



EGE ÜNİVERSİTESİ

YÜKSEK LİSANS TEZİ

**GÜNCEL TÜRKÇE METİNLERDE
KALIPLAŞMIŞ SÖZ BİRLİKLERİNİN TESPİTİ**

Sinem HİNÇALAN

Tez Danışmanı: Prof. Dr. Bahar KARAOĞLAN

Uluslararası Bilgisayar Anabilim Dalı

**Bilim Dalı Kodu: 619.02.04
Sunuş Tarihi: 08.09.2011**

Bornova-İZMİR

2011

EGE ÜNİVERSİTESİ FEN BİLİMLERİ ENSTİTÜSÜ

(YÜKSEK LİSANS TEZİ)

**GÜNCEL TÜRKÇE METİNLERDE
KALIPLAŞMIŞ SÖZ BİRLİKLERİNİN
TESPİTİ**

Sinem HİNÇAALAN

Tez Danışmanı: Prof. Dr. Bahar KARAOĞLAN

Uluslararası Bilgisayar Anabilim Dalı

Bilim Dalı Kodu: 619.02.04

Sunuş Tarihi:08.09.2011

Bornova-İZMİR

2011

Sinem HINÇALAN tarafından **YÜKSEK LİSANS** tezi olarak sunulan “**Güncel Türkçe Metinlerde Kalıplaşmış Söz Birliklerinin Tespiti**” başlıklı bu çalışma E.Ü. Lisansüstü Eğitim ve Öğretim Yönetmeliği ile E.Ü. Fen Bilimleri Enstitüsü Eğitim ve Öğretim Yönergesi’nin ilgili hükümleri uyarınca tarafımızdan değerlendirilerek savunmaya değer bulunmuş ve **08/09/2011** tarihinde yapılan tez savunma sınavında aday oybirliği ile başarılı bulunmuştur.

Jüri Üyeleri:

İmza

Jüri Başkanı : Prof. Dr. Bahar KARAOĞLAN

.....

Raportör Üye : Yrd. Doç. Dr. Cengiz GÜNGÖR

.....

Üye : Doç. Dr. Aylin KANTARCI

.....

ÖZET**GÜNCEL TÜRKÇE METİNLERDE
KALIPLAŞMIŞ SÖZ BİRLİKLERİNİN TESPİTİ**

HİNÇALAN, Sinem

Yüksek Lisans Tezi, Uluslararası Bilgisayar Enstitüsü

Tez Yöneticisi: Prof. Dr. Bahar KARAOĞLAN

Eylül 2011, 67 sayfa

Tez çalışmasında, güncel Türkçe metinlerde geçen kalıplaşmış söz birliklerinin bir alt grubu olan ikilemelerin dilbilimsel özellikleri dikkate alınarak tespit edilmesi amaçlanmaktadır.

İki sözcüğün birleşmesiyle bir anlam bütünü olan söz birliklerine "**ikileme**" denir. Kelime türü tespiti, bilgi çıkarsama, makine çevirisi vb. doğal dil işleme çalışmalarında ikilemeler önemli yere sahiptirler. Tez çalışmasında kural tabanlı yaklaşım uygulanarak metin içinde geçen ikilemelerin tespiti yapılmıştır. Kuralların oluşturulması için "Deyimler ve Atasözleri" sözlüğündeki ikilemelerin özellikleri çıkarılmıştır. Bu kurallar, metinde geçen ikilemeleri bulmak amacıyla yapı ve ses benzerliklerine göre ayrılmıştır. Benzerlikler de; kelimenin türüne, uzunluğuna ve harflerin benzerlik oranlarına yani biçimsel özelliklere bakılmış ve kurallar yazılmıştır. Yazılan kurallar "Orta Doğu Teknik" ve "Sabancı" üniversitelerinin ortaklaşa çalışmaları sonucu geliştirmiş oldukları OSTAD derleminin gövdelenmiş ve yüzeysel formunda çalıştırılıp, yakalanan ikilemelerin duyarlılık (precision) ve anma (recall) değerleri hesaplanarak kıyaslanmıştır. Elde edilen sonuçlara göre kelime uzunluğu, sözcük türü, benzer harf sayısı ve sesli-sessiz harf benzerlik oranının ikilemelerin tespitinde önemli etkisi olduğu görülmüştür. Aynı zamanda ikilemelerin tespitinde gövdelenmiş derlem kullanımının yüzeysel forma göre daha fazla başarılı olduğu saptanmıştır.

Anahtar Sözcükler : Doğal dil işleme, ikilemelerin tespiti, kural tabanlı yöntemler, duyarlılık ve anma.

ABSTRACT**IDENTIFICATION OF COLLOCATION IN CURRENT
TURKISH TEXT**

HINÇALAN, Sinem

MSc in International Computer Institute

Supervisor: Prof. Dr. Bahar KARAOĞLAN

September 2011, 67 pages

In this thesis we aim at identifying reiteratives, subgroup of collocations, in current Turkish language by taking account the linguistic characteristics. Reiteratives are formed when two words come together to make a single meaning. They have very important position in part of speech tagging, information retrieval and extraction, machine translation, natural language processing and similar studies.

In this study rule based approach is used in the identification of reiteratives. The rules are formed by examining the “Idioms and Proverbs” dictionary and classified based on vocal and structural (length of the words, number of similar letters, types of words, etc.) similarities of the words making up the reiteratives. These rules are run on OSTAD corpus which is a joint project of METU (Middle East Technical University) and Sabancı Universities both in stemmed and surface forms. The results are then compared on the basis of precision and recall metrics. The results show that the lengths and types of words, number of similar consonants and vocals in the words are effective in the identification of the reiteratives. Also, the performance results from the stemmed corpus is higher than that of the surface formed one.

Keywords: Natural Language Processing, Rule-based Methods,
Determination of Reiteratives, Precision and Recall.

TEŞEKKÜR

Tez çalışması boyunca bana çalışma imkanı sunan tez danışmanım Prof. Dr. Bahar KARAOĞLAN'a deneyimleri, bilgileri ve önerileriyle araştırmaya yönlendirmesi ayrıca sağladığı kaynaklar ile bu zaman zarfında yanımda olmasından dolayı çok teşekkür ederim.

Tez çalışmam sırasında değerli görüşlerini ve desteğini, kaynakların temini konusunda yardımlarını esirgemeyen Dr. Tarık KIŞLA'ya ve Senem KUMOVA'ya, bana yüksek lisans eğitimi yapma imkanı sunan Uluslararası Bilgisayar Enstitüsü'ne teşekkürlerimi sunarım.

Tez çalışması sürecinde bana maddi ve manevi desteklerini esirgemeyen tüm yaşamım boyunca yanımda olan ve de olacak olan annem Saadet HINÇALAN'a, babam Armağan HINÇALAN'a, kardeşim Arhun HINÇALAN'a, anneannem Fűrüzan EVEZ'e ve bana bu çalışmalarım boyunca moral ve motivasyonumu yüksek tutmamı sağlayan sevgili eşim Özgür YURDUSEV'e ve de hayatımın anlamı olan kızım Melis YURDUSEV'e minnetle teşekkür ederim.

İÇİNDEKİLER

Sayfa

ÖZET	V
ABSTRACT	Vİİ
TEŞEKKÜR	İX
İÇİNDEKİLER.....	Xİ
ŞEKİLLER DİZİNİ	XİV
ŞEKİLLER DİZİNİ (DEVAM).....	XV
ŞEKİLLER DİZİNİ (DEVAM).....	XVİ
ÇİZELGELER DİZİNİ	XVİİ
ÇİZELGELER DİZİNİ (DEVAM)	XVİİİ
ÇİZELGELER DİZİNİ (DEVAM)	XİX
1. GİRİŞ	1
2. KALIPLAŞMIŞ SÖZ BİRLİKLERİ.....	4
2.1 Kalıplaşmış Söz Birliği Nedir?	4
2.2 Kalıplaşmış Söz Birliği Türleri	5
2.2.1 Deyimler	6
2.2.2 İkilemeler	6
2.2.3 Atasözleri	7

2.3	Kalıplaşmış Söz Birliklerinin Genel Özellikleri.....	7
2.3.1	Deyimlerin özellikleri	8
2.3.2	İkilemelerin özellikleri.....	9
2.4	Kalıplaşmış Söz Birliklerinin Yapısal Özellikleri	11
2.4.1	Kalıplaşmış olma	12
2.4.2	Öğelerin dizilişi	12
2.4.3	Söz dizimsel türler	13
2.4.4	Sözcük sayısı	14
3.	KALIPLAŞMIŞ SÖZ BİRLİKLERİ ÜZERİNE YAPILAN ÇALIŞMALAR	15
4.	İKİLEMELERİN TESPİTİ ÇALIŞMASI	19
4.1	Sözlük Veritabanındaki İstatistikler.....	19
4.2	İkilemelerin Tespiti İçin Belirlenen Kurallar	23
5.	İKİLEMELERİN TESPİTİ İÇİN GELİŞTİRİLEN KURALLARIN UYGULANIŞ YÖNTEMLERİ	24
5.1	Derlem.....	24
5.2	İkilemelerin Tespitinde Kullanılan Yöntem.....	25
5.2.1	Yapı benzerliğine göre belirlenen kurallar ve kuralların uygulanış yöntemleri.....	28
5.2.2	Ses benzerliğine göre belirlenen kurallar ve kuralların uygulanış yöntemleri.....	30
6.	BULGULAR VE TARTIŞMA	35

6.1	Yapı Benzerliği Kurallarının Çalıştırılması Sonucu Elde Edilen Değerler	35
6.2	Sessiz Harf Benzerliği Kurallarının Çalıştırılması Sonucu Elde Edilen Değerler.....	42
6.3	Sesli-Sessiz Harf Benzerliği Kurallarının Çalıştırılması Sonucu Elde Edilen Değerler.....	47
7.	SONUÇ VE ÖNERİLER	62
	KAYNAKLAR DİZİNİ	65
	KAYNAKLAR DİZİNİ (DEVAM)	66
	KAYNAKLAR DİZİNİ (DEVAM)	67
	ÖZGEÇMİŞ	68

ŞEKİLLER DİZİNİ

<u>Sekil</u>	<u>Sayfa</u>
Şekil 5.1 İkilemelerin Tespitinde Kullanılan Kural Tabanlı Yöntemin Şematik Gösterimi.....	26
Şekil 5.2 Değerlendirme Ölçütlerinin Şematik Gösterimi.....	27
Şekil 5.3 Yapı Benzerliği Kurallarının Süzgeç Yöntemi ile Uygulama Şeması.....	28
Şekil 5.4 Yapı Benzerliği Kurallarının Bariyer Yöntemi ile Uygulama Şeması.....	29
Şekil 5.5 Sessiz Harf Benzerliği Kurallarının Süzgeç Yöntemi ile Uygulama Şeması.....	30
Şekil 5.6 Sessiz Harf Benzerliği Kurallarının Bariyer Yöntemi ile Uygulama Şeması.....	31
Şekil 5.7 Sesli-Sessiz Harf Benzerliği Kurallarının Süzgeç Yöntemi ile Uygulama Şeması.....	32
Şekil 5.8 Sesli-Sessiz Harf Benzerliği Kurallarının Bariyer Yöntemi ile Uygulama Şeması.....	33
Şekil 6.1 Yapı Benzerliği Kurallarının Çalışması Sonucu Bulunan İkilemelerin Yüzde Olarak Görünümü.....	40
Şekil 6.2 Gövdelenmiş ve Gövdelenmemiş Yapı Benzerliği Kurallarının Süzgeç ve Bariyer Yöntemleriyle Çalışması Sonucu Hesaplanan Duyarlılık Değerleri.....	41
Şekil 6.3 Gövdelenmiş ve Gövdelenmemiş Derlemde Yapı Benzerliği Kurallarının Süzgeç ve Bariyer Yöntemleriyle Çalışması Sonucu Hesaplanan Anma Değerleri	42

ŞEKİLLER DİZİNİ (devam)

<u>Şekil</u>	<u>Sayfa</u>
Şekil 6.4 Sessiz Harf Benzerliği Kurallarının Uygulanması Sonucu Bulunan İkilemelerin Yüzde Olarak Görünümü	45
Şekil 6.5 Gövdelenmiş ve Gövdelenmemiş Derlemde Sessiz Harf Benzerliği Kurallarının Duyarlılık Değerleri.....	46
Şekil 6.6 Gövdelenmiş ve Gövdelenmemiş Derlemde Sessiz Harf Benzerliği Kurallarının Anma Değerleri	46
Şekil 6.7 Sesli-Sessiz Harf Benzerliği Kurallarının Süzgeç Yöntemi Uygulandığında Elde Edilen Sonuçların Yüzde Olarak Görünümü	55
Şekil 6.8 Sesli-Sessiz Harf Benzerliği Kurallarının Bariyer Yöntemi Uygulandığında Elde Edilen Sonuçların Yüzde Olarak Görünümü	55
Şekil 6.9 Gövdelenmiş Derlemde Sesli-Sessiz Harf Benzerliği Kural Değerlerinin Yüzde Olarak Görünümü	56
Şekil 6.10 Gövdelenmemiş Derlemde Sesli-Sessiz Harf Benzerliği Kural Değerlerinin Yüzde Olarak Görünümü	56
Şekil 6.11 Gövdelenmiş Derlemde Sesli-Sessiz Harf Benzerlik Kurallarının Süzgeç Yöntemi Uygulandığındaki Harf Benzerlik Oranlarına Göre Duyarlılık Değerleri	57
Şekil 6.12 Gövdelenmiş Derlemde Sesli-Sessiz Harf Benzerlik Kurallarının Bariyer Yöntemi Uygulandığındaki Harf Benzerlik Oranlarına Göre Duyarlılık Değerleri.....	58

ŞEKİLLER DİZİNİ (devam)

<u>Sekil</u>	<u>Sayfa</u>
Şekil 6.13 Gövdelenmiş Derlemde Sesli-Sessiz Harf Benzerlik Kurallarının Süzgeç Yöntemi Uygulandığındaki Harf Benzerlik Oranlarına Göre Anma Değerleri	59
Şekil 6.14 Gövdelenmiş Derlemde Sesli-Sessiz Harf Benzerlik Kurallarının Bariyer Yöntemi Uygulandığındaki Harf Benzerlik Oranlarına Göre Anma Değerleri	59
Şekil 6.15 Gövdelenmemiş Derlemde Sesli-Sessiz Harf Benzerlik Kurallarının Süzgeç Yöntemi Uygulandığındaki Harf Benzerlik Oranlarına Göre Duyarlılık Değerleri.....	60
Şekil 6.16 Gövdelenmemiş Derlemde Sesli-Sessiz Harf Benzerlik Kurallarının Bariyer Yöntemi Uygulandığındaki Harf Benzerlik Oranlarına Göre Duyarlılık Değerleri.....	60
Şekil 6.17 Gövdelenmemiş Derlemde Sesli-Sessiz Harf Benzerlik Kurallarının Süzgeç Yöntemi Uygulandığındaki Harf Benzerlik Oranlarına Göre Anma Değerleri	61
Şekil 6.18 Gövdelenmiş Derlemde Sesli-Sessiz Harf Benzerlik Kurallarının Bariyer Yöntemi Uygulandığındaki Harf Benzerlik Oranlarına Göre Anma Değerleri	61

ÇİZELGELER DİZİNİ

<u>Çizelge</u>	<u>Sayfa</u>
Çizelge 4.1 Veri Tabanından Çıkarılan Kurallara Uyan İkilemelerin İstatistikleri .	20
Çizelge 4.2 İkilemelerin Tespiti İçin Belirlenen Kurallar ve Özellikleri.....	23
Çizelge 4.3 Sesli-Sessiz Harf Benzerlik Oranı Kuralının Veri Tabanındaki İstatistikleri (Toplam ikileme sayısı gövdelenmiş derlemde 338 ve gövdelenmemiş derlemde 355’dir.).....	22
Çizelge 5.1 Derlemde geçen ikilemelerin sayı listesi	25
Çizelge 6.1 Yapı Benzerliği Kuralları Kullanılarak Bulunan İkilemeler.....	36
Çizelge 6.2 Yapı Benzerliği Kurallarının Gövdelenmiş OSTAD Derleminde Süzgeç Yöntemi Uygulanması Sonucu Bulunan Değerler ve İstatistikler	37
Çizelge 6.3 Yapı Benzerliği Kurallarının Gövdelenmiş OSTAD Derleminde Bariyer Yöntemi Uygulanması Sonucu Bulunan Değerler ve İstatistikler	38
Çizelge 6.4 Yapı Benzerliği Kurallarının Gövdelenmemiş OSTAD Derleminde Süzgeç Yöntemi Uygulanması Sonucu Bulunan Değerler ve İstatistikler	39
Çizelge 6.5 Yapı Benzerliği Kurallarının Gövdelenmemiş OSTAD Derleminde Bariyer Yöntemi Uygulanması Sonucu Bulunan Değerler ve İstatistikler	40
Çizelge 6.6 Sessiz Harf Benzerliği Kuralları Kullanılarak Bulunan İkilemeler.....	43

ÇİZELGELER DİZİNİ (devam)

<u>Çizelge</u>	<u>Sayfa</u>
Çizelge 6.7 Sessiz Harf Benzerliği Kurallarının Gövdelenmiş OSTAD Derleminde Süzgeç ve Bariyer Yöntemlerinin Uygulanması Sonucu Bulunan Değerler ve İstatistikler	44
Çizelge 6.8 Sessiz Harf Benzerliği Kurallarının Gövdelenmemiş OSTAD Derleminde Süzgeç ve Bariyer Yöntemlerinin Uygulanması Sonucu Bulunan Değerler ve İstatistikler	45
Çizelge 6.9 Sesli-Sessiz Harf Benzerliği Kuralları Kullanılarak Bulunan İkilemeler	47
Çizelge 6.10 Sesli-Sessiz Harf Benzerliği Kurallarının Gövdelenmiş OSTAD Derleminde Süzgeç Yönteminin Uygulanması Sonucu Bulunan Değerler ve İstatistikler.....	49
Çizelge 6.11 Sesli-Sessiz Harf Benzerliği Kurallarının Gövdelenmiş OSTAD Derleminde Bariyer Yönteminin Uygulanması Sonucu Bulunan Değerler ve İstatistikler.....	49
Çizelge 6.12 İkilemelerin Tespiti İçin Oluşturulan Harf Benzerlik Kuralının Süzgeç Yöntemi Uygulanarak Gövdelenmiş OSTAD Derlemindeki İstatistikleri.....	50
Çizelge 6.13 İkilemelerin Tespiti İçin Oluşturulan Harf Benzerlik Kuralının Bariyer Yöntemi Uygulanarak Gövdelenmiş OSTAD Derlemindeki İstatistikleri.....	50
Çizelge 6.14 Sesli-Sessiz Harf Benzerliği Kurallarının Gövdelenmemiş OSTAD Derleminde Süzgeç Yönteminin Uygulanması Sonucu Bulunan Değerler ve İstatistikler.....	51

ÇİZELGELER DİZİNİ (devam)

<u>Çizelge</u>	<u>Sayfa</u>
Çizelge 6.15 Sesli-Sessiz Harf Benzerliği Kurallarının Gövdelenmemiş OSTAD Derleminde Bariyer Yönteminin Uygulanması Sonucu Bulunan Değerler ve İstatistikler	52
Çizelge 6.16 İkilemelerin Tespiti İçin Oluşturulan Harf Benzerlik Kuralının Süzgeç Yöntemi Uygulanarak Gövdelenmemiş OSTAD Derlemindeki İstatistikleri	53
Çizelge 6.17 İkilemelerin Tespiti İçin Oluşturulan Harf Benzerlik Kuralının Bariyer Yöntemi Uygulanarak Gövdelenmemiş OSTAD Derlemindeki İstatistikleri	53
Çizelge 6.18 Sesli-Sessiz Harf Benzerliği Kurallarının Süzgeç Yöntemi Uygulandığındaki Sonuçları	54
Çizelge 6.19 Sesli-Sessiz Harf Benzerliği Kurallarının Bariyer Yöntemi Uygulandığındaki Sonuçları	54

1. GİRİŞ

Dil insanlar arasında anlaşmayı sağlayan tabii bir vasıtaadır. Kendi kanunları içerisinde yaşayan ve gelişen canlı bir varlıktır (Ergin, 1996). Kısaca dil; düşünce ve duyguları, doğrudan ya da dolaylı olarak bildirmeye yarayan yavaş ta olsa değişim gösteren bir anlatım aracı olarak tanımlanabilir.

Doğal dil işleme, belli kurallar çerçevesinde oluşmuş ve gelişmiş doğal dillerin işlev ve yapısının sayısal olarak çözümlenip anlaşılmasını ve yeniden üretilmesini amaçlayan yapay zeka ve dilbilim sahalarının alt uğraş alanı olarak görülür (Kışla, 2009). Başka bir deyişle doğal dil işleme, doğal dillerin kurallı yapısının çözümlenerek anlaşılması ve yeniden üretilmesi amacı üzerine çalışan bir bilim dalıdır (Adalı, 2006). Doğal dil işleme; yapay zeka (bilgi gösterimi, planlama, akıl yürütme vb.), dil çözümleme, doğal dil üretme, makine çevirisi, anlam belirsizliğinin giderilmesi, kelime türü tespiti ve bilgisayar destekli dilbilim gibi çok değişik alanlarda geliştirilmiş kuram, yöntem ve teknolojileri bir araya getirmektedir (Ofłazer ve Bozşahin, 2006). Aynı zamanda doğal dil işleme kökleme, gövdeleme, dil birimlerini (kelime, hece, söz birlikleri) ayırt etme, türlerini belirleme, noktalama işaretlerini çıkarma, küçük-büyük harf çevirmelerini yapma gibi birçok metin işleme konularını içermektedir. Hesaplamalı doğal dil işleme çalışmaları genelde doğal dillerin modellenmesini ve sayısal olarak temsilini konu almaktadır.

Kalıplaşmış söz birlikleri, dil anlamada önemli bir yer teşkil etmektedir. Bu tez çalışmasında, doğal dil işlemede önemli problem olan kalıplaşmış söz birliklerinin bir alt grubu olan ikilemelerin tespiti konusu ele alınmaktadır. İkilemelerin özelliklerine bakılarak güncel Türkçe metinlerde geçme olasılıklarının belirlenmesi ve elde edilen sonuçların karşılaştırılıp, yorumlanması hedeflenmektedir.

Genelde gerçek anlamından ayrı ilgi çekici anlam taşıyan ve en az iki sözcükten oluşan sözcük grubuna kalıplaşmış söz birlikleri denmektedir. Kalıplaşmış söz birlikleri; deyimler, ikilemeler ve atasözleri olarak karşımıza çıkmaktadırlar. Kalıplaşmış söz birlikleri, yazım ve konuşmada anlamsal bir etkiye

sahiptirler. Bu nedenle kalıplaşmış söz birliklerinin belirlenmesi doğal dil anlama çalışmalarında önemli olmaktadır.

Kalıplaşmış söz birliklerinin bir alt grubu olan ikilemeler ise geçtiği cümlede anlatımı güçlendirmek, anlamı pekiştirmek için aynı sözcüğün tekrarlanmasıyla veya yakın ya da karşıt anlamlı sözcüklerin yan yana kullanılmasıyla oluşan sözcük grubudur.

Kalıplaşmış söz birliklerini tespit etmek amacıyla istatistiksel ve kural tabanlı yöntemler geliştirilmiştir. İstatistiksel yöntemler derlemde kalıplaşmış söz birliklerini; gözlenme sıklığı, karşılıklı enformasyon, t-testi vb. gibi hipotez testler vasıtasıyla tespit etmektedir (Church ve Hanks, 1990). Kural tabanlı yöntemler ise derlem üzerinde işaretlenmiş verilerle çalışan ve belli bir kural listesinin metin üzerinde işlenmesini gerçekleştiren yöntemlerdir.

Tez çalışmasında derlem olarak “Orta Doğu Teknik” ve “Sabancı” üniversitelerinin ortaklaşa çalışmaları sonucu geliştirmiş oldukları OSTAD derlemi kullanılarak ikilemelerin tespitinde kural tabanlı bir yaklaşım uygulanmıştır. “Deyimler ve Atasözleri Sözlüğü” (Kuşçu H., Kuşçu Ü., 1997)’nde geçen ikilemeler incelenmiş ve özellikleri çıkarılmıştır. OSTAD derleminde ikilemelerin işaretlenmesi elle yapılmış ve sözcük türü, kelime uzunluğu, harf benzerlik oranı vb. özellikler dikkate alınarak kurallar belirlenmiştir. Belirlenen kurallar derlemde çalıştırılmıştır. Elde edilen sonuçlar bilgi geri getirim alanında sıklıkla kullanılan duyarlılık (precision) - anma (recall) değerlendirme ölçütleri ile değerlendirilmiş ve birbiri ile kıyaslanmıştır. İkilemelerin tespiti için derlemin hem gövdelenmemiş hem de gövdelenmiş hali üzerinde çalışma yürütülmüş ve gövdelenmenin ikilemelerin tespitindeki etkisi araştırılmıştır. Ayrıca derlem içinde geçen ikilemelerin hem tekrarları çıkarılmış hem de çıkarılmamış hali arasında karşılaştırma yapılmış ve çalışmanın etkililiğine katkısı olmadığı sonucuna varılmıştır.

Tez, hedefi ve kapsamı doğrultusunda şu bölümlerden oluşmaktadır; giriş, kalıplaşmış söz birlikleri, kalıplaşmış söz birlikleri çalışmaları, ikilemeler üzerine yapılan çalışmalar, bulgular ve tartışmalar, sonuç ve öneriler.

Tezde giriş bölümünü izleyen ikinci bölümünde kalıplaşmış söz birliği kavramından bahsedilmiş ve kalıplaşmış söz birliklerinin türleri, özellikleri ve yapısal özelliklerinden detaylı bir biçimde bahsedilmiştir.

Tezin üçüncü bölümünde kalıplaşmış söz birlikleri üzerine yapılan İngilizce ve Türkçe çalışmalar hakkında bilgi verilmiştir.

Tezin dördüncü bölümünde ikilemelerin tespiti için geliştirilen kurallardan ve kuralların veritabanındaki istatistiklerinden bahsedilmiştir.

Tezin beşinci bölümü ise tezde kullanılan derlem ve yöntemlerin açıklanmasını, ikilemelerin tespiti için geliştirilen kurallar ve açıklamalarını içermektedir.

Altıncı bölüm bulgular ve tartışmaları, yedinci bölüm ise sonuç ve önerileri içermektedir.

2. KALIPLAŞMIŞ SÖZ BİRLİKLERİ

Belli ses öbeklerinin insanlar arasında dâişıklı bir değeri kazanarak birer kavrama karşılık olmaları dilin oluşmasında esas sayılmaktadır. Ses öbeklerinin bir araya gelmesi sözcükleri oluşturmaktadır. Sözcükler, her zaman tek başına yeterli bir anlam veremeyebilirler. Bazen birden fazla sözcük bir araya gelerek kendi anlamlarından farklı bir anlam verebilirler. Bu durumda söz birlikleri dediğimiz gruplar meydana gelmektedir. Söz birliklerinin bir bölümü dilde kullanılarak değıştirilemez bir yapı kazanmaktadırlar. Bu söz birliklerine “Kalıplaşmış Söz Birliği” denir. Bunların en önemlisi deyimler ve ikilemelerdir.

Bu bölümde, tez içerisinde sıkça kullanılan kalıplaşmış söz birliği kavramı anlatılarak, kalıplaşmış söz birlikleri türleri, genel ve yapısal özelliklerinden bahsedilecektir.

2.1 Kalıplaşmış Söz Birliği Nedir?

Bir dilin söz varlığı, çeşitli öğelerden oluşmaktadır. Bu öğelerin bir kısmı, konuşan kişinin her kullanımda özgürce seçebildiği, bağımlı ve bağımsız kullanılabilen, sözlüklerde madde başı olan sözcüklerdir (elma, uzay, cam, söz, ile, gibi, koş-(mak), ara-(mak), vb.). Bunların yanında bir başka grup da, her zaman belirli bir biçimde kullanılan atasözleri, deyimler ve ikilemelerden oluşan kalıplaşmış söz birliklerini içermektedir. Söz varlığının bir bölümünü oluşturan kalıplaşmış söz birlikleri, iletişimde de sıklıkla tercih edildiği görülmektedir (Aksan, 2002; Toklu, 2003).

Kalıplaşmış söz birlikleri yaygın bir şekilde kullanılmaktadır. Bu yaygınlık, kalıplaşmış söz birliklerinin sözlü ve yazılı iletişim sırasında çok fazla çaba gerektirmeden, kısa zamanda söylenip anlaşılabilmesinden kaynaklanmaktadır (Wray, 2001). Özellikle konuşma eylemi yapılırken, daha önceden bir bütün haline getirilip hafızada öylece saklanan kalıplaşmış birimler, konuşanlarca kolaylıkla hatırlanıp hemen kullanılabildiği gibi, dinleyiciler tarafından da çabucak algılanabilmektedir. Kalıplaşmış bir ögenin söylenip anlaşılması için gereken zaman ve enerji, her durumda yeni bir söz birliğinin veya tümcenin meydana

getirilip ağızdan çıkması ve algılanmasından daha az olmaktadır (Wray, 2001). Benzer durum, yazılı iletişimde de görülmektedir. Bundan dolayı kalıplaşmış söz birlikleri, sözlü ve yazılı iletişimde kullanımı kolay ve cazip birimler olarak karşımıza çıkmaktadır.

Aksan'a (1996) göre atasözleri, deyimler ve ikilemeler, “bir toplumun bireyleri arasındaki ilişkiler sırasında kullanılması adet olan birtakım sözlerdir.” Belli durumlarda söylenmesi gelenek halini almış olan bu sözler, Türk kültürünün ayrılmaz bir parçası olarak algılanmaktadır.

Ayrıca kalıplaşmış söz birlikleri, işlevlerine ve sözcüklerin anlamlarına dayanılarak da açıklanmıştır. Çotuksöken (1994), kalıplaşmış söz birliklerini “en az iki sözcükten oluşan, içindeki sözcükleri temel (düz) anlamlarını yitirmeden yeni bir kavramı, durumu, eylemi karşılayan söz birlikleri” olarak tanımlamıştır. Bu tanımdan kalıplaşmış söz birliklerinin mecaz anlam içerdiği anlaşılmaktadır. Başka bir tanımda, kalıplaşmış söz birlikleri “mecazsız ve herkesçe büyük bir sıklıkla kullanıla kullanıla artık tek kavramın karşılığı olan birliktelikler” olduğu belirtilmiştir (Yüceol ve Özezen, 2001). Bu tanım da, anlam ve işlev özelliklerine dayanılarak yapılmıştır.

Bir başka tanıma göre kalıplaşmış söz birlikleri, “önceden belirli bir biçime girip öylece hafızada saklanan, söyleneceği sırada yeniden üretilmeden veya dilbilgisel olarak ayrıştırılmadan olduğu gibi hatırlanarak kullanılan, ardışık veya aralı sözcüklerden oluşan dizidir” denmektedir (Wray, 2001). Kalıplaşmış söz birliklerini birbirinden ayıran yapı, işlev, anlam ve bağlam özellikleri olduğu görülmüştür.

2.2 Kalıplaşmış Söz Birliği Türleri

Kalıplaşmış söz birlikleri doğal dil işleme de anlamın belirlenmesi dolayısıyla makine çevirisi, bilgi getirimi, bilgi çekimi gibi konularda önem taşımaktadırlar. Bu konularda önemli olan kalıplaşmış söz birlikleri türlerine baktığımızda; deyimler, ikilemeler ve atasözleri olarak karşımıza çıkmaktadır.

2.2.1 Deyimler

Bir olay veya durumu ifade etmek için, o olay veya durumu birebir karşılayacak kelimeler ve çağrışım yaptıracak söz birlikleri kullanılmaktadır. Böylece ifademize akıcılık ve anlam zenginliği katmaktadır. Bir insanın telaşlı olduğunu anlatmak için “telaşlıdır” demeyiz de “Etekleri tutuşmuş” denmesi bu anlatıma akıcılık ve anlam zenginliği yüklemektedir.

Birden fazla sözcükten oluşmuş, bir kavramı karşılamak amacıyla kullanılan ve bir durumu en kısa yoldan anlatıp, cümleye çekici anlatım özelliği katan, çoğu mecaz anlamı kalıplaşmış söz birliklerine “*deyim*” denmektedir (Aksoy, 1993). Torun’a (2005) göre, “Deyimler bir ulusun dil kullanımındaki becerisini gösteren en önemli dil göstergelerinden biridir” ve özellikle konuşma dilinde deyimlerin sıkça kullanıldığı görülmektedir. Kısacası deyim; dilbilimde sözcükselleşmiş karmaşık birlikteliklerdir. Asıl anlamlarından uzaklaşarak yeni kavramlar meydana getiren, genellikle gerçek anlamından farklı bir anlamı olan, ilgi çekici bir anlatımı bulunan, ifadeyi daha zengin kılan, iki veya daha fazla sözcükten meydana gelen kalıplaşmış söz birlikleridir.

“ Bindiği dalı kesmek” deyimine bakıldığında kendisine faydalı ve yararlı olan şeyi kendi eliyle yok etmek, ortadan kaldırmak anlamına gelmektedir.

Deyimler, bir toplumun kültürünü, tarihini, ortak dil zevkini yansıtmaları bakımından son derece önemli olmaktadır. Deyimlerin çokluğu Türkçeye ayrı bir güzellik ve anlatım gücü kazandırmaktadır.

2.2.2 İkilemeler

Anlatımı etkili kılmak amacıyla aralarında anlam ya da ses bakımından ilişki bulunan, anlatımı güçlendirip pekiştirmek ve zenginleştirmek amacıyla kullanılan aynı ya da sesleri birbirine benzeyen sözcüklerin art arda yinelenmesiyle oluşan kalıplaşmış sözlerdir.

“*Fokur fokur, kılık kıyafet, yıkık dökük, eciş büçüş, ileri geri, baka baka...*” vb. örnekler geçtiği cümlelerin anlatımını güçlendirip, pekiştirmek amacıyla kullanılmaktadır.

2.2.3 Atasözleri

Atasözleri; geçmişten günümüze gelen, uzun deneyimlerden yararlanarak kısa ve özlü öğütler veren, toplum tarafından benimsenerek ortak olarak kullanılan kalıplaşmış cümlelere denmektedir. Atasözleri bir toplumun duygu, düşünce inanç ve kültür yapısını yansıtmaktadır. Atasözlerinin kim tarafından ne zaman söylendiği bilinmemektedir (Özdemir, 2007). Bu sözler topluma mâl olmuş, toplum tarafından benimsenmiş ve yüzyılların düşünce ve mantık sisteminden geçerek günümüze ulaşmış kısa ve özlü cümleler olmaktadır. Örnek vermek gerekirse “*Emeksiz yemek olmaz.*”, “*Yatan aslandan gezen tilki yeğdir...*” vb.

Tez çalışmasında cümle içinde geçen kalıplaşmış söz birliklerinin bir alt grubu olan ikilemelerin bulunması amaçlanmaktadır. Atasözleri kendi başlarına bir cümle oluşturdukları için atasözlerinin özelliklerine tez çalışmasında değinilmeyecektir.

2.3 Kalıplaşmış Söz Birliklerinin Genel Özellikleri

Kalıplaşmış söz birlikleri, dilbilgisel yapı ve kullanım sıklığına göre de tanımlanmıştır. Sayı ve kişi öğeleri dışında değişmeyen birimler olduğu belirtilmiştir (Tannen, 1981).

Kalıplaşmış söz birliklerinin en önemli özelliği cümleye kattığı anlamdır. Anlam özelliğini bir örnekle açıklamak gerekirse; *Güle güle büyüütün* kalıp sözü sadece bebeği olan kişilere söylenebilir. Aynı sözü küçük bir evcil hayvanı büyüten ya da bir çiçek yetiştiren kişilere söylenmesi uygun değildir.

Kalıplaşmış söz birliklerinin özelliklerini deyimlerin ve ikilemelerin özellikleri olarak iki başlık altında aşağıda açıklanmıştır.

2.3.1 Deyimlerin özellikleri

Deyimler cümlede geçtiği yerde en az iki sözcüğün kalıplaşmasından oluşmaktadır. “*Dillere düşmek, kulak asmamak, etekleri zil çalmak...*” başlıca örnekleridir.

Deyimler kalıplaşmış söz birlikleri olduğu için yapıları bozulamaz, bozulursa anlatım bozukluğu ortaya çıkmaktadır. Örneğin; *Kara gün* yerine *siyah gün* kullanılmamaktadır.

Deyimi oluşturan sözcüklerin yeri değiştirilirse de deyim yapıları bozulur (Aksoy, 1993). “Ona güvenebilirsin, onun *içi dışı birdir*.” cümlesindeki “*içi dışı birdir*” deyimini “*dışı içi birdir*” şeklinde kullanılamaz.

Sözcükler başka sözcüklerle kalıplaşarak çoğu kez bu yapı içinde mecaz anlam kazanarak deyimleri oluşturmaktadırlar (Aksoy, 1993). Örneğin; “Bu sorular hepimize *kök söktürdü*.” cümlesindeki “*kök söktürdü*” mecaz anlamı olarak kullanılmaktadır. Oysa ki “Babam bahçede kalan *kökleri söktürdü*.” cümlesinde ise “*kökleri söktürdü*” gerçek anlamında “bitki köklerini çıkartma” anlamında kullanılmıştır.

Deyimler üç değişik biçimde kalıplaşabilmektedir (Aksoy, 1993). Bunlar;

- **Mastar Biçimiyle :** *İçi götürmemek, koltukları kabarmak, parmak ısırtmak...*
- **Tamlama :** *Ana kuzusu, kör dövüşü, kaçın kurası, ayak bağı...*
- **Cümle Biçimiyle:** *Dostlar alışverişte görsün, gökten zembille mi indi, ağzından yel alsın, armut piş ağzıma düş...*

Genellikle deyimlerin çoğu mastar biçimiyle kalıplaşmıştır. Deyimlerin son sözcükleri “-mek, -mak” ekleri alarak mecaz anlam kazanmaktadırlar. Örneğin; “*İpe un sermek, küplere binmek...*” gibi deyimlerin büyük bir bölümü mecaz anlam taşımaktadırlar. Az sayıda da olsa bazı deyimler gerçek anlamında kullanılmaktadır. “*Alan razı satan razı, azı gitti çoğu kaldı...*” vb.

Deyimler kişisel bir durumu, davranışı ya da duyguyu anlatan sözleri oluşturmaktadır. Bu özellikleriyle de diğer kalıplaşmış sözlerden, atasözlerinden ayrılmaktadır. “*on parmağında on marifet, gözden düşmek, aklını başına toplamak...*” vb.

Bazı deyimler birbirine yakın anlamlıdır. Bunlar aynı davranış, durum ya da duygu için kullanılabilir. “*Ağzı süt kokmak, ana kuzusu*” deyimlerine bakılırsa çok genç, toy anlamındadır.

2.3.2 İkilemelerin özellikleri

Yinelenen ve yan yana kullanılan her sözcük ikileme oluşturmamaktadır. “Üzüm üzüme *baka baka* kararır.” cümlesinde “baka baka” ikileme iken “üzüm üzüme” ikileme değildir. İkilemeyi oluşturan sözcüklerin yerleri değiştirilmemeli ve sözcüklerin arasına virgül ya da kısa çizgi konulmamalıdır. Örneğin; “El, elin eşeğini *güle oynaya* arar.” cümlesinde “güle oynaya” ikilemedir.

İkilemeler anlamsal kuruluşlarına göre farklılıklar göstermektedirler (Sakça, 2009).

- ✓ Aynı sözcüğün yinelenmesi ile oluşan ikilemeler bulunmaktadır. “*Uzun uzun* düşündüm”, “*Çay fokur fokur* kaynıyordu.” örnekleri verilebilir. Doğadaki seslerin taklit yoluyla yazıya geçirilmesiyle elde edilen sözcüklere yansıma denmektedir. *Pat, güm, şap, tak...* bu seslerin iki kez yazılması ile ikileme oluşmaktadır.
- ✓ Eş anlamlı sözcüklerden oluşan ikilemeler bulunmaktadır. “*Akıllı uslu* çocukları severdi.”, “*Güçlü kuvvetli* biriydi.” cümlelerindeki “akıllı uslu” ve “güçlü kuvvetli” anlamları aynı olan sesleri farklı olan iki sözcüğün yan yana gelmesiyle ikileme oluşturmaktadır.
- ✓ Yakın anlamlı sözcüklerden oluşan ikilemeler bulunmaktadır. “Yaptıklarına *akıl fikir* ermez.”, “*Yalan yanlış* konuşur.” cümlelerindeki “akıl fikir” ve “yalan yanlış” yakın anlamlı iki sözcüğün bir araya gelmesi ile oluşan ikilemelerdir.

- ✓ Karşıt anlamlı sözcüklerden oluşan ikilemeler bulunmaktadır. “*Aşağı yukarı* bir saattir bekliyoruz.” cümlesindeki “aşağı yukarı” zıt anlamlı sözcüklerin yan yana gelmesiyle oluşan ikilemedir.
- ✓ Biri anlamlı biri anlamsız sesçe benzer sözcüklerden oluşan ikilemeler bulunmaktadır. “*Çoluk çocuk* evde seni bekliyor.”, “*Kaba saba* insanlardan hoşlanmam.” cümlelerindeki “çoluk çocuk” ve “kaba saba” biri anlamlı biri anlamsız sözcüklerin bir araya gelmesiyle oluşan ikilemelerdir.
- ✓ İki sözcüğü de anlamsız sesçe benzer sözcüklerden oluşan ikilemeler bulunmaktadır. “*İvir zıvır* şeylerle karnını doyurma.”, “*Eften püften* işlerle uğraşma.” cümlelerindeki “ıvır zıvır” ve “eften püften” anlamsız ama sesçe benzer sözcüklerin bir araya gelmesiyle oluşan ikilemelerdir. İki sözcüğü de anlamsız olan ikilemelerde ikilemeyi oluşturan sözcükler tek başına kullanılmamaktadır.
- ✓ Aynı sözcüğün başına “m” getirilmesiyle oluşan ikilemeler bulunmaktadır. “*Şaka maka* okulu da bitirdik.”, “*Çocuk mocuk* dinlemem, alırım ayağımın altına!” cümlelerindeki “şaka maka” ve “çocuk mocuk” kalıplaşmış söz birliklerinde ikinci sözcüğün başına “m” gelmesiyle ikileme oluşturmaktadır.

İkilemeler anlamı güçlendirip pekiştirir ve bazen söze abartma anlamı da katmaktadırlar. “Kırlarda *kucak kucak* çiçek topladık.”, “Arkadaşını *yere göğe* sığdıramıyor.” cümlelerindeki “kucak kucak” ve “yere göğe” ikilemeleri geçtiği cümlede abartma anlamı kazandırmaktadır.

İkilemeyi oluşturan sözcükler çekim veya yapım ekleri almaktadırlar. “el ele, diz dize, arada sırada, sağdan soldan, kaşı gözü, uzaktan uzağa...” örneklerinde çekim eki almış ikilemeler bulunmaktadır. “günlük güneşlik, irili ufaklı, güçlü kuvvetli, kırık dökük, yerli yersiz, dalgın dalgın, dertli dertli...” örneklerinde yapım eki almış ikilemeler bulunmaktadır.

İkilemeyi oluşturan sözcükler ayrı yazılmaktadır. İkilemeyi oluşturan sözcüklerin arasına hiçbir noktalama işareti (virgül, kısa çizgi) konulmamaktadır. “gece gündüz” ikilemesi doğru yazılmıştır. “gecegündüz”, “gece-gündüz” ve “gece, gündüz” yanlış yazımlardır.

Zaman içerisinde kalıplaşarak birleşik sözcük olmuş ikilemeler bitişik yazılmaktadır. Tez çalışması kapsamında kalıplaşmış söz birliklerinden bahsedileceği için “civciv, cızbız, çerçöp, fıfıs, şipşak...” ikilemeleri gibi birleşik yazılan ikilemelere yer verilmeyecektir.

İkilemeler tek başlarına ya da başka sözcüklerle öbekleşerek deyim oluşturmaktadırlar. “*İleri geri* konuşmak sana yakışmıyor.” cümlesinde “ileri geri” ikilemesi “konuşmak” eylemiyle birleşerek deyim oluşturmuştur. “Yabancı basın, Fatih Terim’i *yere göğe* sığdıramıyor.”, “*Elden ayaktan* düşünce kimseler istemez bizi.” gibi örnekler verilebilir.

İkilemeler cümle içinde “*Ad*” göreviyle kullanılabilmektedirler. “Çocukları için *varını yoğunu* harcamıştı. “, “Canlarım benim, *malım mülküm* size feda olsun.” gibi örnekler verilebilir (Sakça, 2009).

İkilemeler cümle içinde “*Sıfat*” göreviyle kullanılabilmektedirler. “Havuzun içinde *irili ufaklı* taşlar vardı.”, “Bu *abuk sabuk* fikirleri kafana kim sokuyor?” vb. örnekleri bulunmaktadır (Sakça, 2009).

İkilemeler cümle içinde “*Zarf*” göreviyle kullanılabilmektedirler. “Sınıfta kaldığını öğrenince *hüngür hüngür* ağladı.”, “Ben gelene kadar burada *akıllı uslu* oturun.” vb. örnekler verilebilir (Sakça, 2009).

2.4 Kalıplaşmış Söz Birliklerinin Yapısal Özellikleri

Yapı, “bir bütünün kurucu parçaları ya da öğelerinin bütünün özel doğasını tanımlayacak biçimde karşılıklı bağıntıları” şeklinde tanımlanmaktadır (Williams, 2005). Dil açısından bakıldığında yapının, dilsel öğeler tarafından oluşturulan, iç bağıntılara ve öğelerin işlevlerine dayanan özerk nitelikli dizge, bütün olduğu söylenebilmektedir (Vardar, 1998). Bu bütünü oluşturan öğeler, çeşitli dizimsel ilişkiler ve anlamsal bağlarla bir araya gelmektedirler. Kalıplaşmış söz birliklerinin yapısal özellikleri aşağıda açıklanmaktadır (Gökdayı, 2008).

2.4.1 Kalıplaşmış olma

Bu birimlerin en belirgin yapısal özelliği, belirli bir biçimde kalıplaşmış olmalarıdır (Çotuksöken, 1994; Yüceol ve Özezen, 2001).

Deyimlerin bir kısmı hiç değişmeyen bir yapıya sahipken, büyük bir kısmı ad çekim, kip ve kişi ekleri gibi çeşitli çekim eklerini alabilmekte, konuşan ve dinleyene göre yeni biçimlere girebilmektedir (Aksoy, 198; Çotuksöken, 1994; Yüceol ve Özezen, 2001). “*Abayı yakmak* → *Abayı yakmış, Abayı yakmışsın, Abayı yakmış, Abayı yakmışsın, Abayı yakmışsın, Abayı yakmışsın*” örnek verilebilir.

Ayrıca deyimlerde, bütünü oluşturan sözcüklerin arasına başka sözcükler getirilebilmektedir. “*Abayı yakmak* → *Abayı fena yakmış*” örnek olarak verilebilir.

Kalıplaşmış bir yapıya sahip olmak, ikilemelerin de genel bir özelliğidir.

- Yinelemelerden oluşmayan ikilemelerin büyük bir kısmında sözcüklerin sırası değiştirilemez. “*Belli başlı, girintili çıkıntılı, aşağı yukarı, zarar ziyan, vb.*” ikilemelerdeki sözcüklerin sırası değiştirilemez.
- Bazı ikilemelerde de sözcüklerin birisi veya her ikisi de tek başına kullanılmamaktadır. “*Çoluk çocuk, yamru yumru, eski püskü, tek tük, ıvır zıvır vb.*” örnekleri verilebilir.
- İkilemelerin ad çekim eklerini alabildikleri görülmektedir. “*Dost düşman* → Bu işteki kararlılığımızı *dosta düşmana* gösterelim.” örnek verilebilir.

2.4.2 Öğelerin dizilişi

Kalıplaşmış söz birliklerini oluşturan birimlerin diziliş biçimleri yapısal özelliklerindendir (Çotuksöken 1994; Yüceol-Özezen 2001).

Deyimler tümce veya kalıplaşmış söz birliği olabilmektedir. “*Yorgan gitti kavga bitti, vb.*” tümce deyimler için; “*Canlı cenaze, abayı yakmak, el vermek, ağız dalaşı, vb.*” ile kalıplaşmış söz birliklerine örnek olarak verilebilir.

İkilemeler çoğunlukla kalıplaşmış söz birlikleri (*gacır gucur, ipsiz sapsız, vb.*) biçiminde bulunmaktadır.

2.4.3 Söz dizimsel türler

Kalıplaşmış söz birliklerinin sözdizimsel yapıları, kendi içinde belirli alt türlerde bulunmaktadır.

Deyimler;

- ✓ Bileşik Eylem : “*fırsat kollamak*”
- ✓ Ad ve Sıfat Tamlaması : “*püf noktası, acemi çaylak*”
- ✓ İlgeç Öbeği : “*damdan düşer gibi*”
- ✓ İkileme : “*sıkı fıkı*”
- ✓ Sonu mastarla biten sözcük grupları : “*ağzından çıkan kulağı işitmemek*”

İkilemeler; çoğunlukla kalıplaşmış söz birlikleri biçiminde bulunmaktadır.

- ✓ Yinelemeler: “*yavaş yavaş*”
- ✓ Eksiz sözcüklerin bir araya gelmesinden : “*köşe bucak*”
- ✓ Çeşitli ekler almış sözcüklerden : “*kucak kucağa, eninde sonunda, oflayıp puflamak*”
- ✓ Ad Tamlamaları : “*suyunun suyu*” vb.

- Kalıplaşmış söz birlikleri, ad tamlaması (*Allah aşkına*) ve ilgeç öbeği (*Allah için*) yapısında da bulunmaktadır.
- Tek sözcükten oluşan veya öbek biçiminde bulunan bazı sözler, “*de-, söyle-*” gibi eylemlerle genişletilip bir tümce haline getirilebilmektedir. Örneğin; “*Allah aşkına*”, “*Allah aşkınamak*” biçiminde yazılamaz. “*Allah aşkına dedi*” biçiminde tamamlanarak bir tümceye çevrilebilir.

2.4.4 Sözcük sayısı

Kalıplaşmış söz birliklerini oluşturan sözcük sayısı, onların yapılarını belirlemektedir. Kalıplaşmış söz birliklerinde bulunan sözcüklerin sayısı, oldukça az olmaktadır. Bunun nedeni de, az sayıda sözcükten oluşan birimleri hafızada saklayıp kullanmanın, çok sözcüklü yapılardan daha kolay olması olarak açıklanmaktadır (Wray, 2002).

Deyimlerin de kısa olduğu görülmektedir. “*el koymak, göz boyamak, ölçüyü kaçırmak, vb.*” örnekleri verilebilir.

En az iki “*gözünü açmak*” ve en fazla on beş “*Bizim tavuk bir yumurta yumurtlar, yedi mahalle duyar, elin kısırağı küheylan doğurur [hiç] sesi çıkmaz*” sözcüğün bulunduğu belirtilmiştir (Çotuksöken, 1994). Böyle örneklere rağmen, deyimlerin büyük bir kısmı iki sözcüklü olarak bulunmaktadır.

İkilemeler ise adı üzerinde çoğunlukla sözcük veya kalıplaşmış söz birlikleri “*aç susuz, saç saça, baş başa*” olan iki birimden oluşan bir yapıya sahiptir.

3. KALIPLAŞMIŞ SÖZ BİRLİKLERİ ÜZERİNE YAPILAN ÇALIŞMALAR

Doğal diller için yapılan çalışmalara baktığımızda daha ziyade kalıplaşmış söz birliklerinin bir üst kümesi olan eşdizim (collocation) kavramı üzerinde durulmuştur. Eşdizim; rastlantısal olamayacak miktarda bir arada görülen sözcükler arası ilişkiye verilen isim olarak tanımlanmaktadır (Hoey, 1991). Eşdizimlerin belirlenmesi amacıyla birçok istatistiksel ve kural tabanlı yöntemler geliştirilmiştir. Bu bölümde kalıplaşmış söz birlikleri üzerine yapılan yabancı ve Türkçe çalışmalardan bahsedilmektedir.

İstatistiksel çalışma olan karşılıklı enformasyon yöntemi (Church ve Hanks, 1990) bir arada görünen iki kelimenin birlikte ve kendi başına görülme olasılığı arasındaki ilişkiyi hesaplamaktadır. Eşdizimlerin, kelimeler arası ilişkisini gösteren yöntem genelde hesaplamalı dil bilim çalışmalarında kullanılmaktadır.

Smadja'ya (1993) göre eşdizimlerin tespiti için eşdizim gücü, dağılımı ve kelimeler arası uzaklık değerleri önemli olmaktadır. İstatistiksel çalışma olan yöntem, bir kelimenin kendinden sonra gelen tüm kelimelerle ilişkisinin belirlenmesi vasıtasıyla hangi kelimenin bu kelimeyle bir eşdizim oluşturacağının tespitini içermektedir.

Ortalama- varyans yöntemi eşdizimlerin saptanmasında kelimeler arasındaki uzaklığın önemli olduğunu göstermektedir (Manning ve Schütze, 1999). İngilizce üzerine yapılan istatistiksel çalışmalarda özellikle aralarına farklı kelimeler alan eşdizimleri yakalamakta başarılı olmaktadır.

Temel bir uygulama olan gözlenme sıklığı yönteminde derlemde bir arada bulunan kelimelerin sıklık değerleri ölçülmekte ve kelime birliklerinin sıklık değerleri azalacak şekilde listelenip, yüksek sıklık değerine sahip olan kelimelerin eşdizim olduğu kabul edilmektedir.

Oflazer (2004); Türkçe'de kelime birliklerinin tespit çalışmasında, kelime ekleri belirlenmiş ve sözcük türleri işaretlenmiş derlem üzerinde kural tabanlı

çözüm önermiştir. Yürütülmüş olan çalışmada morfolojik analiz yapılarak belli kurallar çerçevesinde eşdizimleri metin içerisinde tespit edilmekte ve kelime birliklerinin 4 farklı biçimbilimsel yapıdan oluştuğunu söylemektedir. Bunlar:

- ❖ *Lexicalized Collocations (Kelime biçimbirimlerinin sabit olmadığı söz birlikleri)*; Kelime birliklerinin ek aldığı ifadelerden bahsetmektedir.
Örneğin; *hiç olmazsa (Adverb+Neg. Verb)*, *ipe sapa gelmez (Noun+Noun+Neg. Verb)* gibi.
- ❖ *Semi-lexicalized Collocations (Kelime biçimbirimlerinin yarı sabit olduğu söz birlikleri)*; kelime birliklerinin bir kısmının ek almadığı bir kısmının alabildiği yarı sabit ifadelerden bahsetmektedir.
Örneğin; *kafayı ye- → kafayı yedim, kafayı yiyeceklerdi, kafayı yiyenler, kafayı yediği (Noun+Verb)*, *devam et- → devam ettim, devam ettiler (Noun+Verb)* gibi.
- ❖ *Non-lexicalized Collocations (Kelime biçimbirimlerinin sabit olduğu söz birlikleri)*; kelime birliklerinin tekrarları veya zıtlıkları içeren ifadelerden bahsetmektedir.
Örneğin; *ev ev (Noun+Noun)*, *koşa koşa (Verb+Verb)*, *uyur uyumaz (Verb+Neg. Verb)* gibi.
- ❖ *Named-entities (İsimlendirilmiş varlıklar)*; adlandırılmış varlıklardan bahsetmektedir.
Örneğin; *Türkiye Büyük Millet Meclisi vb.*

Oflazer (2004) çalışmasında kelime birliklerinin tespit edilmesi için yukarıda bahsedilen 4 farklı biçimbilimsel yapı kullanılmış ve 1100 kural oluşturulmuştur. Bu kuralların çalıştırılması sonucu duyarlılık %100 ve anma da %60 olarak belirtilmiştir.

Özkan (2007); “Türkiye Türkçesinde Belirteçlerin Fiillerle Birliktelik Kullanımları ve Eşdizimliliği” üzerine çalışmıştır. Eşdizimlilik kavramı içerisinde yer alan dilbilgisel sözcük birlikteliklerini şu şekilde sıralanmaktadır.

* Belirteç+Fiil → Açıkça söylemek, acı acı konuşmak

* Sıfat+Ad → Güzel sanatlar, derin uyku

* Niteleyici söz+Ad → Bir güzel, bir ara, bir yana

* Fiil+Ad → Bilir kişi, çıkmaz sokak

* Fiil+Fiil → Dönüp bakmak, yakıp yıkmak

* Ad+İlgeç → Bana göre, senin gibi

* ...

Özkan (2007); eşdizimsel yapıların sözcük türü değişimi olduğunu söylemektedir. Örneğin; *bir güzel* → *Söz+Ad* birlikteliği *belirteç* olurken, *bilir kişi* → *Fiil+Ad* birlikteliği ise *isim* olmaktadır.

Özkan (2007) çalışması sonucunda; belirteçlerin fiillerle anlamsal ve dizimsel ilişkisellikleri derlem tabanlı olarak ortaya koymuş ve sadece ilişkisellikler değil aynı zamanda belirteçlerin durağan biçimleri tespit edilmiştir. Aynı zamanda ilişkiselliklerin dağılımları belirlenmiş ve olumluluk-olumsuzluluk açısından yapılarını ortaya koymuştur.

Gökdayı (2008), çalışmasında çeşitli kaynaklarda verilen kalıp söz tanımları incelemiş, bu sözleri tanımlayabilmek ve tanıyabilmek için yapısal, işlevsel, anlamsal ve bağlamsal özellikleri tespit edilmesi üzerine çalışmıştır. Kalıplaşmış söz birliklerinin yapısal özelliklerini Gökdayı (2008) aşağıda dört başlık halinde vermektedir. Bunlar:

- Kalıplaşmış olarak bulunma,
- Öbeklerin dizilişi,
- Sözdizimsel türler,
- Az sayıda sözcükten oluşmadır.

Gökdayı (2008) çalışmasında; kalıp sözlerin tanımlanabilmesi ve tanınabilmesi için, yapı, işlev, anlam ve bağlam özellikleri teker teker yeterli olmadığını ve kalıplaşmış bir dil birimiyle karşılaşıncı, onun kalıp söz olup

olmadığını belirleyebilmek için bu özelliklerin bir arada düşünülmesiyle oluşabileceğini savunmaktadır.

Taşgüzel (2004); “İlköğretim Türkçe Ders Kitaplarında Öğretici Nitelikli Metinlerdeki Eşdizimsel Örüntülemelerin Görünümü” üzerine çalışmıştır. Metinlerin tutarlılık düzeyini gözlemlenebilir kılmak amacıyla çalışmasında, İlköğretim 6, 7, 8. sınıf Türkçe ders kitaplarındaki metinlerin söylem yapısı, sınıflar arası beklenen olası kavramsal gelişimler de göz önünde bulundurularak metin dilbilimsel veriler çerçevesinde çözümlemiştir.

Kumova (2011) çalışmasında, eşdizimlerin tespiti için kelimeler arası bağların ve kelime anlam bütünlüğünün değerlendirilmesine dayanan “eşdizim eğilimi yöntemi” geliştirmiştir. İstatistiksel olan bu yöntem, Türkçe derlem üzerinde uygulanarak eşdizim eğilimi yüksek olan bir kelimenin ardındaki kelimenin en az çabayla tahmin edilmesi gerektiği düşüncesini bir kelimeyi takip eden farklı kelime sayısına bakarak göstermektedir.

4. İKİLEMELERİN TESPİTİ ÇALIŞMASI

Bu bölümde ikilemelerin tespiti için kullanılan kuralların belirlenmesinden ve kuralların “Deyimler ve Atasözleri Sözlüğü” (Kuşçu v.d., 1997)’ndeki ikilemeler üzerinde çalıştırılması ile elde edilen sonuçlardan bahsedilecektir.

4.1 Sözlük Veritabanındaki İstatistikler

Tez kapsamında öncelikle “Deyimler ve Atasözleri Sözlüğü” (Kuşçu v.d., 1997) kullanılarak veri tabanına girilen kalıplaşmış söz birliklerinin gövdelenmesi el ile yapılmış ve ikilemelerin genel özellikleri belirlenmiştir. Veri tabanına girilen sözlükte gövdelenmemiş olarak 355 ve gövdelenmiş olarak 338 tane farklı ikileme bulunmaktadır.

Sözlük veri tabanına girilen ikilemeler incelenmiş ve dilbilgisi kaynaklarından da yararlanılarak bazı özelliklerin ikilemelerin tespitinde etken olacağı düşünülmüştür. Bunlar; birbirini takip eden kelimelerin uzunluğu, sözcük türü, sessiz harf sayıları, sesli harf sayıları ve benzer harflerdir. Bu özellikler doğrultusunda çıkarılan kurallar sözlük veri tabanında yürütülmüş ve geçerlilikleri istatistiksel olarak Çizelge 4.1’de gösterilmiştir. Çizelge 4.1’e bakıldığında kurallar yapı ve ses benzerliklerine göre ikiye ayrılmış ve ses benzerlikleri de kendi arasında sessiz harf benzerliği ve sesli-sessiz harf benzerliği olarak ayrılmış ve çizelgede gövdelenmiş ve gövdelenmemiş veri tabanındaki istatistikleri verilmiştir.

Yapı benzerliğinde sözcük türlerinin aynı olması, kelime uzunluğu, kelimeler arasındaki benzer harf sayıları ve aralarındaki farklar dikkate alınarak kurallar oluşturulmuştur. Sessiz harf benzerliğinde, kelimelerde geçen sessiz harfler dikkate alınarak sessiz harf sayısı ve benzerlikleri üzerinde durulmuştur. Aynı zamanda sesli harflerinde önemli olduğu görülmüş ve sesli harf benzerliği dikkate alınarak sesli-sessiz harf benzerlikleri üzerine kurallar yazılmıştır. Çizelge 4.1’de bu kuralların veri tabanında çalışması sonucunda elde edilen ikileme sayısı ve oranları verilmiştir.

Çizelge 4.1 Veri Tabanından Çıkarılan Kurallara Uyan İkilemelerin İstatistikleri
(Toplam ikileme sayısı gövdelenmiş derlemde 338 ve gövdelenmemiş derlemde 355'dir.)

Kurallar		Gövdelenmiş		Gövdelenmemiş		
		İkileme Sayısı	Oran	İkileme Sayısı	Oran	
Yapı Benzerlikleri	1-Kelime uzunlukları ≤ 1	338	%100	355	%100	
	2-Sözcük türleri aynı	338	%100	355	%100	
	3-Kelime uzunluğu arasındaki fark ≤ 2	322	%95	310	%87	
	4-Kelimeler arasındaki benzer harf sayısı ≤ 0	304	%90	261	%74	
	5-Kelime uzunluğu ≥ 2	321	%95	346	%98	
	6-Kelime uzunluğu=Benzer harf sayısı veya Kelime uzunluğu-Benzer harf sayısı=1	214	%63	153	%43	
Ses Benzerlikleri	Sessiz Harf Benzerliği	7-Kelimelerin toplam sessiz harf sayısı ≤ 0	338	%100	355	%100
		8-Arka arkaya gelen kelimelerin benzer sessiz harf sayısı ≤ 0	288	%85	235	%95
		9-Birinci kelimedeki toplam sessiz harf sayısı=Kelimelerdeki benzer sessiz harf sayısı	190	%56	131	%90
	Sesli-Sessiz Harf Benzerliği	10-Birinci kelime uzunluğu=İkinci kelime uzunluğu veya İkinci kelime uzunluğu-Birinci kelime uzunluğu=1	314	%93	292	%82
		11-Arka arkaya gelen kelimelerin benzer sesli harf sayısı ≤ 0	256	%76	216	%61
		12-Kelimelerdeki benzer sesli ve sessiz harf sayılarının toplamı ile Birinci kelime uzunluğu oranı=%70	209	%62	135	%38

Gövdelemenin ikilemelerin tespitinde etkili olduğu Çizelge 4.1’de görülmektedir. Gövdelenmiş sözlük veri tabanında kuralların uygulanması sonucunda elde edilen ikileme sayısı ile gövdelenmemiş sözlük veri tabanında aynı kuralın uygulanması sonucu elde edilen ikileme sayısı daha azdır. Böylece ikilemelerin ek aldığı anlaşılmaktadır.

Çizelge 4.1’deki kuralların sözlük veritabanındaki istatistikleri incelendiğinde; “Kelime uzunlukları ≤ 1 ”, “Sözcük türleri aynı” ve “Kelimelerin toplam sessiz harf sayısı ≥ 0 ” kurallarının oranları %100’dür. Kısacası tüm ikilemeleri yakalayan kurallar olduğundan ilerleyen bölümlerde eleyici kural olarak kullanılacaktır.

İkilemelerde gözlenen diğer bir özellik ise benzer harflerin bulunmasıdır. Yapı benzerliği kurallarında benzer harflere bakılmıştır ve “sabah akşam, bir iki” gibi benzer harf sayısı az olan ikileme sayısı nadir olduğundan bu ikilemeler Çizelge 4.1’deki Kural-4’ün yazılmasında dikkate alınmamıştır.

Çizelge 4.1’e bakıldığında kelime uzunluğu 2(iki)’den büyük olan ikilemelerin oranı %98’dir ve bu durum ikilemelerin birçoğunun bu kuralı sağladığını gösterir. Yapı benzerliğinde kelime çiftlerindeki, ilk kelimenin uzunluğu ile benzer harf sayısı eşit veya aralarındaki fark 1(bir) olduğunda ikileme olarak yakalanmaktadır.

Yapı benzerliğinde benzer harf sayısına bakılmıştır. “açık seçik, eğrisi doğrusu, ufak tefek” gibi ikilemelerin sessiz harfleri benzer olmasına karşın sesli harfleri farklı olduğu için Çizelge 4.1’deki Kural-6 tarafından yakalanamamaktadır. Bu nedenle yalnızca sessiz harf benzerliğine bakmak için bazı kurallar belirlenmiştir. Bu kurallardan biri olan Çizelge 4.1’deki Kural-8’e bakıldığında arka arkaya gelen kelimelerin benzer sessiz harf sayısı 0(sıfır)’dan farklı olması durumunda kelime çifti olarak yakalamaktadır. Çizelge 4.1’deki Kural-9’da ise birinci kelimenin toplam sessiz harf sayısı ile kelimeler arasındaki benzer sessiz harf sayısı birbirine eşit olduğunda kelime çiftlerini yakalamaktadır. “dökülüp saçılan, bölük pörçük, dayalı döşeli” gibi ikilemelerin yakalanamaması sessiz harf benzerliklerinin az olmasından kaynaklanmaktadır.

Çizelge 4.1'deki Kural-11'in değerlerine bakıldığında sessiz harf kadar sesli harf benzerliğinin de dikkate alınması gerektiği görülmüştür. Örneğin; “dayalı döşeli” gibi ikilemelerde birinci kelimedeki “a” harfi yerine “e” harfi, “ı” harfi yerinede “i” harfi gelmektedir. Bu yüzden sesli harf benzerliğinde “a=e, ı=i, o=ö, u=ü” olarak ele alınmaktadır.

Sesli-sessiz harf benzerliği dikkate alınarak yazılan Çizelge 4.1'deki Kural-12 kelimelerde bulunan benzer sesli-sessiz harflerin toplamı ile kelime uzunluğu arasındaki orana bakmaktadır. Sessiz harf benzerliğinde ise “dayalı döşeli, birer ikişer” gibi ikilemeleri yakalayamamaktadır. Sesli-sessiz harf benzerliğinde ise hem benzer sessiz harflere hem de sesli harflere bakmakta ve kelime uzunluğu ile benzer harf sayısı arasındaki oranı göz önüne almaktadır. Böylece bu oran doğrultusunda “dayalı döşeli, birer ikişer” gibi ikilemeler yakalanmaktadır.

Çizelge 4.1'deki Kural-12'de harf benzerliği oranı bulmak için tek tek denemeler yapılmış ve değerlendirme ölçütlerinde değişiklik gösteren oranlar ele alınmıştır. Çizelge 4.2'de harf benzerliği oranlarına göre elde edilen değerler yer almaktadır.

Çizelge 4.2 Sesli-Sessiz Harf Benzerlik Oranı Kuralının Veri Tabanındaki İstatistikleri
(Toplam ikileme sayısı gövdelenmiş derlemde 338 ve gövdelenmemiş derlemde 355'dir.)

Kelimelerdeki benzer sesli ve sessiz harf sayısı toplamı ile Birinci kelime uzunluğu oranı	Gövdelenmiş		Gövdelenmemiş	
	İkileme Sayısı	Duyarlılık	İkileme Sayısı	Duyarlılık
%33	285	%84	236	%66
%50	263	%78	211	%59
%70	209	%62	150	%42
%76	190	%56	135	%38
%81	155	%46	102	%29

Çizelge 4.2'de görüldüğü üzere benzerlik oranı arttıkça ikileme sayısı azalmakta, azaldıkça ise artmaktadır. Böylece harf benzerlik oranı ile ikileme sayısı arasında ters orantı vardır.

4.2 İkilemelerin Tespiti İçin Belirlenen Kurallar

İkilemelerin tespiti için belirlenen kuralların özellikleri Çizelge 4.3'te gösterilmiştir.

Çizelge 4.3 İkilemelerin Tespiti İçin Belirlenen Kurallar ve Özellikleri

KURALLAR		
Yapı Benzerlikleri	Ses Benzerlikleri	
Uzunluk-Türü-Benzerliği	SesSiz Harf	Sesli-Sessiz Harf Benzerliği Yüzdesi
1-Kelime uzunluğu,	1-Kelimelerdeki sessiz harf sayısı,	1-Kelimelerin uzunluğu,
2-Sözcük türlerinin benzer olması,	2-Kelimeler arasındaki benzer sessiz harf sayısı.	2-Sözcük türlerinin benzer olması,
3-Kelimelerdeki harflerin benzerliği.		3-Benzer sessiz ve sesli harf sayısına,
		4-Benzer harflerin toplamının kelime uzunluğuna oranına.

Çizelge 4.3'ten de anlaşılacağı gibi kurallar kelimelerin yapı ve ses benzerliklerine göre oluşturulmuştur. Yapı benzerliklerinde kelimelerin uzunluğu, sözcük türleri ve harf benzerliklerine bakılmıştır. Ses benzerliğinde ise sesli ve sessiz harf benzerlikleri dikkate alınmıştır.

Kuralların uygulanmasında harf benzerlikleri için kelime sonlarından başlayarak bakılmıştır. Kelimelerin sondan değerlendirilmesindeki amaç ikilemelere genel olarak bakıldığında son harflerin benzerliğinin daha fazla olduğudur.

Örneğin; 1- a-b-u-k s-a-b-u-k 2- k-u-b-a k-u-b-a-s

Yukarıdaki birinci durumda ikilemede ilk harflerinde benzerlik yoktur. Aynı şekilde ikinci, üçüncü vb. harflerinde de benzerlik göstermemektedir. Yukarıdaki ikilemenin kelimelerini ters ele aldığımızda yani ikinci durumda dört tane harf benzerliği bulunmaktadır.

5. İKİLEMELERİN TESPİTİ İÇİN GELİŞTİRİLEN KURALLARIN UYGULANIŞ YÖNTEMLERİ

Bu bölümde ikilemelerin tespitinde kullanılan derlemden, yöntemlerden, geliştirilen kurallardan ve kuralların uygulanış yöntemlerinden bahsedilecektir.

5.1 Derlem

Derlem, “dilbilimsel bilginin koleksiyonudur, yazılı yada kaydedilen konuşmalar şeklinde olabilir” (Crystal, 1991) veya “doğal olarak meydana gelen metinlerden dilin çeşitliliğini ve durumunu belirlemek amacıyla seçilen ve bir araya getirilen metinler topluluğu” şeklinde tanımlanabilir (Sinclair, 1991). Derlemler aslında doğal dil işleme alanında kullanılmak için yazılardan oluşturulmuş özel bir veritabanıdır ve kelimeleri hızlı şekilde bulma ve işleme gibi özel işlemleri yapmaya izin vermektedir (Aktaş, 2006). Farklı bir tanımıyla derlem, doğal dil işleme, bilgi geri getirim, hesaplamalı dil bilim alanlarında ve çeşitli sistemlerin geliştirilmesi veya sınanması sırasında kullanılan metin gruplarını oluşturmaktadır (Kumova, 2005).

Tez çalışmasında “Orta Doğu Teknik” ve “Sabancı” üniversitelerinin ortaklaşa çalışmaları sonucu geliştirmiş oldukları Türkçe ODTÜ derleminin alt kümesinden oluşan OSTAD derlemi kullanılmıştır. OSTAD derlemi ~52179 kelime (yüzeysel form), ~17691 farklı kelime, ~8734 gövde ve ~4480 kökten oluşmaktadır (Ofłazer v.d., 2003). Gövdelenmiş OSTAD derlemi 46532 kelime ve 5665 cümleden oluşmaktadır.

Kelime, bir veya birden çok heceli ses birliklerinden oluşan, aynı dili konuşan kişiler arasında zihinde tek başına kullanıldığında somut veya soyut bir kavrama karşılık olan yahut da somut ve soyut kavramlar arasında geçici ilişkiler kurmaya yarayan dil birimidir. Kök ise kelimenin ekleri ayrıldıktan sonra geriye kalan en küçük anlamlı birimidir. Köklere yapım eklerinin eklenmesiyle oluşturulan ve anlamca bağlandıkları kökle ilişkili bulunan türemiş kelimeye ise gövde denmektedir (*BSTS / Gramer Terimleri Sözlüğü*, 2003). Tez çalışmasında gövdelenmiş ve gövdelenmemiş (yüzeysel form) derlem üzerinde çalışılmıştır.

Böylece ikilemelerin tespitinde dördüncü bölümde görülen gövdelenmenin etkisi bu bölümde OSTAD derleminde araştırılmıştır.

“Deyimler ve Atasözleri Sözlüğü” (Kuşçu v.d., 1997)’nde ve literatür kaynaklarında geçen kalıplaşmış söz birlikleri OSTAD derleminde elle işaretlenmesi yapılmıştır. İşaretleme yapılırken tüm kalıplaşmış söz birliklerine numara verilmiş ve ikileme olanları ise “x” işareti konulmuştur. İşaretleme yapılmış kalıplaşmış söz birliklerinin sayısı 1224’dür. Bunların 287 tanesi ikilemeleri oluşturmaktadır.

İkilemelerin gövdelenmiş ve gövdelenmemiş derlemdeki tüm ikilemeler ve farklı ikilemelerin sayısı Çizelge 5.1’de verilmektedir.

Çizelge 5.1 Derlemde geçen ikilemelerin sayı listesi

Derlemde Geçen İkilemeler	Gövdelenmiş Derlem	Gövdelenmemiş Derlem
<i>Tüm İkilemelerin Sayısı</i>	287	287
<i>Farklı İkilemelerin Sayısı</i>	162	189

Çizelge 5.1’den görüldüğü üzere tüm ikilemelere ve farklı ikilemelere bakıldığında frekans bilgilerinin ikilemelerin bulunmasında fazla katkı sağlayamayacağı anlaşılmaktadır.

5.2 İkilemelerin Tespitinde Kullanılan Yöntem

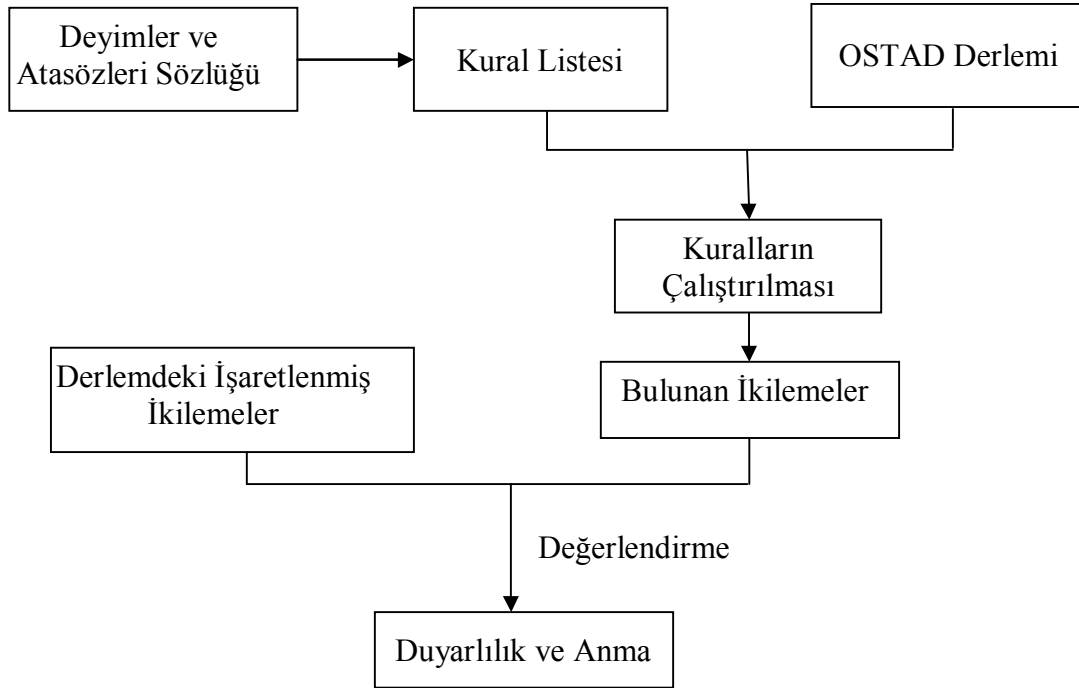
Türkçe köken olarak Ural-Altay diller grubuna girmektedir. Kural tabanlıdır ve kurallarını bozmadan yaşaya gelmiştir. Türkçe bitişken bir dildir; her sözcük bir kök ve bu kökün sonuna eklenmiş olan eklerden oluşmaktadır (Cebiroğlu v.d., 2002).

Doğal dil işleme çalışmalarında kalıplaşmış söz birliklerinin tespitinde hem Kural Tabanlı hem de İstatistiksel Yöntemler kullanılmıştır. İstatistiksel yöntemler

harfin ve kelimenin gözlenme sıklığı bilgisini temel almaktadır. Genellikle uygulanan istatistiksel teknikler; gözlenme sıklığı, karşılıklı enformasyon (Church ve Hanks, 1990), ortalama-varyans, Smadja yöntemi (1993) ve hipotez testleridir (Kumova, 2011). Kural tabanlı yöntem ise belli bir problemin çözümü için belirlenen kuralların çalıştırılması ve elde edilen sonuçların değerlendirilmesidir.

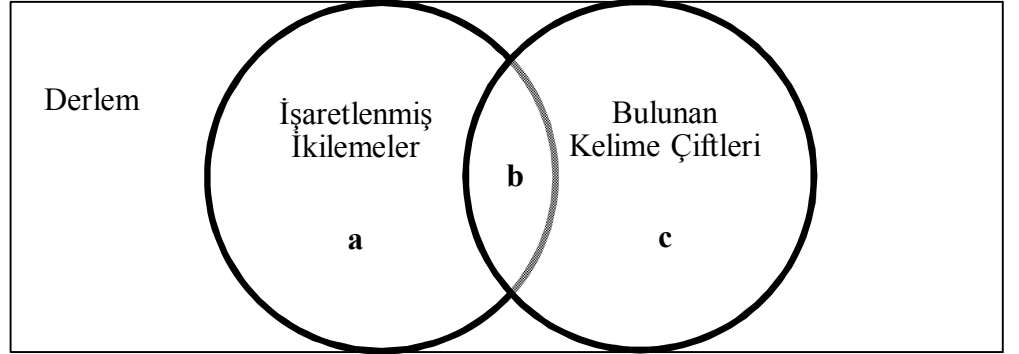
Tez çalışmasında kullanılan kural tabanlı yöntem; kelime türü işaretlenmiş veriler üzerinde çalışıp, sözcük türü, kelime uzunluğu gibi belli kurallar çerçevesinde sorgu kümesi vasıtasıyla derlemde kalıplaşmış söz birliklerinin tespitini amaçlamaktadır (Ofłazer v.d., 2004; Gökdayı, 2008).

Tez kapsamı doğrultusunda belirlenen kurallar OSTAD derlemi üzerinde uygulanmıştır. Kuralların uygulanması sonucunda elde edilen başarımlar, bilgi geri getirim alanında kullanılan duyarlılık (precision) ve anma (recall) ölçütleri ile değerlendirilmiş ve kıyaslanmıştır. İkilemelerin tespitinde gövdelenmiş ve gövdelenmemiş derlem üzerinde çalışılmış ve gövdelemenin ikilemelerin tespitindeki etkisi araştırılmıştır. Tezde kullanılan kural tabanlı yöntemin şematik gösterimi aşağıdaki gibidir.



Şekil 5.1 İkilemelerin Tespitinde Kullanılan Kural Tabanlı Yöntemin Şematik Gösterimi

Yukarıdaki şemadan da anlaşılacağı üzere ikilemelerin tespiti için “Deyimler ve Atasözleri Sözlüğü” veri tabanına girilmiş ve sözlükteki ikilemeler ile literatür kaynakları dikkate alınarak kural listesi oluşturulmuştur. Kurallar OSTAD derlemi üzerinde çalıştırılmış ve ikileme aday listesi elde edilmiştir. Elde edilen ikilemeler derlemde işaretlenmiş ikilemelerle karşılaştırılarak değerlendirilmiştir. Değerlendirme ölçütlerinin şematik olarak Şekil 5.2’de gösterilerek açıklanmıştır.



Şekil 5.2 Değerlendirme Ölçütlerinin Şematik Gösterimi

Şekil 5.2’de görüldüğü gibi duyarlılık= $b/(b+c)$ ve anma= $b/(a+b)$ olarak hesaplanmaktadır. Kısacası *duyarlılık (precision)*; bir yöntemin çekilen ikililer ile ikileme olarak atadığı ikililer içindeki doğru ikilemelerin sayısal oranıdır. *Anma (recall)* ise; bir yöntem ile derlem içinden çekilen doğru ikilemelerin derlem içinde yer alan doğru ikilemelere oranı olarak tanımlanmaktadır.

Tez kapsamında öncelikle “Deyimler ve Atasözleri Sözlüğü” (Kuşçu v.d., 1997) ve literatür kaynakları kullanılarak belirlenen kurallar, yapı ve ses benzerliğine göre ayrılmıştır. Benzerliklere göre belirlenen kuralların uygulanış yöntemleri aşağıda kısaca açıklanmıştır.

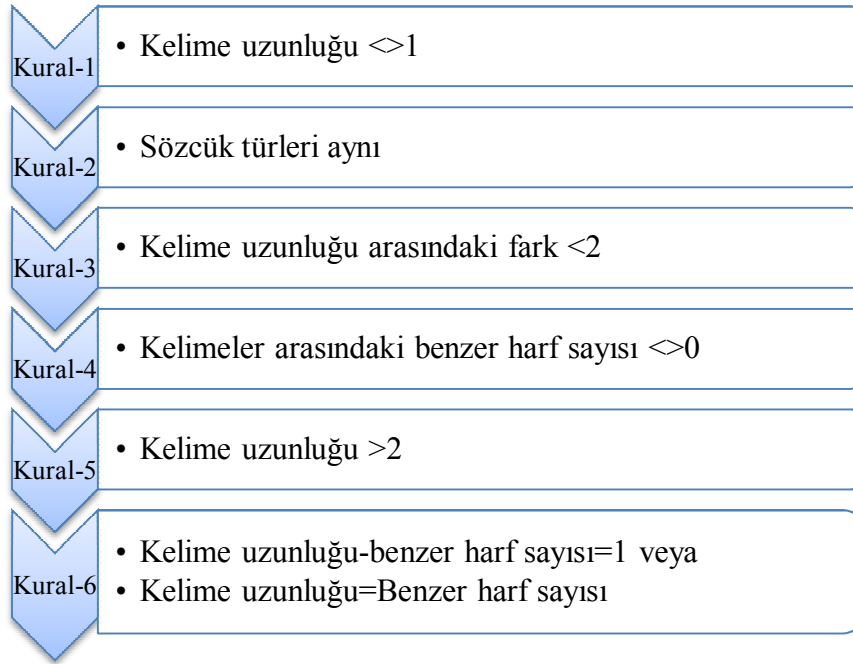
Süzgeç Yöntemi : Bir kuralın uygulanması sonucunda yakalanan kelime çiftlerini bir sonraki kuralın çalışmasında kullanmaktadır. Böylece her kuralın uygulanması sonucu derlemdeki kelime çiftleri elenmiş olur yani kurallar süzgeç görevi görür.

Bariyer Yöntemi : İlk iki kural, derlemdeki tüm ikilemeleri yakaladığından bariyer görevi görmekte ve bu kuralların uygulanması sonucu yakalanan kelime çiftleri üzerinden diğer kurallar çalıştırılmaktadır.

Yukarıda açıklanan süzgeç ve bariyer yöntemleri, kuralların yapı ve ses benzerliğine göre ikiye ayrılır. Ses benzerliği de sesli harf benzerliği ve sessiz-sesli harf benzerliği olarak ikiye ayrılır. Aşağıda benzerlikler ve kuralların uygulanış yöntemleri açıklanmaktadır.

5.2.1 Yapı benzerliğine göre belirlenen kurallar ve kuralların uygulanış yöntemleri

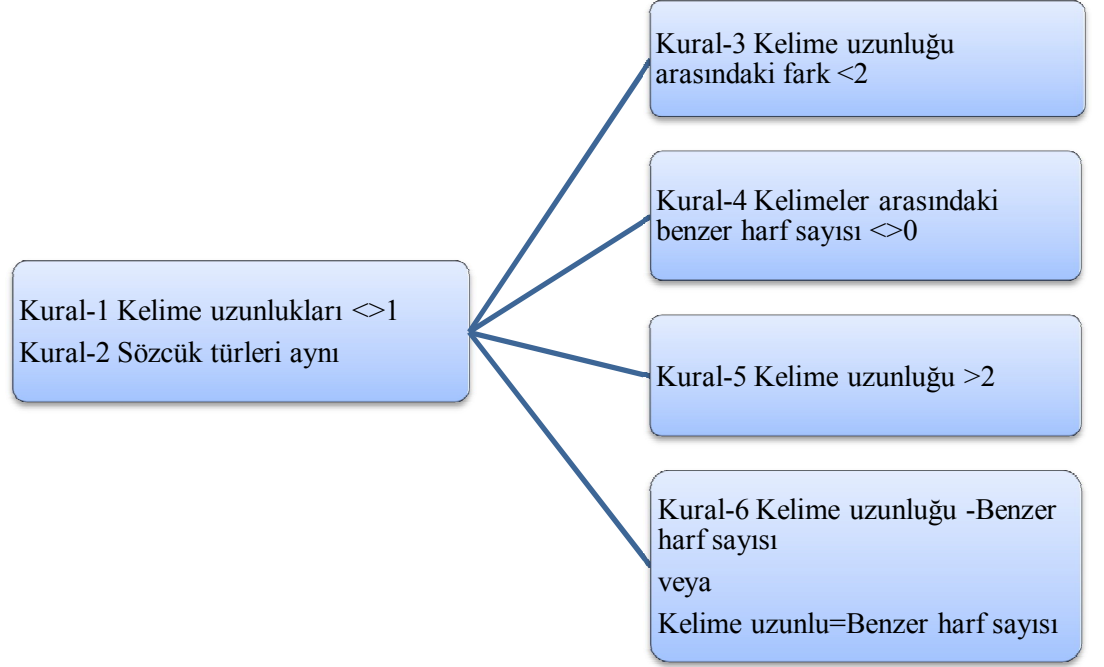
Yapı benzerliği kurallarında; sözcük türü, kelime uzunluğu ve benzer harf sayısı dikkate alınarak yazılmıştır. Yazılan kurallar süzgeç ve bariyer yöntemleri kullanılarak OSTAD derleminde çalıştırılmıştır.



Şekil 5.3 Yapı Benzerliği Kurallarının Süzgeç Yöntemi ile Uygulama Şeması

Şekil 5.3’de yapı benzerliğine göre oluşturulan kuralların süzgeç yöntemi ile uygulaması gösterilmiştir. Süzgeç yapısının mantığı şu şekilde çalışmaktadır. Öncelikle arka arkaya gelen kelime çiftleri alınıp Kural-1 işlenmektedir. Kural-

1'den elde edilen kelime çiftleri alınıp ve Kural-2 için işletilmektedir. Bu şekilde tüm kurallar çalıştırılmakta ve en son elde edilen sonuçlar başarıyı vermektedir.



Şekil 5.4 Yapı Benzerliği Kurallarının Bariyer Yöntemi ile Uygulama Şeması

Şekil 5.4'de yapı benzerliği kurallarının bariyer yöntemi ile uygulaması gösterilmiştir. Kural-1 ve Kural-2'nin çalışması sonucu Çizelge 4.1'de görüldüğü üzere sözlük veri tabanındaki tüm ikilemeleri yakaladığı için bu kurallar bariyer oluşturmaktadır. Kısacası ilk iki kural peş peşe çalıştırıldıktan sonra elde edilen kelime çiftleri diğer kuralların çalışması için derlem olarak görev yapmaktadır. İlk iki kural haricindeki kurallar, ilk iki kuralın yakaladığı kelime çiftlerinde çalıştırılıp ve çıkan sonuçlar ayrı ayrı değerlendirilmektedir. Şekil 5.3'deki süzgeç yönteminde ise çalıştırılan her kuralın sonucunda bulunan kelime çiftleri bir sonraki kuralı çalıştırmak için derlem görevi yapmaktadır.

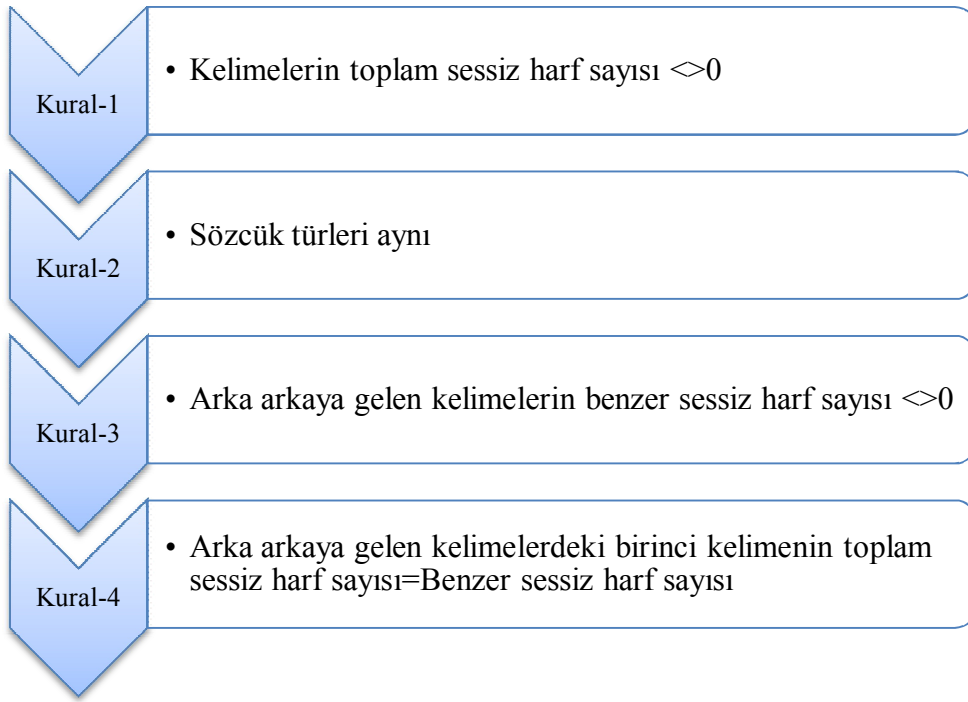
Yakalanan ikilemelere örnek verilirse “çoluk çocuk” ikilemesinde sözcük türleri ve kelime uzunlukları aynıdır. Kelimelerdeki benzer harf sayısı ile kelime uzunluğu arasındaki fark 1(bir) olduğundan bu ikilemeyi yakalamaktadır. Buna karşılık “açık seçik” ikilemesinde harf benzerliği ile kelime uzunluğu arasındaki fark 2(iki) olması nedeniyle bu ikilemeyi kurallar yakalayamamaktadır.

5.2.2 Ses benzerliğine göre belirlenen kurallar ve kuralların uygulanış yöntemleri

Ses benzerlikleri dikkate alınarak yazılan kurallar sessiz harf benzerliği ve sesli-sessiz harf benzerliği olmak üzere ikiye ayrılmaktadır.

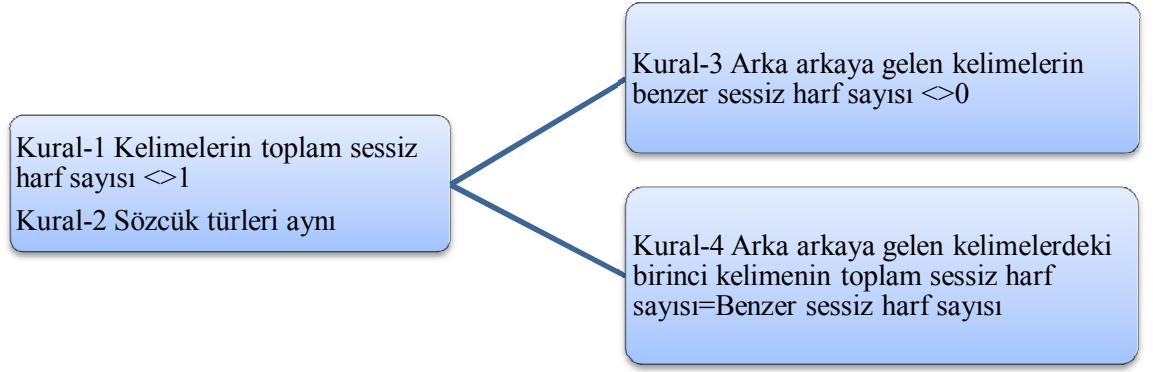
5.2.2.1 Sessiz harf benzerliğine göre belirlenen kurallar ve kuralların uygulanış yöntemleri

Sessiz harf benzerliği dikkate alınarak yazılan kurallar süzgeç ve bariyer yöntemleri kullanılarak OSTAD derleminde çalıştırılmıştır.



Şekil 5.5 Sessiz Harf Benzerliği Kurallarının Süzgeç Yöntemi ile Uygulama Şeması

Şekil 5.5'te sessiz harf benzerliği dikkate alınarak belirlenen kuralların süzgeç yapısında uygulaması gösterilmiştir. Arka arkaya gelen kelime çiftlerini ilk kural tarafından çalıştırıldıktan sonra yakalanan kelime çiftleri Kural-2 tarafından çalıştırılmaktadır. Kural-3 ise Kural-2'den elde edilen kelime çiftlerinde ve aynı şekilde Kural-4 ise Kural-3'te yakalanan çiftleri ile çalıştırılmaktadır. Kural-4'ün çalışması sonucunda yakalanan kelime çiftleri ve ikilemeler başarıyı vermektedir.



Şekil 5.6 Sessiz Harf Benzerliği Kurallarının Bariyer Yöntemi ile Uygulama Şeması

Şekil 5.6’da bariyer yöntemi ilk iki kural Çizelge 4.1’de veri tabanındaki tüm ikilemeleri yakaladığından bariyer görevi görmektedir. İlk iki kuralın derlemde çalışması sonucu yakalanan kelime çiftlerinde ayrı ayrı Kural-3 ve Kural-4 üzerinde uygulanmış ve bulunan sonuçlar değerlendirme ölçütleri ile değerlendirilmiştir.

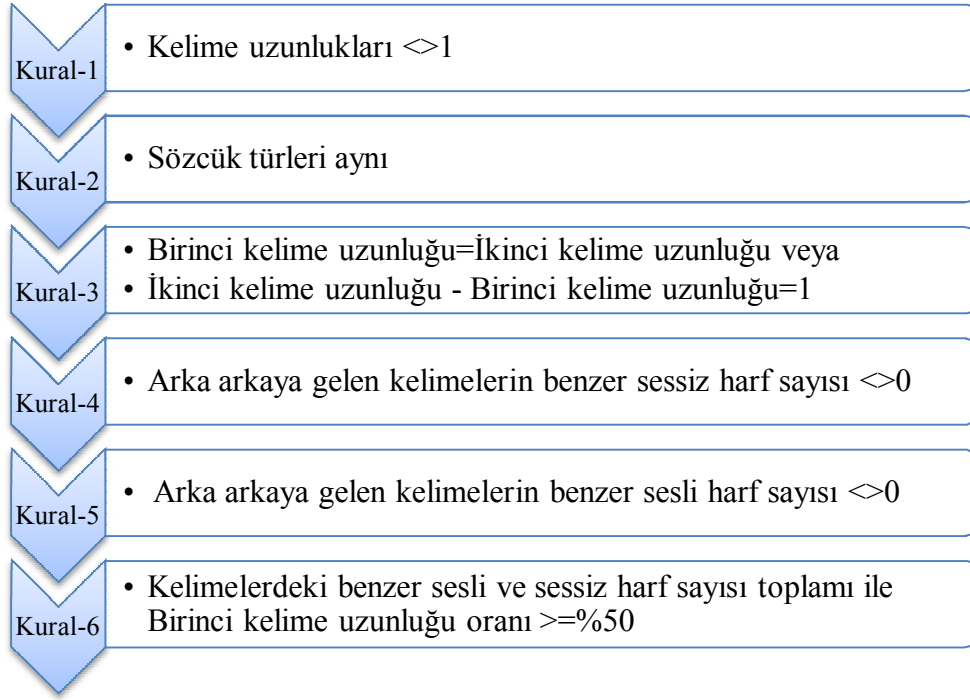
Kuralların uygulanması sonucu yakalanan ikilemelerden “Açık seçik” ikilemesini ele aldığımızda “açık” kelimesindeki toplam sessiz harf sayısı 2(iki)’dir. “açık seçik” ikilemesinde benzer sessiz harf sayısı 2(iki)’dir. İki şarttaki harf sayıları eşit olduğundan “açık seçik” ikilemesini sessiz harf benzerliği kuralları tarafından yakalamaktadır.

Yakalanamayan ikilemelere bakıldığında, örneğin “bayat mayat”. “bayat” kelimesindeki toplam sessiz harf sayısı 3(üç)’dür. “bayat mayat” ikilemesinde benzer harf sayısı 2(iki)’dir. İki şarttaki harf sayıları birbirine eşit olmadığından “bayat mayat” ikilemesini sessiz harf benzerliği kuralları yakalayamamaktadır. Bu da sessiz harf benzerliği kurallarının sınırlılığını göstermektedir.

Yakalanan ama ikileme olmayan kelimelere bakıldığında, örneğin “olan yalan” kelime birliğinde “olan” kelimesinin toplam sessiz harf sayısı 2(iki)’dir. “olan yalan” kelime çiftinde benzer harf sayısı 2(iki)’dir. İki şarttaki harf sayıları benzer olduğundan “olan yalan” kelime çiftini sessiz harf benzerliği kuralları yakalamaktadır. Bu da sessiz harf benzerliği kurallarının zayıf yönünü göstermektedir.

5.2.2.2 Sesli-sessiz harf benzerliğine göre belirlenen kurallar ve kuralların uygulanış yöntemleri

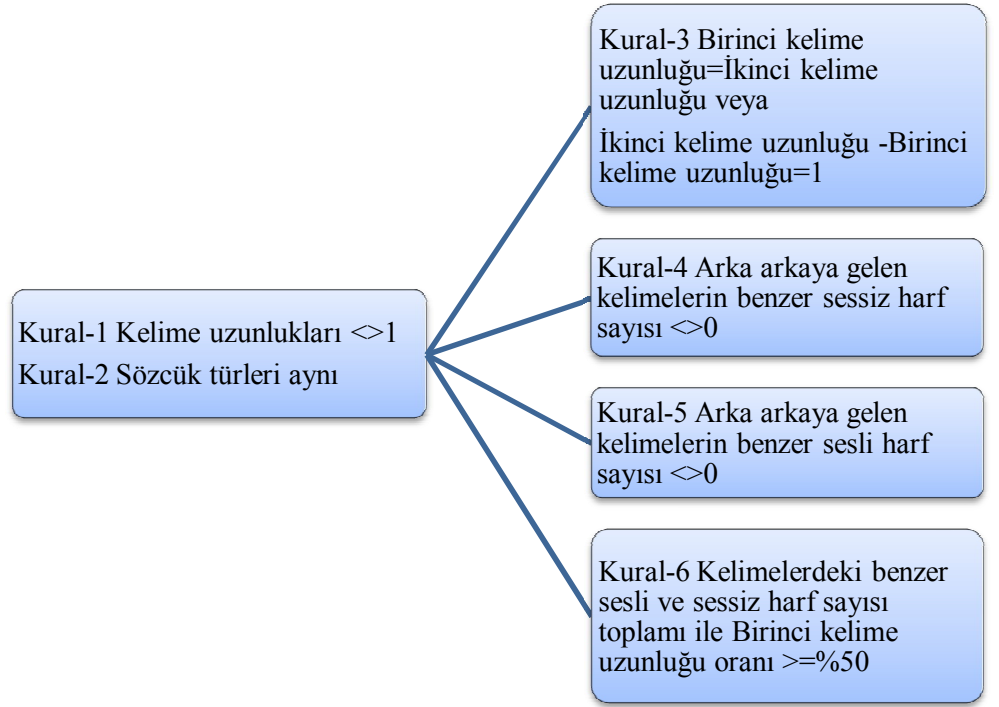
Tezde ses benzerliklerinden sesli ve sessiz harf benzerliği dikkate alınarak geliştirilen kurallar süzgeç ve bariyer yöntemleri kullanılarak OSTAD derleminde çalıştırılmıştır.



Şekil 5.7 Sesli-Sessiz Harf Benzerliği Kurallarının Süzgeç Yöntemi ile Uygulama Şeması

Şekil 5.7’de sesli-sessiz harf benzerliği kurallarının süzgeç yöntemi uygulama şeması verilmiştir. Sesli-sessiz harf benzerliği kurallarına göre, ilk önce Kural-1 çalıştırılmaktadır. Kural-1’de yakalanan kelime çiftleri Kural-2 için çalıştırılır ve aynı şekilde yakalanan kelime çiftleri Kural-3 için çalıştırılıp, diğer kurallara da aynı işlem uygulanır. Derlemde arka arkaya gelen kelimelerin uzunlukları ≤ 1 ve sözcük türleri aynı olan iki kuralla derlemdeki kelime çiftlerinin büyük çoğunluğu elenmektedir. Kelimelerin uzunluk farkı 1 veya birbirine eşit ise kelime çiftlerindeki benzer sessiz harf sayısına bakılmaktadır. 0(sıfır)’dan farklı ise kelime çiftleri arasında benzerlik gösterdiği varsayılarak benzer sesli harflere bakılmaktadır. Veri tabanındaki ikilemelerin incelenmesi sonucu sesli harflerin kendi içinde değişebildiği görülmüştür. ”a→e, ı→i, o→ö, u→ü” değişikliği göz önüne alınarak benzer sesli harf sayısı hesaplanmış ve sayının

0(sıfır)'dan farklı olması sonucunda kelimeler arasında benzerlik olduğu anlaşılmaktadır. Benzer sesli ve sessiz harf toplamının kelime uzunluğuna oranı belli bir yüzdenin üstünde ise kelime çiftlerini yakalamaktadır. Böylece kuralların çalışması sonucu kelimeler arasındaki harf benzerliğini belli bir yüzde değeri kullanarak ikilemeleri bulmayı amaçlamaktadır.



Şekil 5.8 Sesli-Sessiz Harf Benzerliği Kurallarının Bariyer Yöntemi ile Uygulama Şeması

Şekil 5.8'deki şema yapısı incelendiğinde ilk iki kural Çizelge 4.1'den de anlaşılacağı gibi tüm ikilemeleri yakaladığından bariyer görevi görmektedir. İlk iki kuralın derlemde çalışması sonucu yakalanan kelime çiftlerinde ayrı ayrı diğer kurallar için uygulanmış ve bulunan sonuçlar değerlendirme ölçütleri ile değerlendirilmiştir.

Örnek vermek gerekirse, “haşır neşir” ikilemesinde benzer sessiz harflerin toplam sayısı 2(iki) ve benzer sesli harflerin toplam sayısı 2(iki)'dir. İki değerde sıfırdan farklı ve ikilemenin kelime uzunlukları aynıdır. İkilemedeki benzer sesli ve sessiz harf sayısı toplamı kelime uzunluğuna oranı %81 olduğundan yakalanmaktadır.

Yakalanamayan ikilemelere bakıldığında “sabah akşam” ikilemesindeki benzer harf sayısı 2(iki)’dir. İkilemedeki benzer sesli harf sayısı toplamı kelime uzunluğuna oranı %20 olduğundan ve kelime çiftleri arasındaki benzerlikte %33’den fazla olanlara bakıldığından dolayı “sabah akşam” ikilemesini yakalamamaktadır. Bu ikilemedeki %20 oranındaki benzerliğin alınmamasının önemli nedeni ikilemelerin daha fazla olduğu kelime çiftlerini yakalamak ve bu nedenle sesli-sessiz harf benzerliğinde, %33’den fazla olan harf benzerlik oranına bakmamızdaki amaç değerlendirme ölçütü olan duyarlılığı arttırmak içindir.

Yakalanan ama ikileme olmayan kelime çiftlerine bakıldığında “olan yalan” kelime çiftinde benzer harf sayısı 3(üç)’tür. Kelime çiftindeki benzer sesli ve sessiz harf sayısı toplamı kelime uzunluğuna oranı %70 olduğundan dolayı kelime çiftlerini yakalamaktadır.

Not: Sesli-sessiz harf benzerliğinde en iyi sonucu harf benzerliği oranı %70 olduğunda alınmaktadır.

6. BULGULAR VE TARTIŞMA

Tez çalışmasında öncelikle “Deyimler ve Atasözleri Sözlüğü” (Kuşçu v.d., 1997)’ndeki ve literatür kaynaklarındaki kalıplaşmış söz birlikleri incelenip, kurallar oluşturulmuştur. Daha sonra bu kurallar OSTAD derlemi üzerinde süzgeç ve bariyer yapılarında çalıştırılmış ve ikilemelerin yakalanma başarımlarına bakılmıştır. Yakalanan ikilemelerin değerlendirilmesi ve kıyaslanması amacıyla yüzeysel ve gövdelenmiş derlem üzerinde duyarlılık ve anma değerleri hesaplanmıştır. Bu değerlendirme ölçütleri her benzerlik için ayrı ayrı tablolaştırılmış ve grafikler çizilmiştir.

Kuralların çalıştırılması sonucu yakalanan kelime çiftleri aşağıdaki kriterler çerçevesinde değerlendirilmiştir.

- ✓ *Yanlış Olup Doğru Bulunan*; kuralların çalışması sonucunda tespit edilen ama yanlış olan kelime çiftlerini ifade etmektedir.
- ✓ *Doğru Olup Yanlış Bulunan*; kuralların çalıştırılması sonucunda bulunamayıp dolayısıyla yanlış olarak değerlendirilen kelime çiftlerini ifade etmektedir.
- ✓ *Doğru Olup Doğru Bulunan*; kuralların çalıştırılması sonucunda ve doğru tespit edilen kelime çiftlerini ifade etmektedir.

6.1 Yapı Benzerliği Kurallarının Çalıştırılması Sonucu Elde Edilen Değerler

Yapı benzerlikleri dikkate alınarak yazılan kurallar kelime uzunluğu, sözcük türü ve benzer harf sayısına bakmaktadır.

Kuralların çalışması sonucu bulunan ve bulunamayan kelime çiftleri Çizelge 6.1’de verilmiştir. Tekrar eden ve özellikleri bakımından benzerlik gösteren kelime çiftleri çizelgeye konulmamıştır.

Çizelge 6.1 Yapı Benzerliği Kuralları Kullanılarak Bulunan İkilemeler

<i>Yanlış Olup Doğru Veren</i>	<i>Doğru Olup Yanlış Veren</i>	<i>Doğru Olup Doğru Veren</i>
Acı yanı	açık seçik	Adım adım
Bitmeyen bitmeyen	Az buz	Apar topar
Bodrum Bodrum	belli belirsiz	Arada sırada
Gulo gulo	bir iki	Ara sıra
Isı akısı	bitirir bitirmez	Bağıra çağıra
Linn Linn	bölük pörçük	Beyin meyin
Otuz dokuz	dayalı döşeli	Coşku moşku
Sise isine	dökülüp saçılan	Çalı malı
	Haşır neşir	Çıtı pıtı
	Sabah akşam	Çoluk çocuk
	saçım başım	Kanlı canlı
	Ufak tefek	Orasına burasına
	Yaka paça	Özene bezene
<i>Yukarıdaki Kelime Birlikteliklerin Toplam Sayısı</i>		
71	49	238

Çizelge 6.1’deki yanlış olup doğru veren kriteri incelendiğinde, kuralların çalışması sonucu “bitmeyen bitmeyen”, “Bodrum Bodrum” ve “Linn Linn” gibi birbirini tekrar eden kelime çiftleri yakalanmıştır ve bunlar ikileme değildir. “otuz dokuz” gibi yakalanan kelime çiftlerine bakıldığında ilk kelimeyle ikinci kelime arasındaki benzer harf sayısı ile ilk kelimenin uzunluğu arasındaki fark 1 (bir) olduğu için kuralların çalışması sonucu yakalanmıştır ama bu kelime çiftleri de ikileme değildir.

Çizelge 6.1’deki doğru olup yanlış veren kriteri ise derlemde işaretlenmiş olan ikilemelerde kuralların çalışması sonucu yakalanamaması durumudur. “Açık seçik”, “haşır neşir”, “yaka paça” vb. ikilemelere bakıldığında bu kelimelerin benzer harfleri olmasına rağmen ilk kelimenin uzunluğu ile benzer harf sayısı arasındaki fark 1(bir)’den büyüktür. Böylece buna benzer ikilemeler yakalanamamaktadır. “Bir iki”, “sabah akşam”, “dökülüp saçılan” vb. ikilemelerde ise benzer harfler bulunmadığı için yakalanmamaktadır.

Çizelge 6.1’deki doğru olup doğru veren kriteri ise derlemde ikileme olarak işaretlenmiş ve kuralların çalışması sonucunda yakalanan kelime çiftlerini ifade etmektedir. “apar topar”, “arada sırada”, “çıtı pıtı”, “özene bezene” vb. ikilemeler benzer harf sayısı ile ilk kelimenin uzunluğu arasındaki fark 1(bir), kelime türleri

aynı, kelime uzunlukları 2(iki)'den büyük ve farkı 2(iki)'den küçük olduğundan yakalanmaktadır.

Tez kapsamında gövdelenmiş OSTAD derleminde kelime benzerlikleri dikkate alınan kurallar süzgeç ve bariyer yapısı uygulanarak çalıştırılmış, yakalanan kelime çiftleri sayısı, yakalanan ikileme sayıları, duyarlılık ve anma değerleri Çizelge 6.2 ve Çizelge 6.3'de verilmiştir.

Çizelge 6.2 Yapı Benzerliği Kurallarının Gövdelenmiş OSTAD Derleminde Süzgeç Yöntemi Uygulanması Sonucu Bulunan Değerler ve İstatistikler (Derlemde bulunan ikileme sayısı 287'dir.)

Kural No	Kurallar	Yakalanan İkilemeler	Yakalanan Kelime Çiftleri	Duyarlılık	Anma
1	Kelime uzunlukları ≤ 1	285	40130	%0,71	%99,3
2	Sözcük türleri aynı	285	9894	%2,88	%99,3
3	Kelime uzunluğu arasındaki fark ≤ 2	282	7548	%3,74	%98,26
4	Kelimeler arasındaki benzer harf sayısı ≤ 0	273	2086	%13,09	%95,12
5	Kelime uzunluğu > 2	258	1857	%13,89	%89,9
6	Kelime uzunluğu-benzer harf sayısı=1 veya Kelime uzunluğu=Benzer harf sayısı	238	309	%77,02	%82,93

Çizelge 6.2'de kurallar sırasıyla çalıştırılmıştır. İlk kuralın çalışması sonucu yakalanan kelime çiftleri kullanılarak ikinci kural çalıştırılmış ve ikinci kuralda yakalanan kelime çiftleri kullanılarak bir sonraki kural çalıştırılmış ve diğer kurallar da bu şekilde çalıştırılarak, elde edilen sonuçlar Çizelge 6.2'de verilmiştir. Kısacası her kuralın çalışması sonucunda yakalanan kelime çiftlerinin sayısı azalmakta ve kurallar süzgeç olarak görev göstermektedir.

Çizelge 6.3 Yapı Benzerliği Kurallarının Gövdelenmiş OSTAD Derleminde Bariyer Yöntemi Uygulanması Sonucu Bulunan Değerler ve İstatistikler (Derlemde bulunan ikileme sayısı 287'dir.)

Kural No	Kurallar	Yakalanan İkilemeler	Yakalanan Kelime Çiftleri	Duyarlılık	Anma
1	Kelime uzunlukları ≤ 1	285	40130	%0,71	%99,3
2	Sözcük türleri aynı	285	9894	%2,88	%99,3
3	Kelime uzunluğu >2	269	8414	%3,2	%93,73
4	Kelime uzunluğu arasındaki fark ≤ 2	282	7548	%3,74	%98,26
5	Kelimeler arasındaki benzer harf sayısı ≤ 0	273	2574	%10,61	%95,12
6	Kelime uzunluğu-benzer harf sayısı=1 veya Kelime uzunluğu=Benzer harf sayısı	253	547	%46,25	%88,15

Çizelge 6.3'de gri renkle işaretlenmiş ilk iki kuralın arka arkaya çalıştırılması sonucu yakalanan kelime çiftleri diğer kurallar için derlem görevi görmektedir. Bariyer olarak ilk iki kuralın seçilmesinin nedeni bu kurallar çalıştırıldığında derlemdeki tüm ikilemeleri yakalaması ve derlemdeki yakalanan kelime çiftleri sayısının azalmasındandır. Böylece diğer kurallar daha az kelime çiftleri üzerinde çalıştırılmıştır. Gri renkte işaretlenmemiş diğer dört kural işaretlenmiş kuralların yakaladığı kelime çiftlerinde ayrı ayrı çalıştırılmış ve yakalanan kelime çiftleri, ikilemeler, duyarlılık, anma değerleri Çizelge 6.3'e yazılmıştır.

Gövdelenmiş OSTAD derleminde Çizelge 6.2 ve Çizelge 6.3'teki kuralların geçerliliği istatistiksel olarak gösterilmiştir. Kurallara bariyer yöntemi uygulandığında daha fazla ikileme yakalanarak duyarlılık değeri azalmaktadır. Kurallara süzgeç yöntemi uygulandığında ise yakalanan ikileme sayısı bariyer yöntemine göre düşük olmasına rağmen yakalanan kelime çiftleri sayısı daha az olduğu için süzgeç yönteminin duyarlılık değeri bariyer yöntemine göre daha yüksektir.

Tez de yüzeysel OSTAD derleminde kurallara süzgeç ve bariyer yöntemi uygulanmış, yakalanan kelime çiftleri sayısı, yakalanan ikileme sayıları, duyarlılık ve anma değerleri Çizelge 6.4 ve Çizelge 6.5'de verilmiştir.

Çizelge 6.4 Yapı Benzerliği Kurallarının Gövdelenmemiş OSTAD Derleminde Süzgeç Yöntemi Uygulanması Sonucu Bulunan Değerler ve İstatistikler (Derlemde bulunan ikileme sayısı 287'dir.)

Kural No	Kurallar	Yakalanan İkilemeler	Yakalanan Kelime Çiftleri	Duyarlılık	Anma
1	Kelime uzunlukları $\leq$ 1	287	40215	%0,71	%100
2	Sözcük türleri aynı	287	9928	%2,89	%100
3	Kelime uzunluğu $\geq$ 2	280	9607	%2,91	%97,56
4	Kelime uzunluğu arasındaki fark $\leq$ 2	241	6160	%3,91	%83,97
5	Kelimeler arasındaki benzer harf sayısı $\leq$ 0	190	1972	%9,63	%66,2
6	Kelime uzunluğu-benzer harf sayısı=1 veya Kelime uzunluğu=Benzer harf sayısı	161	171	%94,15	%56,1

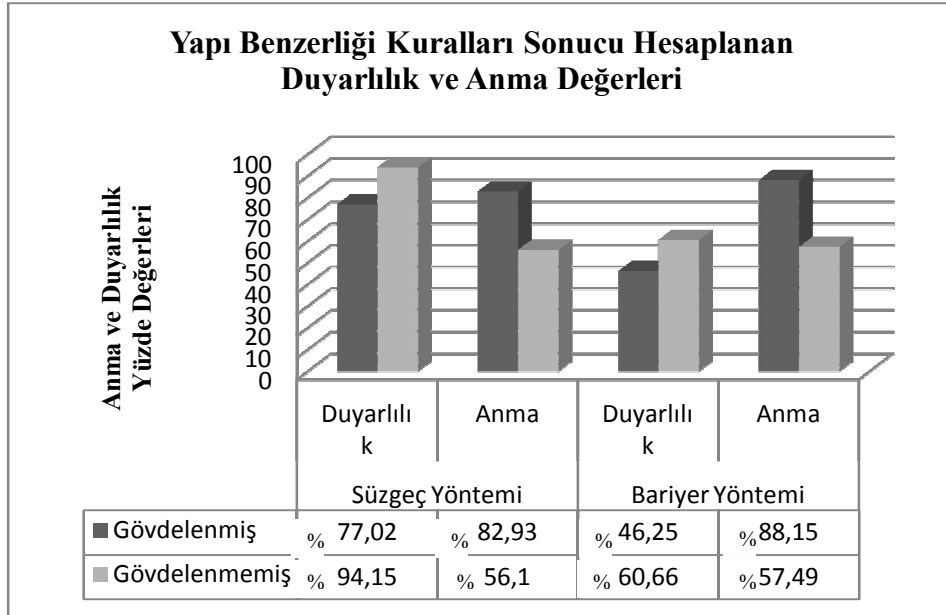
Çizelge 6.4'deki ilk kural gövdelenmemiş OSTAD derleminde çalıştırılmış ve yakalanan kelime çiftleri ikinci kural için kullanılmıştır. Bu şekilde sırayla kurallar çalıştırılmış ve kuralların çalışması sonucu yakalanan kelime çiftleri bir sonraki kural için derlem görevi görmüştür. Çizelge 6.4'de kurallar kelime çiftlerini süzme görevi göstermektedir.

Çizelge 6.5'de gri renkle işaretlenmiş ilk iki kural bariyer görevi göstermektedir. Böylece ilk iki kuralın çalışması sonucu yakalanan kelime çiftleri diğer kurallar için derlem görevi görmektedir. İşaretlenmemiş diğer dört kural işaretlenmiş kuralların yakaladığı kelime çiftlerinde ayrı ayrı çalıştırılmış ve yakalanan kelime çiftleri, ikilemeler, duyarlılık, anma değerleri Çizelge 6.5'e yazılmıştır.

Çizelge 6.5 Yapı Benzerliği Kurallarının Gövdelenmemiş OSTAD Derleminde Bariyer Yöntemi Uygulanması Sonucu Bulunan Değerler ve İstatistikler (Derlemde bulunan ikileme sayısı 287'dir.)

Kural No	Kurallar	Yakalanan İkilemeler	Yakalanan Kelime Birliktelikleri	Duyarlılık	Anma
1	Kelime uzunlukları ≤ 1	287	40215	%0,71	%100
2	Sözcük türleri aynı	287	9928	%2,89	%100
3	Kelime uzunluğu >2	280	9607	%2,91	%97,56
4	Kelime uzunluğu arasındaki fark <2	246	6363	%3,87	%85,71
5	Kelimeler arasındaki benzer harf sayısı ≤ 0	209	2992	%6,99	%72,82
6	Kelime uzunluğu-benzer harf sayısı=1 veya Kelime uzunluğu=Benzer harf sayısı	165	272	%60,66	%57,49

Gövdelenmemiş OSTAD derleminde Çizelge 6.4 ve Çizelge 6.5'teki kuralların geçerliliği istatistiksel olarak gösterilmiştir.

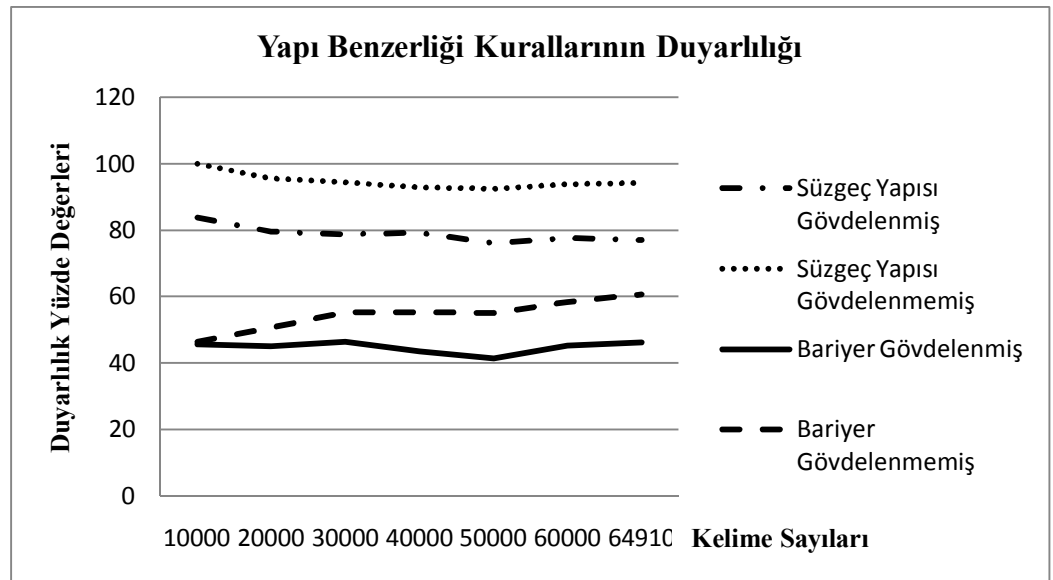


Şekil 6.1 Yapı Benzerliği Kurallarının Çalışması Sonucu Bulunan İkilemelerin Yüzde Olarak Görünümü

Tez kapsamında gövdelenmiş ve gövdelenmemiş derlemde süzgeç ve bariyer yöntemi uygulanarak tüm kuralların çalışması sonucu yakalanan ikilemeler

ve kelime çiftleri için değerlendirme ölçütleri olan duyarlılık ve anma değerleri Şekil 6.1’de gösterilmiştir. Gövdelenmiş derlemde duyarlılık ve anma değerlerinin gövdelenmemiş derleme göre daha iyi sonuç vermesi derlemdeki kelimelerin eklerden arınmış halde bulunması ile ilgilidir. Böylece gövdelenmiş ve gövdelenmemiş derlem arasındaki fark ortaya çıkmaktadır.

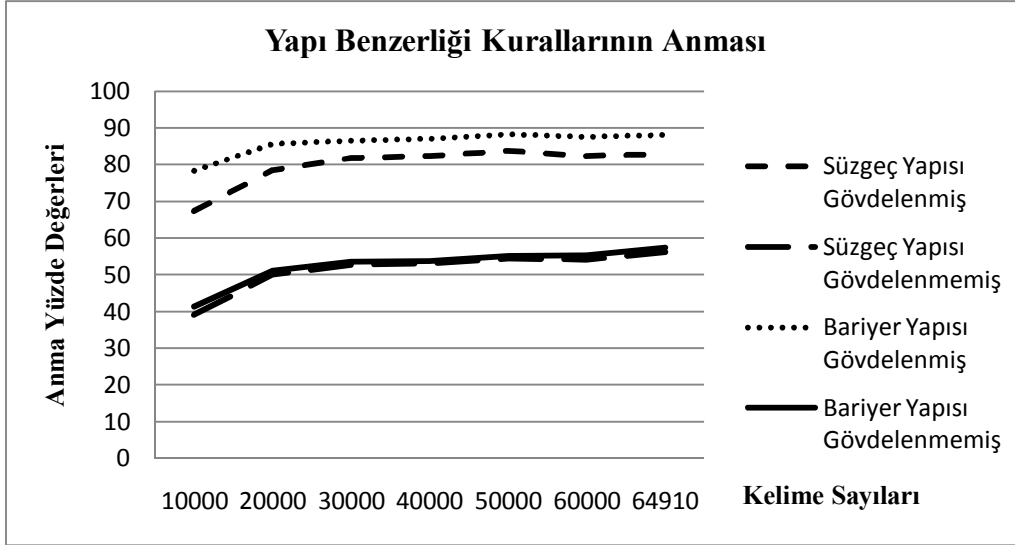
Kuralların süzgeç ve bariyer yapısında çalışması sonucunda derlemdeki ilk 10000 kelime içinden yakalanan ikilemelerin duyarlılık değerini gövdelenmiş ve gövdelenmemiş derlem için vermektedir. Derlemdeki baştan itibaren 20000, 30000, 40000, 50000 ve 60000 kelime içinde yakalanan ikilemelerin duyarlılık değeri Şekil 6.2’de ve anma değeri Şekil 6.3’de verilmektedir. Derlemdeki kelime basamakları için tüm değerler tek tek hesaplanmıştır.



Şekil 6.2 Gövdelenmiş ve Gövdelenmemiş Yapı Benzerliği Kurallarının Süzgeç ve Bariyer Yöntemleriyle Çalışması Sonucu Hesaplanan Duyarlılık Değerleri

Şekil 6.2’deki duyarlılık değerlerine bakıldığında kuralların süzgeç yönteminde çalışması sonucu gövdelenmiş ve gövdelenmemiş derlemdeki değerler bariyer yöntemine göre daha yüksektir. Duyarlılık değerlerinin gövdelenmemiş derlemde daha yüksek çıkmasının nedeni ek almış kelimelerin daha az benzerlik göstermesi ile ilgilidir. Böylece gövdelenmemiş derlemde kuralların çalışması sonucu yakalanan kelimelerin ve ikilemelerin sayısının gövdelenmiş derlemde

yakalananlara göre daha az olması gövdelenmemiş derlemde duyarlılık değerinin yüksek olmasına neden olur.



Şekil 6.3 Gövdelenmiş ve Gövdelenmemiş Derlemde Yapı Benzerliği Kurallarının Süzgeç ve Bariyer Yöntemleriyle Çalışması Sonucu Hesaplanan Anma Değerleri

Şekil 6.3'de görüldüğü gibi anma değerine bakıldığında kuralların her iki yapıda çalışması sonucu bulunan ikilemeler ile derlemde geçen ikilemelerin oranının gövdelenmemiş derlemde az olmasının nedeni yakalanan ikilemelerin sayısının az olması yani ek almış ikilemelerin yakalanamamasından kaynaklandığını göstermektedir. Gövdelenmiş derlemde ise kelimelerin gövde durumunda olmasından dolayı anma değeri daha yüksek çıkmaktadır.

Şekil 6.2'deki duyarlılık değerlerinde en iyi sonucu gövdelenmemiş derleme süzgeç yöntemi uygulandığında alınmaktadır. Şekil 6.3'teki anma değerlerindeki en iyi sonucu ise gövdelenmiş derlemde bariyer yöntemi uygulandığında alınmaktadır.

6.2 Sessiz Harf Benzerliği Kurallarının Çalıştırılması Sonucu Elde Edilen Değerler

Ses benzerliğinden sessiz harf benzerliği dikkate alınarak yazılan kurallarda; sessiz ve benzer sessiz harf sayılarına bakılmıştır. Kuralların çalışması sonucu yakalanan ikilemeler ve kelime çiftleri doğruluk kriterlerine göre üç başlık altında toplanmıştır.

Çizelge 6.6 Sessiz Harf Benzerliği Kuralları Kullanılarak Bulunan İkilemeler

<i>Yanlış Olup Doğru Veren</i>	<i>Doğru Olup Yanlış Veren</i>	<i>Doğru Olup Doğru Veren</i>
Adam sandım	bağıra çağıra	Abuk sabuk
Alanı haline	Bayat mayat	Açık seçik
Bodrum Bodrum	belli belirsiz	Apar topar
Gulo gulo	bir iki	Arada sırada
Linn Linn	bitirir bitirmez	Eğrisi Doğrusu
Olan yalan	Boyu boyuma	Ekonomi mekonomi
Oldu geldi	bölük pörçük	Elektronik melektronik
Süre sıra	Çalı malı	Falan filan
	Çoluk çocuk	İçki miçki
	Çıtı pıtı	İşletme mişletme
	dayalı döşeli	Orasına burasına
	devşir devşirme	Özene bezene
	dökülüp saçılan	Tek tük
	Sabah akşam	Ufak tefek
<i>Yukarıdaki Kelime Birlikteliklerin Toplam Sayısı</i>		
126	51	227

Çizelge 6.6’ da yanlış olup doğru veren kriterinde “Gulo Gulo ”, “Bodrum Bodrum” ve “Linn Linn” gibi kelime çiftleri birbirini tekrar eden kelimeler olduğundan yakalanmaktadır ve ikileme değildir. ”oldu geldi” gibi kelime çiftlerine bakıldığında ilk kelimedeki toplam sessiz harf sayısı ile arka arkaya gelen kelimelerdeki benzer sessiz harf sayısı birbirine eşit olduğu için kurallar yakalanmıştır.

Çizelge 6.6’da doğru olup yanlış veren dediğimizde ise derlemde işaretlenmiş olan ikilemelerin yakalanmadığını ifade etmektedir. “Bağıra çağıra”, “beyin meyin”, “çıtı pıtı”, “dayalı döşeli” vb. bakıldığında bu kelimelerin benzer sessiz harfleri olmasına rağmen ilk kelimedeki sessiz harf sayısı ile benzer sessiz harf sayısı birbirine eşit olmadığından yani sessiz harflerden birisi farklılık gösterdiğinden yakalanamamaktadır. “Bir iki”, “sabah akşam”, “dökülüp saçılan” vb. ikilemelerde ise benzer sessiz harfler bulunmadığı için yakalanmamaktadır. “bitirir bitirmez”, “görür görmez” vb. ikilemeler ise olumsuzluk eki aldığından benzer sessiz harf sayısı eşit değildir ve yakalanamamaktadır.

Çizelge 6.6’da doğru olup doğru veren kriterinde ikileme olarak yakalanan kelime çiftleri ise “abuk sabuk”, “açık seçik”, “elektronik melektronik”, “ufak

tefek” vb. ikilemelerdir. Bu ikilemeler ilk kelimedeki toplam sessiz harf sayısı ile arka arkaya gelen kelimelerdeki benzer sessiz harf sayısı birbirine eşit olduğu için kuralların çalışması sonucu yakalanmıştır.

Tez çalışmasında gövdelenmiş derlemde sessiz harf benzerliği kuralları süzgeç ve bariyer yöntemleri ile çalıştırılmıştır. Çalıştırma sonucunda her iki yöntemde de yakalanan kelime çiftleri ve ikileme sayısı aynıdır. Bunun nedeni kuralların birbirini tamamlamasıdır. Bu yüzden her iki yöntem için yakalanan kelime çiftleri sayısı, yakalanan ikileme sayıları, duyarlılık ve anma değerleri Çizelge 6.7’de verilmiştir.

Çizelge 6.7 Sessiz Harf Benzerliği Kurallarının Gövdelenmiş OSTAD Derleminde Süzgeç ve Bariyer Yöntemlerinin Uygulanması Sonucu Bulunan Değerler ve İstatistikler (Derlemde bulunan ikileme sayısı 287’dir.)

Kural No	Kurallar	Yakalanan İkilemeler	Yakalanan Kelime Çiftleri	Duyarlılık	Anma
1	Kelimelerin toplam sessiz harf sayısı ≥ 0	285	52011	%0,55	%99,3
2	Sözcük türleri aynı	285	9995	%2,85	%99,3
3	Arka arkaya gelen kelimelerin benzer sessiz harf sayısı ≥ 0	265	1445	%18,34	%92,33
4	Birinci kelimedeki toplam sessiz harf sayısı = Arka arkaya gelen kelimelerin benzer sessiz harf sayısı	229	353	%64,87	%79,79

Sessiz harf benzerliğindeki kurallar birbirinin tamamlayıcısı olduğu için süzgeç ve bariyer yöntemleri uygulandığında aynı sonuçlar elde edilmiştir. Gövdelenmiş OSTAD derleminde kuralların geçerliliği istatistiksel olarak Çizelge 6.7’de gösterilmiştir.

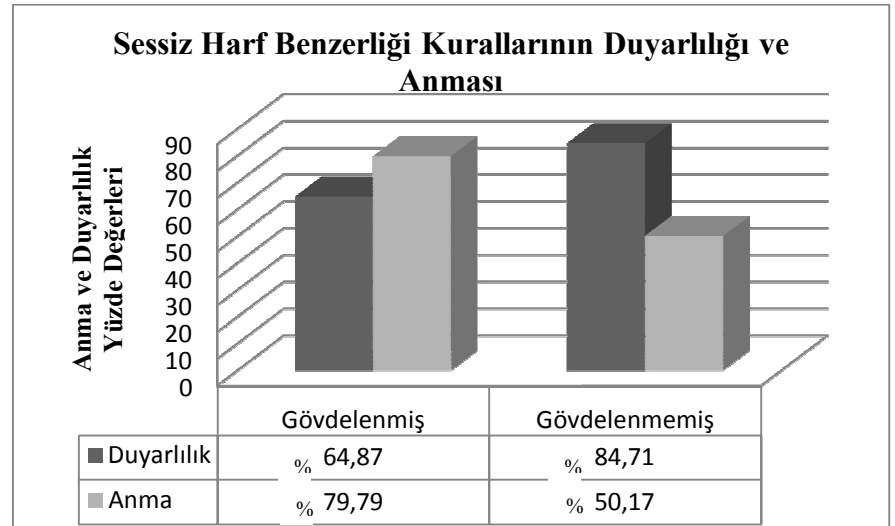
Tezde yüzeysel formda kuralların her iki yöntemde de çalıştırılması sonucu duyarlılık, anma, yakalanan kelime çiftleri ve ikileme sayıları değerleri Çizelge 6.8’de verilmiştir.

Çizelge 6.8 Sessiz Harf Benzerliği Kurallarının Gövdelenmemiş OSTAD Derleminde Süzgeç ve Bariyer Yöntemlerinin Uygulanması Sonucu Bulunan Değerler ve İstatistikler (Derlemde bulunan ikileme sayısı 287’dir.)

Kural No	Kurallar	Yakalanan İkilemeler	Yakalanan Kelime Çiftleri	Duyarlılık	Anma
1	Kelimelerin toplam sessiz harf sayısı ≤ 0	287	52727	%0,54	%100
2	Sözcük türleri aynı	287	10006	%2,87	%100
3	Arka arkaya gelen kelimelerin benzer sessiz harf sayısı ≤ 0	184	1639	%11,23	%64,11
4	Birinci kelimedeki toplam sessiz harf sayısı = Arka arkaya gelen kelimelerin benzer sessiz harf sayısı	144	170	%84,71	%50,17

Çizelge 6.7 ve Çizelge 6.8’deki üçüncü kural olan “benzer sessiz harf sayısı” kuralı, dördüncü kural olan “toplam sessiz harf sayısı eşittir benzer sessiz harf sayısı” kuralının tamamlayıcısıdır. Böylece süzgeç ve bariyer yöntemlerinin uygulandığında her iki yöntemin sonucu aynı çıkmaktadır.

Tez kapsamında belirlenmiş olan sessiz harf benzerliği kuralları yüzeysel ve gövdelenmiş formda uygulanması sonucunda Şekil 6.4’deki grafik elde edilmiştir.

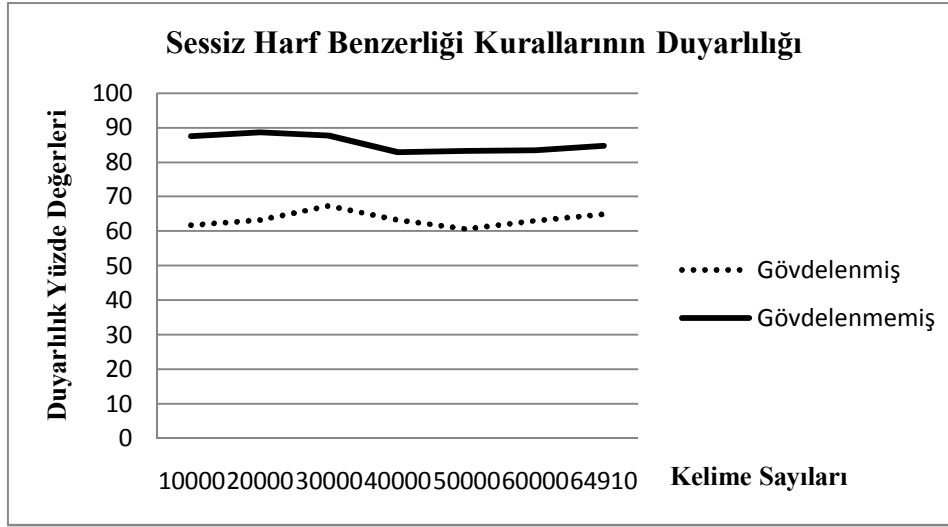


Şekil 6.4 Sessiz Harf Benzerliği Kurallarının Uygulanması Sonucu Bulunan İkilemelerin Yüzde Olarak Görünümü

Şekil 6.4’deki gövdelenmiş derlemde anma değerinin gövdelenmemiş derleme göre yüksek olması derlemdeki kelimelerin eklerden arınmış halde

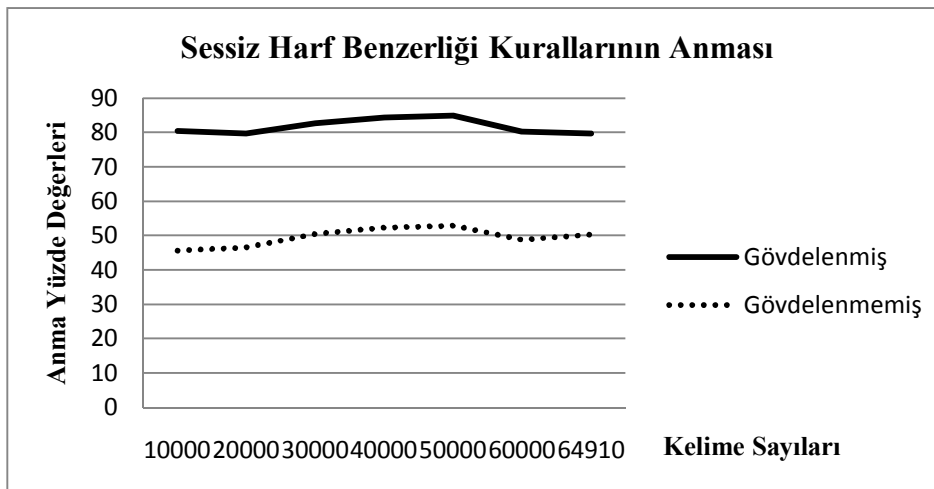
bulunması ile ilgilidir. Böylece gövdelenmiş ve gövdelenmemiş derlem arasındaki fark ortaya çıkmaktadır.

Derlemdeki baştan itibaren ilk 10000, 20000, 30000, 40000, 50000 ve 60000 kelime içinde yakalanan ikilemelerin duyarlılık değeri Şekil 6.5’de ve anma değeri Şekil 6.6’da verilmektedir. Derlemdeki kelime basamakları için tüm değerler tek tek hesaplanmıştır.



Şekil 6.5 Gövdelenmiş ve Gövdelenmemiş Derlemde Sessiz Harf Benzerliği Kurallarının Duyarlılık Değerleri

Şekil 6.5’e bakıldığında duyarlılık değerleri gövdelenmemiş derlemde yüksek çıkmasının nedeni daha az kelime çiftleri yakalanmasındandır.



Şekil 6.6 Gövdelenmiş ve Gövdelenmemiş Derlemde Sessiz Harf Benzerliği Kurallarının Anma Değerleri

Şekil 6.6'daki anma değerlerine bakıldığında kuralların çalışması sonucu bulunan ikilemeler ile derlemde geçen ikilemelerin oranının gövdelenmemiş derlemde gövdelenmiş derleme göre az olduğu görülmektedir. Gövdelenmemiş derlemde anma değerinin düşük olmasının nedeni yakalanan ikilemelerin sayısının az olması ve ek almış ikilemelerin yakalanamamasından kaynaklanmaktadır. Gövdelenmiş derlemde ise kelimelerin gövde durumunda olmasından dolayı anma değeri daha yüksek çıkmaktadır.

6.3 Sesli-Sessiz Harf Benzerliği Kurallarının Çalıştırılması Sonucu Elde Edilen Değerler

Ses benzerliğinden sesli-sessiz harf benzerliği dikkate alınarak belirlenen kurallar; sözcük türleri, benzer sesli-sessiz harf sayısı, kelime uzunluğu ve benzer sesli-sessiz harf sayılarının kelime uzunluğuna oranını göz önünde bulundurularak yazılmıştır. Sesli harf benzerliği kuralında a=e, ı=i, o=ö, u=ü olarak değerlendirilerek ikilemelerdeki sesli harf geçişlerini yakalamayı amaçlamaktadır.

Kuralların çalışması sonucu yakalanan kelime çiftleri ve ikilemeler doğruluk kriterleri göz önüne alınarak üç ana başlık altında toplanmış ve Çizelge 6.9 oluşturulmuştur. Tekrar eden ve benzerlik gösteren kelime çiftleri çizelgeye konulmamıştır.

Çizelge 6.9 Sesli-Sessiz Harf Benzerliği Kuralları Kullanılarak Bulunan İkilemeler

<i>Yanlış Olup Doğru Veren</i>	<i>Doğru Olup Yanlış Veren</i>	<i>Doğru Olup Doğru Veren</i>
anneme babama	sabah akşam	Abuk sabuk
defa daha	dök saç	açık seçik
deyip kesip	belli belirsiz	arada sırada
Linn Linn	Bir iki	bağıra çağıra
olan yalan		Çoluk çocuk
Onlar bunlar		dayalı döşeli
Bodrum Bodrum		ekonomi mekonomi
süre sıra		falan filan
		haşır neşir
		orasına burasına
		bölük pörçük
<i>Yukarıdaki Kelime Birlikteliklerin Toplam Sayısı</i>		
48	44	234

Çizelge 6.9’da yanlış olup doğru veren kriterine bakılacak olursa “anneme babama” gibi kelime çiftlerindeki harf benzerliği %50 olduğundan yakalamaktadır. Yine aynı şekilde “deyip kesip”, “olan yalan” gibi kelime çiftlerini ise %70 harf benzerliği ile yakalamaktadır. Yakalanan kelime çiftleri ikileme özelliği taşımamaktadır. “Linn Linn”, “Bodrum Bodrum” vb. kelime çiftleri ise birbirini tekrar eden kelimeler olduğundan yakalanmıştır.

Çizelge 6.9’da doğru olup yanlış veren kriterinde “sabah akşam”, “dök saç” vb. ikilemelere bakıldığında benzer harfleri olmaması nedeniyle kuralların çalışması sonucunda yakalanamamaktadır.

Çizelge 6.9’da doğru olup doğru veren kriterinde yakalanan ikilemeler ise “apar topar”, “abuk sabuk”, “dayalı döşeli”, “ipe sapa” vb. ikilemelerin benzer sesli ve sessiz harf sayısının toplamı %50’nin ve “bölük pörçük” ikilemesinde ise %33’ün üzerinde olması nedeniyle ikileme olarak kuralların çalışması sonucu yakalanmaktadır.

Tezde yüzeysel ve gövdelenmiş derlemde kurallar çalıştırılmış, harf benzerliği yüzde oranları tek tek denenmiş, oran değerleri hesaplanmış ve değerlendirme ölçütlerinde değişiklik gösteren sonuçlar dikkate alınmıştır. Çıkan sonuçlar ve hesaplanan değerler çizelgelerde gövdelenmiş ve gövdelenmemiş derlem olarak ayrı ayrı verilmiştir.

Gövdelenmiş derlemde sesli-sessiz harf benzerliği kuralları süzgeç ve bariyer yöntemi uygulanarak çalıştırılmış, yakalanan kelime çiftleri sayısı, yakalanan ikileme sayıları, duyarlılık ve anma değerleri Çizelge 6.10 ve Çizelge 6.11’de verilmiştir.

Çizelge 6.10 Sesli-Sessiz Harf Benzerliği Kurallarının Gövdelenmiş OSTAD Derleminde Süzgeç Yönteminin Uygulanması Sonucu Bulunan Değerler ve İstatistikler (Derlemde bulunan ikileme sayısı 287'dir.)

Kural No	Kurallar	Yakalanan İkilemeler	Yakalanan Kelime Çiftleri	Duyarlılık	Anma
1	Kelime uzunlukları ≤ 1	285	40130	%0,71	%99,3
2	Sözcük türleri aynı	285	9894	%2,88	%99,3
3	Birinci kelime uzunluğu = İkinci kelime uzunluğu veya İkinci kelime uzunluğu – Birinci kelime uzunluğu=1	280	3727	%7,51	%97,56
4	Arka arkaya gelen kelimelerin benzer sessiz harf sayısı ≤ 0	265	830	%31,93	%97,33
5	Arka arkaya gelen kelimelerin benzer sesli harf sayısı ≤ 0	251	434	%57,83	%90,29
6	Kelimelerdeki benzer sesli ve sessiz harf sayısı toplamı ile Birinci kelime uzunluğu oranı $\geq$ %70	234	282	%82,98	%84,17

Çizelge 6.11 Sesli-Sessiz Harf Benzerliği Kurallarının Gövdelenmiş OSTAD Derleminde Bariyer Yönteminin Uygulanması Sonucu Bulunan Değerler ve İstatistikler (Derlemde bulunan ikileme sayısı 287'dir.)

Kural No	Kurallar	Yakalanan İkilemeler	Yakalanan Kelime Çiftleri	Duyarlılık	Anma
1	Kelime uzunlukları ≤ 1	285	40130	%0,71	%99,3
2	Sözcük türleri aynı	285	9894	%2,81	%99,3
3	Birinci kelime uzunluğu = İkinci kelime uzunluğu veya İkinci kelime uzunluğu – Birinci kelime uzunluğu=1	280	3727	%7,51	%97,56
4	Arka arkaya gelen kelimelerin benzer sessiz harf sayısı ≤ 0	265	1443	%18,36	%92,33
5	Arka arkaya gelen kelimelerin benzer sesli harf sayısı ≤ 0	259	1704	%15,2	%90,24
6	Kelimelerdeki benzer sesli ve sessiz harf sayısı toplamı ile Birinci kelime uzunluğu oranı $\geq$ %70	234	302	%77,48	%84,17

Sesli-sessiz harf benzerliği kurallarına süzgeç ve bariyer yöntemi uygulanmış ve gövdelenmiş derlemde çalıştırılmıştır. Süzgeç yönteminin uygulanmasından elde edilen sonuçlar Çizelge 6.10’da verilmektedir. Süzgeç yönteminde bir kuralda yakalanan kelime çiftleri bir sonraki kuralın çalışmasında derlem görevi görmektedir. Bu şekilde tüm kurallar sırasıyla çalıştırılmış ve her kural için elde edilen sonuçlar çizelgelere yazılmıştır. Bariyer yöntemi ise ilk iki kuralın çalıştırılması sonucu yakalanan kelime çiftleri diğer kuralların çalışması için derlem görevi görmektedir. Bariyer yönteminin uygulanmasından elde edilen sonuçlar Çizelge 6.11’de verilmiştir. Altıncı kural olan “harf benzerliği yüzde oranı” kuralının yüzde oranları denenerek bulunmuş ve bu oranlar doğrultusunda hesaplanan duyarlılık ve anma değerleri süzgeç yöntemi için Çizelge 6.12’de ve bariyer yöntemi için Çizelge 6.13’te verilmiştir.

Çizelge 6.12 İkilemelerin Tespiti İçin Oluşturulan Harf Benzerlik Kuralının Süzgeç Yöntemi Uygulanarak Gövdelenmiş OSTAD Derlemindeki İstatistikleri (Derlemde bulunan ikileme sayısı 287’dir.)

Harf Benzerlik Oranı	Yakalanan İkilemeler	Yakalanan Kelime Çiftleri	Duyarlılık	Anma
%33	251	430	%58,37	%90,29
%50	248	366	%67,76	%89,21
%70	234	282	%82,98	%84,17
%76	227	269	%84,39	%81,65
%81	207	249	%83,13	%74,46

Çizelge 6.13 İkilemelerin Tespiti İçin Oluşturulan Harf Benzerlik Kuralının Bariyer Yöntemi Uygulanarak Gövdelenmiş OSTAD Derlemindeki İstatistikleri (Derlemde bulunan ikileme sayısı 287’dir.)

Harf Benzerlik Oranı	Yakalanan İkilemeler	Yakalanan Kelime Çiftleri	Duyarlılık	Anma
%33	269	1092	%24,63	%93,73
%50	258	637	%40,5	%89,9
%70	234	302	%77,48	%84,17
%76	227	288	%78,82	%81,65
%81	207	261	%79,31	%74,46

Çizelge 6.12 ve Çizelge 6.13'e bakıldığında harf benzerlik oranı %33 olduğunda çoğu ikilemeyi yakaladığı görülmektedir. Harf benzerlik oranı azaldıkça yakalanan kelime çiftleri ve ikileme sayısı artacağından duyarlılık değeri azalarak ve anma değeri artmaktadır. Harf benzerlik oranı arttıkça duyarlılık değeri artmaktadır. Kısacası harf benzerlik oranı duyarlılık değeri ile doğru orantılı ve anma değeri ile de ters orantılıdır. Çizelge 6.12 ve Çizelge 6.13'de görüldüğü üzere en iyi sonuç harf benzerlik oranı %70 olduğunda alınmaktadır. Çizelge 6.13'e bakıldığında bariyer yapısı uygulanınca duyarlılık değeri düşmektedir. Bunun en önemli nedeni bariyer yapısında yakalanan kelime çiftleri sayısının daha fazla olmasındandır.

Gövdelenmemiş derlemde sesli-sessiz harf benzerliği kuralları süzgeç ve bariyer yöntemi uygulayarak çalıştırılmış, duyarlılık, anma, yakalanan kelime çiftleri ve ikileme sayıları, değerleri Çizelge 6.14 ve Çizelge 6.15'de verilmiştir.

Çizelge 6.14 Sesli-Sessiz Harf Benzerliği Kurallarının Gövdelenmemiş OSTAD Derleminde Süzgeç Yönteminin Uygulanması Sonucu Bulunan Değerler ve İstatistikler (Derlemde bulunan ikileme sayısı 287'dir.)

Kural No	Kurallar	Yakalanan İkilemeler	Yakalanan Kelime Çiftleri	Duyarlılık	Anma
1	Kelime uzunlukları $\diamond 1$	287	40215	%0,71	%100
2	Sözcük türleri aynı	287	9928	%2,89	%100
3	Birinci kelime uzunluğu = İkinci kelime uzunluğu veya İkinci kelime uzunluğu – Birinci kelime uzunluğu=1	218	2486	%8,77	%75,96
4	Arka arkaya gelen kelimelerin benzer sessiz harf sayısı $\diamond 0$	180	593	%30,35	%62,72
5	Arka arkaya gelen kelimelerin benzer sesli harf sayısı $\diamond 0$	166	292	%56,85	%57,84
6	Kelimelerdeki benzer sesli ve sessiz harf sayısı toplamı ile Birinci kelime uzunluğu oranı $\geq$ %70	153	162	%94,44	%53,31

Çizelge 6.15 Sesli-Sessiz Harf Benzerliği Kurallarının Gövdelenmemiş OSTAD Derleminde Bariyer Yönteminin Uygulanması Sonucu Bulunan Değerler ve İstatistikler (Derlemde bulunan ikileme sayısı 287'dir.)

Kural No	Kurallar	Yakalanan İkilemeler	Yakalanan Kelime Çiftleri	Duyarlılık	Anma
1	Kelime uzunlukları ≤ 1	287	40215	%0,71	%100
2	Sözcük türleri aynı	287	9928	%2,89	%100
3	Birinci kelime uzunluğu = İkinci kelime uzunluğu veya İkinci kelime uzunluğu – Birinci kelime uzunluğu=1	218	2486	%8,77	%75,54
4	Arka arkaya gelen kelimelerin benzer sessiz harf sayısı ≤ 0	184	1637	%11,24	%64,11
5	Arka arkaya gelen kelimelerin benzer sesli harf sayısı ≤ 0	194	1914	%10,14	%67,6
6	Kelimelerdeki benzer sesli ve sessiz harf sayısı toplamı ile Birinci kelime uzunluğu oranı $\geq$ %70	153	172	%88,95	%53,31

Gövdelenmemiş OSTAD derleminde sesli-sessiz harf benzerliği kuralları süzgeç ve bariyer yöntemleri uygulanarak çalıştırılmıştır. Süzgeç yönteminin uygulamasında ilk kural çalıştırılır ve bu ilk kuraldan yakalanan kelime çiftleri ikinci kuralın çalışması için derlemi oluşturmaktadır. Tüm kurallar bu şekilde çalıştırılmış ve elde edilen değerler hesaplanıp Çizelge 6.14'e yazılmıştır. Bariyer yönteminin uygulanmasında ise ilk iki kuralın çalıştırılması sonucu yakalanan kelime çiftleri diğer kuralların çalışması için derlem görevi görmektedir. Bariyer yönteminin uygulanmasından elde edilen sonuçlar Çizelge 6.15'de verilmiştir. Altıncı kural olan "harf benzerliği yüzde oranı" kuralının yüzde oranları denenerak bulunmuş ve bu oranlar doğrultusunda hesaplanan duyarlılık ve anma değerleri süzgeç yöntemi Çizelge 6.16'da ve bariyer yöntemi için Çizelge 6.17'te verilmiştir.

Çizelge 6.16 İkilemelerin Tespiti İçin Oluşturulan Harf Benzerlik Kuralının Süzgeç Yöntemi Uygulanarak Gövdelenmemiş OSTAD Derlemindeki İstatistikleri (Derlemde bulunan ikileme sayısı 287'dir.)

Harf Benzerlik Oranı	Yakalanan İkilemeler	Yakalanan Kelime Çiftleri	Duyarlılık	Anma
%33	166	250	%66,4	%57,84
%50	163	186	%87,63	%56,79
%70	153	162	%94,44	%53,31
%76	145	151	%96,03	%50,52
%81	123	129	%95,35	%44,24

Çizelge 6.17 İkilemelerin Tespiti İçin Oluşturulan Harf Benzerlik Kuralının Bariyer Yöntemi Uygulanarak Gövdelenmemiş OSTAD Derlemindeki İstatistikleri (Derlemde bulunan ikileme sayısı 287'dir.)

Harf Benzerlik Oranı	Yakalanan İkilemeler	Yakalanan Kelime Çiftleri	Duyarlılık	Anma
%33	190	641	%29,64	%66,2
%50	178	341	%52,2	%62,02
%70	153	172	%88,95	%53,31
%76	145	156	%92,95	%50,52
%81	123	133	%92,48	%44,24

Çizelge 6.16 ve Çizelge 6.17'ye bakıldığında harf benzerlik oranı arttıkça yakalanan kelime çiftleri ve ikileme sayısı azalacağından duyarlılık değeri artmakta ve anma değeri azalmaktadır. Harf benzerlik oranı duyarlılık değeri ile doğru orantılı ve anma değeri ile de ters orantılıdır. Çizelge 6.16 ve Çizelge 6.17'de görüldüğü üzere en iyi sonuç harf benzerlik oranı %70 olduğunda alınmaktadır. Çizelge 6.16'ya bakıldığında süzgeç yöntemi uygulanınca duyarlılık değeri artmaktadır. Bunun en önemli nedeni süzgeç yönteminde yakalanan kelime çiftleri sayısının daha az olmasındandır.

Çizelge 6.18 Sesli-Sessiz Harf Benzerliği Kurallarının Süzgeç Yöntemi Uygulandığındaki Sonuçları

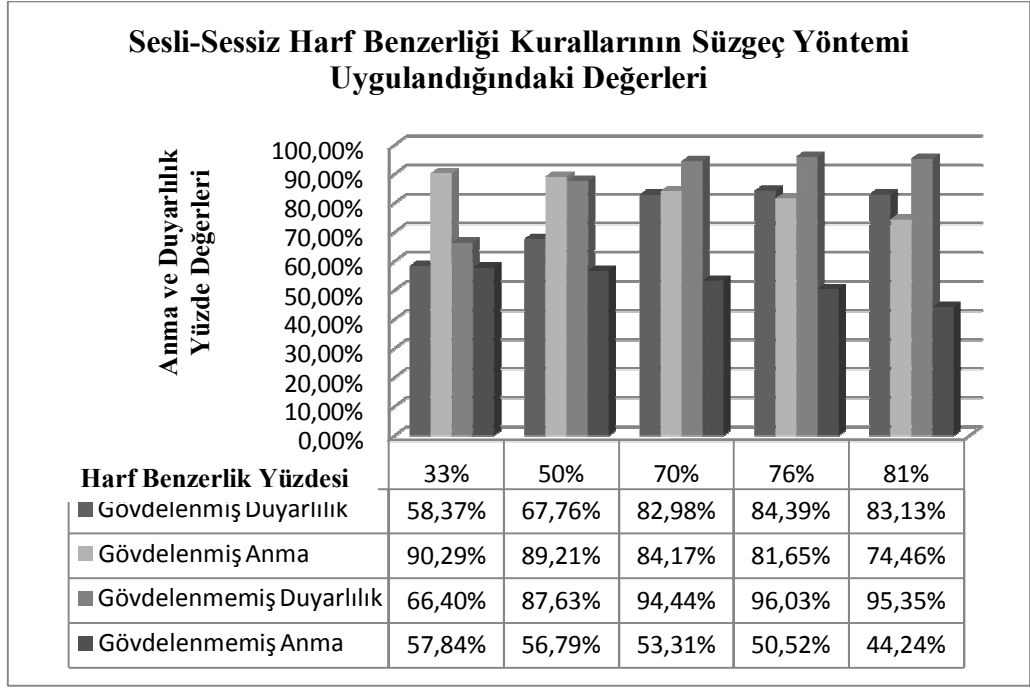
<i>Harf Benzerlik Yüzdeleri</i>	<i>Gövdelenmiş</i>		<i>Gövdelenmemiş</i>	
	<i>Duyarlılık</i>	<i>Anma</i>	<i>Duyarlılık</i>	<i>Anma</i>
%33	%58,37	%90,29	%66,4	%57,84
%50	%67,76	%89,21	%87,63	%56,79
%70	%82,98	%84,17	%94,44	%53,31
%76	%84,39	%81,65	%96,03	%50,52
%81	%83,13	%74,46	%95,35	%44,24

Çizelge 6.19 Sesli-Sessiz Harf Benzerliği Kurallarının Bariyer Yöntemi Uygulandığındaki Sonuçları

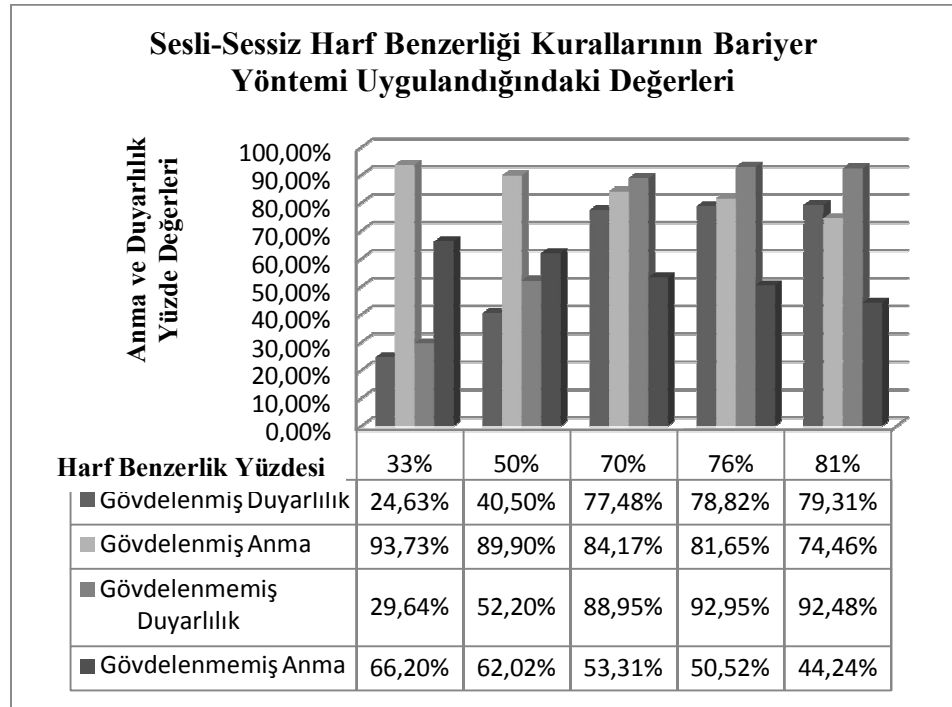
<i>Harf Benzerlik Yüzdeleri</i>	<i>Gövdelenmiş</i>		<i>Gövdelenmemiş</i>	
	<i>Duyarlılık</i>	<i>Anma</i>	<i>Duyarlılık</i>	<i>Anma</i>
%33	%24,63	%93,73	%29,64	%66,2
%50	%40,5	%89,9	%52,2	%62,02
%70	%77,48	%84,17	%88,95	%53,31
%76	%78,82	%81,65	%92,95	%50,52
%81	%79,31	%74,46	%92,48	%44,24

Çizelge 6.18’de gövdelenmiş ve gövdelenmemiş derlemlerde anmanın en yüksek olduğu benzerlik oranı %33 iken anmanın en düşük olduğu benzerlik oranı ise %81’dir. Benzerlik oranı arttıkça yakalanan ikilemeler azalmakta böylece anma değeri de düşmektedir. Anma ile benzerlik oranı arasında ters orantı vardır.

Gövdelenmemiş derlemde duyarlılığın yüksek olmasının nedeni kelimelerin ekleri çıkarılmamış halde ele alınmasından kaynaklanmaktadır. Gövdelenmemiş derlemde yakalanan kelime çiftleri az olacağından dolayı duyarlılık değeri yüksek çıkmaktadır. Anma, yakalanan ikilemeler ile derlemde geçen ikilemelerin oranını verdiği için gövdeli kelime çiftlerinde yakalanan ikilemelerin sayıca fazla olması gövdelenmiş derlemde anmanın yüksek olmasına yol açmaktadır.

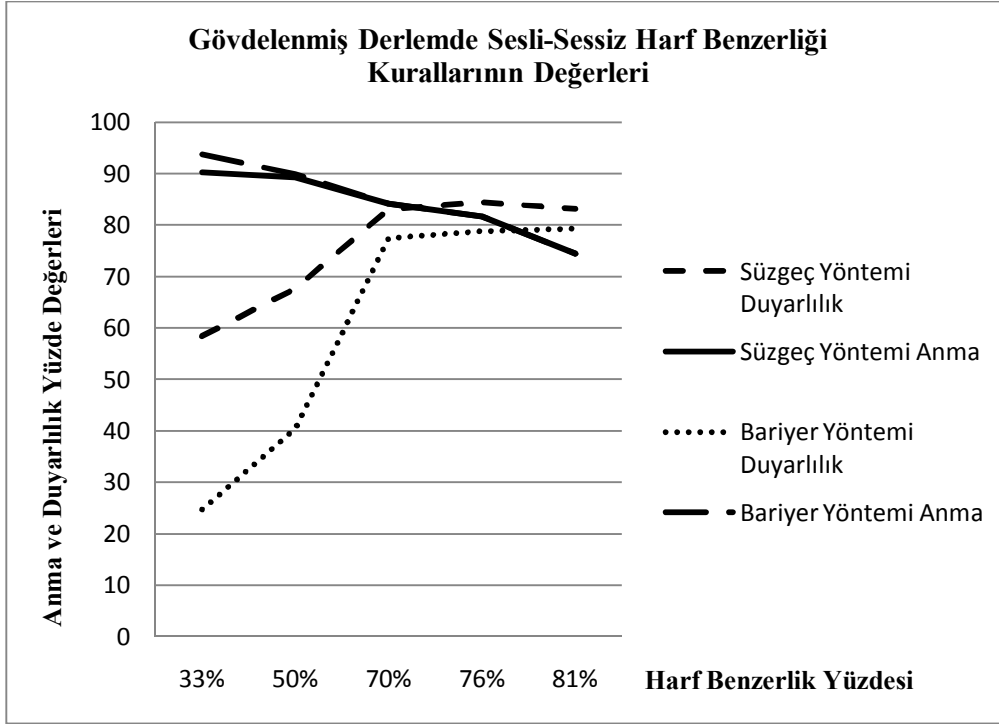


Şekil 6.7 Sesli-Sessiz Harf Benzerliği Kurallarının Süzgeç Yöntemi Uygulandığında Elde Edilen Sonuçların Yüzde Olarak Görünümü

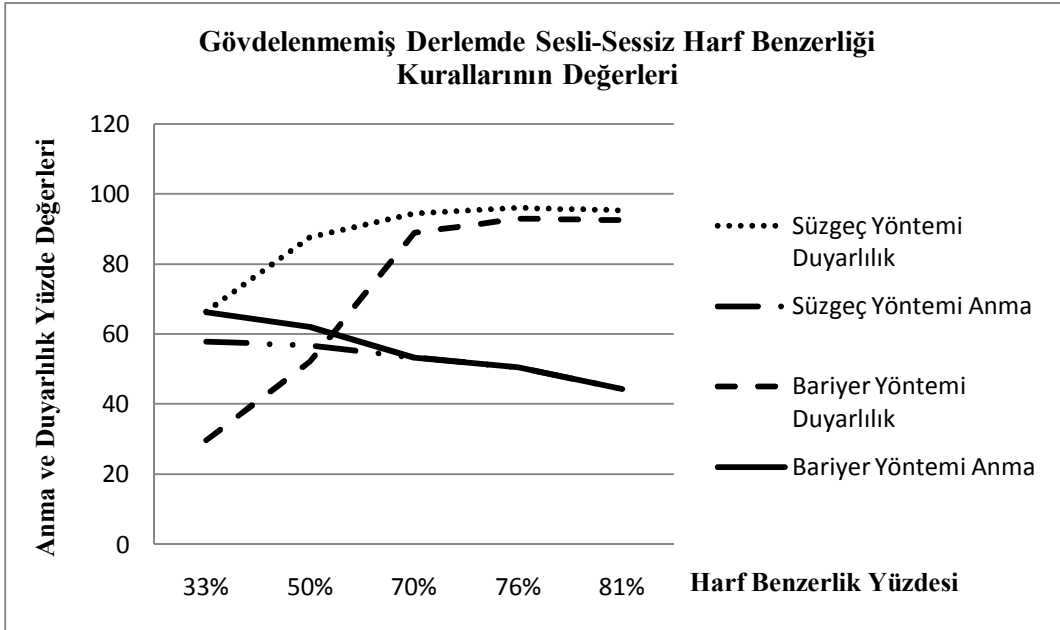


Şekil 6.8 Sesli-Sessiz Harf Benzerliği Kurallarının Bariyer Yöntemi Uygulandığında Elde Edilen Sonuçların Yüzde Olarak Görünümü

Şekil 6.7 ve Şekil 6.8 incelendiğinde en iyi sonucu gövdelenmiş ve gövdelenmemiş derlemde benzerlik oranı %70 olduğunda alınmaktadır.



Şekil 6.9 Gövdelenmiş Derlemde Sesli-Sessiz Harf Benzerliği Kural Değerlerinin Yüzde Olarak Görünümü

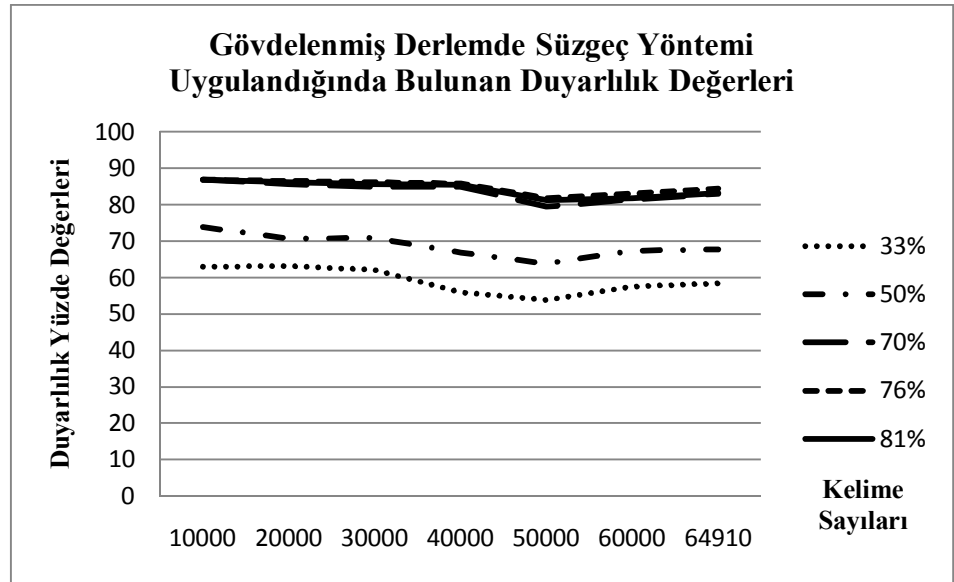


Şekil 6.10 Gövdelenmemiş Derlemde Sesli-Sessiz Harf Benzerliği Kural Değerlerinin Yüzde Olarak Görünümü

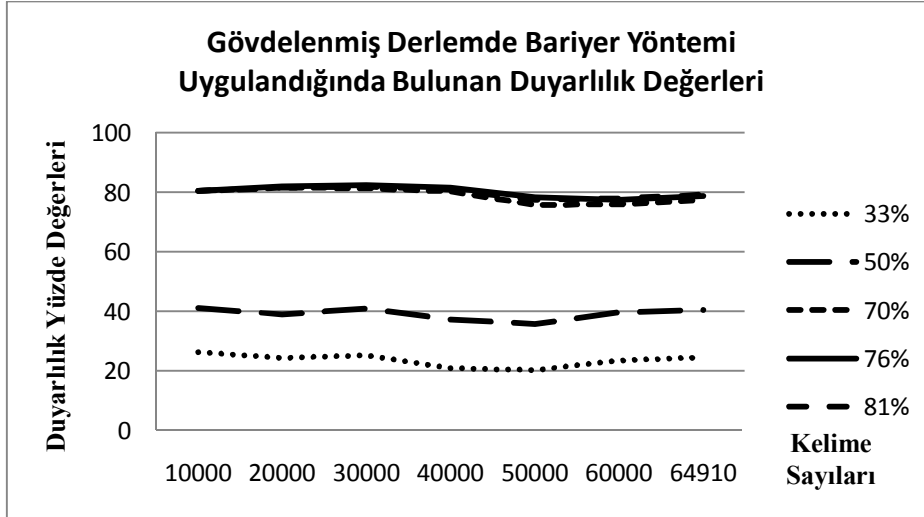
Şekil 6.9 ve 6.10'a bakıldığında her iki derlemde de duyarlılık değeri harf benzerliği ile doğru orantılı olup harf benzerliği oranı arttıkça duyarlılık değeri de artmaktadır. Anma değeri ise harf benzerliği ile ters orantılıdır. Harf benzerliği

oranı arttıkça anma değeri azalmaktadır. Bunun nedeni kelime çiftlerinde geçen harfler birbirlerine ne kadar çok benzerlik gösterirse derlemde geçen ikilemelerin yakalanması daha az olmaktadır. Böylece daha az kelime çiftleri yakalanacağı için duyarlılık değeri artmaktadır. Eğer harf benzerlik oranı azaldığında ise derlemde geçen ikilemelerin çoğu yakalandığından anma değeri yükselmektedir.

Şekil 6.9’da her iki yapıda da harf benzerlik oranı düşük olduğu zaman duyarlılık ve anma değerleri birbirine uzak iken harf benzerlik oranı arttıkça duyarlılık ve anma değerleri birbirlerine yakınlaşmaktadır. Kısacası harf benzerlik oranının artması ile yakalanan kelime çiftleri arasından ikileme olma olasılığı daha yüksektir.



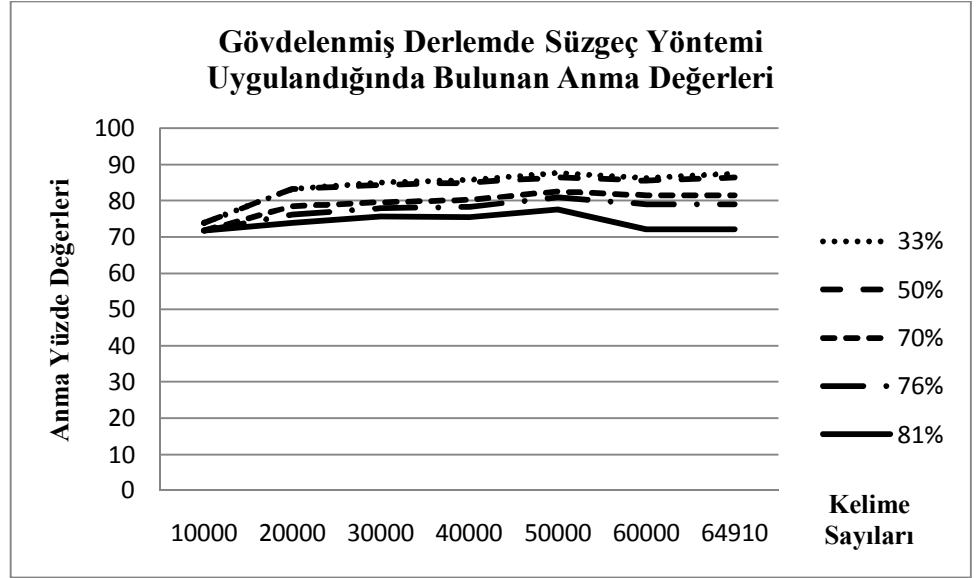
Şekil 6.11 Gövdelenmiş Derlemde Sesli-Sessiz Harf Benzerlik Kurallarının Süzgeç Yöntemi Uygulandığındaki Harf Benzerlik Oranlarına Göre Duyarlılık Değerleri



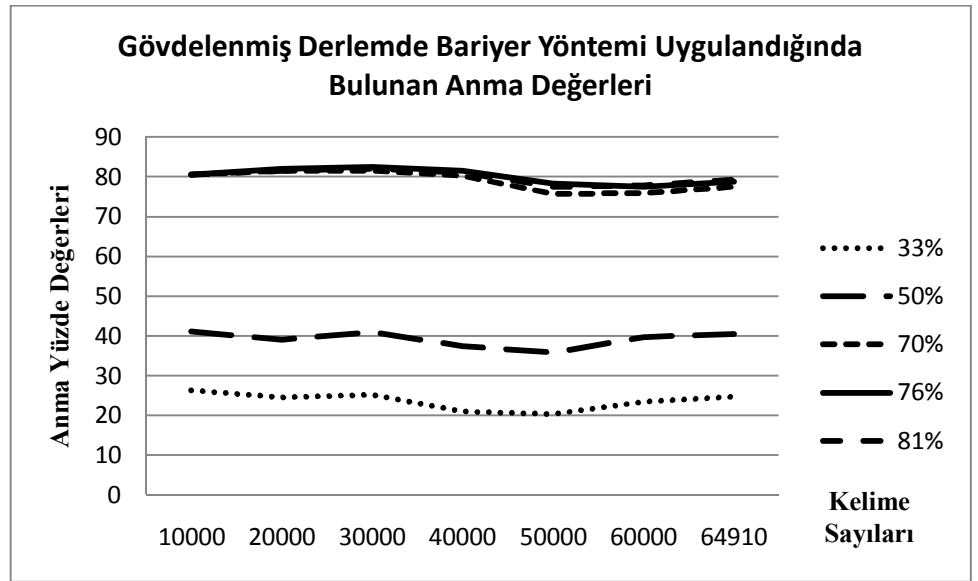
Şekil 6.12 Gövdelenmiş Derlemde Sesli-Sessiz Harf Benzerlik Kurallarının Bariyer Yöntemi Uygulandığında Harf Benzerlik Oranlarına Göre Duyarlılık Değerleri

Şekil 6.11 ve 6.12 incelendiğinde derlemde geçen ilk 10000 kelime alınmış ve duyarlılık değeri hesaplanmıştır. Daha sonra derlemde geçen ilk 20000 kelime alınmış ve böylece 10000 kelime arttırarak tek tek duyarlılık değerleri hesaplanıp yukarıdaki kelime çiftleri bazında grafikler oluşturulmuştur. Şekil 6.11 ve 6.12'ye bakıldığında kurallar iki yapı uygulansa da harf benzerlik oranı %70'den yüksek olduğunda daha az kelime çiftleri yakalanacağından duyarlılık değeri artmaktadır. Bu durumda harf benzerlik oranı %70 olduğunda daha fazla kelime çiftlerinin elendiğini ve sınır teşkil ettiği görülmektedir.

Şekil 6.11'deki duyarlılık değerleri kurallara süzgeç yöntemi uygulandığı için Şekil 6.12'deki duyarlılık değerlerine göre birbirine daha yakındır. Bunun nedeni bir kuralın çalışması sonucu yakalanan kelime çiftleri diğer kuralın çalışması için derlem oluşturmasındandır.

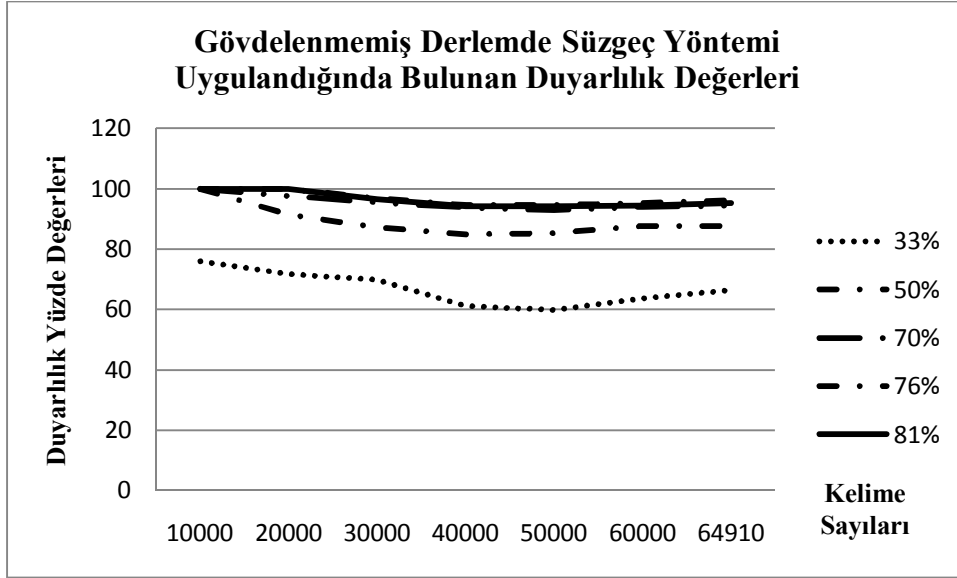


Şekil 6.13 Gövdelenmiş Derlemde Sesli-Sessiz Harf Benzerlik Kurallarının Süzgeç Yöntemi Uygulandığındaki Harf Benzerlik Oranlarına Göre Anma Değerleri

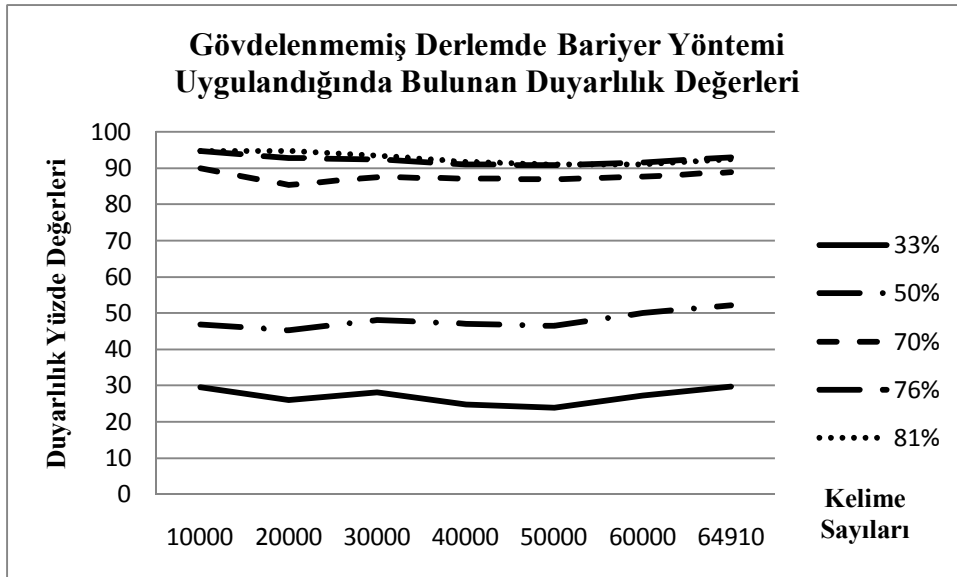


Şekil 6.14 Gövdelenmiş Derlemde Sesli-Sessiz Harf Benzerlik Kurallarının Bariyer Yöntemi Uygulandığındaki Harf Benzerlik Oranlarına Göre Anma Değerleri

Şekil 6.13'e bakıldığında gövdelenmiş derlemde süzgeç yapısı uygulanması sonucunda harf benzerliği oranlarının anma değerleri birbirine yakın çıkmıştır. Şekil 6.14'te ise bariyer yapısı uygulanmış ve bu yapıda kurallar diğer yapıya göre daha geniş derlem üzerinde çalışmasından dolayı harf benzerlik oranları yakın değildir.

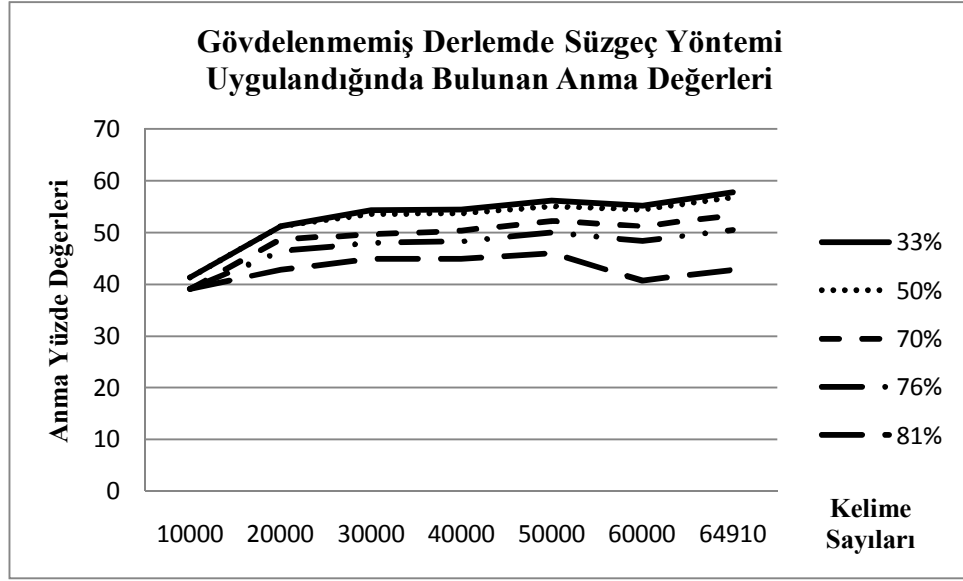


Şekil 6.15 Gövdelenmemiş Derlemde Sesli-Sessiz Harf Benzerlik Kurallarının Süzgeç Yöntemi Uygulandığındaki Harf Benzerlik Oranlarına Göre Duyarlılık Değerleri

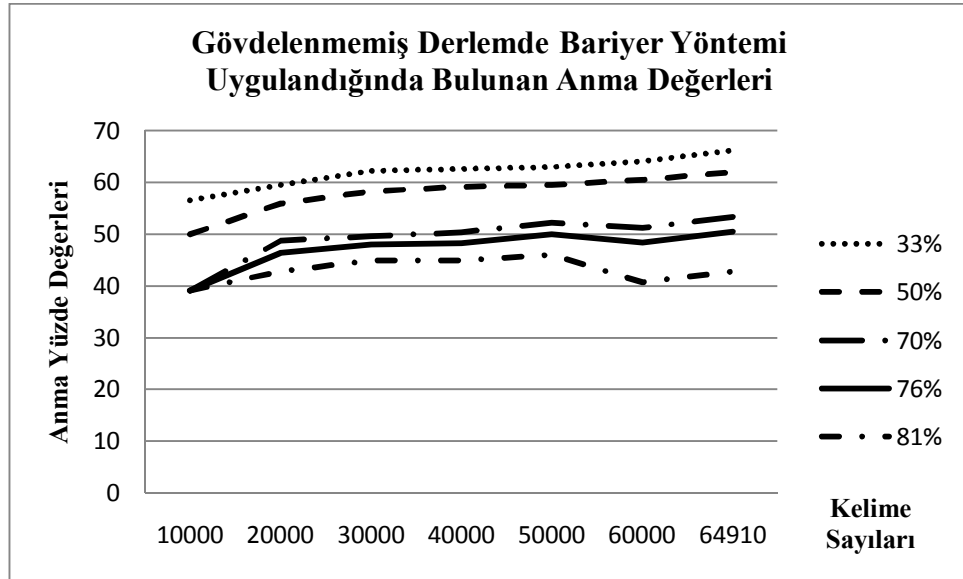


Şekil 6.16 Gövdelenmemiş Derlemde Sesli-Sessiz Harf Benzerlik Kurallarının Bariyer Yöntemi Uygulandığındaki Harf Benzerlik Oranlarına Göre Duyarlılık Değerleri

Şekil 6.15 ve 6.16 incelendiğinde kuralların iki yapıda çalışması sonucu harf benzerlik oranı %70'den yüksek olduğunda daha az kelime çiftleri yakalanacağından duyarlılık değeri artmaktadır. Bu durumda harf benzerlik oranı %70 olduğunda en iyi sonucu aldığımızı ve sınır teşkil ettiği görülmektedir.



Şekil 6.17 Gövdelenmemiş Derlemde Sesli-Sessiz Harf Benzerlik Kurallarının Süzgeç Yöntemi Uygulandığında Harf Benzerlik Oranlarına Göre Anma Değerleri



Şekil 6.18 Gövdelenmiş Derlemde Sesli-Sessiz Harf Benzerlik Kurallarının Bariyer Yöntemi Uygulandığında Harf Benzerlik Oranlarına Göre Anma Değerleri

Şekil 6.17 ve Şekil 6.18'e bakıldığında gövdelenmemiş derlemde kelimeler eklerden ayrılmadığını ve yöntemlerden elde edilen sonuçların birbirine daha yakın olduğu görülmektedir.

7. SONUÇ VE ÖNERİLER

Doğal dil işleme çalışmalarında birlikte gözlenen kelime birlikteliklerine eşdizim denmektedir. Bu çalışmalar kapsamında kelime grupları içinde eşdizimlerin yanı sıra deyimler, ikilemeler ve atasözleri de yer almaktadır. Kısacası kalıplaşmış söz birlikleri eşdizimlerin bir alt kümesini oluşturmaktadır.

Tez kapsamında, doğal dil işlemede önemli yere sahip olan kalıplaşmış söz birliklerinin bir alt kümesi olan ikilemeler özelliklerine bakılarak güncel Türkçe metinlerde tespit edilmesini amaçlayan bir çalışma sunulmuştur. Atasözleri kalıplaşmış söz birlikleri olsa da kendi başlarına bir cümle oluşturdıkları için tez çalışması kapsamında yer almamıştır.

Tezde, kalıplaşmış söz birliği olan ikilemelerin sınanması için ODTÜ-Sabancı ortak çalışması olan OSTAD derlemi kullanılmıştır. Çalışmanın temel üç aşaması şunlardır:

- ✓ *Kuralların belirlenmesi*; “Deyimler ve Atasözleri Sözlüğü” (Kuşçu v.d., 1997) kalıplaşmış söz birlikleri veri tabanına girilmiş ve özellikleri incelenmiştir. Bu özellikler doğrultusunda kalıplaşmış söz birliklerinin bir alt kümesi olan ikilemelerin sınanması için ses ve yapı benzerlikleri doğrultusunda kurallar belirlenmiştir.
- ✓ *Kuralların uygulanması*; belirlenen kurallar OSTAD derleminde çalıştırılarak sonuçlar elde edilmiştir.
- ✓ *Değerlendirme*; Sonuçlar anma ve duyarlılık esasında değerlendirilmiş ve grafikler ile gösterilmiştir. Kelimelerin yüzeysel form ile gövdelenmiş formları arasındaki farklılıklara bakılmıştır.

Çıkan sonuçlar doğrultusunda en iyi başarımlar sesli-sessiz harf benzerliği kurallarının harf benzerlik oranı kuralı %70 olduğunda elde edilmiştir. Kuralların çalışması sonucunda yakalanan ikilemelerin özellikleri aşağıda listelenmiştir. Özelliklerin yanında verilen örnekler derlemden alınmıştır.

- ✓ Aynı sözcüğün yenilenmesi ile oluşan ikilemeler; “fokur fokur”, “yavaş yavaş” vb.
- ✓ Biri anlamlı biri anlamsız sesçe benzer olan ikilemeler; “çoluk çocuk”, “kaba saba” vb.
- ✓ İki sözcükte anlamsız sesçe benzer sözcüklerden oluşan ikilemeler; “apar topar”, “falan filan” vb.
- ✓ Aynı sözcüğe “m” getirilmesiyle oluşan ikilemeler; “beyin meyin”, “şaka maka” vb.
- ✓ İkilemeyi oluşturan kelimeler arasına noktalama işaretleri almamaktadırlar.
- ✓ İkilemeler cümle içinde “*Ad*” göreviyle kullanılırlar; “*Çalı malı, sırık mırık, şeker meker* derken, Ayşe kadın giriverir sahneye.” cümlesinde ad göreviyle kullanılan ikilemeler geçmektedir.
- ✓ İkilemeler cümle içinde “*Sıfat*” göreviyle kullanılırlar; “*Uzak uzak* ülkelerden birinde fakir bir anne ve kızı yaşarmış.” cümlesindeki ikileme sıfat göreviyle kullanılmıştır.
- ✓ İkilemeler cümle içinde “*Zarf*” göreviyle kullanılırlar; “Gerçi ben *kendi kendime* idare ederdim ama şu dersler yok mu, o konuda kesinlikle yardıma ihtiyacım var.” cümlesindeki ikileme zarf göreviyle kullanılmıştır.

Çalışma sırasında kalıplaşmış söz birlikleri el ile OSTAD derlemi üzerinde işaretlenmiş, böylece ikilemelerin dışında kalıplaşmış söz birlikleri tespiti çalışmaları için bir alt yapı oluşturulmuştur.

Tezde, kuralların çalışması sonucu elde edilen sonuçlara bakıldığında ikilemelerin yakalanmasında gövdelenmiş derlemin yüzeysel forma göre daha fazla başarılı olduğu görülmüştür.

Sesli-sessiz harf benzerliği kurallarının çalışması sonucu harf benzerliği oranı ile duyarlılık arasında doğru orantı olduğu görülmüştür. Böylece kelimelerde geçen harflerin benzerliği arttıkça yakalanan kelime çiftleri daha az olacağından duyarlılık değerinin arttığı görülmüştür. Ters durum ise anma için geçerlidir. Harf benzerliği

oranı arttıkça yakalanan kelime çiftlerindeki ikilemelerin sayısı azalacağından dolayı benzerlik ile anma arasında ters orantı olduğu görülmüştür.

Bunların dışında kelimelerde sıkça kullanılan diskriminant analizi yöntemi de uygulanmıştır. Ama daha iyi bir sonuç elde edilmediğinden sonuçları burada verilmemiştir.

İkilemeler tek başlarına ya da başka sözcüklerle öbekleşerek deyim oluşturmaktadırlar. "*İleri geri* konuşmak sana yakışmıyor." cümlesinde "ileri geri" ikilemesi "konuşmak" eylemiyle birleşerek deyim oluşturmuştur. Bu örnekte görüldüğü gibi ikilemeler deyimlerin içinde sıkça görülmektedir. Bu tez çalışması sonucunda elde edilen sonuçların ileriki çalışmalarda deyimlerin tespiti problemine örnek teşkil edeceği düşünülmektedir.

KAYNAKLAR DİZİNİ

- Aktaş, Ö.**, 2006, Türkçe için Verimli bir Cümle Sonu Belirleme Yöntemi, Dokuz Eylül Üniversitesi, İzmir.
- Aksan, Doğan**, 2002, Türkçenin Sözvarlığı, Engin Yayınevi, Ankara.
- Aksan, Doğan**, 1996, Anadilimizin Söz Denizinde, Bilgi Yayınevi, Ankara.
- Aksoy, Ö. A.**, 1993, Atasozleri ve Deyimler Sozluğu I, II, İnkılap Kitabevi, İstanbul.
- Atalay, N.B., Oflazer, K. and Say, B.**, 2003, The Annotation Process in the Turkish Treebank, in Proceedings of the EACL Workshop on Linguistically Interpreted Corpora - LINC, Budapest, Hungary.
- Church, K. W. & Hanks, P.**, 1990, Word Association Norms, Mutual Information, and Lexicography Computational Linguistics.
- Crystal, D.**, 1991, A Dictionary of Linguistics and Phonetics, Blackwell.
- Çotuksöken, Yusuf**, 1994, Deyimlerimiz, Özgül Yayınları, İstanbul.
- Dinçer, T.**, 2004, Türkçe için istatistiksel bir bilgi geri-getirim sistemi, Doktora Tezi, U.B.E.,Ege Üniversitesi, İzmir.
- Ergin, M.**, 1996, Türk Dili, Bayrak Yayınevi.
- Gökdayı, Hürriyet**, “Türkçede Kalıp Sözler”, 2008.
- Gülşen Cebiroğlu ve Eşref Adalı**, 2002, Sözlüksüz köke ulaşma yöntemi, *In* Proceedings of the 19th TBD Bilişim Kurultayı, İstanbul.
- Hoey, M.**, 1991, Patterns of Lexis in Text, Oxford University Press.
- Kışla, T.**, 2009, Türkçe İçin Tümleşik Bir Biçimbirim Çözümleme ve Sözcük Türü Tespit Yöntemi, Doktora Tezi, Uluslar arası Bilgisayar Enstitüsü, Ege Üniversitesi, İzmir.

KAYNAKLAR DİZİNİ (devam)

- Kumova, S.**, 2010, Türkçede Hesaplamalı Metin Analizi, Doktora Tezi, Uluslararası Bilgisayar Enstitüsü, Ege Üniversitesi, İzmir.
- Kumova, S.**, 2005, Derlem Hazırlama Kriterlerinin Oluşturulması, Yüksek Lisans Tezi, Uluslararası Bilgisayar Enstitüsü, Ege Üniversitesi, İzmir.
- Kuşçu Hüseyin ve Ülkü**, 1998, Atasözleri ve Deyimler Sözlüğü, Altın Yayınevi, Ankara.
- Manning, C.D. and Schütze, H.**, 1999, Foundations of Statistical Natural Language Processing.
- Oflazer, Çetinoğlu, Say**, 2004, Integrating Morphology with Multi-word Expression Processing in Turkish.
- Oflazer, K. ve Bozsahin, H. C.**, 2006, Türkçe Doğal Dil İşleme, http://www.turkoloji.cu.edu.tr/DILBILIM/turkce_dogal_dil_isleme.pdf, (Erişim tarihi: 18 Nisan 2011)
- Özdemir, A.**, 2007, Türk Atasözlerinden Seçmeler, İstanbul.
- Özkan**, 2007, Türkiye Türkçesinde Belirteçlerin Fiillerle Birliktelik Kullanımı ve Eşdizimliliği, Adana.
- Sakça, B.**, 2009, “İkilemeler”, <http://www.yenimakale.com/ikilemeler.html> (Erişim tarihi: 28 Mart 2011)
- Smadja, F.A.**, 1993, Retrieving Collocations from Text: Xtract. Computational Linguistics.
- Sarıkaş, F.**, 2006, “Problems in Translating Collocations”, Elektronik Sosyal Bilimler Dergisi.
- Sinclair, J.**, 1991, Corpus Concordance, Collocation.

KAYNAKLAR DİZİNİ (devam)

- Tannen, Deborah ve Öztekin, Piyale C.** , 1981, Formulaic Expressions in Turkish and Greek, Florian Coulmas (der.) içinde, The Hague: Mouton.
- Taşgüzel,** 2004, İlköğretim Türkçe Ders Kitaplarında Öğretici Nitelikli Metinlerdeki Eşdizimsel Örüntülemelerin Görünümü üzerine çalışma, Ankara.
- Toklu, M. Osman,** 2003, Dilbilime Giriş, Akçağ Yayınları, Ankara.
- Torun, Y.,** 2005, Nurullah Ataç'ın Denemelerinde Devrik Yapılar, Doktora Tezi, Adana.
- Yüceol Özezen, Muna,** 2001, Türkçe Deyimler Üzerine Birkaç Söz, Türk Dili.
- Vardar, Berke,** 1988, Açıklamalı Dilbilim Terimleri Sözlüğü, İstanbul.
- Williams, Raymond,** 2005, Anahtar Sözcükler Kültür ve Toplumun Söz Varlığı, (Çev. Savaş Kılıç), İletişim Yayınları, İstanbul.
- Wray, Alison.,** 2001, Formulaic Language and the Lexicon, Cambridge: Cambridge University Press.

ÖZGEÇMİŞ

Sinem HINÇALAN

Adres: 89 Sok. No:4/1 D:8 Güzelyalı/İZMİR

E-Posta: shincalan@yahoo.com

KİŞİSEL BİLGİLER:

Doğum Tarihi : 24.01.1982

Doğum Yeri : İzmir

EĞİTİM BİLGİLERİ

2005 – 2011	Ege Üniversitesi Uluslararası Bilgisayar Enstitüsü <i>Uluslararası Bilgisayar Anabilim Dalı</i>
2000 – 2005	Ege Üniversitesi Eğitim Fakültesi <i>Bilgisayar ve Öğretim Teknolojileri Öğretmenliği</i>
1996 – 2000	Buca Teknik Lisesi

İŞ TECRÜBELERİ

2007 –	Selma Yiğitalp Anadolu Lisesi <i>Bilişim Teknolojileri Dersi Öğretmeni</i>
2006 – 2007	Yunuslar Çok Programlı Lisesi <i>Bilişim Teknolojileri Dersi Öğretmeni</i>
2006 – 2005	Buca Kız Meslek Lisesi <i>Bilişim Teknolojileri Dersi Öğretmeni</i>