

**TÜRKİYE CUMHURİYETİ
ANKARA ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
DİL BİLİM ANABİLİM DALI**

**WALS VERİTABANINDA BİLGİSAYAR DESTEKLİ
TİPOLOJİK VERİ ÇÖZÜMLEMESİ**

Tezli Yüksek Lisans Tezi

Mehmet Umut MUTLU

Ankara, 2021

**TÜRKİYE CUMHURİYETİ
ANKARA ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
DİL BİLİM ANABİLİM DALI**

**WALS VERİTABANINDA BİLGİSAYAR DESTEKLİ TİPOLOJİK
VERİ ÇÖZÜMLEMESİ**

Tezli Yüksek Lisans Tezi

Mehmet Umut MUTLU

Tez Danışmanı

Prof. Dr. Özgür AYDIN

Ankara, 2021

**TÜRKİYE CUMHURİYETİ
ANKARA ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
DİL BİLİM ANABİLİM DALI**

**WALS VERİTABANINDA BİLGİSAYAR DESTEKLİ
TİPOLOJİK VERİ ÇÖZÜMLEMESİ**

YÜKSEK LİSANS TEZİ

Tez Danışmanı: Prof. Dr. Özgür AYDIN

TEZ JÜRİSİ ÜYELERİ

Adı ve Soyadı

İmzası

- 1- Prof. Dr. Özgür AYDIN**
- 2- Prof. Dr. Nadir Engin UZUN**
- 3- Dr. Öğr. Üyesi İpek Pınar UZUN**

Tez Savunması Tarihi

23.06.2021

T.C.
ANKARA ÜNİVERSİTESİ
Sosyal Bilimler Enstitüsü Müdürlüğü'ne,

Prof. Dr. Özgür AYDIN danışmanlığında hazırladığım “WALS Veritabanında Bilgisayar Destekli Tipolojik Veri Çözümlemesi (Ankara, 2021)” adlı yüksek lisans ☒ - doktora/bütünleşik doktora ☐ tezimdeki bütün bilgilerin akademik kurallara ve etik davranış ilkelerine uygun olarak toplanıp sunulduğunu, başka kaynaklardan aldığım bilgileri metinde ve kaynakçada eksiksiz olarak gösterdiğimi, çalışma sürecinde bilimsel araştırma ve etik kurallarına uygun olarak davrandığımı ve aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul edeceğimi beyan ederim.

23.06.2021

Mehmet Umut MUTLU

ÖNSÖZ VE TEŞEKKÜR

Lisans öğrenimim sırasında *yeniden* tanıştığım dil, bana kendisinin bizzat gerçeklik olduğunu defalarca ve farklı şekillerde göstermiştir. Öyle ki deneyimlediğimiz her gerçeklik, dilin süzgecinden geçerek ulaşmaktadır zihinlerimize. Deyim yerindeyse dil her köşe başını tutmuş ve dilbilim ise onun peşinde bir dedektif gibidir. Mesleğim olmasını istediğim bu alan, bu nedenle yalnızca bir bilim dalı olmanın ötesindedir benim için, çünkü dili anlamak dilden fazlasını anlamaktır.

Yeryüzünde varlığını sürdürmeye çalışan 7000'den fazla dil bulunuyor. Bu aynı zamanda 7000'den fazla gerçeklik mi demek öyleyse? Kendimizi bir an için hiç bilmediğimiz bir dilin konuşulduğu bir coğrafyada, kalabalık bir caddenin ortasında bulduğumuzu düşünelim. Dediklerini anlamadığımız, bildiklerimizden çok farklı sözcükler kullanan o insanlar aslında çok yakından tanıdığımız şeylerden söz ediyorlar muhtemelen. Daha da önemlisi çok yakından tanıdığımız ve aynısından bizde de olan bir araç kullanıyorlar bunun için: DİL. Son yüzyılda süregelen çağdaş dilbilim çalışmaları, bu çok çeşitliymiş gibi görünen özel dillerin büyük resimde bir bütünü oluşturan parçalar olduğunu kanıtlar nitelikte. Çeşitli nedenlerle benzer amaçlar için farklı araçlar kullanan dillerin, tesadüfi sayılamayacak bir benzerlikle bir noktada kesişmesi aslında hepsinin ortak bir mekanizma üzerinde çalıştığının göstergesi. Her birimizin genetik olarak doğuştan getirdiği o muhteşem yeti...

Söz konusu çeşitlilik içerisinde bir düzen aramaya çalışan dil tipolojisi araştırmacıları, yüzyıllardır farklı gözlemlerle dillerin dünyasının nasıl görüldüğüne dair resim çizmeye çalışmışlar. Teknolojik olanaklar çerçevesinde bu resmi bilgisayar nasıl çizer merakıyla, insanların düşünüp uyguladığı birkaç işlemi bilgisayara devrettiğimiz bu çalışmanın bir öncü çalışma niteliği taşıdığını, dolayısıyla ele alınabilecek daha pek çok

farklı soru olduğunu belirtmek isterim. Temennim Türkçe alanyazında eksik kalan bu noktanın bir an önce çok daha geniş kapsamlı çalışmalarla doldurulmasıdır.

Bu bölümde isimler anıp teşekkür etmek gelenek haline gelmiştir. Ancak bana hep öyle gelir ki okuyucuların tanımadıkları bu adların arkasında onların gerçek katkıları ve değerleri kaybolur. Bu nedenle, ben farklı bir şey yaparak teşekkür etmek istediğim kişileri adlarıyla değil, benim için olan önemleriyle anmak istiyorum. Başta, bilimin aydınlattığı yolda özgürce hareket etmeme izin veren, ancak hep uzaktan takip ettiğini hissettiğim danışman hocam ve son iki yıldır yüksek lisans öğrenimimi birlikte yürüttüğüm değerli hocalarım olmak üzere, bugüne dek hayatın herhangi bir alanında, herhangi bir şekilde, doğrudan veya dolaylı olarak bilgisini ve deneyimlerini paylaşan tüm öğretmenlerime teşekkürlerimi kuru kuruya değil kariyerim süresince yapacağım çalışmalarımı sunmak istiyorum. Yine, başta yüksek lisans süreci boyunca kurduğumuz dostluklar nedeniyle dönem arkadaşlarım olmak üzere, bugüne dek zor ve sıkıcı yolları daha çekilir ve eğlenceli kılan tüm arkadaşlarıma, dostlarıma teşekkür ediyorum.

Biz insanlar genellikle bir film izlerken kamera arkasındakileri, bir tabloya bakarken ressamın atölyesini, bir roman okurken yazarın odasında neler yaşandığını düşünmeyiz. Çoğu zaman görünen güzel şeylerin üzerinde görünmeyen emekler vardır, farkında olmayız. Ben de bu nedenle tez çalışmamı, araştırmacının odasını sıkça ziyaret eden ve görünmeyen emekleriyle onu yetiştirip her koşulda destekleyen annem Sündüs Mutlu ve babam Ümit Mutlu'ya armağan ediyorum.

Dillerin çeşitliliği içinde gezmek kimi zaman baş döndürücüydü. Her seferinde gördüklerimi insanlığa aktarma hedefiyle ayılıp yeniden yola koyuldum. Yolun sonunda sonuçlarımı sunmanın heyecanı, bilime ve insanlığa katkı sunabilme dileklerimle!

Mehmet Umut MUTLU
Ankara, Haziran 2021

İÇİNDEKİLER

ÖNSÖZ VE TEŞEKKÜR	i
İÇİNDEKİLER.....	iii
KISALTMALAR ve İMLER.....	v
TABLolar.....	vi
ŞEKİLLER.....	vii
1. GİRİŞ.....	1
1.1. Tezin Adı	1
1.2. Tezin Konusu.....	1
1.3. Tezin Amacı.....	1
1.4. Tezin Önemi	2
1.5. Tezin Araştırma Soruları	4
1.6. Tezin Kapsamı ve Sınırlılıkları.....	4
1.6.1. Metodolojik Sınırlılıklar	4
1.6.2. Kuramsal Sınırlılıklar	6
1.7. Tezin Yapılanışı.....	7
2. KURAMSAL ARTALAN	8
2.1. Giriş: Tipoloji Neden Gerekli?	8
2.2. Dil Tipolojisinin Kısa Tarihi	11
2.2.1. Dilbilgisinden Karşılaştırmalı Dilbilgisine ve Tipolojiye	11
2.2.2. Schleicher ve Soyağacı Kuramı.....	13
2.2.3. Schmidt ve Dalga Kuramı	15
2.2.4. Greenberg ve Evrenceleri	16
2.2.5. Chomsky ve Evrensel Dilbilgisi	17
2.2.6. Hesaplamalı Tipoloji	19
2.3. The World Atlas of Language Structures (WALS)	21
2.2.7. WALS'in Düzenlenişi ve Kullanımı	22
2.2.7.1. Diller.....	22
2.2.7.2. Özellikler	24
2.2.7.3. Haritalar.....	25
2.2.8. WALS İçin Karşılaştırma Yöntemi Önerileri.....	26
2.2.8.1. Tipiklik (Comrie, 2010).....	26
2.2.8.2. Tutarlılık (Orgun, 2011)	27
2.2.8.3. Karşılıklı Benzerlik/Benzemezlik (Uzun, 2012a)	28

2.2.9. WALS Üzerine Hesaplamalı Çalışmalar	31
2.4. Ara Sonuç	32
3. YÖNTEM	34
3.1. Veri Önışlemlemesi	34
3.2. Eksik Veri Analizi ve normalize()	36
3.3. Çözümleme Yöntemlerinin Fonksiyonlaştırılması	38
3.4. Çözümleme Adımları ve Raporlama	39
4. BULGULAR	41
4.1. Eksik Veri Analizi ve Normalizasyon	41
4.2. Tipolojik Karşılaştırma Sonuçları.....	45
4.2.1. Tipiklik	46
4.2.2. Tutarlılık	48
4.2.3. Karşılıklı Benzerlik.....	50
5. TARTIŞMA VE SONUÇ	56
5.1. Veritabanına İlişkin Gözlemler.....	56
5.2. Tipolojik Karşılaştırma Sonuçları.....	58
5.2.1. Tipiklik	58
5.2.2. Tutarlılık	59
5.2.3. Karşılıklı Benzerlik.....	59
5.3. Sonuç ve Öneriler	60
KAYNAKÇA.....	63
ÖZET	70
ABSTRACT	71
TÜRKÇE-İNGİLİZCE TERİMLER	72
EKLER	75
EK-1: WALS Özellikler Listesi.....	75
EK-2: Normalizasyon Sürecinde Yapılan İşlemlerin Dökümü	82

KISALTMALAR ve İMLER

*	dilbilgisel olarak bozuk yapı
?	dilbilgisel olarak daha az bozuk yapı
#	anlambilimsel olarak kabuledilemez yapı
>	R komut satırı imi
1TK	1. tekil kişi
BEL	belirtme durumu
CLLD	Cross-Linguistic Linked Data
CLDF	Cross-Linguistic Data Formats
E	eylem
ED	Evrensel Dilbilgisi
GEÇ	geçmiş zaman
HD	hesaplama dilbilim
İYE	iyelik
MÖ	Milattan Önce
N	nesne
NA	<i>not applicaple</i> (mevcut değil)
Ö	özne
REALİS	gerçeklik kipi
TAM	tamlayan durumu
YÖN	yönelme durumu
WALS	the World Atlas of Language Structure

TABLÖLAR

Tablo 1. Comrie'nin (2010) çözümleme örneği.....	27
Tablo 2. Orgun'un (2011) çözümleme örneği.....	27
Tablo 3. Uzun'un (2012a) karşılıklı benzerlik çözümleme örneği	28
Tablo 4. Uzun'un (2012a) ikili özelliklerde karşılıklı benzemezlik çözümleme örneği	29
Tablo 5. Uzun'un (2012a) sıralı özelliklerde karşılıklı benzemezlik çözümleme örneği	29
Tablo 6. Uzun'un (2012a) kategorik özelliklerde karşılıklı benzemezlik çözümleme örneği.....	30
Tablo 7. Uzun'un (2012a) kategorik özelliklerde birleşik değerler için önerdiği karşılıklı benzemezlik çözümleme örneği	30
Tablo 8. Özelliklerin alanlara göre içerdikleri dil ve özellik sayıları.....	42
Tablo 9. Normalizasyon öncesinde ve sonrasında en çok etiket alan ilk 10 dil ve özellik.....	45
Tablo 10. Dillerin kendi başlarına tipiklik oranlarına göre örneklemelerin dağılımı (%).....	46
Tablo 11. Dillerin kendi başlarına tipiklik değer ölçümleri (Daga – Örneklem 100) (%).....	47
Tablo 12. Dil ailesi tutarlılık testi (Altay örneği).....	49
Tablo 13. Ailelerin tutarlılık oranları (Örneklem 200).....	49
Tablo 14. Normalizasyon öncesinde ve sonrasında farklı ailelerden diller arasında en çok tutarlılık gösteren ilk 5 dil (Örneklem 200).....	50
Tablo 15. Dillerin kendi başlarına tipiklik oranlarına göre örneklemelerin dağılımı (%).....	51
Tablo 16. 100 dilli örneklemde benzerlik oranları	51
Tablo 17. 200 dilli örneklemde benzerlik oranları	52
Tablo 18. Alanlar bazında benzerlik tablosu (Tay dili-Vietnam dili örneği) (%).....	52
Tablo 19. Ailelerin karşılıklı benzerlik ortalamaları	55

ŞEKİLLER

Şekil 1. Darwin ve Schleicher'in soy ilişkilerini göstermek için kullandıkları ağaç şemaları.....	14
Şekil 2. Schmidt'in önerdiği Dalga Modeli için temsili çizim (Daval-Markussen, 2011'den Türkçeye çevrilmiştir.).....	15
Şekil 3. WALS üzerinde Türkçeye ait sayfanın görüntüsü	24
Şekil 4. WALS haritaları örneği (1A).....	26
Şekil 5. Önişleme aşamaları şeması.....	34
Şekil 6. Dillere ait vektörlerin temsili görünümü	35
Şekil 7. Normalizasyon işlemi öncesinde ve sonrasında eksik veri görselleştirmesi	43
Şekil 8. Normalizasyon öncesi ve sonrasında etiketli diller	44
Şekil 9. Normalizasyon öncesi ve sonrasında etiketli özellikler	44
Şekil 10. Tipiklik haritası kesiti (Örneklem 200).....	48
Şekil 11. Örneklemelerin normalizasyon sonrası benzerlik görselleştirmesi	53
Şekil 12. 100 dilli örneklemde benzerlik ölçümüne normalizasyonun etkisi.....	54
Şekil 13. 200 dilli örneklemde benzerlik ölçümüne normalizasyonun etkisi.....	54

1. GİRİŞ

1.1. Tezin Adı

WALS Veritabanında Bilgisayar Destekli Veri Çözümlemesi

1.2. Tezin Konusu

Bu tezin konusunu, *World Atlas of Language Structures* (WALS) (Dryer ve Haspelmath, 2013) verisi üzerinden yapılabilecek tipolojik karşılaştırma/sıralama değerlendirmesi için alanyazında önerilmiş olan üç farklı yöntemin (Comrie, 2010; Orgun, 2011; Uzun, 2012a), R dilinde (R Core Team, 2021) işlenerek bir paket¹ haline getirilmesi ve bu sayede bilgisayar destekli tipolojik bir çözümlemenin nasıl yapılacağına ilişkin yöntem ve araçların geliştirilmesi oluşturmaktadır.

1.3. Tezin Amacı

Dil tipolojisi (language typology) veya *dilbilimsel tipoloji* (linguistic typology), başlangıcı karşılaştırmalı dilbilgisi çalışmalarına dek uzanan, dolayısıyla da yazını oldukça geniş bir çalışma alanıdır. Dil tipolojisinin tarihsel gelişimi 2. Bölüm’de sunulmuşsa da burada kısaca belirtilmesi gereken önemli bir nokta, bu tezin ana omurgasını oluşturan kilometre taşının gelişen teknolojiyle birlikte dil tipolojisi

¹ Yazılım terminolojisinde *paket* (package) veya *kütüphane* (library), kişilerin belli amaçlar için yazıp başkalarının kullanımına sunduğu kodlardan oluşan bir bütün yazılıma gönderimde bulunmaktadır.

arařtırmaları iin de oluřan yeni olanaklar olduėudur. Bu kapsamda deėerlendirilebilecek bir alıřma olan WALS (Dryer ve Haspelmath, 2013), dnya dillerinin yapısal zellikleri hakkında bilgiler ieren tipoloji veritabanlarından biridir.

WALS verisinin tipolojik bir incelemede nasıl deėerlendirilmesi gerektiėi, bařka bir deyiřle veritabanında sunulan bilgiler doėrultusunda dillerin nasıl sıralanması gerektiėi, sorusu farklı arařtırmacıların ilgisini eken bir konu olmuř ve bu noktada farklı yntemler nerilmiřtir. Bunlardan Comrie'nin (2010) dilin bir zelliėinin dnya dillerindeki daėılımına dayanan *tipiklik* (typicality), Orgun'un (2011) ortak zellikler sergileyen dillerin bu davranıřlarını diėer zelliklerde de srdrp srdrmediklerine dayanan *tutarlılık* (coherence) ve Uzun'un (2012a) dillerin *karřılıklı benzerlik ve benzemezlikleri* (mutual similarity/dissimilarity) zerinden nerdiėi puanlama nerileri bařat yntemlerdendir. Bu tez alıřması kapsamında, adı geen bu  yntemin R dilinde bir paket haline getirilmesi ve bu  yntem ile uygulanan tipolojik zmlemenin sonularının karřılařtırmalı bir biimde betimlenmesi amalanmaktadır.

1.4. Tezin nemi

Arařtırmacılar, WALS tabanlı olarak yaptıkları tipoloji alıřmalarında iki temel sorun ile karřılařmaktadırlar. Bunlardan ilki veritabanındaki *eksik veri* (missing data) oranının analiz yapmayı zorlařtırıcı lde yksek olmasıdır. Bu sorunun eřitli yntem denemeleriyle ele alan pek ok alıřma (bkz. Daum III ve Campbell, 2007; Daum III, 2009; Lu, 2013; Takamura vd., 2016 vb.) bulunmaktadır. İkinci sorun ise, WALS'teki verilerin karřılařtırmasının web arayznn veya bilgisayar uygulamasının sndėu olanaklar dıřında, hızlı ve doėru bir biimde yapma yolunun neredeyse hi bulunmamasıdır. Bu iki nedenle, alıřmalar ya zgl bir dil ailesi ya da belirli birka dil veya birka zellikle sınırlandırılmak zorunda kalmıřtır (rn. Albu, 2006; Aėbaht, 2014; Tashmurad, 2015 vb.).

Bu tez çalışmasının çözümleme aşaması için tercih edilen ve bir önceki başlıkta da sözü edilmiş olan üç yöntem, R ortamında birer fonksiyon olarak tanımlanıp normalize edilmiş WALS verisiyle birlikte tez sürecinin tamamlanmasının ardından bir paket olarak araştırmacıların kullanımına sunulmaktadır. Bu da iki açıdan önemlidir: İlk olarak, Türkçe alanyazında önerilmiş olan yöntemler R dili evreninde daha geniş bir kitleye ulaştırılmış olacaktır. Ayrıca, çalışmanın çerçevesi her ne kadar R ve WALS ile sınırlıymış gibi görünse de tipoloji çalışmalarında *veri bilim* (data science) ve *veri mühendisliği* (data engineering) tekniklerinin kullanımının dilbilimsel çözümlemeler için de ilerletici bir etkiye sahip olacağını gösterme noktasında Türkçe alanyazında öncü bir rol üstlenmektedir. Dahası, yalnızca WALS için değil aynı zamanda bu gibi tüm veritabanları için de uyarlanabilir ve uygulanabilir bir yöntem denemesi olma potansiyeline sahiptir.

Öte yandan, elde edilecek olan betimsel bulgular, mevcut durumda WALS'teki en yüksek veri olanağına sahip örneklemeler üzerinden dillerin genel görünümüne ilişkin geniş bir bakış sunacaktır. Bu sonuçlar, gelecekte WALS'in sağladığı veri oranının artması, WALS üzerinden seçilebilecek farklı bir örneklem, tercih edilebilecek farklı bir yöntem vb. seçenekler çerçevesinde de tekrar edilebilir bir çözümleme olanağı ile karşılaştırmaya uygun bir biçimde sunulmuş olacaktır. Ayrıca araştırmacılar, kendi topladıkları verileri veya farklı bir veritabanından elde ettikleri verileri WALS veritabanı yapısında derleyerek yine paket içeriğindeki yöntemler ile çözümleme yapabileceklerdir.

Son olarak, çalışma kapsamında sürdürülen eksik veri analizi sırasında WALS verisinin yapısının yöntemleri etkileyen tarafları olduğu gözlemlenmiş ve bu yapıların yeniden düzenlendiği *normalizasyon* adı verilen bir süreç gerçekleştirilmiştir. Normalizasyon adımlarının ve ek materyalde sunulmuş olan normalize edilmiş WALS verisinin de araştırmacılar için kaynak oluşturması hedeflenmektedir.

1.5. Tezin Araştırma Soruları

Yukarıda özetlenmeye çalışılan çerçevede, bu tez çalışmasında şu sorulara yanıt aranmaya çalışılmaktadır:

- i. WALS için önerilen çözümleme yöntemleri, bilgisayar desteği kullanılarak sistematik, hızlı ve kolay bir hale getirilebilir mi?
- ii. WALS verisinde karşılaştırmalı dil incelemeleri için alanyazında önerilmiş olan üç yöntem, WALS örneklemi üzerinde ne gibi betimsel sonuçlar vermektedir?
- iii. WALS verisi, dillerin aile/akraba olması ile yapısal benzerlikleri arasında doğrudan bir ilişki sunmakta mıdır?

1.6. Tezin Kapsamı ve Sınırlılıkları

Tezin kapsamını ve sınırlılıklarını belirlerken örneklem belirleme, çözümleme yöntemi seçme gibi metodolojik sınırlılıklar ve incelemenin yapıldığı betimsel verinin kuramsal yönüne gönderimde bulunan kuramsal sınırlılıklar olarak iki ayrı başlık içerisinde ele alınabilir.

1.6.1. Metodolojik Sınırlılıklar

Bu başlık altında değinilecek ilk konu örneklem seçimi üzerinedir. Çalışmamızın bir tipoloji çalışması olduğu göz önünde bulundurulursa, araştırma evreninin dünya üzerinde yaşayan tüm diller olduğu söylenebilir. Ancak “Yaşayan kaç dil vardır?” sorusuna genel geçer bir yanıt bulmak, araştırmacıların kuramsal önvarsayımlarının çeşitliliği nedeniyle neredeyse olanaksızdır (Anderson, 2010). Örneğin; pek çok çalışma çerçevesinde kabul gören bir kaynak olan *Ethnologue*’un (Eberhard vd., 2021) güncel verilerinde bu sayı 7139 olarak paylaşılmıştır. Diller hakkında yapısal betimlemeleri

derleyen WALS'in içeriğinde ise güncel sürüm itibarıyla deęişikeler de dahil 2662 dil bulunmaktadır.²

Bununla birlikte, WALS'te bulunan her dil için eşit sayıda bilgi bulunmadığından yüksek miktarda da eksik veri bulunmaktadır. WALS editör takımı, bu sorunu göz önünde tutarak 100 ve 200 dilden oluşan iki adet örneklem oluşturulduğunu ve yazarların bu örneklemdeki diller hakkında eksik veri bırakmamaları yönünde motive edildiğini belirtmiştir (Comrie vd., 2013).³ Bu doğrultuda, çalışmamızın örneklemine bu söz konusu iki örneklem oluşturmaktadır.

Benzer bir sınırlılık ise dillerin değerlendirildiği özellikler için geçerlidir. WALS üzerinde 11 ulamda 192 özellik bulunmaktadır, ancak bunların tümü doğrudan ve yalnızca dilbilgisiyle ilişkili değildir. Bu doğrultuda *İşaret Dilleri* ulamındaki tüm özellikler (örneklemdaki hiçbir dil işaret dili olmadığı için), yazı dizgesi ve *dildışı/dilötesi* (para-linguistic) verilerin bulunduğu *Diğer* ulamındaki tüm özellikler olmak üzere toplam 4 özellik çalışma dışında bırakılmıştır, örneklemdaki diller kalan 188 özellik çerçevesinde incelenmiştir. Son durumda, bu tez çalışmasında incelenen veritabanı, 199 dil ve 188 özellikten oluşmaktadır (eksik veri analizi hariç).

² WALS *Giriş* (Introduction) bölümünde (Comrie vd., 2013), sistemde 2676 dile ait veri bulunduğu bilgisi eski sürümden kalmıştır. Günümüzde yayında olan mevcut sürümde, belirtildiği üzere 2662 dile ait veri görünmektedir. Bu bilgi hatası, tez çalışması sürecinde WALS editör takımına iletilmiştir. Bkz. <https://github.com/clldf-datasets/wals/pull/91>

³ WALS'in 100 dillik örneklemini (<https://wals.info/languoid/samples/100>) 99, 200 dillik örneklemini (<https://wals.info/languoid/samples/200>) ise 199 dil içermektedir. Bu durum, WALS takımına tez çalışması sürecinde e-posta yoluyla bildirilmiş olup daha önce de bildirildiği (bkz. <https://github.com/clldf-datasets/wals/issues/72>), dolayısıyla bilgileri dahilinde olduğu ve gelecek güncellemede düzeltileceği yönünde dönüt alınmıştır.

Sözü edilen bu örneklem üzerinde yapılacak hesaplamalı tipoloji incelemeleri için ilerleyen bölümlerde değinilen kullanılmış farklı yöntemler bulunmaktadır. Sözü edilen çalışmaların ortak yanı, kullandıkları yöntemlerin aynı zamanda örneğin görüntü işleme veya herhangi bir niceliksel veri gibi çeşitli evrenler için de kullanılan yöntemler olması ve ilgili çalışmalarda dil verisi üzerine kullanılmış olmasıdır. Ancak çalışmamızda, seçilen yöntemler (Comrie, 2010; Orgun, 2011; Uzun, 2012a) yalnızca tipoloji verisi için önerilmiş dile özgü yöntemlerdir. Dahası, alanyazında daha önce kısıtlı örneklemeler üzerinde sınanmış bu yöntemler henüz hesaplamalı olarak işlenmediğinden ilgili yöntemler ile yapılan analizler elle yürütülmektedir. Bu sebeplerle, daha bütüncül veriler üzerinde, daha hızlı ve sistematik kullanımının sağlanması amacıyla adı geçen yöntemler tercih edilmiştir.

Bununla birlikte, WALS veri yapısının Uzun'un (2012a) önerdiği yöntemlerden karşılıklı benzemezlik çözümleme önerisini gerçekleştirmeyi zorlaştıran ve ayrıntıları ilgili bölümlerde açıklanan bazı yönleri bulunmaktadır. Buna, bu çalışmada önerilmiş olan normalizasyon süreci ve 0 değeri atama da farklı bir boyut kazandırmaktadır. Bu nedenle, karşılıklı benzemezlik analizinin eksiksiz olarak gerçekleştirilmesi veri yapısından kaynaklı nedenlerle ileri çalışmalara bırakılmak üzere kapsam dışı tutulmuştur.

Son olarak, bu çalışmada yürütülen analiz süreci, burada kullanılmayan farklı yöntemlerle de tekrarlanabilir ve sonuçlar arasında karşılaştırmalar yapılabilir.

1.6.2. Kuramsal Sınırlılıklar

WALS herhangi bir dilbilgisi kuramını benimsememiş, kuram bağımsız bir veritabanıdır. Veriler yalnızca farklı yazarların farklı diller için yaptıkları incelemelerden oluşan *betimleyici dilbilgisi* (descriptive grammar) kaynaklarından derlenerek

oluşturulmaktadır.⁴ Dolayısıyla, ilgili kaynakların güncelliğini yitirmesi gibi nedenlerle, bilimsel ilerlemenin de bir gereği olarak, zaman içinde bazı verilerin güncellenmesi gerekebilmektedir. WALS sisteminde kullanıcıların geribildirim yaparak düzenleme önermesi mümkündür ve bu geribildirimler doğrultusunda yapılmış olan tüm iyileştirmeler de web sayfası üzerinde sergilenmektedir.⁵ Ancak kuşkusuz ki mevcut durumdaki tüm eksik ve/veya hatalı bilgilerin tespit edilip düzeltilmesini sağlamak bu çalışmanın amacını ve süresini aşacak bir uğraş olacaktır. Dolayısıyla, yapılan çözümleme, veritabanındaki tüm verilerin tümünün gerçekte örtüştüğü varsayılarak gerçekleştirilmiştir.

1.7. Tezin Yapılanışı

Tezin bir sonraki bölümünde öncelikle “geleneksel” olarak nitelendirilebilecek dil tipolojisi çalışmalarının genel tarihçesine değinilip ardından günümüz olanaklarına, WALS’e ve WALS tabanlı tipoloji çalışmalarına yer verilecektir. Tezin üçüncü bölümünde, çalışmada kullanılan verinin önışlemlenmesi ve çözümleme yönteminin ayrıntıları sunulacaktır. Tezin dördüncü bölümünde, kullanılan üç yöntemin çözümleme bulguları ayrı ayrı sunulacak ve son bölümde, elde edilen bulguların değerlendirilmesi yapılacaktır.

⁴ İlgili kaynakların tam listesi için; <https://wals.info/refdb>

⁵ Bkz. <https://wals.info/changes>

2. KURAMSAL ARTALAN

2.1. Giriş: Tipoloji Neden Gerekli?

Dili bir ses/işaret ve anlam birlikteliği olarak kabul etmek, aynı zamanda her bir dilsel birimin bir anlam/işlev kodlama görevi olduğunu söylemektir. Bununla birlikte, bu birimlerin işleyişlerinin dilden dile çeşitlilik gösterdiği gerçeğiye çok küçük bir veri incelendiğinde bile gözlemlenebilmektedir. Sözgelimi, bir kişi annesinin pilav pişirdiğine ilişkin gerçekliğe ait bir bilgiyi Türkçede, İngilizcede ve Nias dilinde aşağıdaki gibi farklı biçimlerde kodlamak durumundadır:

(1) a. *Türkçe*

Annem pilav pişirdi.

b. *İngilizce*

My mother cooked rice.

ben.1TK.TAM anne pişir.GEÇ pilav.BEL

‘Annem pilav pişirdi.’

c. *Nias⁶*

I-rino vakhe ina-gu

3TK.REALIS-pişir BEL.pilav anne-1TK.İYE

‘Annem pilav pişirdi.’

Üç farklı dilden alınan bu veri, bu üç dilin aynı önermesel anlamı kodlarken bazı benzer ve farklı özellikler sergilediğini göstermektedir. Burada akla ilk gelebilecek fark, ortak kavramlara farklı sözcüklerle gönderimde bulunulması olabilir. Nitekim bu doğrudur, ancak diller arasındaki benzerlik ve farklılıklar bunun ötesinde de incelenebilir ve dilbilgisinin çeşitli düzlemlerinde de kendini göstermektedir. Sözgelimi, (1)’de

⁶ Bu dile ait veri Dryer (2013)’ten alınmıştır.

sunulan tümcelerin hepsinde (Ö)zne, (N)esne ve (E)ylem görevinde bulunan bir birim bulunmakta, ancak bu birimlerin dizilişi Türkçede ÖNE, İngilizcede ÖEN, Nias dilinde ise ENÖ biçimindedir. Bu noktada, (1a)'daki Türkçe verinin (2)'de verilen diğer tüm dizilişlerinin, Türkçe *anadili konuşurları* (native speaker) tarafından *düzgün* (well-formed) ve *kabuledilebilir* (acceptable) bulunacağını, ancak aynı durumun İngilizce veri için geçerli olmayacağını da belirtmek gerekmektedir.

(2)	<i>Türkçe</i>	<i>İngilizce</i>
a.	Annem pilav pişirdi.	* My mother rice cooked.
b.	Annem pişirdi pilav.	My mother cooked rice.
c.	Pişirdi annem pilav.	* Cooked my mother rice.
ç.	Pişirdi pilav annem.	* Cooked rice my mother.
d.	? Pilav annem pişirdi.	? Rice my mother cooked. ⁷
e.	Pilav pişirdi annem.	# Rice cooked my mother.

Gözlemlenebilecek başka bir nokta da her üç tümcedeki eylemin *çekimsel* (inflectional) süreçler için ek aldığı, ancak bu eklenmenin Türkçe ve İngilizce veride eylemin sonunda, Nias dilinde ise eylemin başında gerçekleştiğidir. Bu gözlem, gözlemciyi Türkçe ve İngilizce ile Nias dilini iki ayrı grup olarak görme yanılgısına düşürmemelidir, çünkü Türkçe ve İngilizce veride özne konumlarındaki adların çekimlenişine bakıldığında bu iki dilin de birbirinden farklılaştığını görmek mümkündür. Öyle ki Türkçede adın önünde bulunan ve TAM yüklenen 1TK adlı, yani *benim*,

⁷ Yüzey görünümünde bir *ilgi tümceciği* (relative clause) ile açıklanan bir ad görünmektedir ve yapısal olarak düzgündür ancak bu, diğerleriyle aynı sözdizimsel yapı değildir. Ayrıca İngilizcede nesne konumundaki belirtisiz ad öbeklerinin tümce başına taşınarak konulaştırılmasıyla ilgili bir tartışma için Ward ve Prince (1991) incelenebilir. Her iki durumda da ilgili yapılar a örnekleriyle aynı olmadığından ? imi kullanılmıştır.

seçimlikken İngilizcede zorunludur. Yine, Türkçede *anne* adı, üzerine sessel yükü olan bir iyelik yüklenmişken, başka bir deyişle önündeki adıyla uyuma girmişken, İngilizcede böyle bir durum söz konusu değildir. Ayrıca bu veriye (3)'teki gibi başka bir İngilizce tümce eklendiğinde eylem üzerindeki eklenme davranışının her koşulda Türkçe ve İngilizce için aynı gerçekleşmediği görülebilecektir. Öyle ki GEÇ çekimi sırasında İngilizce eylemlerin kimi *eklenme* (affixation) sürecinden geçerken, kimileri de (3)'teki gibi başka bir biçime dönüşmektedir (*go > went*):

(3) *İngilizce*

My	mother	went	to	Tokat.
ben.1TK.TAM	anne	git.GEÇ	İLG	Tokat
'Annem Tokat'a gitti.'				

Örnekleri ve örnekler hakkında yapılabilecek gözlemleri artırabilmek mümkünse de üç farklı dilden sunulan bu küçük veri bile, araştırmacıları diller arasında dilbilgisi özellikleri çerçevesinde bazı benzerlikler ve farklılıklar olduğu ve bunların karşılaştırmalı yöntemle incelenebileceği çıkarımına ulaştırmada yeterli görünmektedir. İşte, dil tipolojisi veya dilbilimsel tipoloji olarak adlandırılan dilbilim alanının temel araştırma konusu da tam olarak dillerdeki bu benzerlik ve farklılıklar çerçevesinde yapılabilecek incelemelerdir.

Böyle bir uğraş, hiç kuşku yok ki tek tek özel dillerin yapıları hakkında bir bilgi birikimi oluşmasını sağlayacaktır. Oluşan bu alanyazın, hem kuramsal çalışmalar hem de çeviriden dil öğretimine pek çok uygulamalı dilbilim alanı için önemli bir yarar sunmaktadır. Dil tipolojisi çalışmalarının burada önemle belirtilmesi gereken başka bir hedef noktası da dilbilimcilere, sayıları 7000'den fazla olan özel dillerin çeşitliliğinin aslında nasıl birbirini bütünleyen bir yapı oluşturduğunu ve dolayısıyla tümevarımsal bir yol izleyerek *dil yetisine* (language faculty) nasıl ulaşılabileceğini gösterebilme gücüdür. Bu bölümün takip eden başlıklarında özetlenmeye çalışılan tarihsel birikimi ve geline

noktadaki gelecek hedeflerini bu çerçevede değerlendirmenin önemli olduğunu belirtmek gerekmektedir.

2.2. Dil Tipolojisinin Kısa Tarihi

Bu başlık altında dil tipolojisinin tarihsel gelişimine yer verilecek olup teknolojik olanaklar öncesinde sağlanan birikimin ve bugünün genel hatlarıyla özetlenmesi hedeflenmektedir.

2.2.1. Dilbilgisinden Karşılaştırmalı Dilbilgisine ve Tipolojiye

Her ne kadar çağdaş dilbilim tarihi Ferdinand de Saussure'ün çalışmalarıyla başlatılıyor olsa da insanlığın dile olan ilgisi mitsel anlatılara dek uzanan çok daha eski bir meraktır. Çin, Babil, Hint, Mısır ve İbrani uygarlıklarında dilin doğuşunu doğaüstü mitolojik varlıklarla ilişkilendiren pek çok mitolojik anlatı bulunmaktadır (Başkan, 2003, ss. 21–22).

Dile olan ilgi her zaman mitlerle kalmamış, bugünkü dilbilgisi çalışmalarının atası sayılabilecek ilk çalışmalar henüz milattan önceki yüzyıllarda başlamış görünmektedir. Bunların en bilinenlerinden biri Antik Yunan (Grek) uygarlığıdır. Felsefeyle tanınan Grekler, dili de felsefenin aracı olarak görüp onu daha yakından tanımak istemişlerdir. Sözcüklerin anlamlarının doğal mı yoksa uzlaşımsal mı olduğu gibi tartışmaların yapıldığı; ad, eylem, bağlaç vb. sözdizimsel ulamlara ait terimlerin ilk kez kullanıldığı dönem bu dönemdir (Kıran, 1996, ss. 28–29; Başkan, 2003, ss. 23–29).

Felsefe gibi din de dil üzerine düşünmek için önemli bir etken olmuştur tarihte. Hint uygarlığının kutsal metinleri olarak bilinen Veda metinleri üzerine çalışan Pāṇini, bu metinlerin yazıldığı Sanskrit dilini henüz MÖ 300'lü yıllardayken çağdaş betimleme yöntemlerine benzer bir biçimde betimlediği için “ilk betimlemeci dilbilimci” olarak

(François ve Ponsonnet, 2013) nitelendirilmektedir. Benzer biçimde Romalılar da Grekçe üzerine yapılmış çalışmaları kilisenin dili Latinceye uyarlayarak din güdümlü dil çalışmaları sürdürmüşlerdir. Müslüman düşünürler de Kuran'ın yazıldığı dil olan Arapça üzerine çalışmalar sürdürmüşler ve pek çok eser vermişlerdir (Kıran, 1996, ss. 29–30; Başkan, 2003, ss. 29–33).

Türk tarihinde bilinen en eski ve önemli çalışma ise Kaşgarlı Mahmut'a ait olan *Dîvanü Lugati't Türk*'tür. Aslında bir sözlük olan bu eser, Türklerin İslamiyet'i kabulünden sonraki dönemde sıkı ilişkiler kurdukları Araplara Türkçe öğretmek için Arapça yazılmıştır. Yalnızca bir sözcük listesi değil aynı zamanda bazı dilbilgisel bilgileri de içeren bu eser, Türkçe için önemli bir yere sahiptir. Yine Ali Şir Nivai'nin Çağatay Türkçesiyle Farsçayı karşılaştırdığı *Muhakemetü'l-Lugateyn* adlı eseri öncü çalışmalar içerisinde sayılmaktadır (Özgen ve Koşaner, 2020, ss. 49–50).

Çeşitli uygarlıklar kendi dillerini farklı amaçlar için bir araç olarak görüp çalışmış olsa da *karşılaştırmalı dilbilgisinin* (comparative grammar) ve sonrasında da dil tipolojisinin doğuşuna neden olan asıl gelişme dillerin birbirleriyle çeşitli nedenlerle karşılaşması ve bu dillerin dilbilgilerinin karşılaştırılmaya başlanmasıdır. Bu çerçevede ele alınabilecek birkaç tarihsel gelişme bulunmaktadır. Örneğin, Batılı misyonerlerin Sanskritçeyi keşfetmesinin ardından onun Latince, Grekçe, İngilizce, Almanca, Fransızca vb. Avrupa dilleriyle benzerliği araştırılmıştır ve bu da Bopp ve Grimm gibi araştırmacıların da katkısıyla sonraları Hint-Avrupa Dil Ailesi tezinin doğuşuna neden olmuştur (Kıran ve Eziler Kıran, 2018, ss. 49–50). Benzer şekilde, Rönesans ve Reform hareketlerinin ardından Avrupalı ulusların İncil'i kendi dillerinde anlama çabası onları kendi dillerine dönmeye ve bu yönde karşılaştırmalı çalışmalar yapmaya itmiştir (Özgen ve Koşaner, 2020, s. 45).

Bütün bu gelişmeler, çeşitli amaçlar için bir araç olarak görülen dilin yavaş yavaş amaç konumuna taşınmasını sağlamıştır. Bundan sonraki dönemde dile bakış açısı daha pozitivist bir biçim alacaktır.

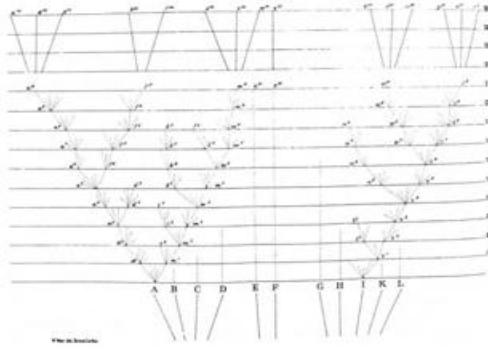
2.2.2. Schleicher ve Soyağacı Kuramı

Hiç kuşkusuz evrim kuramları tarihinde Darwin'in (1859) yeri ve önemi büyüktür. Onun *Türlerin Kökeni* (On the Origin of Species) adlı yapıtında, canlıların evrim yolculuğuna ilişkin işaret ettiği bazı noktalar dil çalışmalarında da yer bulmuştur. Öyle ki Schleicher'in dilleri organik bir canlı olarak görmesi ve bunun üzerine geliştirdiği *Soyağacı Kuramı* (İng. Tree Theory, Alm. Stammbaumtheorie) da bu düşüncenin bir ürünüdür.

Schleicher (1861, akt. Başkan, 2003, ss. 55–59), sözcüklerin önceki biçimlerini takip ederek *kök dile* (proto-language) ulaşmaya çalıştığı *yeniden kurma* (reconstruction) adı verilen bir yöntemi benimsemiş ve bunu başta Hint-Avrupa dil ailesi olmak üzere pek çok dil ailesi için uygulamaya çalışmıştır. Dilleri hayvanlar ve bitkiler gibi canlı organizmalar olarak gören Schleicher, onların zaman içinde bozulduğunu söyleyerek zamanı diller için bir düşman olarak görmüştür. Öyle ki Hint-Avrupa dil ailesi için Sanskritçe kök dile en yakın, en bozulmamış dil olarak mükemmeldir, ancak değişime uğrayan diller onun sözleriyle “bozulmuş” ve “yozlaşmış”tır. Schleicher ayrıca dilleri kökler ve ekler üzerinden kurduğu üç sınıfa ayırmaktadır. Bu sınıflandırmada; Çince gibi sözcükleri ek almayan dilleri *ayrışık diller*, Türkçe gibi ek alan dilleri *yanaşık diller*, Arapça gibi köklerin ve eklerin kaynaştığı dilleri ise *kaynaşık diller* olarak adlandırmaktadır. Aynı zamanda felsefeyle de ilgilenen Schleicher, bu üç sınıfın üzerine Hegel'den öykünerek *tez*, *antitez* ve *sentez* kavramlarını yerleştirmiştir. Çince gibi sözcüklerin tek başına bulunduğu dillerde kökler tezlere karşılık gelmektedir. Türkçe gibi sözcüklerin ek aldığı dillerde kökler tezlere, onlara eklenen ekler antitezlere karşılık

gelmektedir. Arapça gibi köklerin ve eklerin kaynaştığı dillerde ise sözcükler senteze yani tez ve antitezin karışımına karşılık gelmektedir.

Schleicher, ayrıca Darwin'in canlılar ile ilgili olarak savladığı bazı düşünceleri dillere uyarlamıştır. Örneğin; Bask dilinin çevresindeki diğer dillere karşı direnerek yok olmamasını Darwin'in güçlü olanın hayatta kalması ilkesine bağlamıştır. Yine, Amerikan yerli dilleriyle ilgili de bir tespitte bulunan Schleicher, onların çok karmaşık oldukları için hayatta kalamadıklarını savlamıştır. Ona göre, bu diller ya karmaşık yapılarını terk edip bozulacaktı ya da yok olacaktı (Özgen ve Koşaner, 2020, s. 48).



Darwin'in türlerin soy ilişkisini gösterdiği bir çizimi



Schleicher'in dillerin soy ilişkisini gösterdiği bir çizimi

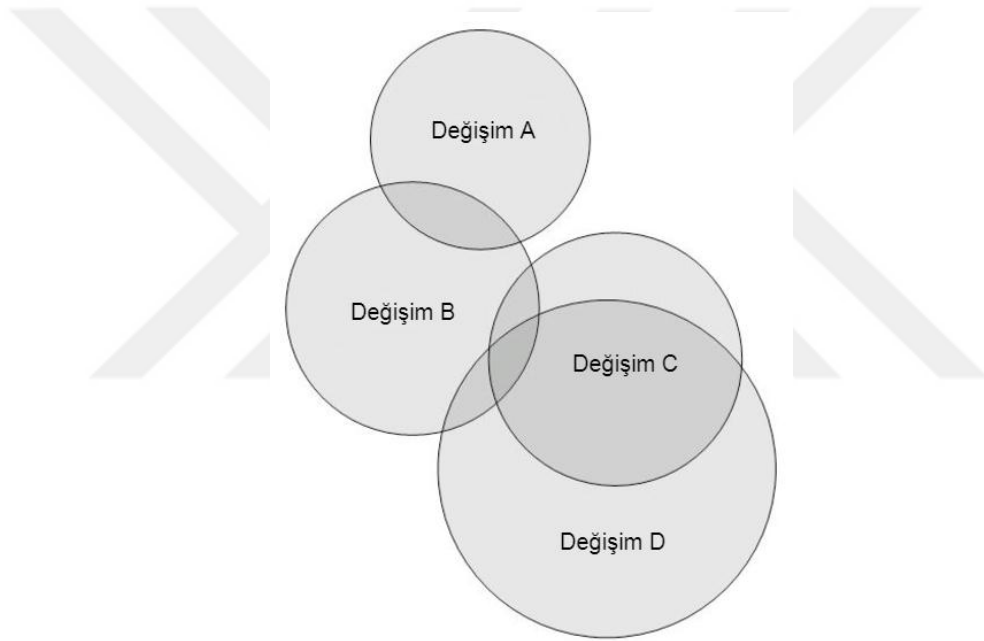
Şekil 1. Darwin ve Schleicher'in soy ilişkilerini göstermek için kullandıkları ağaç şemaları (akt. Richards, 2002).⁸

Görüldüğü üzere, Schleicher dillerin bir kök dilden çıktığını ve bu doğrultuda her dilin bir soyağacının olduğunu ileri sürmüş ve dillerarası benzerlikleri de bu soy ilişkileri çerçevesinde değerlendirmiştir.

⁸ Darwin'in ve Schleicher'in öncülük ettiği bu teknik, bugün *filogenetik* (phylogenetics) ve *filolinguistik* (phylolinguistics) adlarıyla modern anlamda yürütülmektedir. Bu kapsamda ilk çalışma Gray ve Jordan'ın (2000) Avustralya dilleri için yaptığı analizdir (yöntem ile ilgili ayrıntılı bilgi için bkz. Ben Hamed, 2015).

2.2.3. Schmidt ve Dalga Kuramı

Schleicher'in *Soyağacı Kuramı*na bir alternatif ortaya atan Schmidt'in (1872) *Dalga Kuramı* (İng. Wave Theory, Alm. Wellentheorie) dillerdeki değişimi suya atılan bir taş ve onun yarattığı dalgalar üzerinden bir benzetme ile açıklar. Nasıl ki suya bir taş atıldığında, sudaki hareketlilik taşın atıldığı ana merkezden uzaklara dalga dalga yayılıyorsa dillerdeki değişimde aynı şekilde dalga dalga başka dilleri etkileyebilir. Bu, toplumsal ilişkiler sayesinde olmaktadır ve bu şekilde diller birbirleriyle kaynaşabilmektedir (akt. François, 2014).



Şekil 2. Schmidt'in önerdiği Dalga Modeli için temsili çizim (Daval-Markussen, 2011'den Türkçeye çevrilmiştir.)

Aslında Schmidt'in önerdiği modelin tipolojik incelemeler için sunduğu en önemli sav, benzerliğin aile/soy bağıyla ilgili olmadığıdır. Öyle ki bir dil kendi soyundaki dillerin taşımadığı bir özelliği etkileşimler sonucunda başka bir dilden alabilir, dahası bunu soyundan başka bir dile veya dillere de bulaştırabilirdi. Dolayısıyla, bu modeli kabul eden bir araştırmacı dillerin bir soy ilişkisi içinde olmadığını, olsalar bile artık bu bağın izinin sürülemeyeceğini kabul etmiş demektir.

2.2.4. Greenberg ve Evrenceleri

Greenberg o güne dek süregelen tipoloji çalışmalarına evrenselleştirilebilir bir gömlek giydirmeye çalışmış olması nedeniyle, tipoloji araştırmaları için önemli bir dönüm noktasını oluşturmaktadır. Bu noktada, Greenberg'ün çalışmasına geçmeden önce onu daha iyi anlayabilmek için birkaç terim üzerinde durmak gerekmektedir.

Dil tipolojisi için dillerarası sistematik örüntülerin çalışıldığı alan da denilebilir. Tipolojinin bu tanımı aslında dillerarası bir genelleştirme yapmaktır. Bu genelleştirilebilir örüntülere *evrence* (universal) adı verilir. Tipolojik evrencelerin en bilinen örnekleri *sezdirimsel evrenceler* (implicational universal). “Bir dilde gösterme sözcükleri baş olan adlardan sonra geliyorsa, ilgi tümcecikleri de sonra gelir” gibi bir sezdirimsel evrence örneği, bir a özelliğinin olduğu her dilde b özelliğinin de olduğunu söylemekte ve dillerin bu yolla sınıflandırılabilceğini önermektedir (Croft, 2002).

Evrenceler söz konusu olduğunda dile getirilebilecek tek evrence türü sezdirimsel evrenceler değildir. Sözgelimi; en az bir ağız ünlüsü tüm dillerde bulunur, ancak bu özellik o dildeki başka bir özelliğin öngörülmesini sağlamaz. Bu da sezdirimsel evrencenin tam tersi olarak *sezdirimsel olmayan evrenceler* (non-implicational universal) biçiminde adlandırılmaktadır. Bunun dışında; sezdirilen özelliğin kesin olarak bulunduğu dil kümeleri *mutlak evrenceler* (absolute universal) olarak adlandırılırken, bu olasılığın kesin değil ancak büyük çoğunlukla olduğu örüntüler için ise *eğilim* (tendency) denmektedir. Son olarak; dilbilgisinin herhangi bir bileşenine ait ulamlara gönderimde bulunan *tözsel evrenceler* (substantive universal) ve bunların biçimlenme kurallarına karşılık gelen *biçimsel evrenceler* (formal universal) de bulunmaktadır (Comrie, 1989, ss. 15–23).

Greenberg (1966), aralarında Türkçenin de bulunduğu farklı ailelerden 30 dil üzerinde çalışmış ve toplamda 45 adet evrence tanımlamıştır. Örneğin; Greenberg'ün 1 numaralı evrencesi, bir dilde bildirme tümcelerinin öznesinin ve nesnesinin yalın

durumda olması durumunda bu tmcede baskın diziliřin znenin nesneden nce gelmesi ynnde olduėunu belirtmektedir.

Greenberg evrencelerini altı farklı ulama ayırmaktadır. Bunlardan ilki, her dilde olan ancak dilleri tanımlama amacı olmayan evrenceler olan *kısıtsız evrenceler*dir (unrestricted universals). Yukarıda da değinilen sezdirimsel evrenceler iki yapısal zellik arasındaki korelasyona baėlı evrencelerdir. c evrence tr *kısıtlı eřdeėerlik evrenceleri*dir (restricted equivalence universals), ki bu da evrensel olarak her dilde bulunmayan zellikler arası korelatif iliřkilere gnderimde bulunmaktadır. Bir dilde bir a zelliėi bulunuyorsa b zelliėinin de bulunması a yerine c zelliėi bulunan dillere gre daha olasıysa bu durumda istatistiksel baėıntılardan sz edilmektedir. Son olarak ise yeterli byklkteki veri zerinde dillerin gsterdiėi evrensel istatistikler dillere iliřkin evrensel sıklık daėılımlarını sunmaktadır (Greenberg vd., 1963, ss. xix–xxi).

Greenberg’n alıřması evrenselleřtirilebilir bir yapı sunması nedeniyle, aėdař dilbilim aısından nemlidir.

2.2.5. Chomsky ve Evrensel Dilbilgisi

Chomsky, 1957’de yayımlanan *Szdizimsel Yapılar* (Syntactic Structures) adlı eseriyle bugn *retici Dilbilgisi* (Generative Grammar) olarak anılan kuramının ilk adımlarını atmıř ve bunu zaman iinde gncellemiřtir. Bu bařlıkta, kuramı tm ayrıntılarıyla ele almak yerine, yalnızca 1981’de *Ynetme ve Baėlama Kuramı* (Government and Binding Theory) ile kurama dahil edilen ve tipolojik arařtırmalarla iliřkilendirilebilecek *Evrensel Dilbilgisi* (Universal Grammar), *ilke* (principle) ve *deėiřtirgen* (parameter) kavramları zerinde durulacaktır.

Kurama gre, *Evrensel Dilbilgisi* (ED) trmzn tamamı iin ortak ve her bir birey iin deneyim ncesinde de var olan ilkeler btnnden oluřan bir dizgedir (Chomsky, 1981, s. 7). Daha net bir anlatımla, her bir insan genetik olarak ED ile

doğmaktadır ve bu dil yetisini oluşturan temel mekanizmadır. Haegeman (1994, s. 15), ED'nin yapısını oluşturan ilke ve parametreler hakkında şöyle bir genel çerçeve çizmiştir:

(4) a. ED, dilden dile değişmeyen bir dizi mutlak evrenceler, kavramlar ve ilkeler içerir.

b. ED tarafından tam olarak belirlenmeyen, ancak dillerarası değişiklik gösteren dile özgü özellikler bulunur. ED bu gibi ögeler için bir dizi seçenek sunmaktadır.

Şimdi, (4)'teki tanımlar çerçevesinde (1)'deki veri yeniden ele alınacak olursa; Türkçe, İngilizce ve Nias dilinden alınmış tümcelerinin hepsinde ED'ye göre eylemin nesne ile ikili olarak dallanacağı (İkililik İlkesi için bkz. Haegeman, 1994, ss. 138–144) bir ilkeyken (6a), dillerde eylem ile nesnenin eylem-nesne (baş-ilk) ya da nesne-eylem (baş-son) biçimlerindeki dizilişlerden birinin görülmesi (bkz. Haegeman, 1994, ss. 95–97) de parametrik bir durumdur (6b). Bu çerçevede üç farklı birimin kaç farklı biçimde sıralanabileceği düşünülecek olursa, altı olası diziliş var gibi görünmektedir: ÖNE, ÖEN, EÖN, ENÖ, NEÖ, NÖE. Ancak bu dizilişlerin dünya dillerindeki dağılımına bakıldığında ise gerçekten de altı dizilişin de örneklerinin varlığı görülmekle birlikte, bazı dizilişlerin dağılımlarının diğerlerine göre daha fazla olduğu görülebilmektedir (Dryer, 2013). Bu durum da kuramın öbek yapı ile ilgili savlarına koşutluk göstermektedir, ancak konunun akışını bozmamak için öbek yapı ile ilgili tartışmayı burada sonlandırıyoruz.

Görüldüğü üzere, Chomsky'nin görüşü de Greenberg gibi dillere evrenselleştirilebilir bir inceleme alanı çizmiştir. Ancak iki görüşün arasında belli farklar bulunmaktadır. İlk olarak, Chomsky Greenberg gibi bir veritabanı üzerinden somut genellemeler yapmaz. Bunun yerine, daha soyut ve kuramsal bir yapı çizerek dillerden gelen örneklerin bu yapıya uygunluğu üzerinden ilerler. Dolayısıyla, Greenberg tümevarımcı, Chomsky ise tümdengelimci bir yöntem izlemektedir. Aynı şekilde, Greenberg'ün çalışmaları veri güdümlüken Chomsky'nin çalışmaları için ise böyle bir şey söz konusu değildir (Comrie, 1989, ss. 1–15).

2.2.6. Hesaplamalı Tipoloji

Hesaplamalı dilbilim (computational linguistics, HD) en genel anlamıyla, doğal dillerin makineler tarafından anlamlandırılmasının ve üretilmesinin amaçlandığı çalışma alanıdır. Bu çalışmalar, dil dizgesinin daha iyi anlaşılması amacıyla kuramsal tabanlı olarak “bilimsel” bir amaç güttüğü gibi, *doğal dil işleme* (natural language processing) çözümleri başlığı altında değerlendirilen teknik uygulamalarla “mühendislik” amacı da gütmektedir (Grishman, 1994, ss. 4–6).

HD’nin matematik, bilişim vb. başka disiplinlerle sıkı bir ilişki içinde olması, onun adlandırılması noktasında da sorun yaratmıştır. Öyle ki bu durum, insanları, yalnızca bir araç olarak kullanılan matematik veya bilgisayar gibi şeyleri HD’nin ayrılmaz bir parçası olarak görme yanılgısına düşürmüştür. Bu noktada, bundan sonrasında kavramsal bir karışıklık yaşanmaması için *computational linguistics* terimi için bugüne kadar yapılmış olan Türkçe karşılık önerilerine kısaca değinmek yerinde olacaktır.

Öncelikle belirtmek gerekir ki bu uzlaşımsızlığın temel nedeni, İngilizcedeki *to compute* eyleminin Türkçe söz varlığında tam bir karşılığının bulunmamasıdır, bu sebeple de söz konusu eylemden türeyen *computational* sıfatı için de farklı çeviri önerileri sunulmuştur. Örneğin; Bozşahin ve Zeyrek (2000) *ber-* kökünden türettikleri *berimsel* sıfatını ve *berimsel dilbilim* terimini, Sankur (2004) *bilişimsel dilbilimi* önermekte, yöntemsel olarak matematik ve istatistiğe yakın olması nedeniyle bazı çalışmalarda bizim de tercih ettiğimiz *hesaplamalı dilbilim*, sayıca daha fazla çalışmada ise *computation* sözcüğünün *bilgisayar* (computer) sözcüğünü çağrıştırması ve bu tarz çalışmalarda genellikle bilgisayarın bir araç olarak kullanılması nedeniyle bilgisayarlı dilbilim terimi tercih edilmektedir. HD çalışmalarında bilgisayarların bir araç olarak kullanılması HD’yi “bilgisayarla yapılan dilbilim” gibi hatalı bir anlamsal noktaya yerleştirmişse de daha önce de söylendiği gibi bilgisayarlar HD için yalnızca bir araçtır.

Teknolojinin gelişmesiyle birlikte bilgi üretimi ve paylaşımı hızlanmış, bu da bilimsel yayınların hızlanarak artmasına neden olmuştur. Bu gelişim kendisini dilbilim veya daha özgül olarak dilbilgisi çalışmalarında da göstermiş ve pek çok dilin dilbilgisine ait çok sayıda bilimsel bulgu elde edilmiştir. Artan veri sayısı insan gücüyle yapılan çalışmaları gittikçe güçleştirdiğinden, çalışmalar HD'nin bir çalışma kolu olarak bilgisayar ortamına aktarılmaya ve sistematize edilmeye çalışılmaktadır. *Hesaplamalı tipoloji* (computational typology) diyebileceğimiz bu uğraş, henüz gelişim aşamasında olsa da alanyazında farklı yaklaşımları ve yöntemleri benimseyen öncü çalışmalar mevcuttur.

Sözü edilen hesaplamalı tipoloji çalışmalarından biri olan, *LinGO Grammar Matrix* (Bender vd. 2002; Bender vd., 2010) adlı bilgisayar uygulaması bir sormaca yardımıyla araştırmacılardan toplanan verilerle dillerin dilbilgisi matrislerini *Baş Güdümlü Öbek Yapı* (Pollard ve Sag, 1994) kuramı çerçevesinde oluşturmakta ve matrisler arasında karşılaştırma yapma olanağı sunmaktadır. Başka bir çalışma olan *Online Database of Interlinear Text* (ODIN, Lewis, 2006; Lewis ve Xia, 2010; Xia vd., 2016) ise webden topladığı metinleri *satırarası açıklama* (interlinear glossing) yapıp farklı dillere ait veriler arasında karşılaştırma yapma olanağı sunan bir projedir.

Doğrudan tipoloji çalışmaları için yapılmamış olsa da tipoloji çalışmaları için de kullanışlı bir veri kaynağı olabilecek olan bazı dillerarası veri derleme çalışmaları da bu başlık altında ele alınabilir. Örneğin; ilk olarak baskı versiyonu Miller (1995) ve Fellbaum (1998) tarafından yayımlanan *WordNet* projesi *eşanlam kümeleri* (synset) oluşturularak bu kümeler aracılığıyla dillerdeki sözcüklerin anlam ilişkilerinin etiketlendiği ve günümüzde çevrim içi veya çevrim dışı olarak kullanılabilen dijital bir araçtır (<https://wordnet.princeton.edu/>). Aynı mimari kullanılarak Türkçe için de bir sözcük ağı *BalkaNet* projesi (Tufis vd., 2004) kapsamında yapılmış ve Türkçe verinin bu proje içerisinde yer alan Bulgarca, Çekçe, Yunanca, Romence ve Sırpça veriyle

karşılaştırma olanağı doğmuştur. WordNet'e benzer başka bir mimari ise *ConceptNet* (Speer vd., 2017) projesidir. Dillerin ortak kavramları hangi sözcüklerle karşıladığı ve bu sözcüklerin anlambilimsel ve biçimbilimsel ilişkileri konusunda veri sunan bu dijital araç da çevrim içi olarak kullanılabilir (http://conceptnet.io/).

Yukarıdaki örneklerin dışında yapı ve sözlükçe ile ilgili çeşitli bilgilerin dillerarası olarak etiketlendiği veritabanları da şemsiye bir terim olarak CLLD⁹ (Cross-Linguistic Linked Data, Dillerarası Bağlantılı Veri) adıyla anılmaktadır. Aslında bir *çeşitlilik dilbilim* (diversity linguistics) projesi olan CLLD, *Max Planck Society* tarafından fonlanan ve 2013 yılında başlayıp 2016 yılında biten ve bu kapsamda çalışmamızın temel veri kaynağını oluşturan WALS dahil pek çok veritabanının oluşmasını sağlayan bir proje olsa da proje kapsamında oturtulan CLDF (Cross-Linguistic Data Formats, Dillerarası Veri Formatları) standardı (Forkel vd., 2018) sayesinde sürekli olarak geliştirilebilir bir alan olarak ortaya çıkmıştır. Bugün itibarıyla CLLD kapsamında; dünya dillerindeki ödünçleme sözcükleri derlemeyi amaçlayan *The World Loanword Database* (WOLD, Haspelmath ve Tadmor, 2009), benzer biçimde ödünçlenen ekleri derlemeyi amaçlayan *AfBo* (Seifart, 2020) ve sesbilimsel bir derleme olan *PHOIBLE* (Moran ve McCloy, 2019) gibi 22 adet betimsel veritabanı bulunmaktadır.

Buraya kadar tipolojinin dününü ve bugününü ele aldık. Bir sonraki başlıkta bir CLLD veritabanı olan ve bu tez çalışmasının verisini oluşturan WALS ve onunla ilgili çalışmalar üzerinde durulacaktır.

2.3. The World Atlas of Language Structures (WALS)

The World Atlas of Language Structures (Dünya Dil Yapıları Atlası, kısaca WALS), dillere ilişkin yapısal özelliklerin, dillerin konuşulduğu bölgelerin coğrafi

⁹ CLLD ile ilgili ayrıntılı bilgi için; <https://clld.org/>

konumlarıyla birlikte derlenip sunulduğu bir elektronik veritabanıdır. İlk olarak CD-ROM ekli bir kitap olarak basılmıştır (bkz. Haspelmath vd., 2005), ancak 2008'in Nisan ayından itibaren <https://wals.info/> adresi üzerinden çevrim içi bir web servisi olarak kullanımdadır (Dryer ve Haspelmath, 2013). WALS içeriği 55 kişilik bir yazar takımı¹⁰ tarafından ilgili diller üzerine yazılmış dilbilgisi kaynaklarından¹¹ derlenerek oluşturulmaktadır.

Bu başlık altında sırasıyla; WALS verisinin nasıl düzenlendiği ve veritabanının nasıl kullanıldığı, WALS verisinin bir dil tipolojisi çözümlemesi için nasıl kullanılabileceği yönünde alanyazında yer alan ve çalışmamızda izlediğimiz yöntemlerin kuramsal arka planını da oluşturan önerilere, son olarak da WALS üzerinde yapılmış *hesaplamalı* (computational) çalışmalara alt başlıklar biçiminde yer verilecektir.

2.2.7. WALS'in Düzenlenişi ve Kullanımı

WALS'in, üzerine inşa edildiği üç ana ögesinin olduğu söylenebilir: Bunlar diller, özellikler ile değerler ve haritalardır. Takip eden altbaşlıklarda bu üç öge WALS'in *Giriş* (Introduction) bölümündeki (Comrie vd., 2013) bilgiler doğrultusunda ele alınacaktır.

2.2.7.1. Diller

WALS'in mevcut güncel sürümünde 2662 dil kaydı bulunmaktadır. Eğer bir kayıt, bir dilin değişkesiyle ilgiliyse o dilin adında ilgili bilgi yay ayrıç içinde belirtilmektedir. Örneğin; bir Nijer-Kongo dili olan Zulunun kuzey ve güney değişkeleri *Zulu (Northern)* ve *Zulu (Southern)* olarak adlandırılmış ve bu gibi kayıtlar ayrı ayrı etiketlenmiştir.

¹⁰ Yazar listesine <https://wals.info/author> bağlantısı üzerinden erişilebilmektedir.

¹¹ Kaynaklara <https://wals.info/refdb> bağlantısı üzerinden erişilebilmektedir.

Diller ile ilgili özel bir durum da işaret dilleriyle ilgilidir. Veritabanı içinde 40 adet işaret dili bulunmaktadır. Ancak bu diller yalnızca -bir sonraki başlıkta ayrıntılarına değinilecek olan- *İşaret Dilleri* özellik alanındaki iki özellik (bkz. 139A ve 140A) çerçevesinde değerlendirilebildiği için çözümleme dışında bırakılmıştır.

Diller ile ilgili bu özel durumlara değindikten sonra, veritabanında bu dillere ilişkin ne tür bilgilerin olduğuna yakından bakılabilir. WALS'in her bir dil için kullandığı benzersiz bir kod bulunmaktadır. Bunlar çoğunlukla o dilin adındaki ilk üç harftir. Bunun yanında *Ethnologue* gibi veritabanlarının da kullandığı ve uluslararası bir standart olan ISO 639-3 kodları gibi farklı kod standartları da diller üzerine etiketlenmiştir. Bu iki kod bazı diller için aynı olsa da bazı diller için farklılaşmaktadır. Örneğin *Türkçe* için WALS kodu da ISO 639-3 kodu da *tur* iken, *Almanca* için WALS kodu *ger*, ISO 639-3 kodu ise *deu* olarak belirlenmiştir. Her iki kod standardının aynı anda etiketlenmesi, verinin farklı ilişkisel veritabanları üzerinden yapılabilecek başka çözümlemeler için de kullanılabilmesine olanak sağlamaktadır. Bunlara ek olarak, dillerin genetik aile bağları (örn. *Türkçe* için *Altay dil ailesinin Türk dilleri grubu*), baskın olarak konuşuldukları bölgelere ait coğrafi konum, ülke ve bölge adları (örn. *Türkçe* için *Türkiye*) gibi bilgiler de dil tipolojisi çalışmaları kapsamında değerlendirilebilecek türdendir.

Son olarak her dilin kendine ait sayfasında, o dil için WALS ve farklı bazı veritabanlarında kullanılan adlar ve kod karşılıkları, dilin yoğunlukla kullanıldığı bölgenin harita üzerinde konumu, genetik aile ilişkileri gibi ek bilgiler yanında, o dil için kaynaklardan toplanan özellik değerleri ve yararlanılan kaynakların bilgileri bulunmaktadır. *Şekil 3*'te *Türkçe* ile ilgili sayfanın ekran görüntüsü paylaşılmaktadır.

THE WORLD ATLAS OF LANGUAGE STRUCTURES ONLINE

Home Features Chapters Languages References Authors

Family: Altaic / Genus: Turkic

Glottocode: nuch1301 ISO 639-3: tur

Language Turkish

WALS code: tur

Showing 1 to 154 of 154 entries

Fid	Value	Feature	Source	Area
1A	Average	Consonant Inventories	Zimmer and Orgun 1999	Phonology
2A	Large (7-14)	Vowel Quality Inventories	Zimmer and Orgun 1999	Phonology
3A	Average	Consonant-Vowel Ratio	Zimmer and Orgun 1999	Phonology
4A	In both plosives and fricatives	Voicing in Plosives and Fricatives	Zimmer and Orgun 1999	Phonology
5A	None missing in /p t k b d g/	Voicing and Gaps in Plosive Systems	Zimmer and Orgun 1999	Phonology
6A	None	Uvular Consonants	Zimmer and Orgun 1999	Phonology
7A	No glottalized consonants	Glottalized Consonants	Zimmer and Orgun 1999	Phonology
8A	/l/, no obstruent laterals	Lateral Consonants	Zimmer and Orgun 1999	Phonology
9A	No velar nasal	The Velar Nasal	Kornfilt 1997	Phonology
10A	Contrast absent	Vowel Nasalization	Zimmer and Orgun 1999; Kornfilt 1997	Phonology
11A	High and mid	Front Rounded Vowels	Zimmer and Orgun 1999	Phonology

Coordinates: WGS84 39°N, 35°E 39.00, 35.00

Spoken in: Turkey

Alternative names

Ruhlen: Turkish
Other: Osmanli
Ethnologue: Turkish

Sources

Aksu-Koç and Slobin 1986
A Psychological Account of the Development and Use of Evidentials in Turkish

Comrie 1998
Rethinking the typology of relative clauses

Dahl 1985
Tense and Aspect Systems
Info at Google Books

Şekil 3. WALS üzerinde Türkçeye ait sayfanın görüntüsü

2.2.7.2. Özellikler

WALS'in bel kemiği olarak nitelendirilebilecek en önemli ögesi olan özellikler; 20 adet *Sesbilim* (Phonology), 12 adet *Biçimbilim* (Morphology), 29 adet *Adcıl Ulaımlar* (Nominal Categories), 8 adet *Adcıl Sözdizim* (Nominal Syntax), 17 adet *Eylemcil Ulaımlar* (Verbal Categories), 56 adet *Sözcük Dizilişı* (Word Order), 26 adet *Basit Tümcecikler* (Simple Clauses), 7 adet *Karmaşıık Tümceler* (Complex Sentences), 13 adet *Sözlükçe* (Lexicon), 2 adet *İşaret Dilleri* (Sign Languages) ve 2 adet de *Diğer* (Other) olmak üzere 11 farklı *alanda* (area) toplamda 192 adettir. Bu özelliklerin hepsi kategorik değerlere sahip olmakla birlikte, değerlerin sayıları 2'den 28'e kadar çeşitlilik göstermektedir.

Buna ek olarak, değerlerin nitelikleriyle ilgili de bazı farklılıklar söz konusudur. Sözelimi, 10A *Ünlü Genizsilleşmesi* türünden bazı özellikler *var* veya *yok* gibi ya da 129A *El ve Kol* türünden özellikler de *aynı* veya *farklı* gibi *ikili* (binary) değerler alan 17 farklı özellik bulunmaktadır. *Bunun yanında*, ikiden fazla değeri olan özelliklerde ise iki

farklı özellik tipiyle karşılaşılmaktadır. Bunlardan ilki, örneğin *1A Ünsüz Envanteri* özelliğinde olduğu gibi *sıralı* (ordinal) veri tipleridir. Bu örnekte diller sahip oldukları ünsüz sayısına göre *Dar*, *Görece Dar*, *Ortalama*, *Görece Geniş* ve *Geniş* olmak üzere beş değerden birini almaktadır. Görüldüğü gibi bu beş değer arasında hiyerarşik bir ilişki bulunmaktadır. Veritabanında bu türden 18 farklı özellik bulunmaktadır. Son özellik tipiye yalnızca *kategorik* (categorical) olan özelliklerdir. Bunlara bir örnek olarak *81A Özne, Nesne ve Eylem Dizilişi* özelliği verilebilir. Bu özellikte dillerin gösterdiği baskın bir diziliş söz konusuysa bunun ne olduğu, yoksa da baskın bir diziliş olmadığı işaretlenmektedir. Dolayısıyla, diller şu yedi seçenekten biriyle etiketlenmek durumunda olup bu değerlerin kendi aralarında bir hiyerarşi söz konusu değildir: *ÖNE*, *ÖEN*, *EÖN*, *ENÖ*, *NEÖ*, *NÖE*, *Baskın diziliş yok*. Veritabanındaki 157 özellik de bu türdendir.

Özellikler, ilgili oldukları konuyla ilgili betimsel açıklamaların ve kuramsal tartışmaların yer aldığı bölümler içerisinde ele alınır.¹² Her özellik, ait olduğu bölümün numarası ve bu bölümdeki harita sayısına göre bir harf alarak kodlanmıştır. Örneğin; *10B Batı Afrika'da Genizsil Ünlüler* adlı özellik Ünlü Genizsilleşmesi (Hajek, 2013) başlıklı onuncu bölümde yer alan ikinci özellik olduğundan *10B* biçiminde kodlanmıştır.

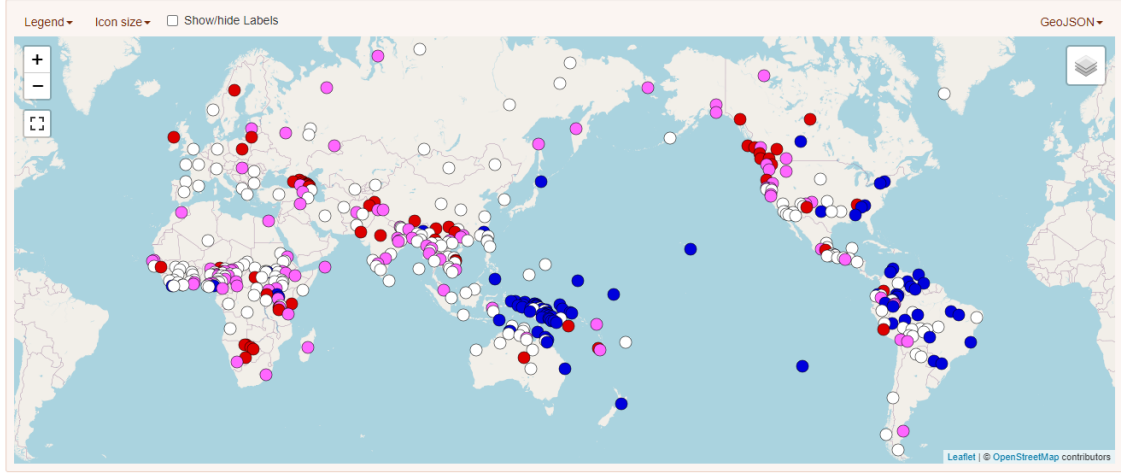
Özellikler ile ilgili ayrıntılı bilgiler *EK-1*'de sunulmuştur.

2.2.7.3. Haritalar

WALS'in önemli bir ögesi de onu bir “atlas” yapan haritalardır. Diller ile ilgili bölümde (2.3.1.1.), dilin konuşulduğu ana merkezin enlem ve boylam bilgilerinin de veritabanına işlendiğinden söz etmiştik. Bu bilgiler kullanılarak her bir özellik için *GeoJSON* haritaları oluşturulmuştur. Bu haritalar sayesinde, hem verilerin

¹² Bölümlere <https://wals.info/chapter> bağlantısı üzerinden erişilebilmektedir.

görselleştirilmesi kolaylaşmakta hem de tipolojik sınıflandırmalarda coğrafi *uzaklık* (distance) etkisi analiz edilebilmektedir.



Şekil 4. WALS haritaları örneği (1A)

2.2.8. WALS İçin Karşılaştırma Yöntemi Önerileri

WALS verisi üzerinden yapılabilecek tipolojik analizler için alanyazında önerilen farklı yöntem önerileri bulunmaktadır. Çalışmamızda, bunlardan Comrie (2010), Orgun (2011) ve Uzun (2012a) tercih edilmiş olup takip eden alt başlıklarda bu yöntem önerileri açıklanacaktır.

2.2.8.1. Tipiklik (Comrie, 2010)

Comrie (2010) bir dilin bir özellik için aldığı değerin dünya dilleri arasındaki dağılım oranı üzerinden bir analiz önermektedir. Eğer bir dil bir özellik çerçevesinde dünya dillerinin çoğunun aldığı değeri alıyorsa bu dil daha “tipik” bir dildir. Aynı şekilde dünya dilleri arasında daha az görülen bir özellik sergiliyorsa “daha az tipik” bir dildir. Yöntemin daha net anlaşılması amacıyla, *Tablo 1*’de Comrie’nin (2010) İngilizce ve Japonca’yı karşılaştırdığı araştırmadan bir örnek paylaşılacaktır.

Tablo 1. *Comrie'nin (2010) çözümleme örneği*

	1A	2A	...	ORTALAMA
<i>İngilizce</i>	Ortalama (32,32)	Geniş (32,5)	...	40,74
<i>Japonca</i>	Görece dar (21,49)	Ortalama (51,15)	...	40,84

(Uzun, 2012a'dan uyarlanmıştır.)

İngilizce 1A özelliği çerçevesinde değerlendirilen dillerin %32,32'si gibi *Ortalama* değeri, Japonca ise bu özellik çerçevesinde değerlendirilen dillerin %21,49'u gibi *Görece küçük* değeri almıştır ve bu durumda 1A için İngilizce Japoncadan daha tipik bir dildir. Aynı yöntemle 2A özelliği ele alındığında ise bu kez Japonca daha tipik bir görünüm sergilemektedir. Bu tüm özellikler için uygulanıp ortalama değer hesaplandığında Japoncanın İngilizceye göre %0,1 oranında daha tipik olduğu gözlemlenmektedir.

2.2.8.2. Tutarlılık (Orgun, 2011)

Orgun (2011) ise diller arasında oynanabilecek bir “oyun” önermektedir. Bu oyunda, eğer iki dil bir özellik çerçevesinde aynı değeri almışsa bu dillerin başka özellikler için de ortak değer alıp almadığına bakılmaktadır. Bu diller benzerliklerini sürdürdükleri her bir özellik için tutarlı davranışları nedeniyle bir puan alırlar. Tüm özellikler ele alınıp puanlama bittikten sonra diller tutarlılıklarına göre sıralanmaktadır. *Tablo 2*'de Orgun (2011)'in çözümlemesine bir örnek sunulmuştur.

Tablo 2. *Orgun'un (2011) çözümleme örneği*

	22A	29A	Puan
<i>Türkçe</i>	Sözcük başına 6-7 ulam	Birleştirimsel değil	-
<i>Arapça</i>	Sözcük başına 6-7 ulam	Birleştirimsel	-
<i>İngilizce</i>	Sözcük başına 2-3 ulam	Birleştirimsel	+
<i>Hintçe</i>	Sözcük başına 2-3 ulam	Birleştirimsel	+
<i>Tagalog</i>	Sözcük başına 2-3 ulam	Sayı/kişi belirleme yok	-

(Uzun, 2012a'dan uyarlanmıştır.)

Yukarıdaki örnekte sadece İngilizce ve Hintçe her iki özellikte de tutarlı bir benzerlik gösterdiklerinden puan almışlar, diğer diller ise alamamışlardır. Orgun (2011) bu yöntem ile Türkçe, İngilizce ve Japonca üzerine gerçekleştirdiği incelemede Japonca'yı en tutarlı dil olarak birinciliğe, Türkçeyi ikinciliğe ve İngilizceyi ise üçüncülüğe yerleştirmiştir.

2.2.8.3. Karşılıklı Benzerlik/Benzemezlik (Uzun, 2012a)

Uzun (2012a), Comrie'nin (2010) ve Orgun'un (2011) yöntemlerini eleştirerek karşılıklı benzerliğe ve benzemezliğe dayanan başka bir çözümleme yöntemi önermektedir. Bu yöntemde, karşılaştırılan iki dilin kaç ortak özellikte değer aldığı ve bu özelliklerin kaçında benzerlik gösterdikleri çözümlenip işlem sonunda karşılıklı benzerlik oranı elde edilir. Orgun (2011) gibi Türkçe, İngilizce ve Japonca'yı inceleyen Uzun (2012a), Türkçenin Japoncayla karşılıklı benzerliğini %54,26, Türkçenin İngilizceyle karşılıklı benzerliğini %40,87 ve İngilizcenin Japoncayla karşılıklı benzerliğini %37,69 olarak hesaplamış ve bunu *Tablo 3*'teki gibi göstermiştir.

Tablo 3. *Uzun'un (2012a) karşılıklı benzerlik çözümleme örneği*

	Türkçe	Japonca	İngilizce	Ortalama Puan	Ortalama Oran	Sıra
<i>Türkçe</i>	-	70/54,26	56/40,87	63	47,56	1.
<i>Japonca</i>	70/54,26	-	49/37,69	59,5	45,97	2.
<i>İngilizce</i>	56/40,87	49/37,69	-	52,5	39,28	3.

(Uzun, 2012a'dan uyarlanmıştır.)

Uzun (2012a), WALs'in veri yapısında ikili, sıralı ve kategorik olmak üzere farklı özellik tipleri olduğu için karşılıklı benzemezliğin hesaplanmasında dikkat edilmesi gereken bazı etmenler olduğunu savlamaktadır. Çözümleme "karşılıklı" olduğu için benzemezliğin hangi dile göre yapıldığı önemli bir etmendir ve bu nedenle yapılması

gereken ilk iş bir kaynak dil seçmektir. Seçilen dile göre üç farklı özellik tipinde izlenecek yöntem sırasıyla paylaşılacak olursa; ikili değer alan özellikler için kaynak dilin tersi değer alan dil *Tablo 4*'te gösterildiği gibi bir benzemezlik puanı alacaktır. Türkçe ile aynı değere sahip İngilizce puan alamazken farklı değere sahip Japonca birer benzemezlik puanı almıştır.

Tablo 4. *Uzun'un (2012a) ikili özelliklerde karşılıklı benzemezlik çözümleme örneği*

Kaynak Dil	Değer (63A)	Puan	Japonca	İngilizce
Türkçe	"ve" "ile"den farklı	0	-	0
	"ve" "ile"yle aynı	1	1	-

(Uzun, 2012a'dan uyarlanmıştır.)

Bu kez, sıralı bir özellik için yapılan benzemezlik hesaplamasında ise incelenen dillerin kaynak dilin değerine olan uzaklığı esas alınmaktadır. *Tablo 5* buna bir örnek olarak sunulmuştur.

Tablo 5. *Uzun'un (2012a) sıralı özelliklerde karşılıklı benzemezlik çözümleme örneği*

Kaynak Dil	Değer (1A)	Puan	Japonca	İngilizce	Bulgarca	Gürcüce	Ermenice
Türkçe	Dar	2	-	-	2	-	-
	Görece dar	1	1	-	-	-	-
	Ortalama	0	-	0	-	-	-
	Görece geniş	1	-	-	-	-	1
	Geniş	2	-	-	-	-	2

(Uzun, 2012a'dan uyarlanmıştır.)

Son özellik türü olan kategorik özelliklerde ise değerlerin birbirlerine uzaklıkları bir anlamlılığa sahip değildir. Daha net bir ifadeyle sıralı özelliklerde olduğu gibi bir değer başka bir değer üzerinde hiyerarşik bir üstünlüğü söz konusu değildir. Bu da her değer birbirine eşit uzaklıkta olması demektir ve bu nedenle kaynak dilden farklı bir değer alan her dil birer benzemezlik puanı alacaktır. Buna bir örnek de *Tablo 6*'da sunulmuştur.

Tablo 6. Uzun 'un (2012a) kategorik özelliklerde karşılıklı benzemezlik çözümleme örneği

Kaynak Dil	Değer (117A)	Puan	Japonca	İngilizce	Tagalog	Svahilice
<i>Türkçe</i>	Yerel	1	1	-	-	-
	Tamlayan	0	-	-	-	-
	Konusal	1	-	-	1	-
	Bağlaçsal	1	-	-	-	1
	“Sahip Olmak”	1	-	1	-	-

(Uzun, 2012a'dan uyarlanmıştır.)

Kategorik değişkenlerde durum her zaman bu kadar ayrıktır değildir. Bazı özelliklerde değerlerden biri, diğer değerlerden en az ikisini kapsamaktadır. Uzun (2012a) bu durum için 124A 'iste-' *Tümleci Özneleri* özelliğini örnek olarak sunmuştur (bkz. Tablo 7). Bu özellikte beş ayrı değerden üçüncüsü ilk ikisini içermektedir. Bu örnekten hareketle, kaynak dilin ilk değere sahip olduğu kabul edilirse üçüncü değer ilkinin kapsadığı için ikinci değere göre daha yakındır.

Tablo 7. Uzun 'un (2012a) kategorik özelliklerde birleşik değerler için önerdiği karşılıklı benzemezlik çözümleme örneği

Değer	Puan
1. Özne örtük bırakılır.	0
2. Özne açık olarak yer alır.	2
3. Her iki yapı da bulunur.	1
4. Eylemcil isteme eki kullanılır.	3
5. İsteme partikeli kullanılır.	3

(Uzun, 2012a'dan uyarlanmıştır.)

Uzun (2012a) ayrıca WALs'teki “diğer (other)”, “hiçbiri (none)”, “karışık (mixed)” gibi değerlerin de belirsizlik içerdiğini belirtmiş, ancak bunları ileri çalışmalara bırakmıştır.

2.2.9. WALS Üzerine Hesaplamalı Çalışmalar

Kuşkusuz ki WALS verisi pek çok tipolojik araştırmaya kaynaklık etmiş, pek çok yapısal çalışmaya veri sunmuştur. Dolayısıyla, WALS ile ilgili alanyazının burada sözü edilenden çok daha geniş olduğunun altını çizerek burada yalnızca WALS üzerine yapılmış hesaplamalı çalışmalara yer verilecektir.

WALS verisinin bilimsel bir araştırma için uygunluğu kimi araştırmacılarca tartışılmış ve veritabanındaki eksik verilerin sonuçların doğruluğunu etkileyebileceği belirtilmiştir. Bizzat kendisi de WALS üzerine istatistiksel çalışmalar yürüten Cysouw (2008) veri eksikliğinin niceliksel sorunlara neden olacağını; Uzun (2012b) ise WALS'ın henüz “sağlam” bir veritabanı olmadığını dile getirmiştir.

Söz konusu eksik verinin giderilmesi ancak bu noktaların bilimsel olarak aydınlatılabilmesiyle mümkündür. Ancak bazı araştırmacılar, eksik verilerin bilgisayar yardımıyla doldurulabilmesi için bazı denemeler yapmışlardır. Bunlardan biri olan Takamura vd. (2016) mevcut özellik değerlerinden hareketle mevcut olmayan özellik değerlerinin saptanıp saptanamayacağını Japonca için ve belli özellikler çerçevesinde denemiştir. Bunun için mevcut olan değerleri bağımsız değişken olarak belirleyip ona bağlı olarak değişkenlik gösterebilecek eksik verileri bağımlı değişken olarak belirlemiş ve Ayırdedici Analiz uygulamıştır. Kısmen doğruluk veren bu yöntemin ileri çalışmalarla geliştirilmesi gerektiği raporlanmıştır.

Daumé III ve Campbell (2007) ile Daumé III (2009) ise WALS verisi üzerinde Bayes teoremini kullanarak hesaplamalı bir model geliştirmişlerdir. Daumé III ve Campbell (2007) bunu özellikler üzerinden sezdirimsel evrenceler analiz etmek için uygulamış ve tipoloji alanyazınında sözü edilmiş ve edilmemiş 45 sezdirimsel evrenceye ulaşmıştır. Daumé III (2009) ise bölgeler ve özellikler arasındaki ilişkinin analizi için Bayes teoremi ile parametrik olmayan bir model geliştirmiş ve modelin geliştirmeye açık yönleri olsa da kısmi bir başarı sağlanmıştır.

Lu (2013) sözcük dizilişinde olasılıksallığı incelemek amacıyla grafik tabanlı bir analiz yürütmüş, farklı parametreler arasındaki bağımlılık ve bağımsızlık ilişkisini betimlemeye çalışmıştır. Ayrıca bu çalışma yalnızca bir sözcük dizilişi evrecesi değil, aynı zamanda bir eksik veri öngörme tekniği olarak da kullanılmaya uygundur.

Bu başlık altında sözü edilecek son çalışma ise Albu'nun (2006) WALs verisini bütünlüklü olarak analiz ettiği doktora tez çalışmasıdır. Çalışmada verinin yeniden üretilmiş bir versiyonunu Hint-Avrupa dil ailesi üzerinde filogenetik teknikleri kullanılarak niceliksel olarak analiz edilmektedir.

2.4. Ara Sonuç

Görüldüğü üzere, dil çok eski zamanlardan beri bir merak konusu olmuştur. Dile karşı bu merakın kaynağı her ne kadar başlangıçta çeşitli amaçlar için bir araç olarak görülmesi olsa da zaman içinde çağdaş dilbilimin de gelişimine koşturarak dil incelemeleri bir araç olmaktan ziyade amaca dönüşmüştür. Bu süreç boyunca değişen tek şey araştırma nesnesine karşı tutum olmamış, aynı zamanda bilimin ve teknolojik olanakların geldiği son nokta itibarıyla kullanılan yöntem ve teknikler de gelişim ve değişim göstermiştir.

İlk çalışmalarda *bütüncül tipoloji* (holistic typology) olarak adlandırılan ve “aile” gibi genel geçer kavramlarla sonuçlara varmayı hedefleyen çalışmalar, günümüzde belli dilbilgisel yapılar özelinde karşılaştırma yapılan ve *parçasal tipoloji* (partial typology) olarak adlandırılan yeni bakış açısına kaymıştır (Comrie, 2013). Bu noktada, “tipoloji”yi bir terim olarak ilk kez kullanan ve çalışmalarında dillerin soy ilişkileriyle yapısal tipolojilerinin farklı şeyler olduğunu vurgulayan Gabelentz'i de anmak gerekmektedir (Ramat, 2011).

Benzer biçimde, teknolojik olanakların yetersiz olduğu yıllarda eldeki sınırlı veriyle ve uzun zaman alan manuel yöntemlerle yürütülen araştırma süreçleri, verinin

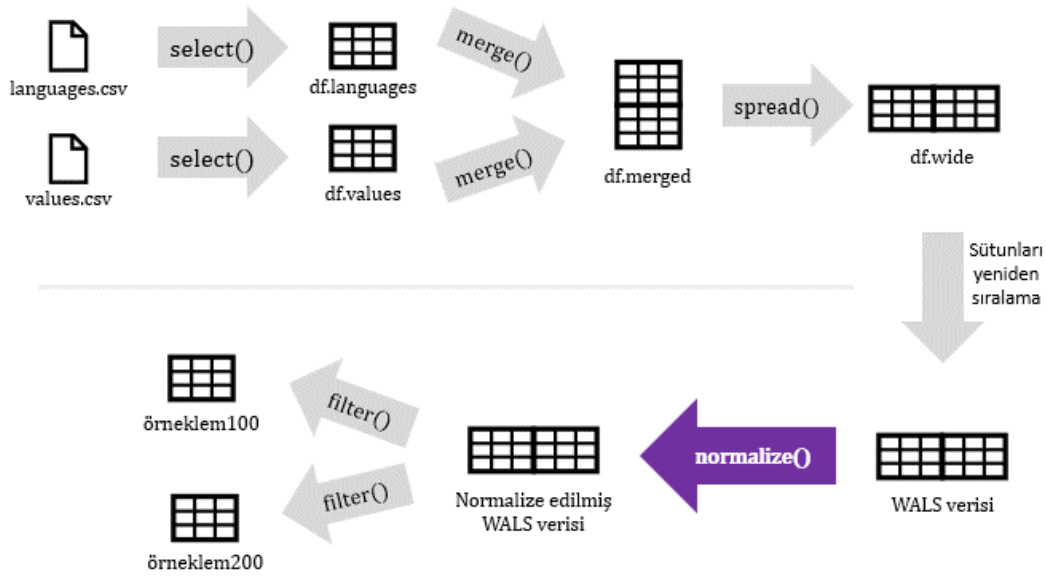
artışıyla birlikte çeşitli biçimlere dönüşmüş ve nihayetinde bilişim teknolojileri yardımıyla yürütülen işlemler olmuştur. Günümüzde, önceki başlıklarda sözü geçen çeşitli teknolojiler yardımıyla dilin çeşitli boyutları için çok sayıda araştırma yapmak mümkündür. Bu araştırmalar, bilgisayarların yüksek işlem kapasitesi ve bellek boyutları gibi olanakları sayesinde bir araştırmacının ulaşamayacağı bir hızda ve kısa bir zamanda gerçekleştirilebilmektedir.



3. YÖNTEM

3.1. Veri Önışlemlemesi

Bu tez çalışmasında nicel veri çözümleme yöntemi ve hesaplamalı tipoloji gözlem tekniği aracılığıyla, WALS veritabanından yararlanılarak, üç farklı nitel/kuramsal yöntem (Comrie, 2010; Orgun, 2011; Uzun, 2012a) çerçevesinde bilgisayar destekli bir R paketi önerisi geliştirilmektedir. Buradan hareketle, WALS'in tez çalışması süreci içerisindeki en güncel sürümü olan v2020.1 numaralı sürümü çalışmamızın verisini oluşturmaktadır.¹³ Veri, çözümlemeye hazır hale getirilmek amacıyla bir dizi önışlemleme sürecinden geçirilmiştir.

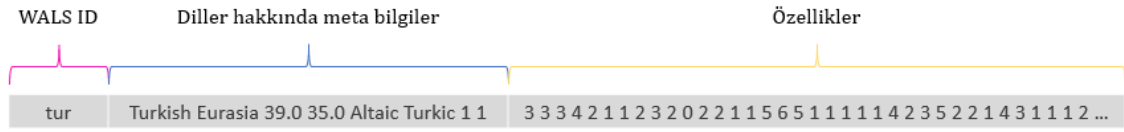


Şekil 5. Önışlemleme aşamaları şeması¹⁴

¹³ Bkz. <https://github.com/clcdf-datasets/wals/releases>

¹⁴ Gri renkle gösterilen süreçlerde harici R paketlerindeki fonksiyonlar kullanılmış, mor renkli süreçlerdeki fonksiyonlar ise tarafımızca geliştirilmiştir.

Şekil 5’te gösterilen bu süreçte sırasıyla *languages.csv* ve *values.csv* dosyalarındaki gerekli sütunlar *tidyverse* (Wickham vd., 2019) paketinin *select()* fonksiyonu yardımıyla seçilmiş, oluşturulan iki veri çerçevesi (data frame) ortak WALS kodu sütunları üzerinden *merge()* fonksiyonu yardımıyla birleştirilmiştir. Ardından, oluşturulan veri çerçevesi aynı paketin *spread()* fonksiyonu yardımıyla yatay veriye dönüştürülmüştür. Sütunların, Şekil 6’daki gibi vektörler oluşturulması amacıyla yeniden sıralanmasının ardından tez çalışması kapsamında tarafımızdan geliştirilen *normalize()* fonksiyonu yardımıyla eksik verilerin bir kısmı doldurulmuştur.¹⁵ Son olarak yine *tidyverse* paketinin bir fonksiyonu olan *filter()* ile 100 ve 200 dil örneklemi ayıştırılmıştır. Bu sürece ait kod dosyası (*preprocessing.R*) tezin ek materyallerinde sunulmuştur.



Şekil 6. Dillere ait vektörlerin temsili görünümü

Yukarıda açıklanan önışleme sonucunda her dil için Şekil 6’daki gibi bir vektör oluşmuştur. Bu vektörün hemen başında ilgili dilin benzersiz WALS kodu, devamında dokuz sütundan oluşan meta bilgiler bölümü (sırasıyla; dilin adı, bölgesi, enlem ve boylam bilgileri, aile, altaile ve soy bilgileri ile örneklemelerde bulunup bulunmama durumları) ve WALS özellik değeri bulunmaktadır.

¹⁵ Eksik verilerin doldurulması için bir önceki bölümde adı geçen istatistiksel yöntemlerin tercih edilmemiş olmasının nedeni, ilgili yöntemlerin yalnızca belli özellikler ve diller üzerinde sınanmış olması ve sonuç olarak üretilen verinin yapay olması nedeniyle bulguları saptırmasını önlemektir.

3.2. Eksik Veri Analizi ve normalize()

WALS'teki eksik veriyle ilgili hesaplamalarda çok basit bir mantık hatası göze çarpmaktadır. Veritabanında 2662 dil bulunduğu ve bunların 192 özellik çerçevesinde değerlendirildiği önkabulüyle veritabanında toplamda $2662 \times 192 = 511104$ veri noktası (data point) olması gerektiği düşünülmektedir. Ancak WALS bu şekilde yapılandırılmamıştır ve belli noktaların eksik olması, daha doğru bir ifadeyle boş olması, aslında olması gereken bir durumdur. Bu nedenle, aşağıda ayrıntıları verilecek olan belli ölçütler çerçevesinde bazı boş veri noktaları eksik verilerden ayırtmak amacıyla 0 değeriyle doldurulmuş, gözlem sırasında karşılaşılan bazı veri etiketleme hataları giderilmiş ve normalize edilmiş bir veri çerçevesi elde edilmiştir.

Boş verilerin doldurulması aşamasında göze çarpan ilk nokta işaret dilleridir. WALS üzerinde 40 işaret dili bulunmaktadır ve bunlar iki özellik (139A ve 140A) çerçevesinde değerlendirilmektedir. İşaret dillerinin bu iki özellik dışındaki 190 özellik için bir değer almasını beklemek, işaret dili olmayan dillerin bu iki özellik için değer almasını beklemek gibi yanlış olur. Dolayısıyla, işaret dili olmayan diller için 139A ve 140A özellikleri 0 değeriyle, işaret dili olan diller için de 139A ve 140A dışındaki diğer tüm özellikler 0 değeriyle doldurulmuştur.

Burada eksik veri ve boş veri arasındaki farka dikkat çekmek amacıyla bir noktayı vurgulamakta yarar var gibi görünmektedir. Yukarıda, veritabanında 40 işaret dili olduğu ve bunların yalnızca iki özellik çerçevesinde değerlendirildiği söylenmişti. Dikkat edilecek olursa 139A özelliğinde 35 işaret dilinin, 140A özelliğinde de 38 işaret dilinin değer aldığı görülmektedir. Dolayısıyla, işaret dilleri için 139A'da 5, 140A'da 2 olmak üzere 7 veri noktası eksik veri oluştururken, işaret dili olmayan 2622 dil için iki özellikteki 5244 veri noktası aslında gerçek anlamda bir eksik veri değildir.

İşaret dillerindekine benzer bir durum, bazı özelliklerde de vardır. Örneğin *10B Batı Afrika'da Genizsil Ünlüler*, başlığından da anlaşılacağı gibi yalnızca Batı Afrika

dillerindeki genizsil ünlüler ile ilgili bir incelemedir. Dolayısıyla, Batı Afrika dili olmayan dillerin bu özellikte bir değer almaması araştırmacıları bir yanılgıya düşürebilir. Öyle ki veritabanında adı bulunan bir Batı Afrika dili için incelemelerin kısıtlılığı nedeniyle herhangi bir değer henüz girilememiş olabilir ve bu dil gelecek güncellemelerle ilgili özellik için bir değer kazanabilir. Ancak örneğin Türkçe gibi Batı Afrika dili olmayan dillerin bu özellik çerçevesinde bir değer alması olasılığı hiç yoktur. Bu gerçek, iki durum arasında bir ayrımı zorunlu kılar ve aynı şekilde bu olasılığı sıfır olan diller için veri noktaları 0 değeriyle doldurulmuştur.

Özellikler ile ilgili başka bir özel durum da iki özelliğin birbiriyle kurduğu koşul ilişkisidir. Örneğin 144D, 144E, 144F, 144G, 144H, 144I, 144J ve 144K özellikleri yalnızca olumsuzluk dizilişinin ÖEN olduğu diller içindir. Bir dilin olumsuzluk dizilişinin ÖEN dizilişli olup olmadığı ise 143A özelliği üzerinde değerlendirilmiştir. Dolayısıyla, 143A’da ÖEN olarak belirtilmemiş olan dillerin 144D-K serisindeki 8 özellik için bir değer alması beklenemez. Böyle bir potansiyel de mümkün değildir. Dolayısıyla, bu türden özellik noktaları da 0 değeriyle doldurulmuştur.

Son olarak, bir dilin belli bir özellik için değeri iki şekilde bulunabilir. Bunlardan ilki alanyazında o dilin ilgili özelliğini betimleyen bir çalışma olup olmadığını taramaktır, ancak bu yöntem *Giriş* bölümünde de söz edildiği gibi bu tez çalışmasının çerçevesini aşan bir girişim olacaktır. Ancak ikinci yol ise WALS üzerinde bir dilin bir özellik için girilen değerinden, girilmemiş olan başka bir özelliği çıkarımlamaktır. Bu aslında yukarıda anlatılan koşullu durumun eksik bir şekilde işlenmesi, dolayısıyla da bir WALS içerik hatası durumudur. Örneğin bir Hoka dili olan Cocopan sabit bir vurgu konumunun olduğu ve bunun son seslem olduğu 14A’da belirtilmiştir. Ancak 14A’da sabit bir vurgu konumuna sahip olduğu belirtilen Cocopa dışındaki dillerin tümü 15A’da da sabit vurgulu olduklarına ilişkin değerle etiketlenmelerine karşın, Cocopa dilinin 15A

veri noktası eksik görünmektedir. Bu durumdaki veri noktaları, daha önce aktarılan özel durumlardaki gibi 0 ile değil gerekli değer ile doldurulmuştur.

Yukarıdaki normalizasyon adımlarının tümü EK-2’de betimlenmiş ve bu sürece ait kod dosyası *R* klasörü içinde (*normalize.R*) tezin ek materyallerinde sunulmuştur.

3.3. Çözümleme Yöntemlerinin Fonksiyonlaştırılması

R dilinde diğer yazılım dillerinde olduğu gibi, bir dizi komutu bir bütün halinde kodlayıp tekrar kullanılabilir hale getirmek *fonksiyonlar* (function) ile mümkündür. Bir dizi işlemi, art arda komutlayıp adlandırdıktan sonra bu ad ile çağırılıp kullanılabilen fonksiyonların işlem yapabilmek için kullanacakları *üyeler* (argument) bulunur. (Venables vd., 2021, ss. 42–50).

Çalışmamız kapsamında Comrie’nin (2010) yöntemi *comrie()* ve *comrie_c()*, Orgun’un (2011) yöntemi *orgun()* ve Uzun’un (2012a) yöntemi de *uzun()* olmak üzere kendi adlarıyla fonksiyonlaştırılmıştır. Burada yalnızca Comrie (2010) için iki adet fonksiyon oluşturulmuştur, çünkü bir dilin tipiklik seviyesi hem kendi içinde hem de karşılaştırmalı olarak hesaplanmak istenmiştir. Bu doğrultuda, *comrie()* dilin kendi içindeki tipikliği, *comrie_c()* ise iki dilin karşılaştırmalı sonucunu sunmaktadır.

Fonksiyonların tümü üç adet üye istemektedir. İlk sırada inceleme yapılan veritabanının değişken adı (data)¹⁶ girilmelidir. İkinci sırada karakter biçiminde inceleme yapılacak dilin WALS kodu (lang veya langs) girilmelidir. Bu üye için yalnızca *comrie()* tek kod alırken *comrie_c()*, *orgun()* ve *uzun()* vektör biçiminde iki kod almalıdır. Son üye ise kullanıcının analiz sonucunu bir veri çerçevesi üzerinde görmek istemesi durumunda “data”, yalnızca bir matematiksel değer olarak görmek istemesi durumunda “value”

¹⁶ Bu paragrafta ayrıca içindeki ifadeler paketteki değişken adlarıdır. Paketin yayımlanması planlandığından kod içerikleri İngilizce yazılmıştır.

seeneklerinden birini girmesi gerektiđi moddur (mode). Sözelimi, kullanıcı Türkenin tipiklik oranını ölçmek istemesi ve bunu veri çerevesi biçiminde görmek istemesi durumunda (5)’teki gibi bir komut vermesi yeterli olacakken sözelimi Türkenin mi Japoncanın mı daha tipik olduđunu görmek isteyen bir kullanıcı kodu (6)’daki gibi düzenleyebilir. Benzer şekilde, Türke ve İngilizce gibi iki dilin tutarlılık ölçümünü matematiksel veri olarak görmek isteyen bir kullanıcı (7) gibi bir komut kullanabilir. Son olarak, kullanıcı İngilizce ile Japoncanın karşılıklı benzeyen özellik deđerlerini incelemek için (8)’deki gibi bir komut verebilir.

(4) > comrie(walsData_n, “tur”, “data”)

(5) > comrie_c(walsData_n, c(“tur”, “jpn”), “value”)

(6) > orgun(walsData_n, c(“tur”, “eng”), “value”)

(7) > uzun(walsData_n, c(“eng”, “jpn”), “data”)

Uzun’un (2012a) önerdiđi karşılıklı benzemezlik analizi ise, yöntemde bazı noktaların belirgin olmaması, belirgin olan noktalar için ise veritabanının iskeletinin yeniden yapılandırılması gerektiđinden bu alışmanın kapsamı dışında bırakılmış olup konuyla ilgili olarak 5. *Tartışma ve Sonuç* bölümünde ileri alışmalarda ele alınmak üzere bir deneme önerisi sunulmuştur. Son olarak adı geen fonksiyonların tümü, aynı anda daha fazla dili özümleyebilecek veya farklı istatistikler sunacak biçimde ileri alışmalar ile geliştirilebilir.

Bu bölüme ilişkin kod dosyası ek materyallerde *R* klasörü içinde *methods.R* adıyla bulunmaktadır.

3.4. özümleme Adımları ve Raporlama

alışmamızda ilk olarak veritabanının tümünde veri doluluk oranları, özellik bazında etiketli diller ve diller bazında etiketli özellikler gibi eşitli ölçeklerde

özmlenmiřtir. Bu verilerin, yukarıda sözü edilen normalizasyon iřlemi ncesindeki ve sonrasındaki durumları da niceliksel bulgular zerinden karřılařtırılmıřtır.

Ardından yukarıda adı geen ve fonksiyonlařtırılan  yntem ayrı ayrı iki rneklem zerinde test edilmiř ve bu ařamada da normalizasyon iřleminin deęerler zerindeki etkisi llmüřtür. Karřılařtırma yntemlerinin analizlerinde eksik veri analizinden farklı olarak sistemde yer alan 40 adet iřaret dili ve iřaret dillerine ait zellikler ile *Dięer* alanındaki zellikler veritabanından ıkarılmıřtır.

alıřmamızda ana ama, bilgisayar desteęi ile daha ok veriyi daha kısa zamanda ve doęru olarak analiz etmektir. Bu doęrultuda, ok kısa bir srede hacimce geniř veri dkmleri iřlenen fonksiyonlar aracılıęıyla elde edilmiřtir ancak burada yalnızca arařtırma soruları erevesinde nemli noktalara yer verilmiřtir. Analiz sonularının tam dkm hem sonuların daha ayrıntılı olarak irdelenebilmesi hem de ileri alıřmalara kaynak oluřturabilmesi amacıyla analiz kodlarıyla birlikte ek materyaller iinde paylařılmıřtır.

Son olarak, zmleme ařamasında ve verilerin raporlanmasında veri maniplasyonu iin bazı temel R fonksiyonları (R Core Team, 2021), *utils* (Nijenhuis ve Wilf, 1978) ve *tidyverse* (Wickham vd., 2019) paketleri, betimsel istatistik iin *psych* (Revelle, 2020) paketi, veri grselleřtirme iin ise *naniar* (Tierney vd., 2020), *ggcorrplot* (Kassambara, 2019) ve *ggplot2* (Wickham, 2016) paketleri kullanılmıřtır.

Bu blme iliřkin kod dosyası ek materyallerde *analysis.R* adıyla bulunmaktadır.

4. BULGULAR

4.1. Eksik Veri Analizi ve Normalizasyon

Eksik veri analizinin bulgularına geçmeden önce bu analizin yalnızca örneklemelerde değil, verinin tümünde yapıldığını anımsatmak gerekmektedir. Böylece çalışmamız kapsamında önerilen normalizasyon sürecinin veritabanının geneli üzerindeki potansiyeli hakkında fikir oluşturmak amaçlanmıştır.

Çözümleme aşamasında, ilk olarak kullanılan ham veritabanındaki eksik veri oranı tüm veritabanında, 100 ve 200 dilli örneklemelerde hem genel hem de diller ve özellikler düzeyinde olmak üzere ayrı ayrı incelenmiştir. Dil vektörlerinin yalnızca özellik değerlerini içeren 192 sütunu esas alındığında, eksik veri, veritabanının tümünde %85,03 oranındayken, 200 dil içeren örneklemin %38,22'sini, 100 dil içeren örneklemin ise %29,61'ini oluşturmaktadır.

Eksik verilere diller özelinde bakılacak olursa mevcut veritabanında, hakkında en fazla bilgi bulunan dil 159 özellikte değer alan İngilizcedir. Onu takip eden 10 dil ise değer aldıkları özellik sayılarıyla birlikte Fransızca (158), Almanca (157), Rusça (156), Fince (155), Modern Yunanca (155), Macarca (155), İspanyolca (155), Türkçe (154), Endonezce (153), Mandarin Çincesi (154) olarak sıralanmaktadır. Veritabanında en çok özellikte etiketlenen dillerin örneklemelerde bulunup bulunmadığına bakıldığında ise en çok etiketli 100 dilden Macarca, Evenki dili, Maori dili, Kanuri dili, Yukagir dili (Kolima değişkesi), Aynu dili, Awa Pit dili, Canela dili, Gilyak dili, Kobon dili, Ermenice (Doğu değişkesi), Letonca, Ketçe ve Epena Pedee dilinin 100 dilli örnekleme; en çok özellikte etiketli ilk 200 dilden Canela dili, Nenets dili, Amhar dili, Kanton dili, Arnavutça, Felemenkçe, Lehçe, Zuni dili, Somalice, Bulgarca, Dyirbal dili, Koyra Chiini dili, Lahu dili ve Macushi dilinin 200 dilli örnekleme bulunmadığı saptanmıştır. Listenin sonuna

ilişkin gözlemlenen önemli bir bulgu da veritabanında yer alan dillerin 253'ünde yalnızca bir özelliğin etiketlenmiş olmasıdır.¹⁷

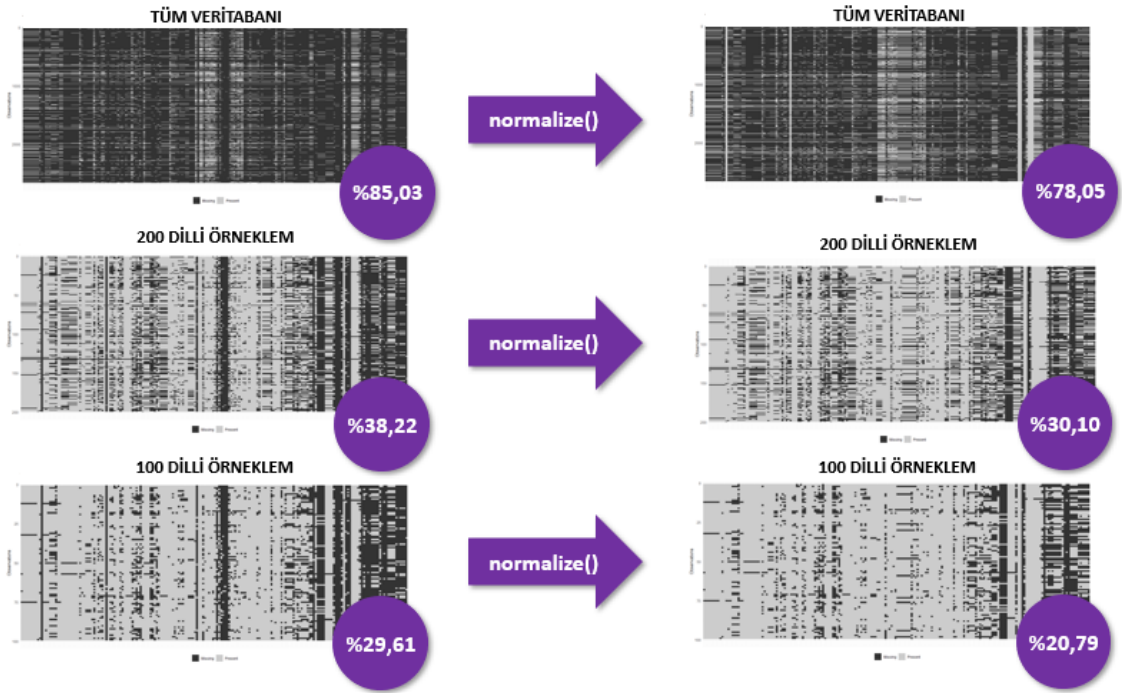
Eksik veriler, özellikler çerçevesinde incelendiğinde ise WALS'in web arayüzünden de görülebileceği üzere en çok dilin etiketlendiği özellik 1518 dil ile 83A *Nesne ve Eylem Dizilişi* iken, en az dilin etiketlendiği özellik 5 dil ile 90G *Çift başlı ilgi tümcecikleri* özelliğidir. Özellikler, alanları çerçevesinde incelendiğinde her bir alanın içerdiği özellik ve bunlarda ortalama kaç dilin etiketli olduğu *Tablo 8*'de sunulmuştur.

Tablo 8. Özelliklerin alanlara göre içerdikleri dil ve özellik sayıları

Alan	Özellik Sayısı	Etiketli Dil (Ort)
<i>Sözcük Dizilişi</i>	56	568,96
<i>Adıl Ulamlar</i>	29	352,83
<i>Basit Tümcecikler</i>	26	352,58
<i>Sesbilim</i>	20	490,65
<i>Eylemcil Ulamlar</i>	17	344,29
<i>Sözlükçe</i>	13	238,92
<i>Biçimbilim</i>	12	275,67
<i>Adıl Sözdizim</i>	8	211,88
<i>Karmaşık Tümceler</i>	7	173,86
<i>İşaret Dilleri</i>	2	74,50
<i>Diğer</i>	2	36,50

Yukarıdaki ham veritabanına ait eksik veri bulguları, 3.2. *Eksik Veri Analizi ve normalize()* başlığında ayrıntılarına değinilen normalizasyon işlemi sonrasında bazı değişimlere uğramıştır. Bu başlığın devamında söz konusu değişimler niceliksel veriler ve görselleştirmeler yardımıyla aktarılacaktır.

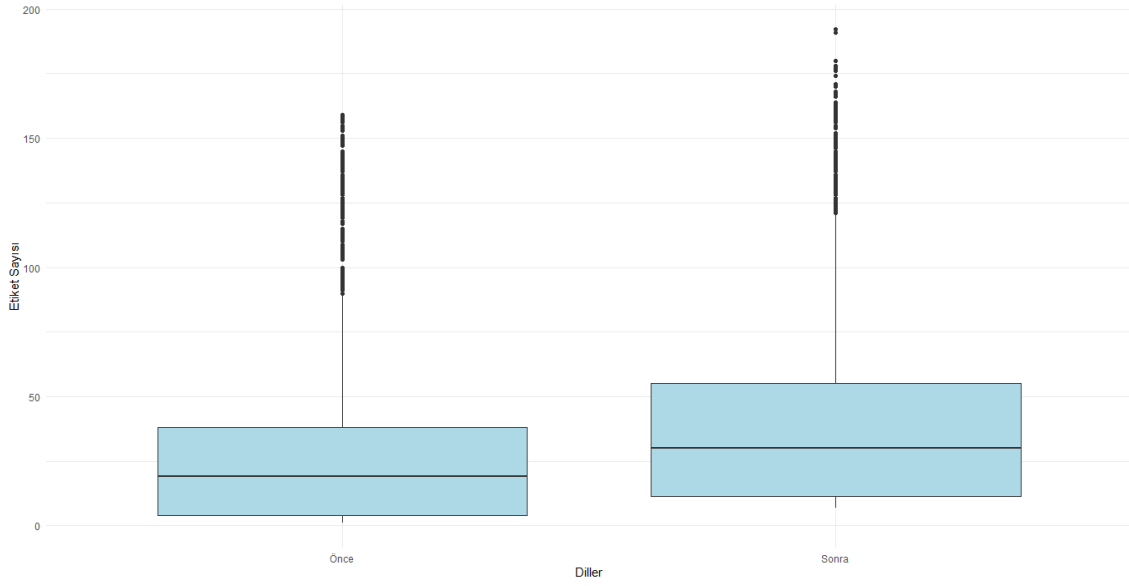
¹⁷ Normalizasyon işlemi sırasında iki dilin 0 etiket aldığı, bu nedenle ham veride görünmediği bulgulanmıştır (bkz. EK-2).



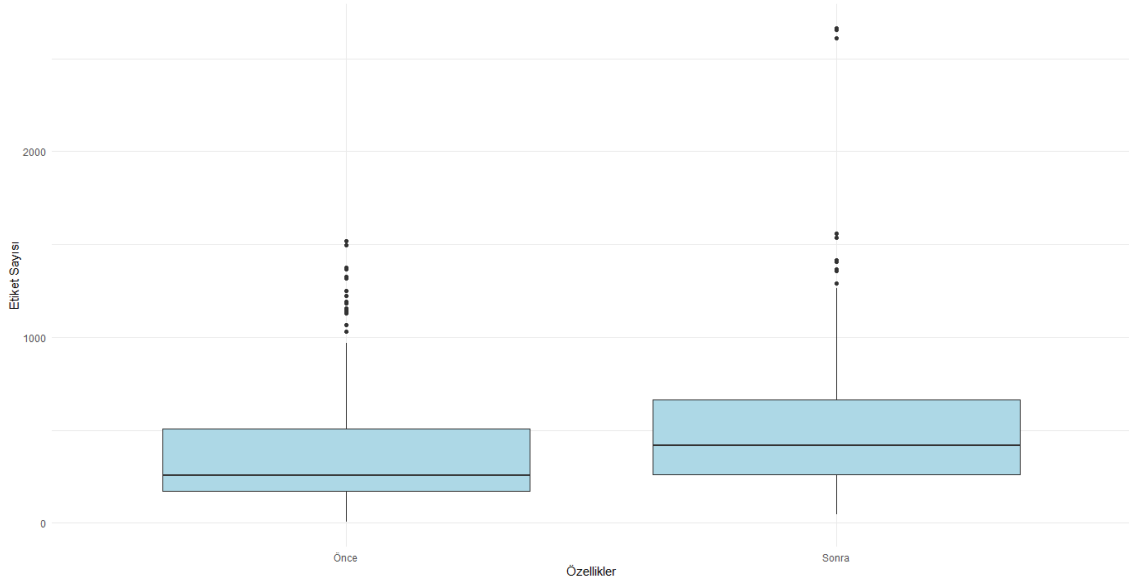
Şekil 7. Normalizasyon işlemi öncesinde ve sonrasında eksik veri görselleştirmesi¹⁸

Karşılaştırmanın anlaşılabilirliğini kolaylaştırmak adına yine genelden özele doğru bir yol izlenecektir. İlk olarak, dil vektörlerinin yalnızca özellik değerlerini içeren 192 sütunu esas alındığında eksik veri, veritabanının tümünde %85,03'ten %78,05'e, 200 dil içeren örneklemde %38,22'den %30,10'a, 100 dil içeren örneklemde ise %29,61'den %20,79'a düşmüştür. Bu değişim, Şekil 7'de görselleştirilerek de sunulmuştur. Bu görselde siyah noktalar eksik verileri, beyaz noktalar ise bir değer ile dolu olan hücreleri göstermektedir. Görüldüğü üzere üç veri çerçevesi için de normalizasyon sonrasında öncesine oranla eksik veride düşüş görülmektedir.

¹⁸ Görselleştirmede *naniar* (Tierney vd., 2020) paketinin *vis_miss()* fonksiyonundan yararlanılmıştır.



Şekil 8. Normalizasyon öncesi ve sonrasında etiketli diller



Şekil 9. Normalizasyon öncesi ve sonrasında etiketli özellikler

Şekil 8’deki ve Şekil 9’deki kutu grafikler ise normalizasyon işlemi sonrasında sırasıyla diller ve özellikler seviyesindeki değişimi özetlemektedir.¹⁹ Her kategori için sol taraftaki kutu normalizasyon öncesini, sağ taraftaki kutu ise normalizasyon sonrasını göstermektedir. Bu çerçevede, dillerin etiketlendikleri özellik sayısına göz atıldığında, normalizasyon öncesinde ortalama 28,75 özellikte etiketlenirken normalizasyon

¹⁹ Görselleştirme için *ggplot2* (Wickham, 2016) paketinden yararlanılmıştır.

sonrasında bu değer 42,14'e yükselmiştir. Aynı durum özelliklerde etiketli dil sayılarında da geçerlidir. Normalizasyon öncesinde özelliklerde ortalama 398,3 dil etiketliken normalizasyon sonrasında ortalama 584,3 dil etiketli duruma geçmiştir.

Tablo 9. Normalizasyon öncesinde ve sonrasında en çok etiket alan ilk 10 dil ve özellik

Diller		Özellikler	
Önce	Sonra	Önce	Sonra
İngilizce (159)	İngilizce (180)	83A (1518)	10B (2662)
Fransızca (158)	Gürcüce (178)	82A (1496)	143B (2662)
Almanca (157)	Bask dili (177)	81A (1376)	143C (2662)
Rusça (156)	Rusça (177)	87A (1367)	143D (2662)
Fince (155)	Fince (176)	143A (1325)	39B (2612)
M. Yunanca (155)	Korece (176)	143E (1325)	83A (1558)
Macarca (155)	İspanyolca (176)	143F (1325)	82A (1536)
İspanyolca (155)	Macarca (174)	143G (1325)	81B (1417)
Türkçe (154)	Endonezce (174)	97A (1316)	81A (1416)
Endonezce (153)	Mandarin (174)	86A (1249)	87A (1407)

Son olarak, *Tablo 9*'da ise normalizasyon işlemi öncesinde ve sonrasında, en çok etikete sahip olan ilk 10 dil ve en çok etikete sahip olan ilk 10 özellik sunulmuştur. Bu tabloda işaret dilleri göz ardı edilmiştir, çünkü İşaret Dilleri özellik alanındaki iki özellik için de değer alan işaret dilleri diğer özellikleri 0 ile doldurulduğundan tüm özellikleri dolu görünmektedir. Oysa bu dillerin diğer dillerden ayrı bir veri çerçevesinde incelenmesi gerekmektedir.

Bu bölüme ilişkin kod dosyası ek materyallerde *missingdataanalysis.R* adıyla bulunmaktadır.

4.2. Tipolojik Karşılaştırma Sonuçları

Bu başlıkta üç farklı yöntemin örneklemeler üzerinde uygulanması sonucunda elde edilen betimsel bulgular alt başlıklar biçiminde sunulacaktır. Karşılaştırmalara ilişkin tüm bulguların dökümü analiz kodları (*analysis.R*) ile birlikte ek dosyada paylaşılmıştır, ancak burada yalnızca araştırma sorularımız çerçevesinde önemli noktalara değinilmektedir.

4.2.1. Tipiklik

3.2. *Eksik Veri Analizi ve normalize()* başlığında ayrıntılı biçimde ele alındığı gibi, normalizasyon işlemi sırasında bazı özelliklerin bazı diller için uygun olmaması durumunda bu dillere 0 değeri atanmıştır. Normalizasyon işlemi sonrasında 10B, 39B, 81B, 90B, 90D, 90E, 90F, 90G, 130B, 143B, 143C, 143D, 144D, 144E, 144F, 144G, 144I, 144L, 144T, 144U, 144V, 144W, 144X, 144Y özelliklerinin en tipik değerleri 0 değerini almış, bu da bu özellikleri etiketli olan veya normalizasyon sonrasında etiketli duruma geçmiş dillerin tipiklik oranını etkilemiştir.

Tablo 10. Dillerin kendi başlarına tipiklik oranlarına göre örneklemelerin dağılımı (%)²⁰

	Örneklem 100		Örneklem 200	
	Önce	Sonra	Önce	Sonra
Ortalama	44,00	47,51	43,40	47,50
Standart Sapma	2,21	2,43	3,03	3,32
Standart Hata	0,22	0,24	0,21	0,24
En Düşük	38,66	41,85	18,64	39,99
En Yüksek	51,03	55,81	51,03	75,02

İlk olarak örneklemdeki her bir dili ayrı ayrı ele alarak sahip oldukları etiket üzerinden tipiklik oranları hesaplanmıştır. Bu hesaplama sonucunda 100 dilli örneklemde normalizasyon öncesinde %51,03 ve sonrasında %55,81 oranıyla en tipik dil Daga iken, en az tipik dil normalizasyon öncesinde %38,66 oranıyla Maricopa ve normalizasyon işleminin ardından %41,85 oranıyla Supyire olmuştur. 200 dilli örneklemde de normalizasyon öncesinde %51,03 oranıyla en tipik dil olan Daga normalizasyondan sonra yerini %75,02'lik tipiklik oranıyla Ngiti diline bırakmış, normalizasyon öncesinde %18,64 oranıyla en az tipik dil olan Nenets (Forest) de normalizasyondan sonra %39,98

²⁰ Betimsel bulgular için *psych* (Revelle, 2020) paketinin *describe()* fonksiyonundan yararlanılmıştır.

oranıyla yerini Ngiti diline bırakmıştır. *Tablo 10* tipiklik oranlarının durumunu göstermektedir.

Bunun dışında, örneklemdeki tüm dillerin hem özellik alanları bazında hem de geneldeki tipiklik oranları normalizasyon öncesi ve sonrası için ayrı ayrı analiz edilip ek materyaller içerisinde paylaşılmıştır (bkz. *typicality100.csv* ve *typicality200.csv*). Örnek olması açısından sadece Dagaya ait kesit *Tablo 11*'de yer almaktadır. Normalizasyon işleminin etkisi tablo üzerinde vurgulanmıştır.

Tablo 11. *Dillerin kendi başlarına tipiklik değer ölçümleri (Daga – Örneklem 100) (%)*

	Normalizasyon	Sesbilim	Biçimbilim	Adıl Uamlar	Adıl Sözdizim	Eylemcil Uamlar	Sözcük Dizilişi	Basit Tümcecikler	Karmaşık Tümcecikler	Sözlükçe	Genel
Daga	Önce	55,45	44,72	43,26	68,23	51,34	51,84	48,00	75,30	64,91	51,03
	Sonra	58,32	44,72	46,62	72,65	51,34	62,45	48,00	75,30	64,91	55,81

Dillerin tipiklik çerçevesinde karşılıklı olarak incelenmesi sonucunda her iki örneklemün önceki ve sonraki durumlarına ilişkin tipiklik haritası oluşturulmuştur. Bu haritaların tamamı veri büyüklüğü nedeniyle ek materyallerde sunulmuştur (bkz. *tmap100.csv*, *tmap100_n.csv*, *tmap200.csv* ve *tmap200_n.csv*).

Şekil 10, tipiklik haritasını okumayı normalizasyon sonrası 200 dilli örneklemden bir kesit üzerinde açıklamaktadır. Tipikliği karşılaştırılmak istenen iki dilin WALS kodları satır ve sütun başlıklarından bulunup kesişim noktasına bakıldığında, bu iki dilden hangisinin daha tipik olduğu bilgisi görünmektedir. Örnekte, Aynu ve Amele dillerinden Amelenin Aynuya göre daha tipik olduğu görülmektedir. Söz konusu iki dilin istatistiksel oranları ise (9)'daki komut yardımıyla alınabilir.

```
(8) > comrie_c(walsData_n, c("ain", "ame"), "value")
```

```
langs    mean
```

```
1  ain 0.4852753
```

```
2  ame 0.4865228
```

```
[1] "ame"
```

	abi	abk	aco	aeg	ain	ala	ame	ana	apu	arm
abi	-	abi	abi	abi	abi	abi	abi	abi	abi	abi
abk	abi	-	aco	aeg	ain	ala	ame	ana	apu	arm
aco	abi	aco	-	aeg	ain	ala	ame	ana	apu	arm
aeg	abi	aeg	aeg	-	ain	ala	ame	ana	apu	aeg
ain	abi	ain	ain	ain	-	ala	ame	ain	ain	ain
ala	abi	ala	ala	ala	ala	-	ala	ala	ala	ala
ame	abi	ame	ame	ame	ame	ala	-	ame	ame	ame
ana	abi	ana	ana	ana	ain	ala	ame	-	apu	ana
apu	abi	apu	apu	apu	ain	ala	ame	apu	-	apu
arm	abi	arm	arm	aeg	ain	ala	ame	ana	apu	-

Şekil 10. Tipiklik haritası kesiti (Örneklem 200)

Tipiklik haritasında her bir dilin kaç dilden daha tipik görüldüğü hesaplanmış ve Daganın her iki örnekleme de hem normalizasyon öncesi hem de sonrasında en tipik dil olduğu bulgulanmıştır. Bu sayım ile ilgili döküm ise yine ek materyallerde (bkz. *tmap100Count.csv* ve *tmap200Count.csv*) paylaşılmıştır.

4.2.2. Tutarlılık

Tutarlılık testi, iki veya daha fazla eleman sayısına sahip herhangi bir dil kümesi ile gerçekleştirilebilir, ancak çalışmamız kapsamında burada yalnızca aynı ailenin üyesi olan dillerin tutarlılık puanları üzerinde durulacaktır. Öyle ki aynı dil ailesinde bulunan dillerin üzerinde eğer bu soy bağı etkiliyse, o zaman bu diller kendi aralarında bir tutarlılığa sahip olmalıdır.

Tablo 12. *Dil ailesi tutarlılık testi (Altay örneği)*

Dil	Tutarlılık (Puan)	
	Önce	Sonra
<i>Türkçe</i>	101	114,5
<i>Khalkha</i>	97,5	111
<i>Evenki</i>	93,5	106,5
Ortalama	97,33	110,67

Tutarlılık çözümlemesi aileler bazında, 100 dilli örneklemdaki tüm dilleri içerdiği ve hacimce daha geniş olduğu için yalnızca 200 dilli örneklem üzerinden yapılmıştır. İlk olarak aile üyelerinin kendi aralarındaki tutarlılıkları hesaplanıp sonrasında her bir aile üyesinin ailenin diğer üyeleriyle ortalama tutarlılığına göre sıralı tablolar oluşturulmuştur. Örnek olması amacıyla *Tablo 12*'de Türkçenin de yer aldığı ve *Tablo 13*'te görülebileceği üzere hem normalizasyon öncesinde hem de normalizasyon sonrasında en yüksek tutarlılığı gösteren Altay dil ailesinin verileri paylaşılmıştır. Örneklemdaki tüm ailelere ilişkin analiz dökümü ek materyallerde *coherences* klasöründe paylaşılmıştır.

Tablo 13. *Ailelerin tutarlılık oranları (Örneklem 200)*

Aile	Dil Sayısı	Tutarlılık (Puan)	
		Önce	Sonra
<i>Altay</i>	3	97,33	110,67
<i>Hint-Avrupa</i>	11	73,36	86,93
<i>Oto-Menguen</i>	3	57,00	79,00
<i>Avustronezya</i>	15	60,48	74,24
<i>Nah-Dağıstan</i>	4	57,17	67,00
<i>Uto-Aztek</i>	4	50,67	66,17
<i>Karayip</i>	2	58,00	66,00
<i>Pama-Nyungan</i>	4	54,83	66,00
<i>Na-Dene</i>	3	49,67	68,00
<i>Afro-Asya</i>	8	51,39	64,00
<i>Nijer-Kongo</i>	13	52,06	63,87
<i>Avusturo-Asya</i>	6	49,87	63,27
<i>Tupiyen</i>	2	53,00	60,00
<i>Trans-Yeni Gine</i>	10	46,42	57,64
<i>Çin-Tibet</i>	8	45,07	55,36
<i>Algonkin</i>	3	40,33	48,67
<i>Ural</i>	3	33,00	44,33
<i>Penutiyen</i>	3	34,33	44,00
<i>Chibchan</i>	3	34,33	42,67
<i>Orta Sudan</i>	2	33,00	40,00

200 dilli örneklemda 20 ailenin tutarlılık ortalamaları ise normalizasyon sonrası verilere göre en tutarlıdan en az tutarlıya doğru sıralandığında *Tablo 13*'teki gibi görünmektedir. Ayrıca bulgular dil sayılarının tutarlılık ölçümündeki etkisini de

sorgulatmaktadır. Hem normalizasyon öncesinde hem de sonrasında en tutarlı dil ailesi olan Altay dil ailesinin örnekleme yalnızca 3 üyesi bulunurken, yine normalizasyon öncesi ve sonrasında ikinci olan Hint-Avrupa dil ailesi 11 dil ile temsil edilmektedir.

200 dilli örnekleme kendi ailesinden en az bir üye daha bulunan tüm diller çıkartılıp geriye her bir aileden birer üye bırakıldığında (toplamda 89 dil) ve tutarlılık testi bu örneklem üzerinden yinlendiğinde normalizasyon sürecinden önce ortalama 39,50; normalizasyonun ardından ortalama 48,30 tutarlılık görülmüştür. Diğer diller ile en çok tutarlılık gösteren ilk 5 dil normalizasyon sonrası sıralamaya göre *Tablo 14*'teki gibidir. Burada dikkat çeken bir nokta ise en çok tutarlılık gösteren bu dillerden dördünün sınıflandırılmayan/izole dillerden olmasıdır.

Tablo 14. Normalizasyon öncesinde ve sonrasında farklı ailelerden diller arasında en çok tutarlılık gösteren ilk 5 dil (Örneklem 200)

Dil	Tutarlılık (Puan)	
	Önce	Sonra
<i>Bask</i>	50,45	62,89
<i>Korece</i>	50,32	62,18
<i>Alamblak</i>	49,08	61,11
<i>Ainu</i>	48,11	60,15
<i>Tiwi</i>	48,43	59,66

4.2.3. Karşılıklı Benzerlik

Karşılıklı benzerlik analizinde dillerin benzerlikleri çiftler halinde hesaplanıp ardından buna göre bir sıralama yapılmaktadır. Bu sıralamayı yapabilmek için *utils* (Nijenhuis ve Wilf, 1978) paketinden *combn()* fonksiyonu yardımıyla $(a, b) = (b, a)$ biçimindeki kopya çiftlere izin verilmeden 99 dil içeren 100 dilli örnekleme 4851, 199 dil içeren 200 dilli örnekleme ise 19701 çift oluşturulmuştur. Daha sonra bu çiftlerin hem normalizasyon öncesindeki hem de sonrasındaki karşılıklı benzerlik oranları

hesaplanmıştır. *Tablo 15*'te her iki örneklem için normalizasyon öncesi ve sonrası bulgular paylaşılmıştır.

Tablo 15. *Dillerin kendi başlarına tipiklik oranlarına göre örneklemelerin dağılımı (%)*²¹

	Örneklem 100		Örneklem 200	
	Önce	Sonra	Önce	Sonra
Ortalama	44,31	47,72	42,96	47,58
Standart Sapma	7,47	7,18	9,18	8,54
Standart Hata	0,11	0,10	0,07	0,06
En Düşük	20,00	25,55	0	20,34
En Yüksek	84,09	86,00	100	100

Bu noktada, sınırlı sayıda veri nedeniyle tam benzerlik gösteren diller %100, hiç ortak değer göstermeyen diller %0 ve hiçbir ortak özellikte değerlendirmemiş diller NA (not applicable) sonucu döndürmekte ve bu durum betimsel bulguları etkilemektedir.

Bu çerçevede, *Tablo 16*'da 100 dilli örneklem, *Tablo 17*'de ise 200 dilli örneklemde normalizasyon sonrası en yüksek karşılıklı benzerliğe sahip ilk 10 dil sıralamasına göre sunulurken %100 benzerlik gösteren çiftler göz ardı edilmiştir.

Tablo 16. *100 dilli örneklemde benzerlik oranları*

Diller	Benzerlik (%)	
	Önce	Sonra
<i>Tay Dili-Vietnam Dili</i>	84,09	86,00
<i>Hmong Njua-Tay Dili</i>	82,73	85,16
<i>Maung Dili-Tiwi Dili</i>	79,33	82,73
<i>Hmong Njua-Vietnam Dili</i>	79,49	82,22
<i>Svahili Dili-Zulu Dili</i>	79,23	80,28
<i>Luvale Dili-Svahili Dili</i>	74,78	76,19
<i>Khalka Dili-Türkçe</i>	73,61	75,64
<i>Endonezce-Vietnam Dili</i>	72,54	75,62
<i>Luvale Dili-Zulu Dili</i>	74,79	75,38
<i>Bagirmi Dili-Sango Dili</i>	71,17	75,19

²¹ Betimsel bulgular için *psych* (Revelle, 2020) paketinin *describe()* fonksiyonundan yararlanılmıştır.

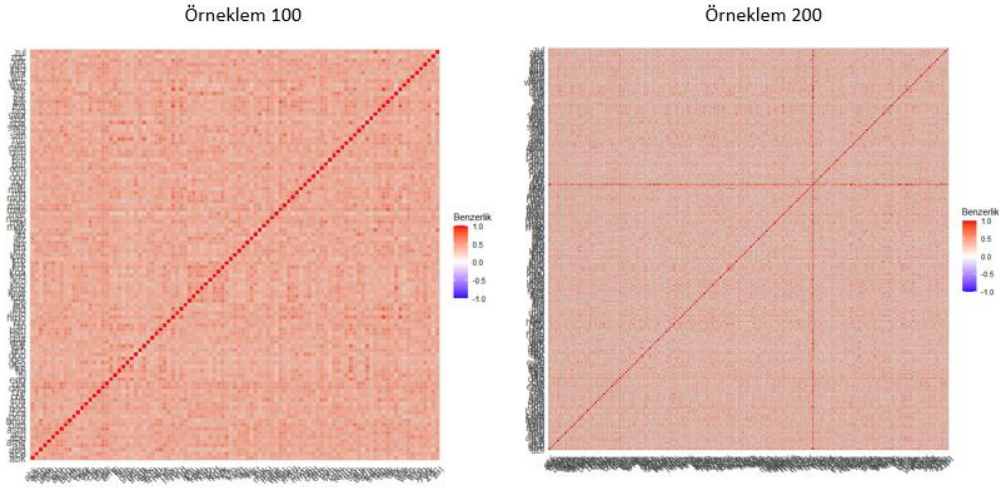
Tablo 17. 200 dilli örnekleme benzerlik oranları

Diller	Benzerlik (%)	
	Önce	Sonra
<i>Tay Dili-Vietnam Dili</i>	84,09	86,00
<i>Alamblak Dili-Nenets (Forest)</i>	50,00	85,71
<i>Bambara-Nenets (Forest)</i>	50,00	85,71
<i>Çukçi Dili-Nenets (Forest)</i>	50,00	85,71
<i>Komançi Dili-Nenets (Forest)</i>	50,00	85,71
<i>Diola Fogny-Nenets (Forest)</i>	50,00	85,71
<i>Dani (Aşağı Büyük Vadi)-Nenets (Forest)</i>	50,00	85,71
<i>Drehu Dili-Nenets (Forest)</i>	50,00	85,71
<i>Garö Dili-Nenets (Forest)</i>	50,00	85,71
<i>Almanca-Nenets (Forest)</i>	50,00	85,71

Dillerin genel benzerlikleri ile özellik alanları bazındaki benzerlikleri de her zaman birbirine koşut değildir. Bununla ilgili veri dökümü ek materyallerde sunulmuştur (bkz. *similarityArea100.csv* ve *similarityArea200.csv*) ve *Tablo 18*'de en yüksek benzerlik gösteren Tay dili-Vietnam dili çiftine ilişkin bulguları gösteren bir kesit sunulmuştur.

Tablo 18. Alanlar bazında benzerlik tablosu (Tay dili-Vietnam dili örneği) (%)

Tay-Vietnam		Normalizasyon	Sesbilim	Biçimbilim	Adıl Uamlar	Adıl Sözdizim	Eylemcil Uamlar	Sözcük Dizilişi	Basit Tümcecikler	Karmaşık Tümcecikler	Sözlükçe	Genel
	Önce	73,33	66,67	81,82	62,50	88,24	96,42	88,46	100	100	84,09	
	Sonra	75,00	66,67	82,61	62,50	88,24	97,73	88,46	100	100	86,00	



Şekil 11. Örneklerin normalizasyon sonrası benzerlik görselleştirmesi²²

Şekil 11’de ise benzerlik değerlerinin daha kolay görülebilmesi için normalizasyon sonrası veriler üzerinden bir görselleştirme uygulanmıştır. Benzerlik seviyesi ne kadar yüksek ise o kadar koyu kırmızıyla gösterilen bu görselleştirme, genel resimde akrabalık ilişkisine bağlı olmaksızın dillerarası benzerliğin belli bir oranda olduğunu göstermektedir. 200 dilli örnekte neredeyse tüm diller ile yüksek benzerlik gösteren Nenets (Forest) dili ise yukarıda değinildiği gibi sınırlı veri nedeniyle ölçümü saptırmaktadır.

Şekil 12 ve Şekil 13 her iki örnek için değişim oranını ve değişimin yönünü göstermektedir. Eğer değişim negatif yönlüyse yani benzerlik oranı düşüyse mavi ve tonlarıyla, değişim pozitif yönlüyse yani benzerlik oranı yükseldiyse kırmızı ve tonlarıyla, herhangi bir değişim olmadıysa gri olarak renklendirilmiştir.²³

²² Görselleştirme için *ggcorrplot* (Kassambra, 2019) paketinden yararlanılmıştır.

²³ Görselleştirme için *ggcorrplot* (Kassambra, 2019) paketinden yararlanılmıştır.

Son olarak, yukarıdaki görselleştirmeyi desteklemek amacıyla *Tablo 19* ise tutarlılık sınamasında olduğu gibi 200 dilli örneklem üzerinde aile olan ve olmayan dillerin karşılıklı benzerlik ortalamalarını normalizasyon sonrasındaki sıralamaya göre sunmaktadır.

Bu bölüme ilişkin veri dökümleri ek materyallerde *results* klasörü içinde, görselleri ise *figures* klasörü içinde bulunmaktadır.

Tablo 19. Ailelerin karşılıklı benzerlik ortalamaları

Aile	Dil Sayısı	Benzerlik (%)	
		Önce	Sonra
<i>Karayip</i>	2	50,88	55,35
<i>Altay</i>	3	53,93	56,82
<i>Ural</i>	3	69,05	71,02
<i>Nah-Dağıstan</i>	4	56,83	59,93
<i>Oto-Menguen</i>	3	58,14	60,42
<i>Na-Dene</i>	3	71,08	73,03
<i>Trans-Yeni Gine</i>	10	43,04	45,35
<i>Avustronezya</i>	15	48,91	53,18
<i>Avusturo-Asya</i>	6	57,75	59,67
<i>Hint-Avrupa</i>	11	56,65	61,29
<i>Pama-Nyungan</i>	4	68,02	70,66
<i>Uto-Aztek</i>	4	56,29	59,22
<i>Nijer-Kongo</i>	13	63,97	70,56
<i>Algonkin</i>	3	58,42	59,61
<i>Çin-Tibet</i>	8	46,62	51,09
<i>Afro-Asya</i>	8	52,82	56,00
<i>Tupiyen</i>	2	57,22	61,00
<i>Chibchan</i>	3	57,45	55,14
<i>Penutiyen</i>	3	22,52	70,95
<i>Orta Sudan</i>	2	54,71	59,57
Aile Olmayan Diller	89	44,13	48,26

5. TARTIŞMA VE SONUÇ

5.1. Veritabanına İlişkin Gözlemler

WALS'e dair eksik veri sorununun alanyazında farklı araştırmacılar tarafından dile getirildiğini ve farklı çözüm önerileri bulunduğunu belirtmiştik. Gerçekten de yapısal özellikleri hakkında bilgi sahibi olunmamış ve yeterince araştırılıp yeterli oranda betimlenmemiş pek çok dil bulunmakla birlikte, bu sorunun WALS'in sorunu olmaktan öte tipoloji ve dilbilgisi araştırmacılarının sorunu olduğunu söylemek gerekmektedir. Elbette ki bir dilin konuşucu sayısı, konuşulduğu coğrafi bölge(ler) ve o dilin pratik yaşamdaki karşılığı gibi pek çok etmen çalışmaların eğilimini ve şiddetini yönlendirmektedir. Özellikle konuşur sayısı oldukça az olan ve merkezi şehir yaşantılarından uzak bölgelerde konuşulan dilleri araştırmak fazladan çaba, emek ve bütçe gerektirmektedir. Bu çerçeveden bakıldığında, WALS'in 2662 dil için sunduğu ve toplam veritabanının yaklaşık olarak beşte birini dolduran bu veri, bu haliyle bile oldukça değerli bir kaynak niteliği taşımaktadır.

Bununla birlikte, veri önışleme ve eksik veri analizi adımlarında yapılan gözlemler ve elde edilen bulgular göstermektedir ki WALS'in veri yapısından kaynaklanan bazı etmenler veritabanının eksik veri oranını gerçekte olduğundan daha fazla görünmesine neden olmaktadır. 3.2. *Eksik Veri Analizi ve normalize()* başlığında ayrıntılarıyla açıklanan süreç sonucunda elde edilen yeni verinin başlangıçtaki veriye göre daha az eksik veri içerdiği bulgulanmıştır. Ayrıca geliştirdiğimiz *normalize()* fonksiyonu, dilbilimin farklı uzmanlık alanlarına hakim kişilerin de katkısıyla ileri çalışmalarda daha da geliştirilebilir ve bu da eksik veri oranının daha da düşmesine neden olacaktır.

Eksik verinin doğru analiz edilmesi, tipolojik çözümlemeler üzerinde de önemli bir etkiye sahiptir. Bu noktada, sözgelimi *10B Batı Afrika'da Genizsil Ünlüler* başlıklı

özellik çerçevesinde bir değerlendirme yapılacak olunursa, Batı Afrika dillerine özel olan bu özelliğin Batı Afrika dillerinden olmayan diller için etiketlenme biçimi, sonuçları etkilemesi bakımından önemlidir. Bu özellik çerçevesinde Türkçe ve İngilizce gibi Batı Afrika dili olmayan iki dil alınacak olursa, WALS verisinde olduğu gibi iki dili de bu özellik açısından boş bırakmaktansa ikisi için de bu özelliğin geçersiz olduğunu etiketlemek bu iki dilin karşılıklı benzerliği hesaplanırken daha anlamlı bir bulguya işaret etmektedir. Yine, sözelimi bir Batı Afrika dili olan Doyayo diliyle Batı Afrika dili olmayan Türkçe arasındaki farklılığı göstermesi açısından da iki dilin de değer alması gerekmektedir. Oysa yayında olan mevcut sürümdeki verinin etiketleme yöntemi Batı Afrika dili olmayan dilleri etiketsiz bırakarak karşılaştırmayı da olanaksız bırakmaktadır.

WALS'in genel yapısıyla ilgili diğer bir nokta da 2.3.1.2. *Özellikler* başlığında da belirttiğimiz gibi özelliklerin kategorik, sıralı ve ikili olmak üzere farklı biçimlerde yapılandırılmış olmasıdır. Bu durum, özellikle 2.3.2.3. *Karşılıklı Benzerlik/Benzemezlik* başlığında aktarılan Uzun'un (2012a) karşılıklı benzemezlik analizi çerçevesinde önemli bir yere sahiptir. Bu analizin, tüm özellikler için yapılabilmesi, ancak veritabanının *bir noktalı kodlama* (one-hot encoding) gibi farklı yöntemlerle yeniden düzenlenmesi sonucunda olanaklı olabilecektir. Çalışmamızın kapsamını ve süresini aşan bu analize, ileri çalışmalarda ele alınmak üzere burada yer verilmemişse de bir yöntem önerisi olarak özelliklerin birbiriyle ilişkisinin betimlendiği bir sözlük yapısı oluşturulması önerilebilir. Böylece çözümleyici, işlem sırasında oluşturulan sözlüğü bir kılavuz olarak kullanıp değerlendirmesini bu çerçevede gerçekleştirebilecektir. Bununla birlikte böyle bir yöntem önerisinin somutlaşabilmesi, WALS'in çok alanlı bir veritabanı olması nedeniyle ancak WALS'in veri sunduğu farklı alanlarda çalışan dilbilimcilerin ortak çalışmasıyla mümkün olabilecektir.

5.2. Tipolojik Karşılaştırma Sonuçları

Bu başlıkta üç farklı yöntem çerçevesinde ulaşılan sonuçlar altbaşlıklar biçiminde ayrı ayrı sunulacaktır.

5.2.1. Tipiklik

Tipiklik, bir dilin bir özellik çerçevesinde aldığı değerin dünya dilleri arasındaki dağılımına göre hesaplanan bir işlem olduğunu anımsanacak olursa ham verinin bu analiz için yanıltıcı bir yönünün olduğu kesindir. Örneğin; *81B İki Baskın Özne, Nesne ve Eylem Dizilişli Diller* özelliği dünya dillerinde sayıca az dilde görünen bir özelliktir. Dillerin büyük bir bölümü *81A Özne, Nesne ve Eylem Dizilişi* özelliğinde belirtildiği gibi baskın bir dizilişe sahiptir. Bu doğrultuda, *81B* için en çok değere sahip olan ÖNE veya ÖEN değerini alan dillere “tipik” demek *81A*’da baskın bir diziliş alan dilleri görmezden gelmek demektir. Normalizasyon işlemi sonucunda *81B*’ye veritabanının kendisinde bulunmayan 0 değerleri eklendiğinde tipiklik değeri daha doğru bir veri üzerinden hesaplanabilmiştir. Bu durum normalize edilmiş pek çok özellik için de geçerlidir.

Ayrıca tipikliğin her durumda $\square\%45$ seviyesinde seyretmesi, daha büyük örneklem çerçevesinde de aynı sonuca ulaşılması koşuluyla, tipiklik ölçümünün tipolojik sınıflandırma için anlamlı bir ölçek olmadığını göstermektedir. Hatta tam aksine tipiklik, dillerin bütünsel tipoloji çerçevesinde sınıflandırılmayacağını gösteren bir ölçek gibi davranmaktadır çünkü her dil aralarında küçük farklar olsa bile neredeyse aynı sonuçlar vermektedir.

5.2.2. Tutarlılık

Tutarlılığın bir dilin kendisine benzer özellikler gösteren diller ile benzerliğini sürdürme kararlılığı olduğu da söylenebilir. Bu durumda 100 dillik bir örnekleme bile çok sayıda altkümeyle analiz yapılabilir, nitekim tezde sunduğumuz çözümleme aracı da bu işlem için kullanılabilir. Ancak bu devasa bir veri dökümünü ve bunun üzerinden farklı bir analizi gerektirir, ki bu bambaşka bir çalışma konusu da olabilir. Belirtilen nedenlerle, çalışmamızda aile/akraba olmanın etkisi öncelikli ölçüm değişkeni olduğu için, tutarlılık analizi sadece akraba olan diller ve akraba olmayan diller üzerinde yapılmıştır.

Elde edilen bulgular, aile üyelerinin kendi aralarındaki tutarlılığın $\square 111$ ile 40 puan arasında seyretmesi nedeniyle bu noktada bir anlamlılık yokmuş gibi bir izlenim oluştursa da dillerin veritabanındaki etiket oranlarının da ölçümü etkilediği düşünülmektedir. Öte yandan; akraba olmayan ve hatta izole dil olarak değerlendirilen dillerin yüksek tutarlılık göstermesi de incelenmesi gereken başka bir soru olarak ortaya çıkmıştır. İlgili sınamalar, sunduğumuz araç yardımıyla etiket ve dil sayıları gibi değişkenler sabitlenerek yinelenip bu yönde ileri çalışmalar yapılabilir.

5.2.3. Karşılıklı Benzerlik

Karşılıklı benzerlik açısından bakıldığında, veritabanındaki örneklemelerde tüm diller için belli bir oranda benzerlik görüldüğü ve bu benzerlik üzerinde aile/akrabalık durumunun belirleyici bir etkisinin olmadığı rahatlıkla söylenebilir.

Buna ek olarak, hakkında az veri girilmiş diller bir başka dille aynı özellikler çerçevesinde benzerlik gösterdiğinde yöntemin aynı özelliklerin sayısına oranlama yönergesi gereği %100 gibi bir karşılıklı benzerlik oranı vermektedir. Ancak bu durum çalışmamızda olduğu gibi, büyük örneklemeler üzerinden çalışıldığında istatistiksel olarak sapmalara ve yanıltıcı sonuçlara neden olmaktadır. Sözelimi 100 özellikte etiketlenen

herhangi bir A dili ile yalnızca A dilinin etiketlendiği özelliklerden ikisi için etiketlenen ve A ile aynı değeri alan B dili göz önünde tutulursa –ki yöntem böyle önermektedir– bu iki dil birbirine %100 benziyor gibi görünmektedir. Ancak A dilinin diğer 98 özelliğinin ölçümde bir yeri olmamalı mıdır? Bu sorunu çözmek için, ya karşılaştırılan her iki dilin tüm özelliklerinin veya WALS’ın tüm özellik sayısının formüle katılması ya da daha doğru bir çözüm yolu olarak iki dili karşılaştırmak için belli bir veri doluluk oranı aranması önerilebilir. Bu eşik oranı ise standart bir ortalama hesabı olmanın ötesinde özelliklerin etkisi ve dağılımı dikkate alınarak ileri çalışmalar ile belirlenmelidir.

Normalizasyon işlemi kapsamında eklenen 0 değerinin, tipiklik için olduğu kadar karşılıklı benzerlik analizi için de sonucu etkileyen bir değişken olduğu düşünülmektedir. Konunun daha kolay anlaşılması için tartışmayı aynı özellik üzerinden yürütecek olursak, *81A Özne, Nesne ve Eylem Dizilişi* özelliğine oranla *81B İki Baskın Özne, Nesne ve Eylem Dizilişli Diller* özelliği dünya dillerinde sayıca daha az dilde görülmektedir. Bu durumda, bu özelliğin görülmediği herhangi iki dil de aslında özelliğin görülmemesi açısından benzemektedir. Ancak WALS’ın mevcut yapısında bu özelliğin herhangi bir değerine sahip olmayan bir dil etiketsiz bırakıldığından bu diller karşılıklı benzerlik açısından hesaplanamaz durumda kalmıştır. Bu durum da normalize edilmiş pek çok özellik için geçerlidir.

5.3. Sonuç ve Öneriler

Çalışmamızın ilk adımında, WALS verisinin ham hali üzerine yapılan gözlemler sonucunda tipoloji ölçümlerini etkileyen bazı yapısal sorunlar olduğu gözlemlenmiş ve bunlar normalizasyon adı verilen bir işlem ile düzeltilmeye çalışılmıştır. Ardından alanyazındaki tipolojik sıralama yöntemleri, R dilinde fonksiyonlaştırılmış ve son olarak örneklemeler üzerinde ilgili yöntemler sınanmıştır.

Elde ettiğimiz tüm bulgular doğrultusunda çalışmamız kapsamında ele aldığımız araştırma sorularıyla ilgili olarak ulaşılmış yanıtlara sırasıyla yer vermek gerekirse;

i. WALS için alanyazında önerilmiş olan çözümleme yöntemleri, R dilinde fonksiyonlara kısmen dönüştürülebilmüş ve bilgisayar desteği kullanılarak çözümlemenin daha sistematik, hızlı ve kolay bir hale getirilmesi sağlanmıştır. Öte yandan, WALS veri yapısından kaynaklanan bazı sorunlar nedeniyle karşılıklı benzerlik analizi gibi kimi süreçler işlenememiştir. Bunlar daha uzun süreli ve ayrıntılı ele alınmak üzere ileri çalışmalara bırakılmıştır.

ii. Alanyazında önerilmiş olan üç yöntem, WALS'in 100 ve 200 dil içeren örneklemi üzerinde sınanmış ve elde edilen bulgular önceki bölümdeki başlıklarda ayrıntılı olarak paylaşılmıştır. Elde edilen betimsel bulgular veya çözümlemeye dahil edilebilecek olan potansiyel yeni yöntemlerden edinilecek olan bulgular, ileri çalışmalarda karşılıklı olarak incelenip yine bilgisayar desteği kullanılarak birtakım korelasyonlar bulgulanabilir. Dolayısıyla, kullanılan üç yöntemin birbirinin yerine bir alternatif olarak değil, birlikte kullanılacak araçlar olarak değerlendirilmesinin daha doğru olacağı düşünülmektedir.

iii. Dillerin aynı aileden/akraba olması ile yapısal benzerlikleri arasında doğrudan bir ilişki olabilmesi için söz konusu akraba dillerin yöntemlerin her birinde benzer davranışlar göstermesi gerekir, ancak bulgular göstermektedir ki aile/akraba olmayan diller tipiklik, tutarlılık ve karşılıklı benzerlik çerçevesinde aile/akraba olan diller kadar ortak davranış sergileyebilmektedir. Dolayısıyla, bu iki durum arasında anlamlı ve doğrudan bir ilişki söz konusu değildir.

Çalışmamız, Türkçe alanyazında önerilen ve kullanılan WALS çözümleme yöntemlerini bilgisayar ortamına taşıyan ve bu sayede hesaplamalı tipoloji incelemesi yapan öncü bir çalışma niteliği taşımaktadır. Bu yönüyle, üç yöntem ile iki örneklem

üzerinde yapılmış çözümlerler genel bir çerçeve çizse de ileri çalışmalar için de pek çok araştırma sorusu da bırakılmıştır. Örneğin;

- WALS üzerinden farklı örneklemeler çekilip çözümleme yinlendiğinde de aynı doğrultuda bulgulara ulaşılacak mıdır?
- Üç yöntem ve dahil edilebilecek potansiyel yeni yöntemlerin bulguları arasında korelatif ilişkiler saptanabilir mi?
- WALS veri yapısı çözümlerler için ne kadar uygundur? Ne gibi geliştirme ve dönüştürme işlemleri yapılabilir?

soruları bunlardan yalnızca birkaçıdır.

Buna ek olarak, aynı yöntemlerin diğer CLLD veritabanlarında veya araştırmacıların kendilerinin oluşturduğu veritabanlarında da uygulanması hesaplamalı tipoloji çalışmaları açısından büyük bir öneme sahiptir. Bu çalışmalar, belli bir aşamaya ulaştıktan sonra, yalnızca bir veritabanı üzerinde değil birden çok veritabanı arasında ilişkili analizler de yapılabilecektir.

Son olarak, bu tez çalışması kapsamında geliştirilen fonksiyonların tez sürecinin tamamlanmasının ardından, gerekli dokümantasyon işlemleri yapılarak bir paket biçiminde R evrenine katılması sağlanacaktır. Böylece, tipoloji araştırmacıları analiz aracını kendi çalışmalarında kullanabilme olanağı kazanmış olacaklardır.

KAYNAKÇA

- Ağbaht, M. (2014). *Hatay yerleşik Arap diyalektleri: WALS'e göre dil özellikleri* (Tez No. 368460) [Yüksek lisans tezi, Ankara Üniversitesi]. YÖK Ulusal Tez Merkezi.
- Albu, M. (2006). *Quantitative analysis of typological data* [Doktora tezi, Leipzig Üniversitesi]. MPG Publication Repository. <http://hdl.handle.net/11858/00-001M-0000-0012-8C09-C>
- Anderson, S. R. (2010). *How many languages are there in the world?* [Broşür]. Linguistic Society of America. <https://www.linguisticsociety.org/sites/default/files/how-many-languages.pdf>
- Başkan, Ö. (2003). *Lengüistik metodu* (3. baskı). Multilingual Yayınları.
- Ben Hamed, M. (2015). Phylo-linguistics: Enacting Darwin's linguistic image. T. Heams, P. Huneman, G. Lecointre, M. Silberstein (haz.) *Handbook of evolutionary thinking in the sciences* (ss. 825–852) içinde. Springer. https://doi.org/10.1007/978-94-017-9014-7_39
- Bender, E. M., Drellishak, S., Fokkens, A., Poulson, L. ve Saleem, S. (2010). Grammar customization. *Research on Language and Computation*, 23–72.
- Bender, E. M., Flickinger, D. ve Oepen, S. (2002). The grammar matrix: An open-source starter-kit for the rapid development of crosslinguistically consistent broad-coverage precision grammars. *International Conference on Computational Linguistics 19 (Workshop on Grammar Engineering and Evaluation)*. 8–14. <http://www.aclweb.org/anthology/W02-1502>
- Bozşahin, C. ve Zeyrek, D. (2000). Dilbilgisi, bilişim ve bilişsel bilim. *Dilbilim Araştırmaları*. Simurg Yayınevi. <http://users.metu.edu.tr/bozsahin/dilaras99.pdf> adresinden erişilmiştir.
- Chomsky, N. (1957). *Syntactic structures*. Mouton.
- Chomsky, N. (1981). *Lectures on government and binding*. Mouton de Gruyter.

- Comrie, B. (1989). *Language universals and linguistic typology: Syntax and morphology* (2. baskı). The University of Chicago Press.
- Comrie, B. (2010). Japanese and the other languages of the world. *NINJAL Project Review*, 1, 29–45.
- Comrie, B. (2013). 15. Holistic versus partial typologies. F. Kiefer ve P. van Sterkenburg (haz.), *Eight decades of general linguistics* (ss. 439–452) içinde. Brill.
- Comrie, B., Dryer, M. S., Gil, D. ve Haspelmath, M. (2013). Introduction. M. S. Dryer ve M. Haspelmath (haz.), *The World Atlas of Language Structures Online* içinde. Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/s1>
- Croft, W. (2002). *Typology and universals* (2. baskı). Cambridge University Press.
- Cysouw, M. (2008). Using the World Atlas of Language Structures. *Introduction to the Special Issue of Sprachtypologie und Universalienforschung (STUF)*, 60(2), 181–185
- Darwin, C. (1859). *On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life*. John Murray.
- Daumé III, H. (2009). Non-parametric Bayesian areal linguistics. M. Ostendorf, M. Collins, S. Narayanan, D. W. Oard ve L. Vanderwende (haz.), *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (ss. 593–601) içinde. Association for Computational Linguistics. <https://www.aclweb.org/anthology/N09-1067>
- Daumé III, H. ve Campbell, L. (2007). A Bayesian model for discovering typological implications. A. Zaenen ve A. van den Bosch (haz.), *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics* (ss. 65–72) içinde. Association for Computational Linguistics. <https://www.aclweb.org/anthology/P07-1009>

- Daval-Markussen, A. (2011). Of networks and trees in contact linguistics: new light on the typology of creoles [Yüksek lisans tezi, Aarhus Universitet].
- Dryer, M. S. (2013). Order of subject, object and verb. M. S. Dryer ve M. Haspelmath, (haz.), *The World Atlas of Language Structures Online* içinde. Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/81>
- Dryer, M. S. ve Haspelmath, M. (haz.). (2013). *The World Atlas of Language Structures Online*. Max Planck Institute for Evolutionary Anthropology, <https://wals.info/>
- Eberhard, D. M., Simons, G. F. ve Fennig, C. D. (haz.). (2021). *Ethnologue: Languages of the World* (24. baskı). SIL International. <http://www.ethnologue.com>
- Fellbaum, C. (haz.). (1998) *WordNet: An electronic lexical database*. MIT Press.
- Forkel, R., List, J. M., Greenhill, S. J., Rzymiski, C., Bank, S., Cysouw, M., Hammarström, H., Haspelmath, M., Kaiping, G. A. ve Gray, R. D. (2018). Cross-Linguistic Data Formats, advancing data sharing and re-use in comparative linguistics. *Sci Data*, 5, 180205. <https://doi.org/10.1038/sdata.2018.205>
- François, A. (2014). Trees, waves and linkages: Models of language diversification. Bower, C. ve Evans, B. (haz.), *The Routledge handbook of historical linguistics* (ss. 161–189) içinde. Routledge.
- François, A. ve Ponsonnet, M. (2013). Descriptive linguistics. J. R. McGee ve R. L. Warrms (haz.), *Theory in social and cultural anthropology: an encyclopedia* (ss. 184–187) içinde. SAGE Publications, Inc.
- Gray, R. D. ve Jordan, F. M. (2000). Language trees support the express train sequence of Austronesian expansion. *Nature*, 405, 1052–1055.
- Greenberg, J. H. (1966). Some universals of grammar with particular reference to the order of meaningful elements. J. H. Greenberg (haz.), *Universals of grammar* (ss. 77–173) içinde. MIT Press.

- Greenberg, J. H., Osgood, C. E. ve James, J. J. (1963). *Memorandum concerning language universals*. MIT Press.
- Grishman, R. (1994). *Computational linguistics: an introduction*. Cambridge University Press.
- Haegeman, L. (1994). *Introduction to Government and Binding Theory* (2. baskı). Blackwell.
- Hajek, J. (2013). Vowel nasalization. M. S. Dryer ve M. Haspelmath (haz.) *The World Atlas of Language Structures Online* içinde. Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/10>
- Haspelmath, M., Dryer, M. S., Gil, D. ve Comrie, B. (2005). *The world atlas of language structures*. Oxford University Press.
- Haspelmath, M. ve Tadmor, U. (haz.). (2009). World Loanword Database. Max Planck Institute for Evolutionary Anthropology. <http://wold.clld.org>
- Kassambara, A. (2019). *ggcorrplot: Visualization of a Correlation Matrix using 'ggplot2'*. R package version 0.1.3. <https://CRAN.R-project.org/package=ggcorrplot>
- Kıran, Z. (1996). *Dilbilim akımları*. Onur Yayınları.
- Kıran, Z. ve Eziler Kıran, A. (2018). *Dilbilime giriş* (5. Baskı). Seçkin Yayıncılık.
- Lewis, W. D. (2006). ODIN: A model for adapting and enriching legacy infrastructure. *IEEE International Conference on E-Science*, 2, 137.
- Lewis, W. D. ve Xia, F. (2010). Developing ODIN: A multilingual repository of annotated language data for hundreds of the world's languages. *Journal of Literary and Linguistic Computing*, 25, 303-319.
- Lu, X. (2013). Exploring word order universals: a probabilistic graphical model approach. A. Dey, S. Krause, I. Nikolova, E. Vecchi, S. Bethard, P. I. Nakov ve F. Xu (haz.), *51st Annual Meeting of the Association for Computational Linguistics*

- Proceedings of the Student Research Workshop* (ss. 150–157) içinde. Association for Computational Linguistics. <https://www.aclweb.org/anthology/P13-3022>
- Miller, G. A. (1995). WordNet: A Lexical Database for English. *Communications of the ACM*, 38(11), 39-41.
- Moran, S. ve McCloy, D. (haz.). (2019). PHOIBLE 2.0. *Max Planck Institute for the Science of Human History*. <http://phoible.org/>
- Nijenhuis, A. ve Wilf, H. S. (1978). *Combinatorial algorithms for computers and calculators*. Academic Press.
- Orgun, O. (2011). How typologically coherent is Turkish?. E. Erguvanli-Taylan ve B. Rona (haz.), *Puzzles of Language: Essays in Honour of Karl Zimmer - Turcologica 86* (ss. 157–168) içinde. Harrassowitz.
- Özgen, M. ve Koşaner, Ö. (2020). *Dünya dillerinden örneklerle dilbilimsel tipoloji* (5. baskı). Pegem Akademi Yayıncılık.
- Pollard, C. ve Sag, I. A. (1994). *Head-Driven Phrase Structure Grammar*. University of Chicago Press.
- R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Viyana, Avusturya. <https://www.R-project.org/>
- Ramat, R. (2011). The (early) history of linguistic typology. J. J. Sung (haz.), *The Oxford handbook of linguistic typology* (ss. 9–24) içinde. Oxford University Press.
- Revelle, W. (2020) *psych: Procedures for Personality and Psychological Research*, Northwestern Üniversitesi, Evanston, Illinois, Amerika Birleşik Devletleri, <https://CRAN.R-project.org/package=psychVersion=2.1.3>
- Richards, R. J. (2002). The linguistic creation of man: Charles Darwin, August Schleicher, Ernst Haeckel, and the missing link in nineteenth-century

- Evolutionary Theory. M. Dörries (haz.), *Experimenting in tongues* (ss. 21–175) içinde. Stanford University Press.
- Sankur, B. (2004). *İngilizce-Türkçe ansiklopedik bilişim sözlüğü*. Pusula Yayıncılık.
- Schmidt, J. F. H. (1872). *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen*. Weimar Hermann Böhlac.
- Schleicher, A. (1861). *Compendium der vergleichenden Grammatik der indogermanischen Sprachen*. Weimar.
- Speer, R., Chin, J. ve Havasi, C. (2017). ConceptNet 5.5: An open multilingual graph of general knowledge. *Proceedings of AAAI*, 31, 4444–4451.
- Seifart, F. (2020). AfBo: A world-wide survey of affix borrowing. Max Planck Institute for Evolutionary Anthropology. <https://doi.org/10.5281/zenodo.3610155>
- Takamura, H., Nagata, R. ve Kawasaki, Y. (2016). Discriminative analysis of linguistic features for typological study. N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk ve S. Piperidis (haz.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (ss. 69–76) içinde. European Language Resources Association (ELRA). <https://www.aclweb.org/anthology/L16-1011>
- Tashmurad, S. (2015). *Türk dillerinin tipolojik karşılaştırılması: WALS veri tabanına göre Türk dillerindeki benzerlikler ve farklılıklar* (Tez No. 393356) [Yüksek lisans tezi, Ankara Üniversitesi]. YÖK Ulusal Tez Merkezi.
- Tierney, N., Cook, D., McBain, M. ve Fay, C. (2020). *naniar: Data Structures, Summaries, and Visualisations for Missing Data*. R package version 0.6.0. <https://CRAN.R-project.org/package=naniar>
- Tufis, D., Christea, D. ve Stamou, S. (2004). BalkaNet: aims, methods, results and perspectives. a general overview. *Romanian Journal of Information Science and Technology*, 7(1–2), 9–43.

- Uzun, N. E. (2012a, 18–21 Eylül). WALS-tabanlı tipolojik sıralamalar için bir öneri [Bildiri sunumu]. 16. *Uluslararası Türk Dilbilim Kurultayı*, Orta Doğu Teknik Üniversitesi, Ankara, Türkiye.
- Uzun, N. E. (2012b, 24-26 Mayıs). WALS Yapabilir Miyiz? [Bildiri sunumu]. 26. *Ulusal Dilbilim Kurultayı*, Süleyman Demirel Üniversitesi, Isparta, Türkiye.
- Venables, W. N., Smith, D. M. ve R Core Team. (2021). *An introduction to R*. <https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf>
- Ward, G. L. ve Prince, E. F. (1991). On the topicalization of indefinite NPs. *Journal of Pragmatics*, 16, 167-177.
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Golemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K. ve Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686, <https://doi.org/10.21105/joss.01686>
- Xia, F., Lewis, W. D., Goodman, M. W., Slayden, G., Georgi, R., Crowgey, J. ve Bender, E. M. (2016). Enriching a massively multilingual database of interlinear glossed text. *Language Resources and Evaluation*, 50, 321-349.

ÖZET

Dil tipolojisi, geçmişî karşılaştırmalı dilbilgisi çalışmalarına dek uzanan ve zaman içinde hem kuramsal hem de yöntembilimsel olarak değışim gösteren bir dilbilim çalışma alanıdır. Hiç kuşkusuz ki teknolojik olanakların artması ve veriye erişimin kolaylaşması bu alanı da olumlu yönde etkilemiştir. Bu çerçevede sayılabilecek WALS, dillerin yapısal özelliklerinin etiketlendiğı bir elektronik veritabanıdır. Alanyazında WALS verisi kullanılarak yapılabilecek sıralamalar için çeşitli yöntem önerileri bulunmaktadır. Çalışmamızda bu yöntemlerden üçü (Comrie, 2010; Orgun, 2011; Uzun, 2012a) R ortamında kullanılabilecek bir paket biçimine getirilerek örneklemeler üzerinden tipolojik analiz yürütölmüştür. Araştırma sorularımız çerçevesinde; (i) WALS veri yapısının bilgisayar destekli çalışmaları zorlaştıran bazı yönlerinin bulunduğı ve ileri çalışmalarla ele alınması gerektiğı, (ii) üç yöntemin ve potansiyel yeni yöntemlerin bulgularının karşılaştırmalı olarak ele alınmasının tipolojik araştırmalar çerçevesinde yeni sonuçlara ulaştırabileceğı, (iii) aile/akraba olmanın “tipiklik”, “tutarlılık” ve “karşılıklı benzerlik” çerçevesinde doğrudan ve baskın bir etkisinin olmadığı sonuçlarına ulaşılmıştır. Türkçe alanyazında öncü bir hesaplamalı tipoloji çalışması niteliğı taşıyan bu çalışmanın bulguları ileri çalışmalarla desteklenip geliştirilebilir.

Anahtar Sözcükler: betimsel analiz, hesaplamalı dilbilim, R, tipoloji, WALS

ABSTRACT

Language typology is a field of study in linguistics that dates back to comparative grammar studies, and has changed both theoretically and methodologically over time. It is surely beyond doubt that, the increase in technological opportunities and the ease of access to data have also positively affected this field. WALS, which can be considered within this framework, is an electronic database in which the structural features of languages are tagged. In the literature, there are various method suggestions for preparing rankings using WALS data. In our study, three of these methods (Comrie, 2010; Orgun, 2011; Uzun, 2012a) were brought into a package format that can be used in the R environment, and typological analysis was carried out on the samples. Within the scope of our research questions, it has been concluded that; (i) WALS data structure has some aspects that complicate computer-aided studies and should be addressed with further studies, (ii) comparative analysis of the findings of the three methods and potential new methods can lead to new results within the framework of typological studies, (iii) there is no direct and dominant effect of being family/relative within the framework of “typicality”, “coherence” and “mutual similarity”. The findings of this study, which is a pioneering computational typology study in Turkish literature, can be supported and improved by further studies.

Keywords: *computational linguistics, descriptive analysis, R, typology, WALS*

TÜRKÇE-İNGİLİZCE TERİMLER

absolute universal	mutlak evrence
acceptable	kabuledilebilir
affixation	eklenme
area	alan
argument	üye
binary	ikili
coherence	tutarlılık
comparative grammar	karşılaştırmalı dilbilgisi
complex sentences	karmaşık tümceler
computational	hesaplamalı
computational linguistics	hesaplamalı dilbilim
computational typology	hesaplamalı tipoloji
Cross-Linguistic Data Formats	Dillerarası Veri Formatları
Cross-Linguistic Linked Data	Dillerarası Bağlantılı Veri
data engineering	veri mühendisliği
data frame	veri çerçevesi
data point	veri noktası
data science	veri bilim
descriptive grammar	betimleyici dilbilgisi
distance	uzaklık
diversity linguistics	çeşitlilik dilbilim
formal universal	biçimsel dilbilim
function	fonksiyon
Generative Grammar	Üretici Dilbilgisi

holistic typology	bütüncül tipoloji
implicational universal	sezdirimsel evrence
inflectional	çekimsel
interlinear glossing	satırarası açıklama
language faculty	dil yetisi
language typology	dil tipolojisi
lexicon	sözlükçe
library	kütüphane
linguistic typology	dilbilimsel tipoloji
missing data	eksik veri
morphology	biçimbilim
mutual similarity/dissimilarity	karşılıklı benzerlik/benzemezlik
native speaker	anadili konuşuru
natural language processing	doğal dil işleme
nominal categories	adıl ulamlar
nominal syntax	adıl sözdizim
non-implicational universal	sezdirimsel olmayan evrence
one-hot encoding	bir noktalı etiketleme
ordinal	sıralı
other	diğer
package	paket
para-linguistic	dıldışı/dilötesi
parameter	değiştirgen
partial typology	parçasal tipoloji
phonology	sesbilim
phylogenetics	filogenetik

phylolinguistics	filolinguistik
principle	ilke
proto-language	kök dil
reconstruction	yeniden kurma
relative clause	ilgi tümceciği
restricted equivalence universals	kısıtlı eşdeğerlik evrenceleri
sign languages	işaret dilleri
simple clauses	basit tümcecikler
substantive universal	tözel evrenceler
synset	eşanlam kümeleri
tendency	eğilim
Tree Theory	Soyağacı Kuramı
typicality	tipiklik
universal	evrence
Universal Grammar	Evrensel Dilbilgisi
unrestricted universals	kısıtsız evrenceler
verbal categories	eylemcil ulamlar
Wave Theory	Dalga Kuramı
well-formed	düzgün
word order	sözcük dizilişi

EKLER

EK-1: WALS Özellikler Listesi

ID	Alan	Özellik	Değer	Türü	Dil
1A	Sesbilim	Ünsüz Envanterleri	5	Sıralı	563
2A	Sesbilim	Ünlü Nitelik Envanterleri	3	Sıralı	564
3A	Sesbilim	Ünsüz-Ünlü Oranı	5	Sıralı	564
4A	Sesbilim	Patlamalılarda ve Sürtünmelilerde Ötümlüleşme	4	Kategorik	567
5A	Sesbilim	Patlamalı Dizgelerde Ötümlüleşme ve Boşluklar	5	Kategorik	567
6A	Sesbilim	Küçükdilsil Ünsüzler	4	Kategorik	567
7A	Sesbilim	Gırtlaksıl Ünsüzler	8	Kategorik	567
8A	Sesbilim	Yan Ünsüzler	5	Kategorik	567
9A	Sesbilim	Artdamaksıl Genizsil	3	Kategorik	469
10A	Sesbilim	Ünlü Genizsilleşmesi	2	İkili	244
10B	Sesbilim	Batı Afrika'da Genizsil Ünlüler	5	Kategorik	40
11A	Sesbilim	Ön Yuvarlak Ünlüler	4	Kategorik	562
12A	Sesbilim	Seslem Yapısı	3	Sıralı	486
13A	Sesbilim	Ton	3	Sıralı	527
14A	Sesbilim	Sabit Vurgu Konumları	7	Kategorik	502
15A	Sesbilim	Etkiye Duyarlı Vurgu	8	Kategorik	500
16A	Sesbilim	Etkiye Duyarlı Vurgu Dizgelerinde Etki Etmenleri	7	Kategorik	500
17A	Sesbilim	Ritim Tipleri	5	Kategorik	323
18A	Sesbilim	Yaygın Ünsüzlerin Yokluğu	6	Kategorik	567
19A	Sesbilim	Ender Ünsüzlerin Varlığı	7	Kategorik	567
20A	Biçimbilim	Seçilmiş Çekim Biçimlerinin Kaynaşması	7	Kategorik	165
21A	Biçimbilim	Seçilmiş Çekim Biçimlerinin Oluşumu	5	Kategorik	162
21B	Biçimbilim	Zaman-Görünüş-Kip Çekiminin Oluşumu	6	Kategorik	160
22A	Biçimbilim	Eylemin Çekimsel Birleşimi	7	Sıralı	145
23A	Biçimbilim	Tümcecikte Belirleme Konumu	5	Kategorik	236
24A	Biçimbilim	İyelik Ad Öbeklerinde Belirleme Konumları	5	Kategorik	236
25A	Biçimbilim	Belirleme Konumları: Tüm Dil Tipolojisi	5	Kategorik	236
25B	Biçimbilim	Edici ve Etkilenen Üyelerin Sıfır Belirlemesi	2	İkili	235
26A	Biçimbilim	Çekimsel Biçimbilimde Önekleme- Sonekleme	6	Sıralı	969
27A	Biçimbilim	İkileme	3	Kategorik	368
28A	Biçimbilim	Durum Bağdaşması	4	Kategorik	198
29A	Biçimbilim	Eylemcil Kişi/Sayı Belirlemesinde Bağdaşma	3	Kategorik	198

30A	Adcıl Ulamlar	Cinsiyet Sayısı	5	Sıralı	257
31A	Adcıl Ulamlar	Biyolojik Temelli Olan ve Olmayan Cinsiyet Dizgeleri	3	Kategorik	257
32A	Adcıl Ulamlar	Cinsiyet Atama Dizgeleri	3	Kategorik	257
33A	Adcıl Ulamlar	Adcıl Çoğulluğun Kodlanması	9	Kategorik	1066
34A	Adcıl Ulamlar	Adcıl Çoğulluğun Oluşum Sıklığı	6	Kategorik	291
35A	Adcıl Ulamlar	Bağımsız Kişi Adıllarında Çoğulluk	8	Kategorik	261
36A	Adcıl Ulamlar	Çağrışimsal Çoğulluk	4	Kategorik	236
37A	Adcıl Ulamlar	Belirli Tanımlıklar	5	Kategorik	620
38A	Adcıl Ulamlar	Belirsiz Tanımlıklar	5	Kategorik	534
39A	Adcıl Ulamlar	Bağımsız Adıllarda Kapsayıcı/Dışlayıcı Farkı	5	Kategorik	200
39B	Adcıl Ulamlar	Pama-Nyungan'da Kapsayıcı/Dışlayıcı Farkı	2	İkili	71
40A	Adcıl Ulamlar	Eylem Çekiminde Kapsayıcı/Dışlayıcı Farkı	5	Kategorik	200
41A	Adcıl Ulamlar	Göstericilerde Uzaklık Karşıtlığı	5	Sıralı	234
42A	Adcıl Ulamlar	Pronominal ve Adnominal Göstericileri	3	Kategorik	201
43A	Adcıl Ulamlar	Üçüncü Kişi Adılları ve Göstericileri	6	Kategorik	225
44A	Adcıl Ulamlar	Bağımsız Kişi Adıllarında Cinsiyet Ayrımı	6	Kategorik	378
45A	Adcıl Ulamlar	Adıllarda Kibarlık Ayrımları	4	Kategorik	207
46A	Adcıl Ulamlar	Belirsiz Adıllar	5	Kategorik	326
47A	Adcıl Ulamlar	Pekiştirici ve Dönüşlü Adıllar	2	İkili	168
48A	Adcıl Ulamlar	Edatlarda Kişi Gösterimi	4	Kategorik	378
49A	Adcıl Ulamlar	Durumların Sayısı	9	Sıralı	261
50A	Adcıl Ulamlar	Bakışsız Durum Gösterimi	6	Kategorik	261
51A	Adcıl Ulamlar	Durum Eklerinin Konumu	9	Kategorik	1031
52A	Adcıl Ulamlar	Birliktelikler ve Araçlar	3	Kategorik	322
53A	Adcıl Ulamlar	Sıra Sayıları	8	Kategorik	321
54A	Adcıl Ulamlar	Dağılımsal Sayılar	7	Kategorik	251
55A	Adcıl Ulamlar	Sayı Sınıflandırması	3	Kategorik	400
56A	Adcıl Ulamlar	Bağlaçlar ve Evrensel Niceleyiciler	3	Kategorik	116

57A	Adcıl Ulamlar	Adılsı İyelik Eklerinin Konumu	4	Kategorik	902
58A	Adcıl Sözdizim	Zorunlu İyelik Çekimi	2	İkili	244
58B	Adcıl Sözdizim	İyelik Adlarının Sayısı	4	Sıralı	243
59A	Adcıl Sözdizim	İyelik Sınıflandırması	4	Sıralı	243
60A	Adcıl Sözdizim	Tamlayanlar, Sıfatlar ve İlgi Tümcecikleri	6	Kategorik	138
61A	Adcıl Sözdizim	Adsız Sıfatlar	7	Kategorik	124
62A	Adcıl Sözdizim	Hareket Adı Yapıları	8	Kategorik	168
63A	Adcıl Sözdizim	Ad Öbeği Bağlacı	2	İkili	234
64A	Adcıl Sözdizim	Adcıl ve Eylemcil Bağlaç	3	Kategorik	301
65A	Eylemcil Ulamlar	Bitmişlik/Bitmemişlik Görünüşü	2	İkili	222
66A	Eylemcil Ulamlar	Geçmiş Zaman	4	Sıralı	222
67A	Eylemcil Ulamlar	Gelecek Zaman	2	İkili	222
68A	Eylemcil Ulamlar	Bitmişlik	4	Kategorik	222
69A	Eylemcil Ulamlar	Zaman-Görünüş Eklerinin Konumu	5	Kategorik	1131
70A	Eylemcil Ulamlar	Biçimbilimsel Buyrum	5	Kategorik	547
71A	Eylemcil Ulamlar	Yasaklayıcı	4	Kategorik	495
72A	Eylemcil Ulamlar	Buyrum-Öğüt Dizgeleri	4	Kategorik	375
73A	Eylemcil Ulamlar	İstek	2	İkili	319
74A	Eylemcil Ulamlar	Durumsal Olasılık	3	Kategorik	234
75A	Eylemcil Ulamlar	Bilgisel Olasılık	3	Kategorik	240
76A	Eylemcil Ulamlar	Durumsal ve Bilgisel Kip Gösteriminin Örtüşmesi	3	Kategorik	207
77A	Eylemcil Ulamlar	Kanıtsallığın Anlambilimsel Ayrımları	3	Kategorik	418
78A	Eylemcil Ulamlar	Kanıtsallığın Kodlanması	6	Kategorik	418
79A	Eylemcil Ulamlar	Zaman ve Görünüşe Göre Değişik Biçimlenme	4	Kategorik	193
79B	Eylemcil Ulamlar	Buyrum ve Öğütte Değişik Biçimlenme	5	Kategorik	193
80A	Eylemcil Ulamlar	Sözel Sayı ve Değişik Biçimlenmesi	5	Kategorik	193
81A	Sözcük Dizilişi	Özne, Nesne ve Eylem Dizilişi	7	Kategorik	1377
81B	Sözcük Dizilişi	İki Baskın Özne, Nesne ve Eylem Dizilişli Diller	5	Kategorik	67

82A	Sözcük Dizilişi	Özne ve Eylem Dizilişi	3	Kategorik	1497
83A	Sözcük Dizilişi	Nesne ve Eylem Dizilişi	3	Kategorik	1519
84A	Sözcük Dizilişi	Nesne, Yalındışı Nesne ve Eylem Dizilişi	6	Kategorik	500
85A	Sözcük Dizilişi	İlgeç ve Ad Öbeği Dizilişi	5	Kategorik	1183
86A	Sözcük Dizilişi	Tamlayan ve Ad Dizilişi	3	Kategorik	1249
87A	Sözcük Dizilişi	Sıfat ve Ad Dizilişi	4	Kategorik	1366
88A	Sözcük Dizilişi	Gösterici ve Ad Dizilişi	6	Kategorik	1224
89A	Sözcük Dizilişi	Sayı ve Ad Dizilişi	4	Kategorik	1153
90A	Sözcük Dizilişi	İlgi Tümcecigi ve Ad Dizilişi	6	Kategorik	824
90B	Sözcük Dizilişi	Ad-öncesi ilgi tümcecikleri	5	Kategorik	191
90C	Sözcük Dizilişi	Ad-sonrası ilgi tümcecikleri	4	Kategorik	620
90D	Sözcük Dizilişi	İçten başlı ilgi tümcecikleri	7	Kategorik	63
90E	Sözcük Dizilişi	Bağıntılı ilgi tümcecikleri	7	Kategorik	23
90F	Sözcük Dizilişi	Bitişik ilgi tümcecikleri	2	İkili	10
90G	Sözcük Dizilişi	Çift başlı ilgi tümcecikleri	3	Kategorik	5
91A	Sözcük Dizilişi	Derece Sözcüğü ve Sıfat Dizilişi	3	Kategorik	481
92A	Sözcük Dizilişi	Kutupsal Soru Parçacıklarının Konumu	6	Kategorik	884
93A	Sözcük Dizilişi	İçerik Sorularındaki Soru Öbeklerinin Konumu	3	Kategorik	902
94A	Sözcük Dizilişi	Belirteç Bağlayıcı ve Tümcecik Dizilişi	5	Kategorik	659
95A	Sözcük Dizilişi	Nesne ve Eylem Dizilişi ile İlgeç ve Ad Öbeği Dizilişi Arasındaki İlişki	5	Kategorik	1142
96A	Sözcük Dizilişi	Nesne ve Eylem Dizilişi ile İlgi Tümcecigi ve Ad Arasındaki İlişki	5	Kategorik	879
97A	Sözcük Dizilişi	Nesne ve Eylem Dizilişi ile Sıfat ve Ad Dizilişi Arasındaki İlişki	5	Kategorik	1316
98A	Basit Tümcecikler	Tam Ad Öbeklerinin Durum Gösterimlerinin Sıralanışı	6	Kategorik	190
99A	Basit Tümcecikler	Adılların Durum Gösterimlerinin Sıralanışı	7	Kategorik	172
100A	Basit Tümcecikler	Eylemsel Kişi Gösteriminin Sıralanışı	6	Kategorik	380
101A	Basit Tümcecikler	Adıl Öznelerin Söylenişi	6	Kategorik	711
102A	Basit Tümcecikler	Eylemsel Kişi Gösterimi	5	Kategorik	378
103A	Basit Tümcecikler	Eylemsel Kişi Gösteriminin Üçüncü Kişi Sıfırı	6	Kategorik	380

104A	Basit Tümcecikler	Eylemde Kişi Göstericisinin Dizilişi	5	Kategorik	379
105A	Basit Tümcecikler	Çift Geçişli Yapılar: 'ver-' Eylemi	4	Kategorik	378
106A	Basit Tümcecikler	İşteş Yapılar	4	Kategorik	175
107A	Basit Tümcecikler	Edilgen Yapılar	2	İkili	373
108A	Basit Tümcecikler	Anti-edilgen Yapılar	3	Kategorik	194
108B	Basit Tümcecikler	Anti-edilgen Yapının Üretkenliği	4	Sıralı	186
109A	Basit Tümcecikler	Uygun Yapılar	8	Kategorik	183
109B	Basit Tümcecikler	Uygulamalı Nesnelerin Diğer Roller	5	Kategorik	183
110A	Basit Tümcecikler	Dolaylı Ettirgen Yapılar	3	Kategorik	118
111A	Basit Tümcecikler	Dolaysız Ettirgen Yapılar	4	Kategorik	310
112A	Basit Tümcecikler	Olumsuzlayıcı Biçimbirimler	6	Kategorik	1157
113A	Basit Tümcecikler	Bakışimli ve Bakışimsız Ölçünlü Olumsuzlama	3	Kategorik	297
114A	Basit Tümcecikler	Bakışimsız Ölçünlü Olumsuzlamanın Alt Tipleri	7	Kategorik	297
115A	Basit Tümcecikler	Olumsuzlayıcı Belirsiz Adıllar ve Yüklemcil Olumsuzlama	4	Kategorik	206
116A	Basit Tümcecikler	Kutupsal Sorular	7	Kategorik	955
117A	Basit Tümcecikler	Yüklemcil İyelik	5	Kategorik	240
118A	Basit Tümcecikler	Yüklemcil Sıfatlar	3	Kategorik	386
119A	Basit Tümcecikler	Adcıl ve Konumsal Yüklemleme	2	İkili	386
120A	Basit Tümcecikler	Yüklemcil Adlarda Sıfır Koşaç	2	İkili	386
121A	Basit Tümcecikler	Karşılaştırma Yapıları	4	Kategorik	167
122A	Karmaşık Tümceler	Öznelerde İlgilendirme	4	Kategorik	166
123A	Karmaşık Tümceler	Dolaylı Nesnelerde İlgilendirme	5	Kategorik	112
124A	Karmaşık Tümceler	'iste-' Tümleci Özneleri	5	Kategorik	283
125A	Karmaşık Tümceler	Niyet Tümcecikleri	3	Kategorik	170
126A	Karmaşık Tümceler	'-dİğİndA' Tümcecikleri	3	Kategorik	174
127A	Karmaşık Tümceler	Neden Tümcecikleri	3	Kategorik	169
128A	Karmaşık Tümceler	Sözce Tümleci Tümcecikleri	3	Kategorik	143
129A	Sözlükçe	El ve Kol	2	İkili	617
130A	Sözlükçe	Parmak ve El	2	İkili	593

130B	Sözlükçe	'Parmak' ve 'El'in Kimliğiyle Dillerin Kültürel Ulamları	3	Kategorik	72
131A	Sözlükçe	Sayı Tabanları	6	Kategorik	196
132A	Sözlükçe	Türememiş Temel Renk Ulamlarının Sayısı	7	Sıralı	119
133A	Sözlükçe	Temel Renk Ulamlarının Sayısı	7	Sıralı	119
134A	Sözlükçe	Yeşil ve Mavi	7	Kategorik	120
135A	Sözlükçe	Kırmızı ve Sarı	5	Kategorik	120
136A	Sözlükçe	M-T Adılları	3	Kategorik	230
136B	Sözlükçe	Birinci Tekil Kişide M	2	İkili	230
137A	Sözlükçe	N-M Adılları	3	Kategorik	230
137B	Sözlükçe	İkinci Tekil Kişide M	2	İkili	230
138A	Sözlükçe	Çay	3	Kategorik	230
139A	İşaret Dilleri	İşaret Dillerinde Düzensiz Olumsuzlayıcılar	4	Sıralı	35
140A	İşaret Dilleri	İşaret Dillerinde Soru Parçacıkları	3	Sıralı	38
141A	Diğer	Yazı Dizgeleri	6	Kategorik	6
142A	Diğer	Şaklamaların Dil-ötesi Kullanımı	3	Kategorik	143
143A	Sözcük Dizilişi	Olumsuzlayıcı Biçimbirim ve Eylem Dizilişi	17	Kategorik	1324
143B	Sözcük Dizilişi	Zorunlu Çift Olumsuzlama	15	Kategorik	119
143C	Sözcük Dizilişi	Seçimlik Çift Olumsuzlama	22	Kategorik	81
143D	Sözcük Dizilişi	Seçimlik Üçlü Olumsuzlama	6	Kategorik	6
143E	Sözcük Dizilişi	Eylem-öncesi Olumsuzlayıcı Biçimbirimler	4	Kategorik	1324
143F	Sözcük Dizilişi	Eylem-sonrası Olumsuzlayıcı Biçimbirimler	4	Kategorik	1324
143G	Sözcük Dizilişi	Olumsuzluğu işaret etmenin küçük biçimbilimsel yolları	4	Kategorik	1324
144A	Sözcük Dizilişi	Özne, Nesne ve Eyleme İlişkin Olumsuzlayıcı Sözcük Dizilişi	21	Kategorik	1189
144B	Sözcük Dizilişi	Olumsuzlayıcı sözcüklerin tümcecğin başı veya sonuna göre ve eyleme bitişikliğine göre konumu	6	Kategorik	609
144C	Sözcük Dizilişi	Olumsuz tümceciklerde farklı sözcük dizilişli diller	17	Kategorik	28
144D	Sözcük Dizilişi	ÖEN Dillerinde Olumsuzlayıcı Biçimbirimlerin Konumu	15	Kategorik	463
144E	Sözcük Dizilişi	ÖEN Dillerinde Çoklu Olumsuzlayıcı Yapılar	28	Kategorik	48
144F	Sözcük Dizilişi	ÖEN Dillerinde Zorunlu Çift Olumsuzlama	22	Kategorik	56
144G	Sözcük Dizilişi	ÖEN Dillerinde Seçimlik Çift Olumsuzlama	18	Kategorik	35
144H	Sözcük Dizilişi	OlmszÖEN Dizilişi	4	Kategorik	420
144I	Sözcük Dizilişi	ÖOlmszEN Dizilişi	8	Kategorik	421
144J	Sözcük Dizilişi	ÖEOlmszN Dizilişi	7	Kategorik	446

144K	Sözcük Dizilişi	ÖENOİmsz Dizilişi	4	Kategorik	446
144L	Sözcük Dizilişi	ÖNE Dillerinde Olumsuzlayıcı Biçimbirimlerin Konumu	23	Kategorik	572
144M	Sözcük Dizilişi	ÖNE Dillerinde Çoklu Olumsuzlayıcı Yapılar	27	Kategorik	54
144N	Sözcük Dizilişi	ÖNE Dillerinde Zorunlu Çift Olumsuzlama	17	Kategorik	45
144O	Sözcük Dizilişi	ÖNE Dillerinde Seçimlik Çift Olumsuzlama	21	Kategorik	31
144P	Sözcük Dizilişi	OlmszÖNE Dizilişi	4	Kategorik	408
144Q	Sözcük Dizilişi	ÖOlmszNE Dizilişi	4	Kategorik	408
144R	Sözcük Dizilişi	ÖNOİmszE Dizilişi	8	Kategorik	411
144S	Sözcük Dizilişi	ÖNEOlmsz Dizilişi	11	Kategorik	489
144T	Sözcük Dizilişi	Eylem-Baş Dillerde Olumsuzlayıcı Biçimbirimlerin Konumu	23	Kategorik	152
144U	Sözcük Dizilişi	Eylem-baş dillerde çift olumsuzlama	15	Kategorik	17
144V	Sözcük Dizilişi	Eylem-öncesi Olumsuzlamalı Eylem-Baş	9	Kategorik	152
144W	Sözcük Dizilişi	Hemen Eylem-sonrası veya Özne ve Nesne Arası Olumsuzlamalı Eylem-Baş	6	Kategorik	151
144X	Sözcük Dizilişi	Tümcecik-Sonu Olumsuzlamalı Eylem-Baş	4	Kategorik	151
144Y	Sözcük Dizilişi	Nesne-Baş Dillerde Olumsuzlayıcı Biçimbirimlerin Konumu	11	Kategorik	16

EK-2: Normalizasyon Sürecinde Yapılan İşlemlerin Dökümü

Doldurulan Özellik	Doldurulan Dil	Girilen Değer	Açıklama
-	Fiote ve Jiarong	Meta bilgiler	WALS'in web sürümünde adı geçen iki dilin yalnızca meta bilgileri geçmekte, ancak bu diller hiçbir özellik için değer almamıştır. Bu nedenle, veri çerçevemizde bulunmadıklarından dillere ait vektörler, meta bilgileriyle eklenmiştir.
10B	10B'de değer almayan tüm diller	0	10B özelliği sadece Batı Afrika dilleri içindir ve bu diller de ilgili özellik için etiketlenmiştir.
15A	Cocopa ²⁴	8 - Fixed stress (no weight-sensitivity)	Dil 14A'da sabit vurgulu olarak etiketlenmesine karşın 15A'da etiketlenmemiştir.
16A	Cocopa	1 - No weight, or weight factor unknown	Dil 14A'da sabit vurgulu olarak etiketlenmesine karşın 16A'da etiketlenmemiştir.
16A	Navajo	1 - No weight, or weight factor unknown	Dil 14A'da sabit olmayan bir vurgu ile etiketlenmiştir, ancak vurgunun neye göre belirlendiği bilinmediğinden 16A'da ilgili seçenek ile doldurulmuştur.

²⁴ Bu hata WALS takımıyla da paylaşılmıştır: <https://github.com/cldf-datasets/wals/issues/93>

25B	25A'da ilk üç değerden birini alıp 25B'de etiketlenmeyen diller [Kui (Endonezya değişkesi) ve Ngaanyatjarra] ²⁵	2 - Non-zero marking	Adı geçen iki dil dışında, 25A'da sıfır imlemeli olarak işaretlenmeyen her dil 25B'de de buna koşut olarak aynı değerle etiketlenmiştir. İlgili diller bu çerçevede doldurulmuştur.
39B	Pama-Nyungan dili olmayan tüm diller	0	39B özelliği sadece Pama-Nyungan dilleri içindir ve bu diller de ilgili özellik için etiketlenmiştir.
58A	58B'de bir değer alan, ancak 58A'da etiketlenmeyen diller [Kui (Hindistan değişkesi)] ²⁶	1 - Obligatorily possessed nouns exist	58B'de bir değer alan her dil 58A'da zorunlu olarak ilgili değeri alması gerektiğinden doldurulmuştur.
58B	58A'da ilk değeri alıp 58B'de etiketlenmeyen diller [Kui (Endonezya değişkesi) ve Tzutujil] ³	1 – None reported	Adı geçen iki dil dışında, 58A'da zorunlu iyeliğin bulunduğu işaretlenen her dil 58B'de de buna koşut olarak bir değer almıştır. İlgili dillerin iyelik adlarının sayısına dair bir bilgi bulunmadığından diller bu çerçevede doldurulmuştur.
81B	81A'da değer alan ancak 81B'de değersiz bırakılan her dil	0	81A'da belli bir baskın sözcük dizilişi olduğu belirtilen her dil, aynı anda iki baskın dizilişi olamayacağı için ve 81A'da bir baskın sözcük

²⁵ Bu hata WALS takımıyla da paylaşılmıştır: <https://github.com/cldf-datasets/wals/issues/94>

²⁶ Bu hata WALS takımıyla da paylaşılmıştır: <https://github.com/cldf-datasets/wals/issues/95>

			dizilişi olmadığı belirtilen her dil 81B’de bir değer mantiken alamayacağı için ilgili diller doldurulmuştur.
90B	90A’da ikinci değeri almayan ve 90B’de değer almayan her dil	0	90C, ilgi tümceciğinin addan önce geldiği diller için uygun olduğundan bu sınıfa girmeyen diller ilgili değerle doldurulmuştur.
90C	90A’da ilk değeri almayan ve 90C’de değer almayan her dil	0	90C, ilgi tümceciğinin addan sonra geldiği diller için uygun olduğundan bu sınıfa girmeyen diller ilgili değerle doldurulmuştur.
90D	90A’da üçüncü değeri almayan ve 90D’de değer almayan her dil	0	90D, ilgi tümceciğinin içten başlı diller için uygun olduğundan bu sınıfa girmeyen diller ilgili değerle doldurulmuştur.
90E	90A’da dördüncü değeri almayan ve 90E’de değer almayan her dil	0	90E, ilgi tümceciğinin bağıntılı olduğu diller için uygun olduğundan bu sınıfa girmeyen diller ilgili değerle doldurulmuştur.
90F	90A’da beşinci değeri almayan ve 90F’de değer almayan her dil	0	90F, ilgi tümceciğinin bitişik olduğu diller için uygun olduğundan bu sınıfa girmeyen diller ilgili değerle doldurulmuştur.
90G	90A’da altıncı değeri almayan ve 90G’de değer almayan her dil	0	90G, ilgi tümceciğinin çift başlı olduğu diller için uygun olduğundan bu sınıfa girmeyen

			diller ilgili deęerle doldurulmuştur.
95A	85A’da ilgeci olmadığı belirtilen tüm diller	5 - Languages not falling into one of the preceding four types	85A’da ilgeci olmadığı belirtilen diller, 95A’daki herhangi bir sınıfa giremeyeceğı için ilgili deęerle doldurulmuştur.
130B	130A’da “el” ve “parmak” birimlerinde fark olduğı bildirilen diller	0	130B “el” ve “parmak” birimlerinin aynı olduğı dillere özel olduğı için, farklı olan diller için bu özellik geçerli değıldir.
143B	143A’da zorunlu çift olumsuzlamaya sahip olmadığı belirtilen (14 ve 16) dışındaki diller	0	143B yalnızca 143A’da zorunlu çift olumsuzlamaya sahip olduğı belirtilen diller için olduğundan, bu özellığı olmayan diller doldurulmuştur.
143C	143A’da seçimlik çift olumsuzlamaya sahip olmadığı belirtilen (15 ve 17) dışındaki diller	0	143C yalnızca 143A’da seçimlik çift olumsuzlamaya sahip olduğı belirtilen diller için olduğundan, bu özellığı olmayan diller doldurulmuştur.
143D	143A’da seçimlik üçlü olumsuzlamaya sahip olmadığı belirtilen (16 ve 17) dışındaki diller	0	143D yalnızca 143A’da seçimlik üçlü olumsuzlamaya sahip olduğı belirtilen diller için olduğundan, bu özellığı olmayan diller doldurulmuştur.
144C	144A’da deęer alan, ancak 144C’de deęer almayan diller	0	144C olumsuzluk dizilişı baskın sözcük dizilişinden farklı olan diller için olduğundan, her iki dizilişı de koşut olan diller doldurulmuştur.

144D-K	Baskın sözcük dizilişi ÖEN olmayan ve değer almamış diller	0	Bu özellik serisi baskın sözcük dizilişi ÖEN olan diller için olduğundan, baskın sözcük dizilişi farklı bir değer alan diller doldurulmuştur.
144L-S	Baskın sözcük dizilişi ÖNE olmayan ve değer almamış diller	0	Bu özellik serisi baskın sözcük dizilişi ÖNE olan diller için olduğundan, baskın sözcük dizilişi farklı bir değer alan diller doldurulmuştur.
144T-X	Eylem baş olmayan ve değer almamış diller	0	Bu özellik serisi eylem baş olan diller için olduğundan, baskın sözcük dizilişi farklı bir değer alan diller doldurulmuştur.
144Y	Nesne baş olmayan ve değer almamış diller	0	Bu özellik serisi nesne baş olan diller için olduğundan, baskın sözcük dizilişi farklı bir değer alan diller doldurulmuştur.