

Sınıf Tabanlı N-Gram Modeli ile Türkçe Anlamlı Metin Üretme

Turkish Meaningful Text Generation with Class Based N-Gram Model

Mehmet Ali KUTLUGÜN

Bilgisayar Bilimleri ve Mühendisliği
İstanbul Sabahattin Zaim Üniversitesi
İstanbul, Türkiye
mehmet.kutlugun@std.izu.edu.tr

Yahya ŞİRİN

Bilgisayar Bilimleri ve Mühendisliği
İstanbul Sabahattin Zaim Üniversitesi
İstanbul, Türkiye
yahya.sirin@izu.edu.tr

Özetçe—Metin üretme sistemleri, mevcut cümlelerden yeni bilgi çıkarımları yapabilmek, bir bilgiden yola çıkarak bununla bağlantılı bilgileri elde etmek, internet üzerinden yapılan aramalarda girilen cümlelerin türetilerek kullanıcılara aramalarla ilgili daha detaylı sonuçlar döndürebilmek gibi kolaylıklar sağlarlar. Bu çalışmada, kaynak veri kümesindeki cümlelerden sınıf tabanlı n-gram modeli kullanılarak anlamlı Türkçe dil yapısına uygun cümleler oluşturulması hedeflenmiştir. Cümle üretimini gerçekleştirmek için trigram modeli önerilerek, cümle içindeki kelime ya da kelime gruplarından ilişkili olduğu grup sayısı kadar cümle üretilmiştir. Böylece oluşturulan cümlelerin hiçbirisi diğerleri ile aynı olmayan yeni cümleler elde edilmiştir.

Anahtar Kelimeler —doğal dil işleme; doğal dil üretme; metin işleme; metin üretme; istatistiksel dil modelleme;

Abstract— Text generation systems provide facilities such as making new information deductions from the existing ones, getting information related to them by going out to a knowledgeable way, and generating more detailed results about the calls to the users by generating the codes entered on the internet. In this study, it is aimed to generate meaningful new Turkish sentences using class-based n-gram model from the sentences in the source data set. In order to realize sentence production, a trigram model is proposed and sentences are generated from the word or word groups in the sentence to the number of groups related to it. Thus, new sentences are generated, none of which were identical to the others.

Keywords —natural language processing; natural language generating; text processing; text generation; statistical language modeling.

I. GİRİŞ

A. Doğal Dil İşleme (DDİ)

Doğal dil işleme, doğal bir dili çözümlenme, anlama, yorumlama ve **üretme** gibi işlemlerin bilgisayar sistemleri ile tasarlanması ve gerçekleştirilmesidir. 1950 ve 1960'larda yapay zekânın bir alt konusu olarak görülen bu alan, araştırmacıların elde ettiği başarılar sonunda bilgisayar bilimlerinin temel bir disiplini olarak kabul edilmeye başlanmıştır[1,2].

Doğal dil işlemenin uygulama alanlarından bazıları ; Karakter belirleme, metin doğrulama, okuma, otomatik metin çevirisi, konuşma tanıma, yazılan bir kelimeye göre bir sonraki kelimenin tahmin edilmesi, bilgi bulup çıkarma, metin özetleme, soru yanıtlama, **anlamlı doğal dilde cümle ve metin üretme**, görüşme yapma, büyük boyuttaki metin içinde ifade arama ve yabancı dil eğitimi şeklinde sıralanabilir[2].

B. Metin Üretme

Metin Üretme (Text Generation), doğal dil işlemenin alt dallarından olan doğal dil üretme (Natural Language Generation-NLG) alanı ile ilgili bir teknolojidir. Bir metinden farklı yeni metinler üretme, farklı bir konuda benzer metinler yazma, aynı başlığı içeren yeni bir metin oluşturma gibi amaçlar için kullanılmaktadır[3]. Bu sayede yapılandırılmış veri üzerinden yorum çıkarma işlemleri gerçekleştirilebilmektedir.

Doğal dil üretme teknolojileri ile örneğin, finansal işlemleri ve sistemleri tarayarak istisna ve anomalileri tespit eden bir yapı oluşturulabilmektedir. Bu araçlarla riskler gerçek zamanlı olarak tespit edilebilmekte, etkilenen taraflar zamanında bilgilendirilebilmekte ve bu risklerin ortadan kaldırılması için çözüm önerileri sunulabilmektedir.[4]

Bu alandaki ilk çalışmalar, sabit ya da çok az değişiklik gösteren basit ve standart metinlerin üretiminden ibaret iken, yıllar geçtikçe, daha fazla dilsel anlayış içeren ve çeşitlilik gösteren metinler üretmek amacıyla yeni teknikler geliştirilmiştir [4].

Bir metin üretici geliştirmek için, doğal dilin kaynaklarını tanımlayacak bir dilbilim teorisine ve bu kaynakları bilgisayar ortamında gösterecek ve işleyecek bir yazılım aracına ihtiyaç vardır[2]. Farklı diller arasındaki dilbilgisi ve gramer yapısı büyük ölçüde farklı olduğundan oluşturulacak sistem, temsil edilen dili belirgin hale getirmelidir. İlgisiz sonuçları filtrelemek için ilgi düzeyinin doğrulanması gereklidir. Metin üretme işlemi söz dizim kuralları ve morfolojiye uygun, veriler yapı olarak tutarlı olmalıdır.

Bu çalışmada üretilen metinlerin tutarlılığını sağlamak amacıyla n-gram modelinden yararlanılmıştır. İki veya daha

fazla sözcüğün birlikte gözlenme sayısını temel alan bu model uygulanırken derlemede **yan yana bulunan** (*n-gram*) sözcüklerin **gözlenme sıklıkları** ölçülmektedir. Aynı zamanda sözdizimsel kurallar da dikkate alınarak kurallı bir yapı ile yeni cümleler üretilmektedir. Özetle, kural tabanlı yaklaşım ile istatistiksel yöntemler belli ölçülerde birleştirilerek kullanılmıştır.

II. ÖNCEKİ ÇALIŞMALAR

Bu uygulama alanında **istatistiksel** ve **kural tabanlı** birçok yöntem mevcuttur. Kural tabanlı yaklaşımlar, sözcüğün, cümlenin ve metnin özelliklerini kullanırlar. Bu şekilde dile ait biçimbilimsel, sözdizimsel ve anlamsal özellikler kullanılarak kurallar elde ederler[1,5,7,8,18]. İstatistiksel yöntemler ise bir çeşit **gözlenme sıklığı** bilgisini temel alarak verilen bir derlemedeki eşdizimleri belirleyen yöntemlerden oluşmaktadırlar[6].

Adalı ve Erenler yapmış oldukları çalışmada Türkçe için otomatik metin üretim ve konuşma üretim sistemi üzerinde çalışmışlar ve bir hasta veri tabanından alınan verileri Türkçe dilinin kurallarına uygun şekilde dizerek anlamlı ve dilbilgisi kurallarına uygun raporlar oluşturmuşlardır. Bunun için gramer yapısının oluşturulmasında ilişkisel veri tabanından yararlanılmışlardır[7]. Ancak bu çalışmada kullanılan yöntem, oluşturulan gramer veri tabanı ile sınırlı kalmaktadır.

Özdemir ve Amasyalı kullandıkları kaynak veri tabanındaki bilgi parçaları arasından anlamsal ilişkilerin tutulduğu bir veri tabanının birbiriyle kelime bazında benzerliği olmayan ancak anlamca birbirine yakın olan metinlerin bulunması ve yeni metinler oluşturulması üzerine çalışmışlardır[8]. Bu çalışma gramer yapısı esnek olmayan sadece eşanlamlı veya benzer kelimelerin yer değiştirmesi ile sonuçlanmaktadır.

Mocan, ilköğretim düzeyindeki öğrencilerin okuduklarını anlama düzeylerinin geliştirilmesi amacıyla, metin işleme ve anlamaya dayalı soru soran bir sistem tasarımı üzerinde çalışmıştır[1]. Eğitim alanında, ilköğretim düzeyindeki öğrencilerin anlama düzeylerinin ölçülmeye çalışıldığı bu çalışma, sisteme verilen cümlelerden sadece öğelere ayırma ve bu öğelerin türlerine göre belli sorular üretme üzerine kurulmuştur.

Metin ve Karaoğlan Türkçe bir derlemede eşdizim şeklindeki sözcük grupları üzerinde çalışarak, otomatik olarak eşdizim belirlenmesi için istatistiksel yöntem tekniğini uygulamışlardır[6]. Bu uygulamada tam bir cümle oluşturmaktan ziyade, deyim, bileşik fiiller, söz öbekleri ve tamlamalar şeklindeki eşdizimler üzerinde çalışılmıştır.

İngilizce alanında ise; Brown ve arkadaşları istatistiksel teknikler ile sınıf-bazlı *n-gram* modeli kullanarak İngilizce kelimeleri farklı sınıflara sözdizimsel veya semantik olarak gruplama çalışması gerçekleştirmişlerdir[9]. Bu çalışma da tam ve anlamlı bir cümle oluşturmaktan ziyade, sözdizimsel veya anlamsal olarak kelime gruplarını bir küme altında toplama ile sınırlandırılmıştır.

Mairesse ve arkadaşları sözdizimsel açıklama içermeyen, semantik kavramlardan bir biçim oluşturmak için, alana özgü veriler kullanarak dinamik Bayes ağını kullanan istatistiksel bir İngilizce dil üretici oluşturmuşlardır[10]. Anlamsal olarak

ilişkilendirilen yığınlardan yararlanarak oluşturulan cümleler her ne kadar farklı olsa da, cümle sıralaması dikkate alınmadığından önemli ölçüde benzerlikler içermektedir.

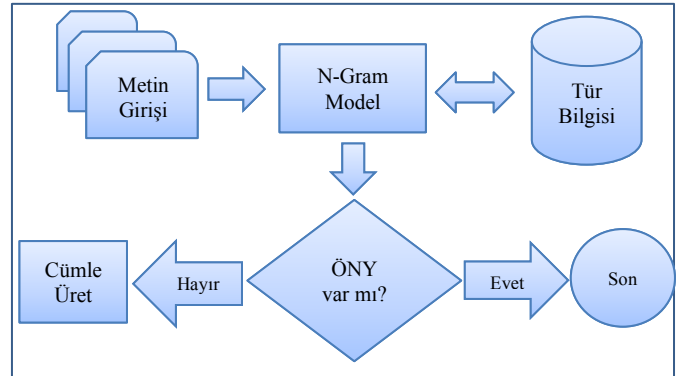
Bauer ve arkadaşları, farklı veri tabanlarından gelen verileri birbirine bağlayan bir algoritma ile veri ilişkilerini, korelasyonları, etkileri ve özel bazı özellikleri tanımlayarak, kural tabanlı farklı türlerde doğal dil metinleri üreten bir sistem geliştirmişlerdir[13]. Bu sistemde kural temelli yaklaşımın doğal dil üretimi için uygun olduğunu ve eldeki problemleri güvenilir ve hızlı bir şekilde çözebileceğini ileri sürmüşlerdir. Ancak, dillerin sözdizimini açıklayan farklı kural setleri aracılığıyla ifade edilebileceği vurgulanmış, bununla birlikte, farklı dillerde kullanılabilen ve cümleleri birden çok dilde ifade edebilecek kadar esnek bir makine oluşturmamanın temel teknik zorluk olduğu belirtilmiştir.

Bu konuda yapılan çalışmalarda sadece bir yöntemi kullanarak doğal dil çözümlemesi yapmak yeterli değildir. Örneğin tek başına kural tabanlı yöntem kullanmak, dilbilgisi kurallarını yazmanın zorluğu nedeniyle kabul görmemektedir. Bunun yanında dilin biçimbilimsel yapısının genel olarak sistematik olmasından ötürü, bu konuda kural tabanlı yaklaşım ağırlık kazanmaktadır. Bu çalışmada, iki yöntem birlikte kullanılarak Türkçe metin üretme işlemi gerçekleştirilmiş, bu sayede sadece tek bir kalıba bağlı kalmadan, değişken uzunluklarda cümleler elde edilmiştir. Oluşturulacak cümleler için *n-gram* modelinin sağladığı kelime gruplarının birlikte bulunma avantajı ile çok detaylı kural tanımlamaları yapılmasına gerek kalmamıştır. Kelimeleri ayrı ayrı ele almak yerine, kelime gruplarını birbirleri ile belirli kurallar çerçevesinde birleştirmek sonuca ulaşmak için yeterli görülmüştür. Bunun yanında kullanılan derlemin boyutuyla kıyaslandığında oluşturulacak yeni cümleler sayıca daha az olabilmektedir.

III. SUNULAN ÇALIŞMA

A. Amaç

Çalışmanın amacı, kaynak veri kümesindeki cümleleri *n-gram* şeklinde kelime gruplarına ayırmak ve istatistiksel yaklaşım ile kural tabanlı yaklaşımı uygun biçimde sentezleyerek, Türkçe anlamlı yeni cümleler oluşturmaktır. Bunun için, her üçlü kelime grubunun cümle öğeleri sınıf-bazlı olarak ele alınmış ve kaynak cümleler ile aynı olmayacak şekilde, yeni anlamlı cümleler elde edilmiştir. Önerilen çalışmaya ait süreç diyagramı Şekil.1'de gösterilmiştir.



Şekil. 1. Süreç Blok Diyagramı

Oluşturulan metni gramatik olarak ifade edebilmek için, DDİ alanının *biçimbilim* ve *sözdizim* özelliklerinden faydalanılmıştır. Aynı zamanda istatistiksel olarak kelime sıklığına dayanan n-gram modeli farklı biçimlerde işlenmiştir.

N-Gram Modeli

İstatistiksel DDİ’de iki veya daha fazla sözcüğün birlikte gözlenme sayısını temel alan bu yöntem uygulanırken derlemde **yan yana bulunan** (*n-gram*) sözcüklerin **gözlenme sıklıkları** ölçülmektedir. Derlemdeki bu sözcük grupları sıklık değerleri azalan sırada listelenir, bu liste oluşturulacak cümleler için aday kelime gruplarını içermektedir. Listede yüksek sıklık değerine sahip olan adayların uygun eşdizimler olduğu kabul edilmektedir[6]. Çoğunlukla n-gramlar 1-gram (unigram), 2-gram (bigram), 3-gram (trigram) şeklinde kullanılmaktadır. Trigram modeli, bir alandaki olası cümlelerin büyük bir bölümünü kaplayan yeterli eğitim verisi varsa oldukça güçlüdür[14]. Bu yöntemdeki en büyük problem gerçek eşdizimlilikleri, diğer sözcük birliklerinden ayıran eşik değerin belirlenmesi aşamasıdır. Yöntemin dezavantajı ise çok sıklıkla gözlenen eşdizim adayları içinde işlevsel bazı kelimelere (*bu, ve, şey, vb.*) rastlanmasıdır[15]. Bu sebeple sıklık değerine bağlı olarak hazırlanan listelerde **sözcük türü** filtresi gibi çeşitli filtreler uygulanır[16].

Türkçe Cümle Yapısı

Cümle öğelerinin sıralanışı açısından Türkçe bir Özne-Nesne-Yüklem (ÖNY) sıralı dil olarak kabul edilebilir. İsimler durum sonekleri almadığında ÖNY dizimi varsayılr. Ancak Türkçe cümle öğelerinin sırasını değiştirmemize olanak tanımaktadır[2]. Bu çalışmada gramer kuralı olarak ÖNY sırası dikkate alınarak cümleler üretilmiştir.

Kelime Türünün Belirlenmesi (Part Of Speech- POS Operation)

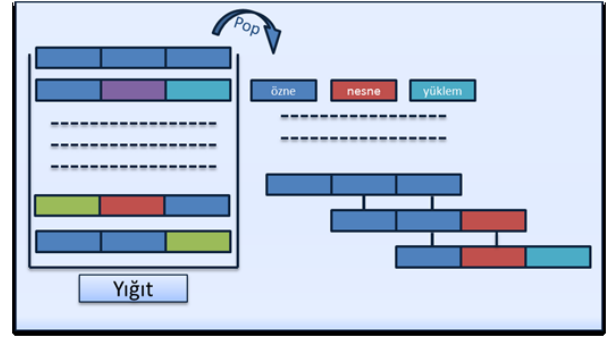
Bir kelimenin türü ya da diğer bir deyişle **ait olduğu sınıf**; İsim (Noun), Fiil (Verb), Zamir (Pronoun), Sıfat (Adjective), Zarf (Adverb), Edat (Preposition), Bağlaç (Conjunction) olabilir. Bazı kelimeler bu sınıflardan sadece birine ait iken, bazıları birden fazlasına ait olabilir. Kelimenin türünün belirlenmesi işlemi, bir kelimenin cümle içindeki konumu incelenerek, o kelime için doğru POS tespit edilerek gerçekleştirilebilir[1].

B. Yöntem / Mimari / Algoritma

Sistem, üçlü kelime gruplarının sıklığı hesaplanarak azalan sırada bir yığında toplanması ve her bir kelimenin ilişkili olduğu sınıfın belirlenerek yeni cümleler üretilmesi şeklinde gerçekleştirilmektedir. Bu yeni cümlelerin hiçbirisi bir diğeriyle aynı olmayacak şekilde tasarlanmıştır.

TABLO I. BULUNAN TRIGRAM İFADE ÖRNEĞİ

Sayı	Trigram	Sıklık
1	Meydana gelen depremden	4
2	gelen depremden sonra	4
3	depremden sonra tsunami	2
4	sonra tsunami uyarısı	2
5	tsunami uyarısı yapılmıştı.	2



Şekil 2. Öğe Sıralı Birleştirme

TABLO II. TESPİT EDİLEN KELİME TÜRLERİ ÖRNEĞİ

Sayı	Kelime	Tür
1	hesaplamak	Eylem
2	hesap	Ad
3	ziyaret	Ad
4	dilek	Ad
5	getirdi	Eylem (çekilmiş)

Bu yöntem ile kaynak cümlelerdeki üçlü kelime grupları mevcut bütün alternatifleri ile hesaba katılarak doğru çözümlenmeler elde edilebilmeye çalışılmıştır. Tüm uygun trigram ifadeler bir yığında toplanarak, en üstte, en sık geçen yer alacak şekilde sıralanmıştır. En üstten başlamak suretiyle yığından çıkarılan ifadeye bağlanabilecek en uygun alternatif sondan iki bağlı trigram veya sondan tek bağlı trigramlar rasgele seçilerek farklı cümleler üretilmiştir. Bu işleme yığında bulunan uygun trigram ifadeler bitene kadar devam edilmiştir. Böylece yığındaki uygun ifadeler kadar cümle oluşturulmuştur.

Üretilen bazı cümleler gramer olarak doğru olmasına rağmen, anlam olarak yanlış çözümlenmeler de oluşmuştur. Örneğin “Rusya dışişleri bakanı Sergey Lavrov NATO ilişkileri konularında açıklamalarda bulundu.” kaynak cümlesinden, “Bulgaristan dışişleri bakanı Sergey Lavrov NATO ilişkileri konularında açıklamalarda bulundu.” şeklinde bir cümle üretilmektedir. Bu cümle her ne kadar gramer kuralları açısından doğru bir cümle olsa da anlam bakımından doğru değildir. Bu ve benzeri üretilen yanlış çözümlenmeler anlambilim ile ilgilidir, mevcut çalışma kapsamında değildir.

1) Kelimenin Cümle İçerisindeki Konumunun İncelenmesi

Bu aşamada üçlü kelime gruplarından tam bir kurallı cümle oluşturulabilmesi için, cümlelerin konum olarak neresinde yer aldığına bakılmaktadır. Örneğin trigram olarak alınan bir ifadenin en başındaki kelimenin ilk harfi büyük ise, bu bir cümle başı şeklinde değerlendirilerek işleme alınmıştır. Benzer şekilde nokta ile biten trigram ifadeleri cümle sonu şeklinde ele alınmıştır.

2) Cümlelerin Öğelerinin Belirlenmesi ve Cümle Üretimi

Trigram olarak alınan bir ifade kurallı, tam bir cümle olabilir ve bu durumda yeni bir cümle üretilemez. Örneğin; “Başbakan Ankara’ya döndü.” trigram ifadesi, öznesi nesnesi ve yüklemi olan tam bir cümle oluşturmaktadır. Trigram

ifadeler ile ögesi eksik olan seçeneklerden farklı alternatiflerle yeni cümleler oluşturulmaktadır. Örneğin; “Yarın hava oldukça” trigram ifadesi ögesi eksik bir ifadedir ve tam bir cümle oluşturmak için, sonrasında gelebilecek diğer alternatif trigram ifadeler ile tamamlanmaktadır. Bu tamamlayıcı ifadeler “hava oldukça **sisli**” ya da “hava oldukça **sıkıcı**” şeklinde devam edebilir.

IV. DENEY

A. Veri Kümesi

Veri kümesi için, YTÜ Kemik Doğal Dil İşleme Grubunun sağladığı ‘42 bin haber’ isimli veri kümesinin ‘Genel’ kategorisindeki 6673 adetten oluşan haber içerikli metinleri kullanılmıştır[17].

B. Uygulama Gerçekleme

Uygulama Java platformunda gerçekleştirilmiş olup, veri kümesinden elde edilen metinler içinden trigram şeklinde kelime grupları elde edilmiştir. Bu gruplar azalan frekans sırasında bir yığılda tutularak, öğelerin kontrolü gerçekleştirilmiştir. Öğeleri tam olan trigram ifadeler yığından atılmıştır. Öğelerinden bir veya birkaçı eksik olan ifadelerden sayıca en fazla geçenler ile yeni cümleler oluşturulmaya başlanmıştır. Cümle başları için, büyük harf ile başlayanlar seçilmiş, nokta ile biten en uygun tamamlayıcı trigram ifadeler ile cümleler sonlandırılmıştır. Trigram ifadelerin birbirini ne ölçüde tamamladığına göre değişkenlik gösteren kısa veya uzun cümleler üretilebilmektedir. Örneğin; “Meydana gelen depremden etkilendi.” şeklinde kısa formda cümleler üretilebildiği gibi, “Meydana gelen depremden sonra tsunami uyarısı yapılmıştı.” şeklinde daha uzun formda cümleler de üretilebilmektedir.

C. Başarım Oranının Hesaplanması

Üretilen cümleler sınıflama ölçeği (nominal scale) ile gramatik açıdan “doğru/yanlış” ve anlam açısından “uygun/uygun değil” ölçütleri ile katılımcılara iki seçenekli anket uygulaması şeklinde sunulmuştur. Sistemin başarımı, birbirinden farklı katılımcıların verdiği ‘doğru’ ve ‘uygun’ sayılarının ortalaması alınarak, elde edilen sonuçlardan tutarlılık yüzdesi hesaplanarak ölçülmüştür.

TABLO III. BAŞARIM DEĞERLENDİRME TABLOSU

Yöntem	Üretilen Cümle Sayısı	Başarım Oranı (%)
2-gram	186	23
3-gram (sondan iki bağlı)	27	71
3-gram (sondan tek bağlı)	64	39

V. SONUÇ VE DEĞERLENDİRME

Bu çalışmada kullanılan yöntem ile, diğer kullanılan yöntemlerden farklı olarak, üçlü kelime grupları şeklinde ayrılan gruplardan belirlenen kurallar çerçevesinde, art arda ekleme yöntemiyle anlamlı ve farklı cümleler üretilmesi gerçekleştirilmiştir. Bunun için kural tabanlı yaklaşım tek başına ele alınmamış, farklı biçimlerde istatistiksel yaklaşım da kullanılarak yeni cümleler elde edilmiştir.

2-gram ifadeler ile yapılan üretme işleminde sayıca daha fazla yeni cümle üretilmesine rağmen, anlam açısından uygunluk seviyesinin daha düşük olduğu görülmüştür. Sonda iki bağlı 3-gram ifadeler kullanılarak daha az sayıda cümle üretilmesine rağmen, daha anlamlı cümleler üretilmiş, bu yöntem ile sistem başarımının en yüksek seviyede olduğu tespit edilmiştir.

KAYNAKLAR

- [1] Mocan, Z., “Metin İşleme:Soru Soran Bir Sistem Tasarımı”, Yüksek Lisans Tezi, İTÜ, İstanbul, 2005.
- [2] Oflazer, K. ve Bozşahin, H.C., “Türkçe Doğal Dil İşleme”, Çukurova Üniversitesi Türkoloji Araştırmaları Merkezi, 2006.
- [3] Gündoğdu, Ö.E. ve Duru, N., “Türkçe Metin Özetlemede Kullanılan Yöntemler”, 18. Akademik Bilişim Konferansı, 2016.
- [4] Narin, B., “Uzman Görüşleri Bağlamında Haber Üretiminde Otomatikleşme:Robot Gazetecilik”, Galatasaray Üniversitesi İletişim Dergisi, 27, 79-108, 2017.
- [5] Kazkılınç, S., “Türkçe Metinlerin Etiketlenmesi”, Yüksek Lisans Tezi, İTÜ, İstanbul, 2012.
- [6] Metin, S.K., Karaoğlan B., “Türkiye Türkçesinde Eşdizimlerin İstatistiksel Yöntemlerle Belirlenmesi”, Bilig Türk Dünyası Sosyal Bilimler Dergisi, 78, 253-286, 2016.
- [7] Adalı, Ş. ve Erenler, Y., “Türkçe için Okuma Fonksiyonlu Otomatik Metin Oluşturma Sistemi”, Elektrik-Elektronik-Bilgisayar Mühendisliği 10. Ulusal Kongresi, 2003.
- [8] Özdemir, C.B. ve Amasyalı, M.F., “Hayat Bilgisi Veritabanı Kullanılarak Otomatik Cümle Üretimi”, XV. Türkiye’de İnternet Konferansı, 2010.
- [9] Brown, P. F. Ve ark., “Class-Based N-Gram Models of Natural Language”, In Proceedings of the IBM Natural Language ITL, pp. 283–298. Paris, 1990.
- [10] Mairesse, F. ve ark., “Phrase-based Statistical Language Generation using Graphical Models and Active Learning”, In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL), Uppsala, 2010.
- [11] Uchimoto, K. ve ark., “Text Generation from Keywords”, COLING ‘02 Proceedings of the 19th International Conference on Computational linguistics, Taipei, 2002.
- [12] Tan, Jiwei ve ark., “From neural sentence summarization to headline generation: a coarse-to-fine approach”, Proceedings of the 26th International Joint Conference on Artificial Intelligence. AAAI Press, 2017.
- [13] Bauer, A. Ve ark., “Rule-based Approach to Text Generation in Natural Language-Automated Text Markup Language (ATML3).”, Challenge+ DC@ RuleML, 2015.
- [14] Erdogan, H., “Speech Recognition for a Travel Reservation System”, IC-AI, 2001.
- [15] Manning, Chris D. ve Schütze, H., “Foundations of Statistical Natural Language Processing”. The MIT Press, 1999.
- [16] Justeson, J.S. ve Katz, S.M., “Principled Disambiguation: Discriminating Adjective Senses with Modified Nouns”, Computational Linguistics, 1995.
- [17] Yıldırım, O. ve Atık, F., “Kişisel Gazete”, YTÜ Bilgisayar Mühendisliği Bölümü, Bitirme Projesi, 2013.
- [18] Korkmaz, T., “Turkish Text Generation with Systemic Functional Grammar”, Yüksek Lisans Tezi, Bilkent Üniversitesi, 1996.