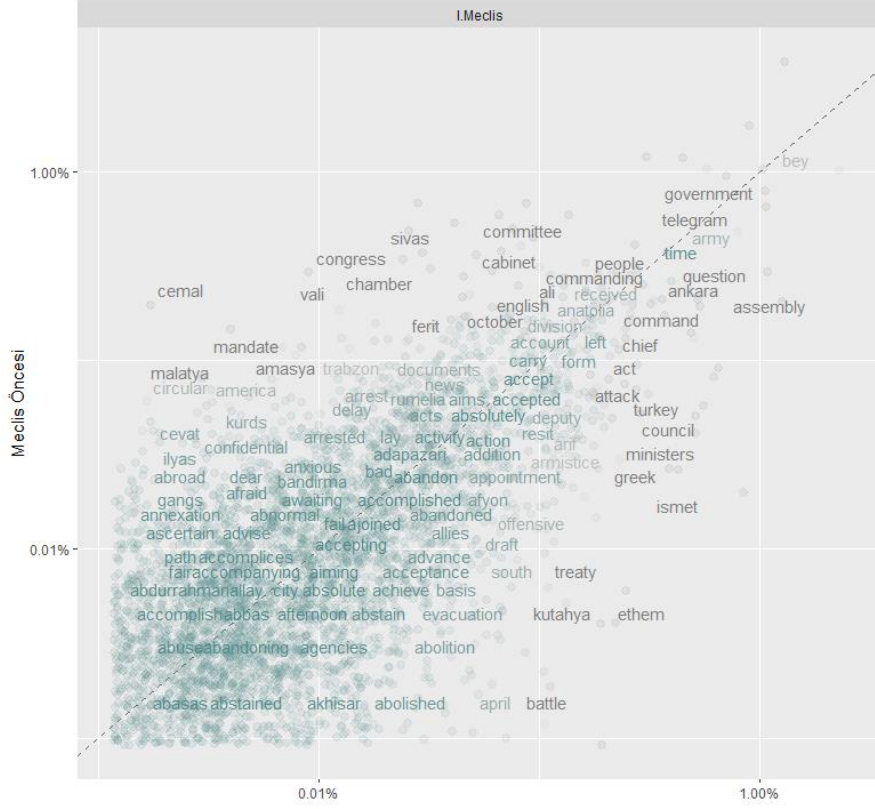
¹⁹ Bu konu hakkında iki farklı makale ile karşılaşmıştır. Fahri Kılıç, “Nutuk'un Analitik Bir İncelemesi”, *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 3 (1), 2016, s. 137. Hakan Uzun, “Atatürk'ün Nutuk'unun İçerik Analizi”, *CTAD: Hacettepe*

madenciliği ve doğal dil işleme yöntemleriyle metinde geçen kişiler ve frekansları tespit edilebilir. Ancak 1919-1927 yılları arasında soyadı kanunu olmadığından ve çok fazla kişi bulunduğundan aynı isimli farklı kişilerin karıştırılması da mümkündür. Bu yüzden tarihsel metinlerdeki kişilerin tespitinde daha detaylı bir analiz çalışması yapılmalıdır.

Kelime frekansları bizlere konu hakkında genel bilgi sunmaya çalışır. Diğer kelimeler incelendiğinde Milli Mücadele sürecinde ve kitapta büyük rol alan çeşitli konular çıkarılabilir. İstanbul Hükümeti ile olan sorunlar ve ilişkiler, kongreler, meclis, ordu, ulus gibi konuların sıklıkla dile getirildiği bu grafikte görülebilir.

3.3. Kelimeler ve Dönemler Arasındaki İlişki

Nutuk içerisinde karşılaştırma yapmak için dönemlere ayrıştırılmıştır. Dönemler arasında bir benzerlik olup olmadığına bakmak için dönemler ikili gruplar halinde incelenmiştir ve karşılaştırmalar Şekil 7'deki gibidir. Şekil 7'de örnek olarak Meclis Öncesi ve I.Meclis dönemleri ele alınmıştır. Grafikteki doğrusal kesikli çizgiye yakın olan kelimeler, her iki metin grubunda da benzer frekanslara sahiptir. Kelimeler doğrusal çizgiden uzaklaştıkça iki dönemden birinde ortak görünürlüğü azalır.



Şekil 7: Meclis Öncesi ve I.Meclis karşılaştırması.

Meclis Öncesi ve I.Meclis Dönemleri'nin karşılaştırılmasında doğrusal çizgiye yakın olan kelimeler iki dönem içerisinde de ortak olarak gözlenirken doğrusal çizgiden uzaklaştıkça dönemler arasındaki farklılıklar gözükmemektedir. Örneğin Meclis Öncesi Dönem'de "Sivas", "Amasya", "Congress (Kongre)", "mandate (manda)" gibi kelimeler, I.Meclis Dönemi'nden farklı olarak ortaya çıkarken, I.Meclis Dönemi'nde "İsmet", "Ethem", "Kütahya", "Greek (Yunan)", "Battle (Savaş)" gibi kelimeler ortaya çıkmaktadır.

Bu grafik dönemler arasındaki benzerlikleri ve farklılıkları göstermektedir. Örneğin İsmet İnönü, Meclis Öncesi Dönem'de çok fazla bahsi geçmemekle birlikte, yükselişinin I.Meclis Dönemi ile başladığı sonucuna varılabilir. Ayrıca şehirlere bakıldığında Meclis Öncesi Dönem'de Sivas ve Amasya Kongre şehirleri bulunurken I.Meclis Dönemi'nde Kurtuluş Savaşı'nın gerçekleştiği şehirlerden biri olan Kütahya şehri göze çarpmaktadır.

Bu analiz aynı zamanda Meclis Öncesi - II.Meclis Dönemleri ve I.Meclis - II.Meclis Dönemlerin karşılaştırılması için de yapılmıştır. Ayrıca dönemler arasındaki ilişkiyi bilimsel olarak gösterebilmek için korelasyon analizleri de yapılmıştır.

3.4. Kelimeler ve Dönemler Arasındaki İlişki: Korelasyon Analizi

Korelasyon, nedensellik bağlamı dışında iki rassal değişken arasındaki doğrusal ilişkinin yönünü ve gücünü açıklar.

Kelime frekans kümelerinin birbirlerine olan benzerlikleri korelasyon analizi ile ölçülmüştür. Bu ölçüm ile dönemlerdeki geçen kelime ilişkilerinin istatistiksel olarak anlamlı bir fark içerip içermediği hipotez testleri yardımıyla incelenmiştir.

Dönemler arasındaki ilişkinin doğrusallığına bakmak için Pearson korelasyon katsayısı kullanılmıştır. Parametrik olmadığı düşüncesi ile bu probleme yaklaşırsa Spearman korelasyon katsayısı veya kategorik veri analizi uygulamalarındaki diğer korelasyon katsayıları da kullanılabilir.

3.4.1. Pearson Korelasyon Katsayısı

- Meclis Öncesi ve I. Meclis Dönemi Arasındaki İlişki

H0: Meclis Öncesi Dönemi ve I.Meclis Dönemi arasında bir ilişki yoktur.

H1: Meclis Öncesi Dönemi ve I.Meclis Dönemi arasında bir ilişki vardır.

Pearson's product-moment correlation

```
data: proportion and Meclis öncesi
t = 102.59, df = 3747, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.8501050 0.8669275
sample estimates:
      cor
0.8587475
```

Şekil 8: Pearson korelasyon katsayısına göre Meclis Öncesi ve I.Meclis Dönemleri arasındaki ilişki.

Şekil 8'de H0 hipotezi 0.95 güven aralığında reddedilir. Hipotez testinin reddedilmesi, Meclis Öncesi ve I. Meclis Öncesi Dönemleri arasında bir ilişki olduğunu işaret etmektedir. Korelasyon değerine bakıldığında 0.8587

değeri ile Meclis Öncesi Dönem ve I.Meclis Dönemi arasında çok yüksek düzeyde bir ilişki bulunduğu söylenebilir.

Ayrıca I.Meclis ve II.Meclis Dönemleri arasındaki korelasyon değeri 0.7580, Meclis Öncesi Dönemi ve II.Meclis Dönemi arasındaki korelasyon değeri 0.6657 olarak çıkmaktadır.

3.4.2. Spearman Korelasyon Katsayısı

Spearman's rank correlation rho

```
data: proportion and Meclis Öncesi
S = 2.815e+09, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.6794571
```

Şekil 9: Spearman korelasyon katsayısına göre Meclis Öncesi ve I.Meclis Dönemleri arasındaki ilişki

Pearson korelasyon katsayısında olduğu gibi Şekil 9'da da H_0 hipotezi 0.95 güven aralığında reddedilir. Meclis Öncesi Dönemi ve I. Meclis Öncesi Dönemi'nde istatistiksel olarak bir ilişki olduğu görülmektedir. Korelasyon değerine bakıldığında 0.6794 değeri ile Meclis Öncesi Dönem ve I.Meclis Dönemi arasında orta düzeyde bir ilişki bulunduğu söylenebilir.

Ayrıca I.Meclis ve II.Meclis Dönemleri arasındaki korelasyon değeri 0.5621, Meclis Öncesi Dönemi ve II.Meclis Dönemi arasındaki korelasyon değeri 0.5134 olarak çıkmaktadır.

Pearson korelasyon katsayısının sonuçlarına bakıldığında, Nutuk'ta olaylar kronolojik bir sıra içinde ve ardı ardına sürekli bir akış içinde anlatıldığında dönemler arasında korelasyonun yüksek çıkması beklenen bir sonuçtur. Aynı zamanda konu anlam bütünlüğü açısından da Nutuk kendi içerisinde çelişmemektedir. Meclis Öncesi - I.Meclis Dönemleri ve I.Meclis - II.Meclis Dönemleri arasında korelasyonun yüksek çıkması konu anlam bütünlükleri açısından beklenen bir durum iken Meclis Öncesi - II.Meclis Dönemleri arasındaki korelasyonun orta düzeyde çıkması Milli Mücadelenin başlangıcı ve sonu arasında farklılık gösterdiği yani bir değişimin olduğu anlaşılmaktadır. Spearman korelasyon katsayısına göre dönemler arası ilişkiler orta düzeyde olduğu görülmektedir.

Meclis Öncesi Dönem'de Milli Mücadele'nin tasarısı yapılmıştır ve ilk adım gerçekleştirilmiştir. I.Meclis Dönemi'nde ise Milli Mücadele'nin fiilen harekete geçirildiği, Anadolu'nun işgaline karşı aktif savaşların yapıldığı ve

savaşlar sonucunda Kuva-yı Milliye hareketi başarılı olduğu anlatılmıştır. Ayrıca saltanatın da kaldırılmasıyla yeni ve bağımsız bir Türk devletinin kurulmasına temel atılmıştır. II.Meclis Dönemi'nde ise başarılı olan Milli Mücadele sonucunda yeni Türk Devleti'nin kurulması, Cumhuriyet'in ilanı ve meclis içerisindeki muhalefet ele alınmıştır.

3.4.3. Phi Korelasyon Katsayısı

N-gram metin madenciliğindeki ideal araçlardan bir tanesidir. Metni kelimelere parçalamak ve kelime çiftlerini keşfetmek çok kolaydır. Ek olarak n-gram yardımıyla kelimeler birbirlerinin yanında bulunmasalar bile, belirli belgeler veya belirli bölümlerde birlikte olma eğiliminde olan kelimeler ile de ilgilenme imkanı tanımaktadır.

Bazı kelime çiftleri en çok rastlanılan kelimelerdir. Ancak bu kelimeler aynı zamanda en yaygın bireysel kelimeler oldukları için özellikle anlamlı değildir. Bunun yerine, birbirleriyle ne sıklıkta görüldüklerine bağlı olarak kelimeler arasındaki ilişki incelenmelidir.

Özellikle burada ikili korelasyon için Phi katsayısına odaklanmak gerekir. Phi katsayısının odak noktası, X ve Y kelimelerinin her ikisinin de diğeri olmadan görüldüğünden ne kadar fazla ortaya çıktığıdır.

Örneğin, n_{11} X ve Y kelimesinin bulunduğu yerin frekansını, n_{00} X ve Y kelimesinin bulunmadığı yeri, n_{10} veya n_{01} ise X ve Y kelimelerinden birinin bulunduğu yerin frekansını verir. Phi katsayısı da Şekil 10'daki formülde ifade edilmiştir.

$$\phi = \frac{n_{11}n_{00} - n_{10}n_{01}}{\sqrt{n_{1.}n_{0.}n_{.0}n_{.1}}}$$

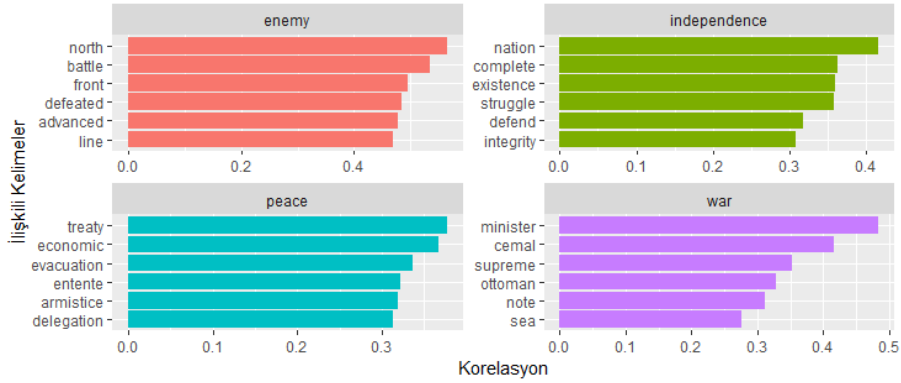
Şekil 10: Phi katsayısı formülü.

Düzenli veri, değişkenler arasında karşılaştırma yapmak veya satırlara göre gruplandırmak için faydalı bir yapıdır. Ancak satırlar arasında karşılaştırma yapmak zor olabilir. Örneğin, aynı belgede iki kelimenin kaç kez tekrar ettiğini ve bunların nasıl ilişkili olduğunun görülmesi. Kelime eşlerinin sayımı veya korelasyonu bulmak için öncelikle verileri geniş bir matrise dönüştürmek gerekir. Aşağıdaki tablo elde edilen korelasyon değerine göre büyükten küçüğe sıralanmıştır ve ikili kelimelerin birbiriyle olan ilişkisini göstermektedir. İlk 10 gözlem incelendiğinde özel isimler ortaya çıkmaktadır.

	item1	item2	correlation
1	nadi	yunus	1.0000000
2	yunus	nadi	1.0000000
3	celalettin	arif	0.9683556
4	arif	celalettin	0.9683556
5	flying	columns	0.9679416
6	columns	flying	0.9679416
7	kaptan	yahya	0.9297209
8	yahya	kaptan	0.9297209
9	sami	bekir	0.8940670
10	bekir	sami	0.8940670

Şekil 11: Bigramlar (İkişer kelimeler) arasındaki korelasyon.

Veri görselleştirme yardımıyla kelimeler arası ilişkiler daha net ortaya çıkartılabilir. Aşağıdaki grafikte Nutuk içerisinde geçen dört farklı kelime seçilmiştir. Grafik, kelimelerin diğer kelimeler ile arasındaki ilişkiyi göstermektedir. En yüksek ilişkiye sahip ilk 6 kelime grafikte gösterilmiştir. “enemy” (düşman), “independence” (bağımsızlık), “peace” (barış) ve “war” (savaş) kelimelerinin en yüksek korelasyona sahip oldukları diğer kelimeler Şekil 12’deki gibidir.



Şekil 12: Bigramlar (İkişer kelimeler) arasında korelasyon.

Savaş içerisinde bulunduğundan “enemy” (düşman) kelimesine bakıldığında savaş ile alakalı kelimeler arasında bir korelasyon olduğu görülmektedir. “independence” (bağımsızlık) kelimesinde ise bağımsız “nation” (ulus), tam bağımsızlık, bağımsız yaşamak, mücadele, savunma ve bütünlük gibi kelimeler ile korelasyona sahiptir. “peace” (barış) kelimesi

antlaşma, ekonomik, itilaf, ateşkes gibi kelimeler ile korelasyon içerisindedir. “war” kelimesinde ise Savunma Bakanı Cemal Paşa ve Osmanlı kelimeleri ön plandadır.

Korelasyon analizi metin içerisindeki ilişkileri istatistiksel olarak anlamlandırabilir. Genel açıdan korelasyon analizi bir metin verisi üzerine uyarlandığına araştırmacıya pek çok ön bilgi sağlamaktadır.

3.5. Düzenli Metin ile Duygu Analizi

Metin madenciliğinde metinlerin içerisindeki anlamları daha net bir şekilde ortaya çıkartmak için duygu analizi yapılmaktadır. Duygu analizi metin içerisindeki kelimeleri olumlu ya da olumsuz olarak sınıflandırır. Ayrıca metinlerde kullanılan dili anlamaya yardım eder.

Metin içerisindeki duyguları ortaya çıkarmak için duygu sözlükleri kullanılır. Duygu sözlüklerinin veri ile içsel birleşimleri sağlanarak duygu analizi yapılmış olur. Genel olarak kullanılan üç sözlük vardır:

- AFINN (Finn Arup Nielsen),
- Bing (Bing Liu ve ortakları),
- NRC (Saif Mohammad ve Peter Turney).

AFINN sözlüğü, olumlu-olumsuz şeklinde her bir kelimenin -5 ile +5 arasında sayısal skoru bulunmaktadır. Bing sözlüğünde olumlu-olumsuz şeklinde iki kategoride bulunur. NRC sözlüğünde ise pozitif, negatif, öfke, beklenti, tiksinti, korku, sevinç, üzüntü, şaşkınlık ve güven (positive, negative, anger, anticipation, disgust, fear, joy, sadness, surprise ve trust) gibi kategoriler bulunur. Bu üç duygu sözlüğü haricinde başka duygu sözlükleri de bulunmaktadır.²⁰

Mustafa Kemal Atatürk Nutuk'ta Milli Mücadele süresince içinde bulunulan durumları sıklıkla dile getirmiştir.

Mustafa Kemal 19 Mayıs 1919'da Samsun'a çıkarken ülkenin genel durumu ve görünüşü şöyleydi:

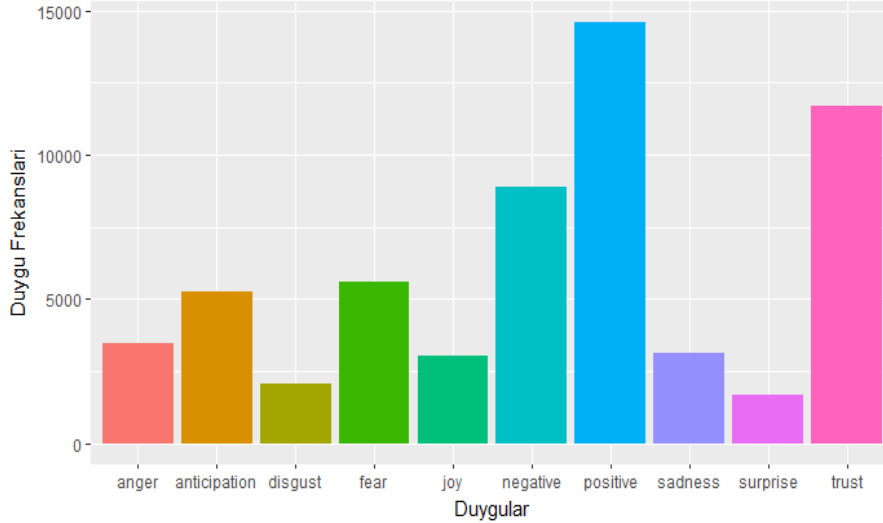
I.Dünya Savaşı'nda mağlup olan taraftan Osmanlı Devleti, ordusu büyük zarar görmüş, ateşkes antlaşmasıyla dağıtılmıştır. Uzun yıllar süren savaş nedeniyle Anadolu halkı fakir ve yorgun bir duruma düşmüştür.

²⁰ Julia Silge, vd., “tidytext: Text Mining and Analysis Using Tidy Data Principles in R.” **JOSS 1 (3). The Open Journal**, 2016.

İstanbul Hükümeti, bütün bu gelişmelere karşı aciz kalmıştır. Anadolu'nun birçok toprağı İtilaf Devletleri tarafından işgal edilmiştir.²¹

Kısaca Atatürk, ülkenin genel durumunu Nutuk'ta bu şekilde özetlemiştir. Ayrıca Nutuk'un ilerleyen sayfalarında bu durumla ilgili Atatürk, o dönemde Osmanlı Halkı içerisinde düşünülen kurtuluş çarelerinin insanlar tarafından üç türlü görüşe sahip olduklarını belirtmiştir. Birinci görüş, İngiliz Mandasını isteyenler; ikinci görüş Amerikan Mandasını isteyenler ve üçüncü görüş, bölgesel kurtuluş mücadelesine başvurmak isteyenlerden oluştuğunu belirtmiştir. Atatürk bu görüşlerin temellerinin çürük olduğuna ve Anadolu halkının kurtuluşu için bu görüşleri reddedip milli egemenliğe dayanan, kayıtsız, şartsız, bağımsız yeni bir Türk Devleti kurma fikrinin gerçek ve tam bağımsızlığa ulaşmanın asıl görüş olmasına karar vermiştir.

Duygu analizinde yani Şekil 13'te de görüldüğü gibi Nutuk içerisinde en fazla olumlu kelimeler ortaya çıkmaktadır. Grafikte "positive" (olumlu) ve "trust" (güven) sütunlarının olumsuz duygulara göre daha fazla olduğu görülmektedir.



Şekil 13: Nrc sözlüğüne göre Nutuk'un duygu analizi.

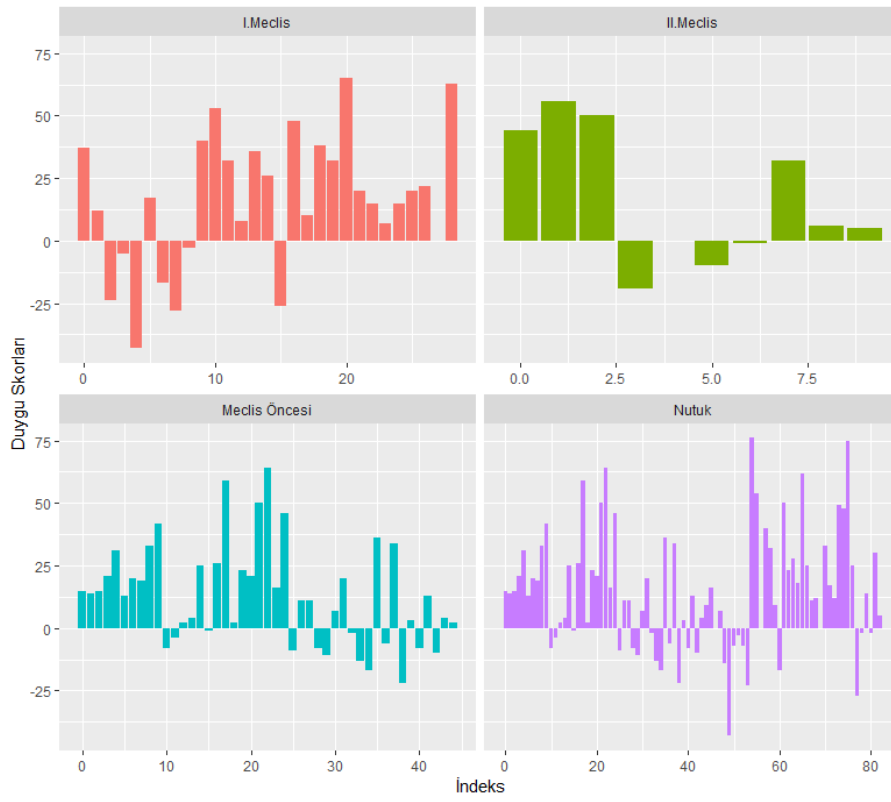
Atatürk'ün 1919'da Milli Mücadele'ye olan inancı ve tüm olumsuz şartlara rağmen Milli Mücadele'nin başarısı burada görülmektedir. Atatürk,

²¹ Mustafa Kemal Atatürk, **Nutuk**, Alfa Yayınları, s. 14-15.

Milli Mücadele'nin amacına, durumların çözülmeye çalışılmasına karşı her zaman olumlu ve güven içerisinde yaklaşmıştır.

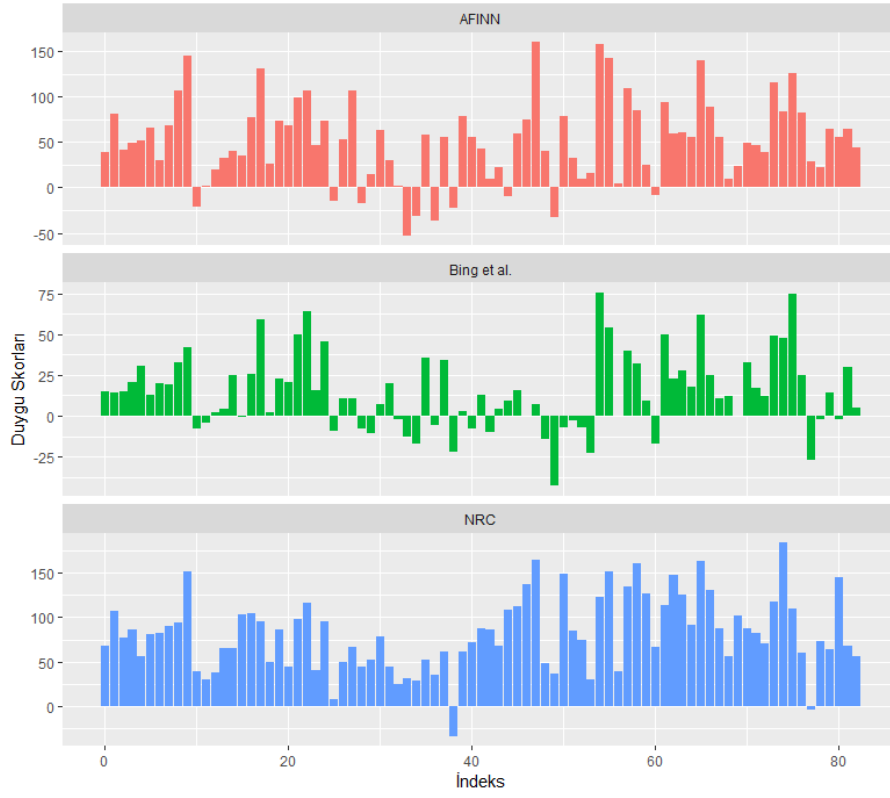
Şekil 13'teki grafiği doğrulamak adına farklı duygu sözlükleriyle Nutuk içerisinde yer alan akışın olumlu olup olmadığına aşağıdaki grafikler yardımıyla bakılmıştır.

Bing sözlüğüne göre (Şekil 14) Nutuk ve dönemlere bakıldığında genel itibariyle Nutuk olumlu bir metin olarak gözükmektedir.



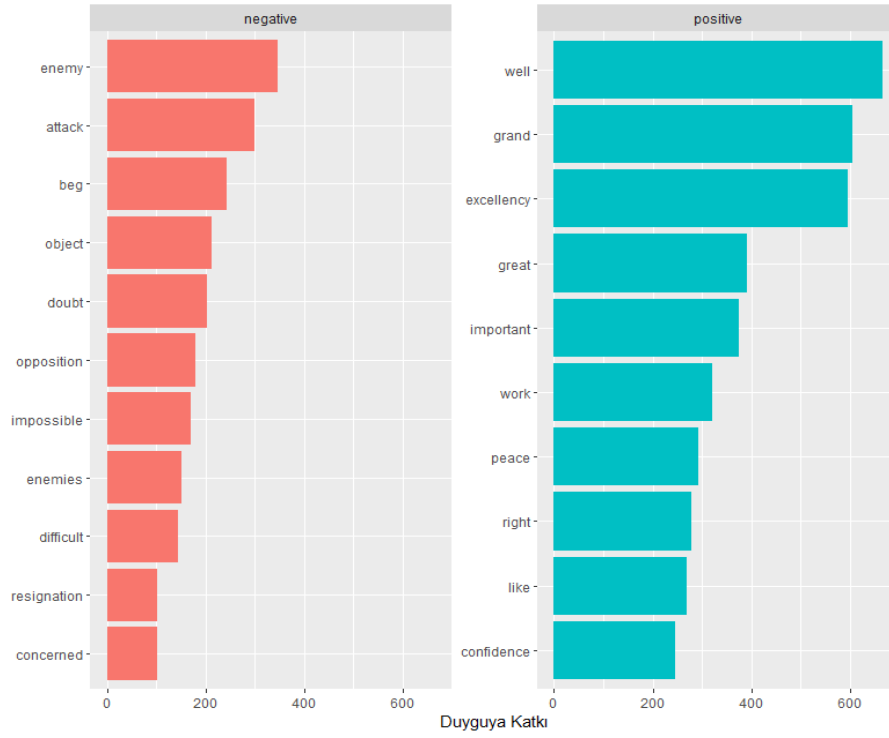
Şekil 14: Bing sözlüğüne göre Nutuk'un ve dönemlerin duygu analizi.

Ayrıca Nutuk, Afinn, Bing ve Nrc duygu sözlükleri ile ele alındığında üç duygu sözlüğünde de çoğunlukla olumlu bir metin olduğu Şekil 15'te gözükmektedir ve üç duygu sözlükleri arasında da benzerlikler görülmektedir.



Şekil 15: Üç sözlüğünün Nutuk üzerinde kıyaslanması.

Nutuk içerisinde en fazla tekrar eden olumlu olumsuz kelimeler ise Şekil 16'da görülmektedir.



Şekil 16: Nutuk'ta en fazla tekrar eden olumlu-olumsuz kelimeler.

Olumsuz kelimelerine bakıldığında “enemy”, “attack” gibi kelimeler ön plandadır ve bir savaş içerisinde bulunduğu anlaşılabilir. “opposition” (muhalafet), “doubt” (şüphe), “impossible” (imkansız), “difficult” (zor), “resignation” (istifa), “concerned” (endişeli) gibi kelimeler ortaya çıkmıştır. Nutuk'a bakıldığında Milli Mücadele sırasında yapılan olumsuz muhalafetin, Milli Mücadele'nin kazanılmasına şüphe ve endişe ile bakıldığı görülmektedir.

Olumlu kelimelere bakıldığında “grand” (büyük) ve “excellency” (hazretleri) görülmektedir. “grand” kelimesi muhtemelen “Grand National Assembly” yani “Türkiye Büyük Millet Meclisi'nin” çok fazla tekrar edilmesinden ön plana çıkmıştır. “excellency” kelimesi ise sürekli bir telgraflaşma halinde olduğu için Mustafa Kemal'e gönderilen telgraflarda “Excellency Mustafa Kemal” yazılmasından kaynaklı ortaya çıkmıştır. Ayrıca olumlu kelimelere bakıldığında “peace” (barış), “right” (doğru-hak) ve “confidence” (güven) gibi kelimeler görülmektedir. Mustafa Kemal Milli Mücadele sırasında ve Nutuk'ta Milli Mücadele'nin haklılığını ortaya koymaya çalışmıştır ve Milli Mücadele Türk toprakları üzerinde barışı

sağlamak üzerine yapıldığı da bilinmektedir. Ayrıca Mustafa Kemal her zaman Milli Mücadele'nin başarılı olacağına inanmıştır.

3.5.1. Kelime Bulutu

Veri görselleştirme, algılanabilirliği arttırmak için verinin etkileşimli ve bilgisayar desteği ile görsel olarak temsil edilmesidir.²² Etkileşimi arttırmak için farklı bir görselleştirme türü olan kelime bulutu da kullanılmıştır. Kelime bulutu, kelime frekanslarına bağlı olarak, toplu bir şekilde en fazla tekrar eden kelimelerin görselleştirilmesinde kullanılır. Ayrıca duygu analizi kelime bulutu üzerinde de gösterilebilir. Nutuk içerisindeki negatif ve pozitif kelimelere genel bir bakış Şekil 17'de görülmektedir.



Şekil 17: Kelime Bulutu ile Şekil 13'ün görselleştirilmesi.

Kelime bulutları herhangi bir görsel üzerinde de yapılabilir. Milli Mücadele lideri Mustafa Kemal üzerinde Nutuk'un görselleştirilmesi Şekil 18'deki gibidir.

²² Metin Vatansver, vd., “Using Visual Data Mining Techniques in Clustering Analysis And An Application”, Sigma 27, 83-104, 2009, **Journal of Engineering and Natural Sciences, Mühendislik ve Fen Bilimleri Dergisi**.

Şekil 19'a göre Nutuk'un en olumsuz duyguları içeren bölümü 0.0269 oran ile sekizinci bölüm olarak gözükmektedir. Meclis Öncesi Dönem'de en olumsuz duyguyu içeren bölüm beşinci bölüm olmakla birlikte I.Meclis Dönemi'ndeki en olumsuz bölüm Nutuk'taki sekizinci bölüme karşılık gelmektedir. II. Meclis Dönemi'nde ise üçüncü bölüm en olumsuz bölüm olarak çıkmıştır.

Nutuk'taki sekizinci bölüm ise “Düzenli orduya geçme kararı” bölümüne karşılık gelmektedir. Bu bölümde yaşanan olaylar şu şekildedir:

- Batı cephesinde, orduda ve halk arasında yapılan propagandalar sonucu “ordudan bir fayda yoktur, dağılsın! Hepimiz Kuva-yı Milliye olalım...” sloganları atılmaya başlanmıştır.
- Batı Cephesi birlikleri arasına Çerkez Ethem ve kardeşleri herkes üstünde otorite kurmaya başlamışlardır.
- Batı Cephesi Komutanı Ali Fuat Paşa, Genel Kurmay Başkanlığı'na Ethem ve Tevfik kardeşlerin etkisiyle olduğu sanılan bir teklifte bulundu: “Yunan ordusunun Gediz yakınında bulunan bağımsız bir tümenine saldırmak.”
- Genel Kurmay Başkanlığı, bu teklifi kabul etmedi. Çünkü düşman ordusu Türk ordularından daha kuvvetli idi ve Türk ordusu henüz tam olarak kurulmamış ve düzene sokulmamıştı.
- Hali hazırda bulunan kuvvet ve yapılan hazırlıklar genel olarak başarı getirmeye elverişli görünmemektedir. Bu nedenle sınırlı ve geçici bir başarı elde etmek için Türk Ordusu yıpratılmış olacaktır.
- Genel Kurmay Başkanlığı, Gediz saldırısının yapılmamasında ısrar etmesine rağmen Batı Cephesi Komutanlığı saldırı kararı almıştır.
- Meclis içerisinde Gediz saldırısı lehinde ve her tarafta propaganda yapılmaktadır.
- Sonunda, Batı Cephesi Komutanı 61. ve 11. Tümenler ve Kuvve-i Seyyareler ile 24 Ekim 1920'de Gediz'de Yunan Ordusuna saldırdı ve Türk Ordusu Gediz'de yenildi.
- Yunan Ordusu, Gediz saldırısından sonra 25 Ekim 1920'de harekete geçti ve Yenişehir ve İnegöl işgal edildi. Türk Ordusu Dumlupınar sırtlarına kadar çekildi.
- Bu başarısızlık sonucunda başta Çerkez Ethem ve kardeşleri bütün suçu cephe komutanına ve düzenli ordu tümenlerine atarak propaganda yapmışlardır.²³

²³ Mustafa Kemal Atatürk, **Nutuk**, Alfa Yayınları, s. 357-362.

Bu olaylar dışında meclis içerisindeki bazı aykırı eğilimlere ve milletvekillerinin seçimi üzerine Nutuk'ta yer verilmiştir ve daha sonrasında Ali Fuat Paşa'nın Batı Cephesi'ne komuta edemeyeceğine karar verilip, Moskova Büyükelçiliği'ne ataması yapılmıştır. Daha sonrasında Batı Cephesi iki cepheye ayrılmıştır. Önemli kesimleri Genel Kurmay Başkanı İsmet Paşa'ya ve güney kesimleri de Refet Paşa verilmiştir. Bu süreçte de Genelkurmay Başkanlığı'na da Fevzi Paşa'nın vekaleti uygun görülmüştür. Böylelikle 1920 yılı Kasım ayının sekizinci günü “düzensiz teşkilat fikir ve siyasetini yıkma kararı” faaliyet ve uygulama alanına konulmuş olup düzenli orduya geçme kararı alınmıştır.

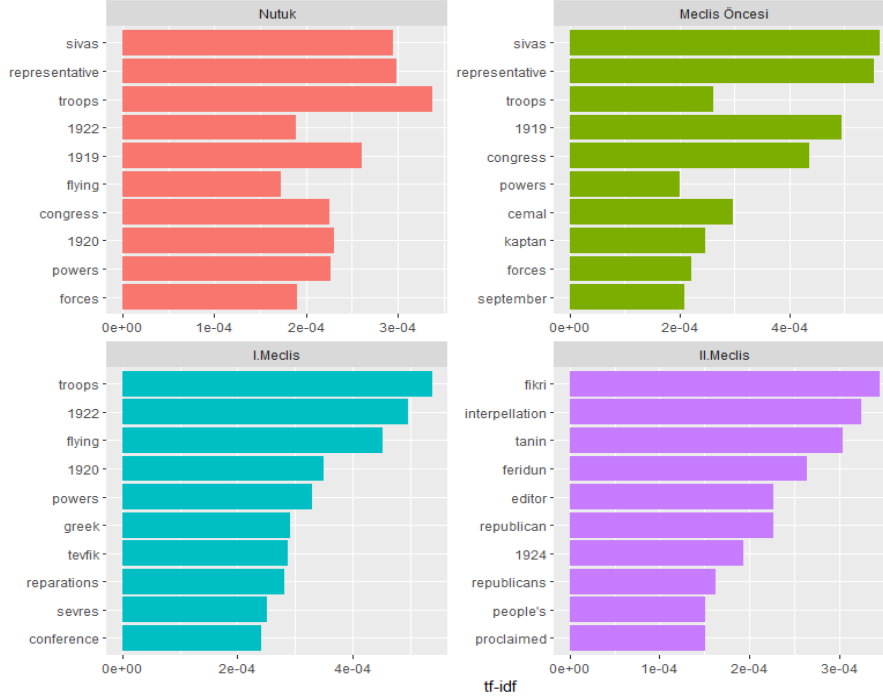
3.6. Terim Frekansı (TF), Terim Frekansı – Ters Doküman Frekansı (TF-IDF)

3.6.1. Kelime ve Doküman Frekansları

Doğal dil işleminin ve metin madenciliğinin ortak bir sorusu, bir belgenin (doküman) ne ile alakalı olduğudur. Bir kelimenin ne kadar önemli olabileceğinin bir ölçüsü, terim (kelime) sıklığıdır. (tf-term frequency)

Bir diğer yaklaşım ise, terimin yaygın olarak kullanılan kelimelerin ağırlığını azaltan ve bir doküman koleksiyonunda çok fazla kullanılmayan kelimelerin ağırlığını arttıran bir terimin ters belge frekansına (inverse document frequency-idf) bakmaktır. İdf, tf (terim frekansı) ile çarpılarak birleştirilebilir.

İstatistiksel tf-idf, bir kelimenin bir belgenin bir koleksiyonunda (veya bir belgede), örneğin bir roman koleksiyonunda veya bir web sitesi koleksiyonundaki bir belgenin ne kadar önemli olduğunu ölçmek için tasarlanmıştır.



Şekil 20: Nutuk ve dönemlerin tf-idf'e göre ön plana çıkan kelimelerin görselleştirilmesi.

Nutuk ve dönemlere göre Şekil 20'ye bakıldığında, dört grafikte de yıllar ön plana çıkmıştır. Dönemlere göre bakıldığında her bir yılın ait oldukları dönemlerde gözükmesi beklenen bir sonuçtur. Genel itibariyle Nutuk içerisinde geçen olaylar 1919-1922 yılları arasındadır. “Sivas” ve “congress” (kongre) kelimeleri ise Nutuk içerisinde önemli yer kaplayan kelimelerdir. Milli Mücadele'nin ateşlenmesinde rol oynayan kongreler ve Milli Mücadele'nin tüm kararlarının alındığı Sivas Kongresi Milli Mücadele'nin yol haritasını çıkarmıştır. Bu kelimeler haricinde “troops”, “powers” ve “forces” kelimeleri askeri birlikleri ve güçleri temsil ettiğinden bir savaş içerisinde bulunduğu anlaşılabilir.

Meclis Öncesi Dönemi, 1919 yılı içerisinde ve kongreler dönemi olarak nitelendirmek mümkündür.

I.Meclis Dönemi'ni savaşlar dönemi olarak adlandırmak doğru sayılabilir. “troops”, “powers”, “Greek”, “flying” gibi kelimeler özellikle meydan muharebeleri dönemlerinde çok fazla yer edinmiştir. Ayrıca “sevres” yani Sevr Antlaşması'nın ön plana çıkmasıyla birlikte aynı

zamanda I.Meclis Dönemi içerisinde gerçekleşen “conference” yani konferanslar ve I.Meclis Lozan Antlaşması ile son bulmaktadır.

II.Meclis Dönemi’nde ise meclis içerisinde bulunan “Fikri”, “Feridun” gibi kelimeler kişileri belirtmektedir. Ayrıca “republican / republicans” ve “people’s” Cumhuriyet Halk Partisi’ni ön plana çıkarmıştır. Ayrıca “proclaimed” (ilan etmek) kelimesi Cumhuriyet’in İlanı ile alakalı olabilir ve “interpellation” (gensoru) kelimesi de meclis içerisinde bir gensoru verilmesine işaret edilmiş olabilir.

3.6.2. Yıllara Göre Terim Frekansı – Ters Doküman Frekansı (TF-IDF)

Nutuk’un dönemlere parçalanması gibi yıllara göre de tf-idf’e bakıldığında biraz daha detaya inilebilmektedir. Burada sadece birer kelimeye parçalama işlemi yapılmıştır. Ayrıca yıllara göre ikişer ve üçer kelimelere parçalayarak tf-idf’lere bakılması daha doğru yorumlamaya yardımcı olacaktır.

Nutuk içerisinde en fazla tekrar eden 9 ülke Şekil 21’deki dağılımları sergilemiştir.



Şekil 21: Nutuk'ta geçen 9 ülkenin yıllara göre dağılımı.

Nutuk'un giriş kısmında memleket için düşünülen kurtuluş çarelerinden bahsedilirken İstanbul Hükümeti, kurtuluşu İngiliz Mandası'nda, İstanbul'daki aydın kimseler ise kurtuluşu Amerikan Mandası'nda ve bir kesim de kurtuluşun bölgesel mücadeleler ile kazanılabileceğini düşünmektedirler. Mustafa Kemal ise bu üç kararı reddedip yeni ve bağımsız bir Türk Devleti kurmak ile kurtuluşun gerçekleşeceğini savunmuştur.

İngiltere ve Amerika'nın birbirine olan benzerlikleri ve birlikte hareket etmelerinden iki ülkenin siyasi anlamda benzer rollerde oldukları anlaşılabilir. Meclis Öncesi Dönem'de Amerika'nın ve İngiltere'nin mandacılık konusunda siyasi baskıları söz konusudur. I.Meclis Dönemi'ne doğru gelindiğinde Amerika'nın ve İngiltere'nin bu siyasi baskılarının neredeyse sıfıra yaklaştığı görülmektedir.

Savaş dönemi içerisinde Yunanistan'ın en aktif olduğu dönem İzmir'in işgal edilmesiyle başlayarak artmıştır. Lozan Antlaşması sonucunda hızlı bir düşüş olduğu görülmektedir.

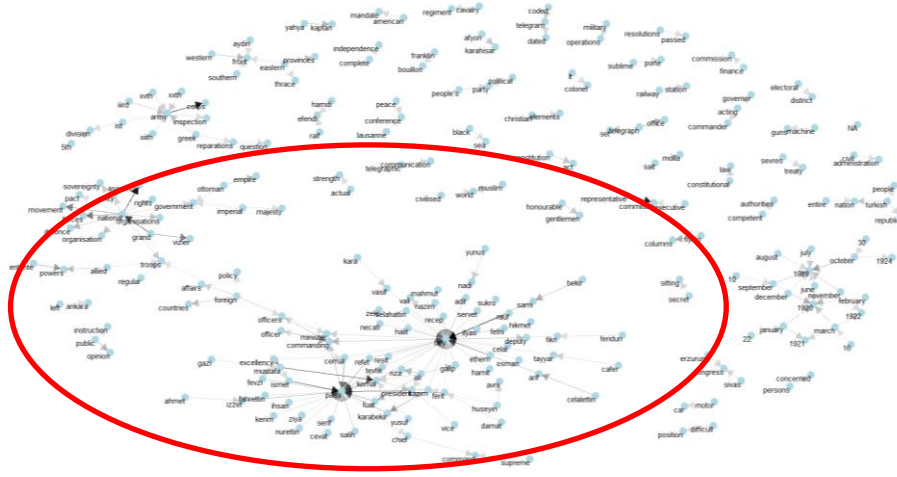
Osmanlı ve Türkiye kelimelerinin de birbirlerine benzer hareket sergiledikleri görülmektedir.

Bu grafikten yola çıkarak ülkeler hakkında da başka çeşitli yargılara da ulaşılabilir.

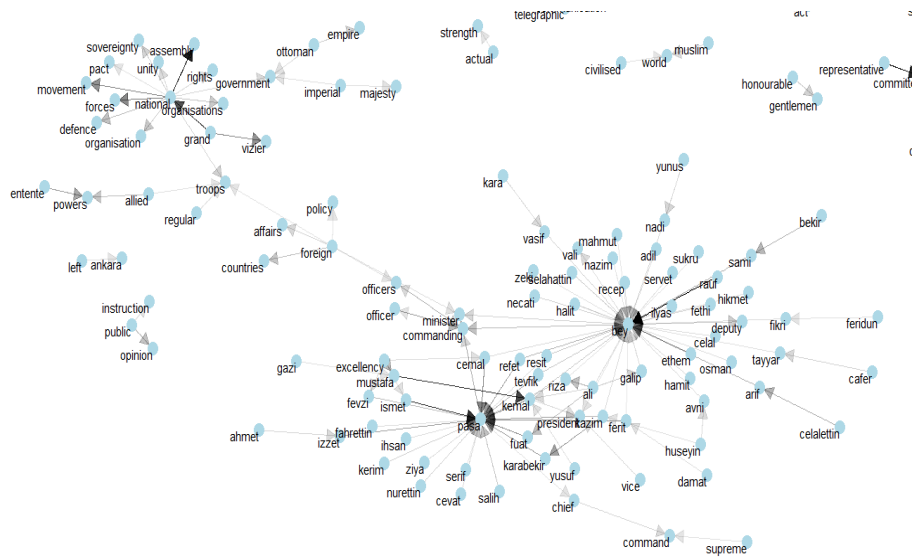
3.7. Metin Madenciliğinde Ağ Analizi

Kelimeler arasındaki tüm ilişkileri aynı anda görselleştirmek için ağ analizi kullanılmaktadır. Ağ analizinde kelimeler bir ağa veya grafiğe yerleştirebilir.

Ağ yapısı ilişkisel düzenli verileri görselleştirmek için kullanışlı bir yoldur. Metin ikişer kelimelere yani bigramlara parçalandıktan sonra ağ analizi yapılmıştır.



Şekil 22: Bigramların frekanslara göre ağ analizi.



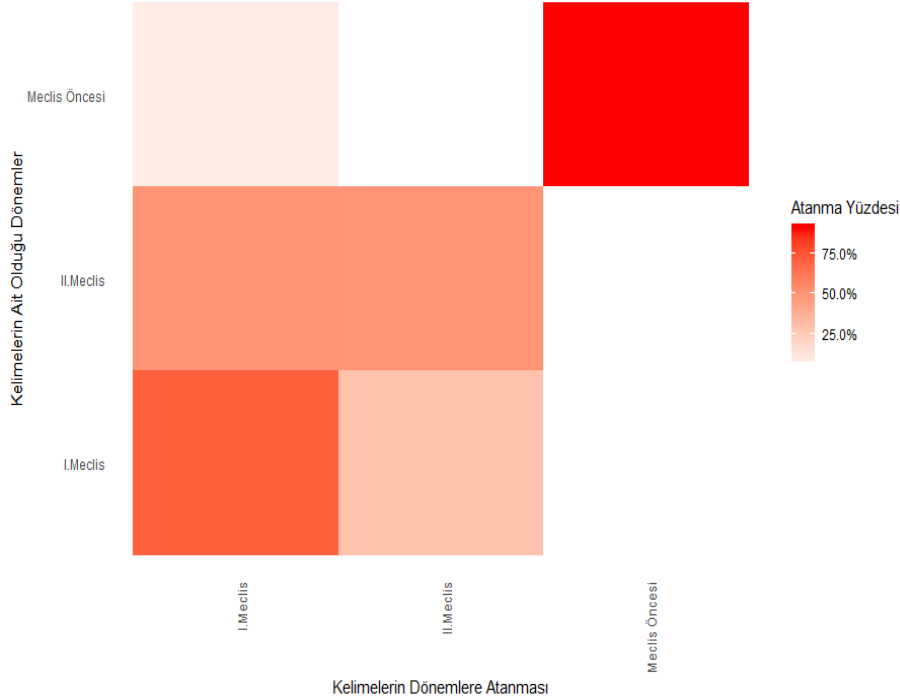
Şekil 23: Bigramdan alınan bir kesit.

Burada iki nokta arasındaki çizgiler iki kelime yani iki nokta arasında bir bağlantı olduğunu gösterir. Ayrıca oklar ile kelime çiftlerinin metin içerisinde nasıl kullanıldığını gösterir. Örneğin “Bekir Sami Bey” ele alındığında ilk önce “Bekir” kelimesi “Sami” kelimesi ile bağlantıda olur;

daha sonra “Bey” kelimesi ile bu ağ tamamlanır. Bigramların ağ analizinde ortaya çıkan ağlara bakıldığında “bey” ve “pasa” kelimelerinin etrafında toplanmış isimler, “army” kelimesi etrafında Türk ordu birlikleri çoğunlukta olmakla beraber “Greek” kelimesi de bu ağa katılmıştır. Görselin sol tarafına bakıldığında “government” kelimesi etrafında “ottoman”, “empire”, “imperial” ve “majesty” kelimeleri toplanmıştır ayrıca “government” kelimesi de “national” kelimesine bağlandığı görülmektedir. Burada Büyük Millet Meclisi, ulusal haklar, ulusal egemenlik, ulusal hareket, ulusal güç, ulusal savunma gibi kelime çiftleri ile bir aradadır. Bir diğer göze çarpan ağ ise cepheleri gösteren ağdır. Burada “thrace”, “western”, “southern” ve “eastern” kelimeleri batı, güney, doğu cephelerini ve Trakya’yı göstermektedir.

3.8. Metin Sınıflandırma Yaklaşımı

Çalışma kapsamında Nutuk dönemlere parçalanmıştır. Korelasyon bölümünde de görüldüğü üzere dönemler arasında yüksek düzeyde ilişki vardır. Korelasyondan farklı olarak dönemler sınıflandırma yöntemi ile karşılaştırması yapılabilir.

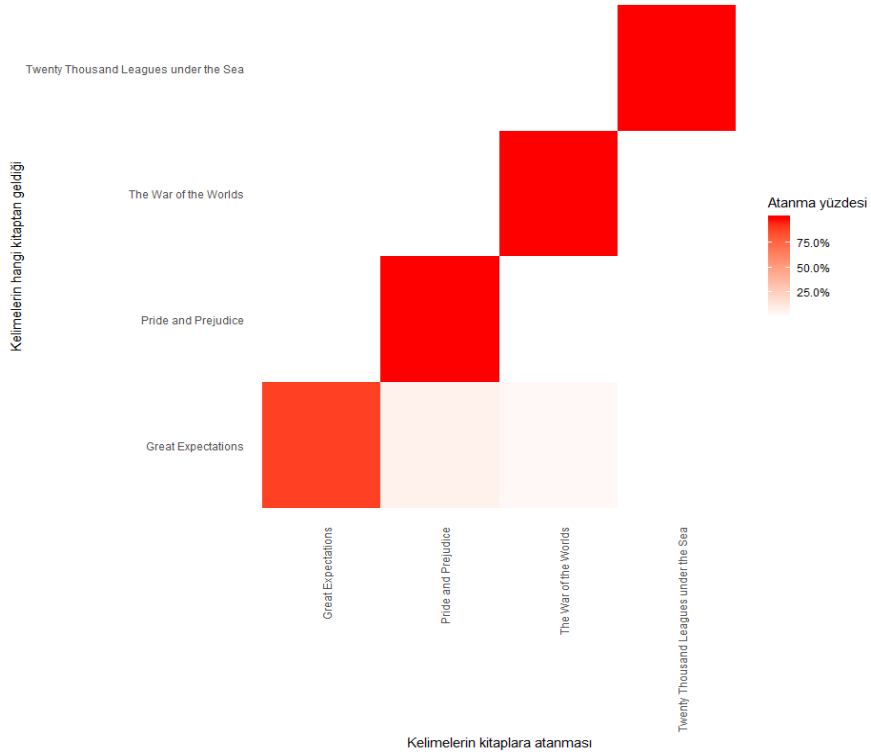


Şekil 24: Dönemlerin sınıflandırılması.

Sınıflandırmaya yani Şekil 24'e bakıldığında dönemler arasındaki benzerlikler ve farklılıklar görülmektedir. Nutuk bütünüyle kişilere ve benzer olaylara dayandığından kendi içerisinde karşılaştırma yapıldığında doğru sınıflandırmanın yapılamadığı görülmektedir.

Nutuk ile alakası olmayan dört farklı kitabın karşılaştırılmasına bakıldığında ise farklı kitapların sınıflandırılması daha doğru sonuçlar vermektedir. Bu dört kitap, sırasıyla Charles Dickens'ın Great Expectations (Büyük Umutlar) kitabı, Jane Austen'ın Pride and Prejudice (Gurur ve Önyargı) kitabı, H.G. Wells'in The War of The Worlds (Dünyalar Savaşı) kitabı ve Jules Verne'den Twenty Thousand Leagues Under The Sea (Denizler Altında 20.000 Fersah) kitabı ele alınarak sınıflandırma işlemi Şekil 25'te yapılmıştır.

Şekilde görüleceği üzere dört farklı kitabın benzerliklerinin olmadığı kolaylıkla gözlenecektir. Metin sınıflandırılması ile metinlerin ayrımı kolaylıkla yapılabilir. Farklı metinler üzerinde sınıflandırma işlemi başarıyla gerçekleşmektedir.



Şekil 25: Farklı dört kitabın sınıflandırılması.

Sonuç

Bu çalışma ile tarihsel olayları ve tarihteki önemli liderleri bir istatistiksel yöntem ile açıklanabilirliğini ortaya koymak ve tarih bilimine istatistiksel bakış açıları ile katkıda bulunulması amaçlanmıştır. Bu amaçların gerçekleştirilmesi için tarihi olay olarak Türk İstiklal Harbi ele alınmış olup bu tarihi olayı birincil kişi olarak anlatan Mustafa Kemal Atatürk ve Nutuk tercih edilmiştir.

Ön araştırmalar sonucunda, tarihsel bir metin içerisinde anlamlı bilgiler çıkarmak için metin madenciliğinin doğru bir araç olduğuna karar verilmiştir. Çalışma kapsamında doğal dil işleme uygulamaları ele alınmamış olup tüm çalışma düzensiz metin formatının düzenli metin formatına dönüştürülmesinden sonra yapılmıştır. Çalışmada kelime frekansları, tf-idf değerleri, kelime bulutu, kelimeler arasındaki korelasyon, ağ analizi ve metin sınıflandırma işlemleri yapılmıştır.

Metin madenciliği, metinler üzerinde analiz ve çıkarımlarda bulunmak için etkili bir yöntemdir. Tarihsel metinler üzerinde ve tarihteki önemli liderleri açıklamada tarafsız ve etkin kullanılması tarihi olayları anlamada ve öğrenmede iyi bir araç olabilir.

Tarih bilimine istatistiksel yöntemler ile katkıda bulunmak, tarihsel olayların daha kolay yorumlanabilmesine olanak sağlayacaktır. Tarihsel olaylar arasında neden-sonuç ilişkilerinin kurulması, tarihsel olayın duygu analizi ile dönemi ve içinde bulunulan şartları, kelime frekansları ile önemli kişi, olay, mekan gibi bilgilerin tespit edilmesi, ağ analizi ile bir ilişki haritasının ortaya çıkarılması, sınıflandırma ile metinlerin karşılaştırılması, benzerliklerinin ve farklılıklarının ortaya konmasının sağlanması gibi çalışmalar başarılı olabilmektedir.

Bu çalışma sonucunda görülüyor ki, Milli Mücadele süreci başarılı bir şekilde temel hatlarıyla ve detaylarıyla ortaya konabilmektedir. Tarihçiler ile istatistikçilerin beraber çalışmaları daha kapsamlı çalışmalar ortaya çıkarmalarına yardımcı olacaktır. Böylece tarihsel olaylar daha anlaşılır bir şekilde açıklanabilir.

Çalışma içerisinde de Atatürk'ün Milli Mücadele'ye olan inancı ve güveni görülmektedir. İstanbul Hükümeti ile olan anlaşmazlıklar, meclis içerisindeki muhalefet ve savaş dönemleri olumsuz olaylar olarak tespit edilmiştir. Ülkelerin kıyaslanması ile Birinci Dünya Savaşı sonrası ülkelerin Türkiye üzerindeki etkisi de görülmektedir. Dönemler arasında karşılaştırmalar yapıldığında ise dönemlerin birbiri ile farklı olduğu gözlenmekte ve her bir dönemi temsil eden özellikler ortaya çıkmıştır.

Özetle kelime frekansları ve tf-idf yöntemleri kullanılarak metin hakkında genel bilgi elde edilmiştir. Bir metin içerisinde geçen kelimelerin sıklıkla kullanılması olayları açıklamada ve kişileri tespit etmede bir gösterge olabileceği ortaya çıkmıştır. Korelasyon analizi ile bilimsel olarak dönemler arasında bir ilişki olduğu ispatlanmış ve bu ilişkinin gücü hesaplanmıştır. Ayrıca korelasyon analizi yardımıyla kelimelerin diğer kelimelerle gözlenme durumları ve ilişkileri ortaya çıkartılmıştır. Metin içerisinden rasgele seçilen kelimelerin diğer kelimelerle olan birlikteliği incelenebilir. Duygu analizi yardımıyla bir metin objektif bir şekilde bilgi verir. Yazarın hangi duyguları taşıdığı, konu ve kapsamın hangi duygularla ele alındığı bu yöntemle kolayca tespit edilebilir. Duygu analizi sonucunda Nutuk olumlu bir metin olarak değerlendirilmiştir. Metin sınıflandırma konusunda tek bir tarihsel metnin ele alınarak kendi içerisinde dönemlerinin sınıflandırılması Nutuk gibi örneklerde pek mümkün görünmemektedir. Bunun nedeni ise kronolojik olarak olayların neden-sonuç bağlamında birbirine dayandığı metinlerde konuların benzerlikler taşıdığı gözlemlenmiştir. Bu araştırma kapsamında tek bir metin ele alındığından sınıflandırma yöntemini bir korelasyon analizi gibi düşünmek daha doğru olacaktır. Sınıflandırma analizi Nutuk'taki dönemlerin birbirlerine benzerliklerini ve ilişkilerini ortaya koymaktadır.

Teşekkür

Bu araştırma, TÜBİTAK 2209-A Üniversite Öğrencileri Araştırma Projeleri Destekleme Programı 2018/1 Dönemi, 1919B011800174 başvuru numaralı, “Tarihsel Olaylara Modern Bir İstatistiksel Yaklaşım: Metin Madenciliği ve Sosyal Ağ Analizi” başlıklı projeden üretilmiştir. Araştırmanın gerçekleştirilmesine katkıda bulunan TÜBİTAK'a teşekkür ederiz.

KAYNAKÇA

Kitaplar

Atatürk, Mustafa Kemal, **Nutuk**, Alfa Yayınları, 2017.

Meydan, Sinan, **1923**, İnkılap Yayınları, İstanbul, 2017.

_____, **Nutuk'un Deşifresi**, Truva Yayınları, İstanbul, 2006.

Şimşir, Bilal N., **Atatürk'ün Büyük Söylevi Üzerine Belgeler**, Ankara, TTK, 1991.

Makale ve Bildiriler

Acun, Fatma, “Dijital Tarih ve DijitalTarihçiliğin Tarihyazımına Etkisi Üzerine”, **Tarihyazımı** , 2 (1), 2020, Journal of Historiography.

- Acun, Ramazan, “Tarihi Metinlerden Cevher Çıkarmak: Metin Madenciliği”, **Avrupa Tarihinde Türk Eli Doç. Dr. Gümeç Karamuk Armağanı**, Ankara, 2017.
- Ağca, Hüseyin, “Atatürk’ün Bütük Nutuk’unda Geçen Milli Kimlikle İlgili Kavram, Terim ve Söz Gruplarının Analizi”, **75. Yılında Nutuk’u Anlayarak Okumak, Bilgi Şöleni**, 17-18 Ekim 2020.
- Arar, İsmail, “Büyük Nutuk’un Kapsamı, Niteliği, Amacı”, **Atatürk’ün Büyük Söylevi’nin 50. Yılı Semineri – Bildiriler ve Tartışmalar**, TTK, Ankara, 1980.
- Aşlıyan, Rifat, Korhan Günel, “Metin İçerikli Türkçe Doküman Sınıflandırılması”, **Akademik Bilişim Konferansı Bildirileri**, Muğla Üniversitesi, 10-12 Şubat, 2010.
- Cicioğlu, Hasan, Atatürk’ün Büyük Nutuk’ta Adları Geçen Muhalifleri, Muhalefet Gerekçeleri, Atatürk’ün Cevabı 111-155, **75. Yılında Nutuk’u Anlayarak Okumak, Bilgi Şöleni**, 17-18 Ekim 2002.
- Gürkan, Hasan, “Atatürk’ün Söylevi’i (Nutuk) Üzerine Bir Bibliyografya Denemesi”, **Çağdaş Türkiye Tarihi Araştırmaları Dergisi**, XIII/27, 2013-Güz.
- Kılıç, Fahri, “Nutuk’un Analitik Bir İncelemesi”, **Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi**, 3 (1), 2016.
- Kılıç, Mehmet, “Büyük Nutuk’un Cumhuriyet Tarihindeki Yeri ve Önemi”, **Türk Dünyası Araştırmaları**, 121, 2019.
- Korkmaz, Zeynep, Ergün Aybars, “Nutuk”, **70. Yılında Ulusal ve Uluslararası Boyutlarıyla Atatürk’ün Büyük Nutku ve Dönemi Sempozyumu**, Yay. Haz. Gül E. Kundakçı, Ankara 1999, s. 8.
- Silge, Julia, David Robinson, “tidytext: Text Mining and Analysis Using Tidy Data Principles in R”, **JOSS 1 (3) The Open Journal**, 2016.
- Şeker, Şadi Evren, Banu Diri, “TimeML and Turkish Temporal Logic”, **Proceedings of International Conference on Artificial Intelligence**, ICAI 2010.
- Tütengil, Cavit Orhan, “Söylev’in Öğretisi”, **Türk Dili Dergisi**, (Söylev Özel Sayısı), Yıl: 27, Cilt: XXXVI, Sayı: 314, 1 Kasım 1977, s. 373.
- Uzun, Hakan, “Atatürk’ün Nutuk’unun İçerik Analizi”, **CTAD: Hacettepe Üniversitesi Atatürk İlkeleri Ve İnkılâp Tarihi Enstitüsü Cumhuriyet Tarihi Araştırmaları Dergisi**, Yıl 1, 2, 2005.
- Vatansever, Metin, Ali Hakan Büyüklü, “Using Visual Data Mining Techniques in Clustering Analysis”, **Journal of Engineering and Natural Sciences Mühendislik ve Fen Bilimleri Dergisi**, Sigma 27, 2009.

Tezler

Tunalı, Volkan, **Metin Madenciliği için İyileştirilmiş Bir Kümeleme Yapısının Tasarımı ve Uygulaması**, Doktora Tezi, Marmara Üniversitesi, Fen Bilimleri Enstitüsü, 2011.

İnternet Kaynakları

Wikipedia, Metin Madenciliği, (https://tr.wikipedia.org/wiki/Metin_madencili%C4%9Fi), (Erişim Tarihi: 14 Mayıs 2019).

Wikipedia, Türk İstiklal Harbi, (https://tr.wikipedia.org/wiki/T%C3%BCrk_Kurtulu%C5%9F_Sava%C5%9F%C4%B1). (Erişim Tarihi: 14 Mayıs 2019).

Kara Harp Okulu, Nutuk Bölümleri, (http://www.kho.edu.tr/hakkinda/harbiyeli_ataturk/nutuk/nutuk.html). (Erişim Tarihi: 14 Mayıs 2019).